Empirical Coordination with Multiple Descriptions

Michail Mylonakis\textsuperscript{1}, Photios A. Stavrou\textsuperscript{1} and Mikael Skoglund\textsuperscript{1}

Abstract—We extend the framework of empirical coordination to a distributed setup where for a given action by nature, multiple descriptions of the action of the decoder are available. We adopt the coding strategy applied by El Gamal and Cover in [1] to get a lower bound of the coordination region. Then, we improve this region by applying the coding scheme applied by Zhang and Berger in [2].

I. INTRODUCTION

The development of information technology and the need for wireless connectivity in networks have enabled a plethora of new transformative technologies and applications. On the level of information theory, this development necessitates the need for distributed coding between the nodes of a network depending on the action taken whereas at the same time, it is desirable for some applications, such as, parallel computing across large scale networks, a network of computers to be interconnected under coherent strategies while distributing the work load across the participating machines (i.e., machine-type communication).

In this paper, we apply ideas from a distributed source coding setup called the multiple description problem (MDP) [3] to the framework of empirical coordination using a fidelity criterion that was recently introduced in [4].

The idea of MDP was formalized in [3] and it is concerned with lossy encoding of information for transmission over an unreliable (and possibly digital) multi-channel communication system. The receiver knows which subset of the channels is working whereas the transmitter does not. The problem boils down to the design of an MD system which, for given channel rates, minimizes the distortions due to reconstruction of the source using information from any subsets of the channels. The first achievability rate-region for MDP using two channels was established in [1] and further generalized in [2]. A detailed and simple exposition of these results is given in [5]. The rate region in [1] was shown to be optimal for memoryless Gaussian sources and mean squared error distortion in [6]. This result was later extended to create high-rate bounds for stationary and smooth sources in [7] and for stationary Gaussian processes in [8]. Bounds to the optimal rate distortion region for MDP for more than two channels are studied in [9]–[12].

The notion of empirical coordination was introduced in [13] and it is achieved if the joint type, measured by total variation distance, of the actions (or nodes) in a network is close to the desired distribution, in probability. This kind of coordination has been studied in various settings, see e.g., [14] and merged with ideas from other research fields, such as game theory [15] and networked control systems [16]. The framework of empirical coordination of [13] was recently extended to the more general framework of empirical coordination subject to fidelity also termed “imperfect empirical coordination” by [4] who was inspired by the work of [17]. Following [4], imperfect empirical coordination is established if the total variation between the joint type of the actions in a network comes close, on average, to a desired distribution within distance pre-specified by a threshold $\Delta$. Clearly, if $\Delta \rightarrow 0$, then, we obtain the empirical coordination in the sense of [13].

In this work, we extend the idea of MDs to the framework of imperfect empirical coordination, using the model of Fig. 1. In the system model of Fig. 1 we are given the action of a random variable $X$ (possibly by nature) that induces the distribution $p_0(x)$. Moreover, the action of $Y$ is produced using two available rate-limited noiseless communication links. We seek, the amount of the communication that is required such that the average distance between the joint type of the actions and the desired joint distribution $p_0(x)p_{Y|X}(y|x)$ to be smaller than a certain threshold $\Delta$, for $i = 1,2,12$, depending on whether the first, the second, or both indices arrive at the decoder, respectively. The system in Fig. 1 can be used in practice. For instance, consider a network where due to packet loss some part of the message does not arrive at the receiver. A traditional coordinated system will fail, while a multiple description system can still coordinate the nodes, with less accuracy.

II. DEFINITIONS

In this section, we describe formally our problem. We begin with some mathematical preliminaries and the definition of the $\Delta$-neighborhood.

Definition 1 (Joint type): The joint type $P_{x^n,y^n}$ of a tuple of sequences $(x^n, y^n)$ is the empirical probability mass
function, given by
\[ P_{X^n,Y^n}(x,y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{ (x_i, y_i) = (x, y) \}, \]
for all \((x,y) \in \mathbb{X} \times \mathbb{Y}\), where \(\mathbf{1}\) is the indicator function.

**Definition 2 (Total variation):** The total variation between two probability mass functions (PMF) is given by
\[ \| p(x,y) - q(x,y) \|_{TV} = \frac{1}{2} \sum_{x,y} | p(x,y) - q(x,y) |. \]

**Definition 3 (Δ-neighborhood):** The Δ-neighborhood of a PMF \(p(x,y)\) is defined as
\[ N_\Delta(p(x,y)) = \{ q(x,y) : \| p(x,y) - q(x,y) \|_{TV} \leq \Delta \}. \]

**Definition 4 (Coordination code):** The coordination code for our set-up consists of five functions-two encoding functions
\[ i_i : \mathbb{X}^n \rightarrow \{1, \ldots, e^{nR_i}\}, \quad i = 1, 2, \]
and three decoding functions
\[ q_i^n : \{1, \ldots, e^{nR_i}\} \rightarrow \mathbb{Y}^n, \quad i=1,2, \]
\[ q_{12} : \{1, \ldots, e^{nR_1}\} \times \{1, \ldots, e^{nR_2}\} \rightarrow \mathbb{Y}^n. \]

**Definition 5 (Achievability):** A desired PMF \(p_{XY}(x,y)\) is achievable for \(\Delta_1\Delta_2\Delta_{12}\)-empirical coordination with the rate-pair \((R_1, R_2)\) if there is an \(N\) such that for all \(n > N\), there exists a coordination code \((e^{nR_1}, e^{nR_2}, n)\) such that
\[ \mathbb{E}\{ \| P_{X^n,Y^n}(x,y) - p_0(x,y) p_{Y|X}(y|x) \|_{TV} \} \leq \Delta_i, \]
for \(i = 1, 2, 12\), depending on whether the first, the second, or both indices arrive at the decoder, respectively.

**Definition 6 (Multiple-rate-distortion-coordination region):** The multiple-rate-distortion-coordination region \(R^T_{p_0}(p_{Y|X}(y|x))\) for every source PMF \(p_0(x)\) and for every conditional PMF \(p_{Y|X}(y|x)\) is defined as:
\[ R^T_{p_0}(p_{Y|X}(y|x)) \triangleq \text{CI} \left\{ \begin{array}{l}
(R_1, R_2, \Delta_1, \Delta_2, \Delta_{12}) : \quad p_0(x) p_{Y|X}(y|x) \text{ is achievable for } \Delta_1\Delta_2\Delta_{12}\text{-empirical coordination at rates } (R_1, R_2) \end{array} \right\}, \]

where
\[ R \left( p_0 p_{\tilde{Y}^{(1)}(1), \tilde{Y}^{(1)}(2), \tilde{Y}^{(12)}(12)} | X \right) \]
\[ \triangleq \left\{ \begin{array}{l}
(R_1, R_2, \Delta_1, \Delta_2, \Delta_{12}) : \\
R_1 \geq I(X; \tilde{Y}^{(1)}, U), \\
R_2 \geq I(X; \tilde{Y}^{(2)}, U), \\
R_1 + R_2 \geq I(X; \tilde{Y}^{(1)}, \tilde{Y}^{(2)}, U) + I(X; \tilde{Y}^{(1)}, \tilde{Y}^{(2)}, \tilde{Y}^{(12)}, U) \\
\end{array} \right\}. \]

**Theorem 2:** The following region is a subset of the Multiple-rate-distortion-coordination region \(R^T_{p_0}(p_{Y|X}(y|x))\) for the source PMF \(p_0(x)\) and for every conditional PMF \(p_{Y|X}(y|x)\):
\[ R^T_{p_0}(p_{Y|X}(y|x)) \supseteq \bigcup \left\{ R \left( p_0 p_{\tilde{Y}^{(1)}(1), \tilde{Y}^{(1)}(2), \tilde{Y}^{(12)}(12)} | X \right) : \right. \]
\[ \left. (R_1, R_2, \Delta_1, \Delta_2, \Delta_{12}) : \\
R_1 \geq I(X; \tilde{Y}^{(1)}, U), \\
R_2 \geq I(X; \tilde{Y}^{(2)}, U), \\
R_1 + R_2 \geq I(X; \tilde{Y}^{(1)}, \tilde{Y}^{(2)}, U) + I(X; \tilde{Y}^{(1)}, \tilde{Y}^{(2)}, \tilde{Y}^{(12)}, U) \right\}. \]

IV. PROOFS

**A. Proof of Theorem 7**

First, we give the proof of Theorem 7 and, then, we sketch the proof of Theorem 2.

- **Setup:** Choose \(\varepsilon_1, \varepsilon_2 > 0\), a PMF \(p_{\tilde{Y}^{(1)}}(y^{(1)}), p_{\tilde{Y}^{(2)}}(y^{(2)}), p_{\tilde{Y}^{(12)}}(y^{(12)})\) and compute the marginals \(p_{\tilde{Y}^{(1)}}(y^{(1)}), p_{\tilde{Y}^{(2)}}(y^{(2)}), p_{\tilde{Y}^{(12)}}(y^{(12)})\) as marginal distributions of \(p_{X}\) for \(p_{\tilde{Y}^{(i)}}(y^{(i)}), p_{\tilde{Y}^{(12)}}(y^{(12)})\) for \(i = 1, 2\).

- **Codebook design:** Generate \(|\mathcal{E}^{n(H_i + \varepsilon_1)}|\) length-\(n\) code-words \(Y^{(i)}(w_i)\), \(w_i = 1, \ldots, |\mathcal{E}^{n(H_i + \varepsilon_1)}|\), by choosing each of the \(|\mathcal{E}^{n(H_i + \varepsilon_1)}|\) symbols \(Y^{(i)}(w_i)\) independently at random according to \(p_{X}(x)\) for \(i = 1, 2\).

- **Encoder Design:** For a given sequence \(x\), the encoders try to find a pair \((w_1, w_2)\) such that
\[ \left( x, Y^{(1)}(w_1), Y^{(2)}(w_2), Y^{(12)}(w_1, w_2) \right) \in \mathcal{E}^{n(H_1 + \varepsilon_1)}(x, Y^{(1)}(w_1), Y^{(2)}(w_2), Y^{(12)}(w_1, w_2)), \]
If they find several possible choices, they pick the first. If they find none, they choose \( w_1 = w_2 = 1 \). The first encoder \( i_1 \) puts out \( w_1 \) and the second encoder \( i_2 \) puts out \( w_2 \).

- **Decoder Design:** The decoder \((y_1^n, y_2^n, y_{12}^n)\) consists of three different decoding functions, depending on whether \( w_1, w_2, \) or both are received. It puts out \( Y(i)(w_i) \) if \( w_i \) is received for \( i=1,2,12 \).

- **Performance Analysis:** We define \( \epsilon' \triangleq \sqrt[2n]{\frac{x}{2n}} \) and partition the space into three disjoint cases: (a) \( x \not\in \mathcal{A}^{s(n)}(p_X) \) in which case we for sure cannot find a pair \((w_1, w_2)\) such that (1) is satisfied (b) \( x \in \mathcal{A}^{s(n)}(p_X) \) and for every pair \((w_1, w_2)\), (1) is not satisfied (c) \( x \in \mathcal{A}^{s(n)}(p_X) \) and exists a pair \((w_1, w_2)\) such that (1) is satisfied. Note that the reduced \( \epsilon' \) is only necessary in order to be able to rely on the lower bound in Lemma 3. Obviously, if \( X \in \mathcal{A}^{s(n)}(p_X) \), then \( X \in \mathcal{A}^{s(n)}(p_X) \).

We obtain in each of the three scenarios \( i = 1, 2, 12 \):

\[
\begin{align*}
&\mathbb{E}\left[ \left\| P_{x^n,y^n}(x, y) - p_0(x) p(y|x) \right\|_{TV} \right] = \\
&\left[ \mathbb{E}\left[ \left\| P_{x^n,y^n}(x, y) - p_0(x) p(y|x) \right\|_{TV} \mid \text{Case a} \right] \times \Pr(\text{Case a}) \right]_{\leq \epsilon} \\
&+ \left[ \mathbb{E}\left[ \left\| P_{x^n,y^n}(x, y) - p_0(x) p(y|x) \right\|_{TV} \mid \text{Case b} \right] \times \Pr(\text{Case b}) \right]_{\leq \epsilon} \\
&+ \left[ \mathbb{E}\left[ \left\| P_{x^n,y^n}(x, y) - p_0(x) p(y|x) \right\|_{TV} \mid \text{Case c} \right] \times \Pr(\text{Case c}) \right]_{\leq 1}
\end{align*}
\]

By Lemma 2 (in Appendix), we can bound the probability of Case (a) as \( \Pr(\text{Case a}) \leq \delta_i(n, \epsilon', X) \). To bound the expected distortion in Case (c), we note that if \((x, \hat{Y}(1), \hat{Y}(2), \hat{Y}(12))\) is jointly typical, then, each pair \((x, \hat{Y}(1)), (x, \hat{Y}(2)), (x, \hat{Y}(12))\) is jointly typical so

\[
\left\| P_{x^n,y^n}(x, y) - p_0(x) p(y|x) \right\|_{TV} \leq \left\| P_{x^n,y^n}(x, y) - p_0(x) p_{\hat{Y}(i)|X}(y|x) \right\|_{TV} + \|p_0(x) p_{\hat{Y}(i)|X}(y|x) - p_0(x) p(y|x)\|_{TV} \leq \epsilon/2 + \Delta_i,
\]

by choosing \( p_{\hat{Y}(i)|X}(y|x) \) such that \( p_0(x) p_{\hat{Y}(i)|X}(y|x) \in N_{\Delta_i} \) for \( i = 1, 2, 12 \).

Given some \( x \in \mathcal{A}^{s(n)}(p_X) \), define \( F(w_1, w_2) \) to be the event that \( w_1 \) and \( w_2 \) give a good choice of codewords, i.e.,

\[
F(w_1, w_2) \triangleq \left\{ \left( \hat{Y}(1)(w_1), \hat{Y}(2)(w_2) \right), \hat{Y}(12)(w_1, w_2) \in \mathcal{A}^{s(n)}(p_X, \hat{Y}(1), \hat{Y}(2), \hat{Y}(12)|X) \right\}.
\]

We can write \( \Pr(\text{Case b}) = \bigcap_{w_1, w_2} \left( F_c(w_1, w_2) \right) = \Pr[\hat{K} = 0] \) with \( K \triangleq \sum_{w_1, w_2} 1(F(w_1, w_2)) \). Note, that if \( K = 0 \), then \( |K - E[K]| = |E[K]| = E[K] \geq \frac{E[K]}{2} \). So, using this bound, we get \( \Pr(\text{Case b}) = \Pr[K = 0] \leq \Pr\left[ |K - E[K]| \geq \frac{E[K]}{2} \right] \leq \text{Var}[E[K]] / (\frac{E[K]}{2})^2 \)

\[
= \frac{\left( E[K^2] - \left( E[K] \right)^2 \right)}{(E[K])^2},
\]

where the second inequality follows from the Chebyshev inequality. We write down,

\[
\Pr(\text{Case b}) \leq \frac{4 \left( E[K^2] - \left( E[K] \right)^2 \right)}{(E[K])^2}.
\]

Therefore, it remains to derive some bounds on \( E[K] \) and \( E[K^2] \). Firstly, \( E[K] \):

\[
E[K] = \mathbb{E}\left[ \sum_{w_1, w_2} 1(F(w_1, w_2)) \right]
= \sum_{w_1, w_2} \mathbb{E}\left[ 1(F(w_1, w_2)) \right]
= \sum_{w_1, w_2} \text{Pr}(F(w_1, w_2)) + \delta_i + \sum_{w_1, w_2} \sum_{\hat{Y}(1), \hat{Y}(2), \hat{Y}(12)} p_{\hat{Y}(1)^{12}}(\hat{Y}(1), \hat{Y}(2)) p_{\hat{Y}(12)}(\hat{Y}(12)|\hat{Y}(1), \hat{Y}(2)) \cdot \delta_i.
\]

Here, we introduce the shorthand, \( \mathcal{A}^{s(n)}(p_X) \) for \( \mathcal{A}^{s(n)}(p_X, \hat{Y}(1), \hat{Y}(2), \hat{Y}(12)|X) \). By repeating application of Lemma 2 and Lemma 3 (in Appendix), we bound the previous as follows:

\[
E[K] \geq \mathbb{E}\left[ \sum_{w_1, w_2} \sum_{\hat{Y}(1), \hat{Y}(2), \hat{Y}(12)} -n(H(\hat{Y}(1)) + \epsilon_m) \cdot e^{-n(H(\hat{Y}(1))^+ - \epsilon_m)} -n(H(\hat{Y}(2)) + \epsilon_m) \cdot e^{-n(H(\hat{Y}(2)) + \epsilon_m)} \cdot e^{-n(H(\hat{Y}(12)) + \epsilon_m)} \cdot e^{-n(H(\hat{Y}(12)) + \epsilon_m)} \cdot e^{-n(H(\hat{Y}(12)) + \epsilon_m)} \cdot e^{-n(H(\hat{Y}(12)) + \epsilon_m)}
\]

\[
\geq \sum_{w_1, w_2} \left( 1 - \delta_i \right) e^{-n(H(\hat{Y}(1)) + H(\hat{Y}(2)) + H(\hat{Y}(12)) + 3\epsilon_m)} \cdot e^{-n(H(\hat{Y}(1)) + H(\hat{Y}(2)) + H(\hat{Y}(12)) + 3\epsilon_m)} \cdot e^{-n(H(\hat{Y}(12)) + 3\epsilon_m)}
\]

\[
= [e^{n(R_1 + \epsilon_m)}] [e^{n(R_2 + \epsilon_m)}] [1 - \delta_i].
\]

\[
\geq \left( \frac{e^{n(R_1 + \epsilon_m)} e^{n(R_2 + \epsilon_m)}}{e^{n(R_1 + \epsilon_m)} e^{n(R_2 + \epsilon_m)}} \right) [1 - \delta_i],
\]

\[
\geq \left( \frac{e^{n(R_1 + \epsilon_m)} e^{n(R_2 + \epsilon_m)}}{e^{n(R_1 + \epsilon_m)} e^{n(R_2 + \epsilon_m)}} \right) [1 - \delta_i].
\]
\begin{align*}
\cdot \exp \left( n \left( H \left( \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)} | X \right) - H \left( \hat{Y}^{(1)} \right) \right)
- H \left( \hat{Y}^{(2)} | \hat{Y}^{(1)} \right) - H \left( \hat{Y}^{(12)} | \hat{Y}^{(1)}, \hat{Y}^{(2)} \right)
+ H \left( \hat{Y}^{(2)} | \hat{Y}^{(1)} \right) - H \left( \hat{Y}^{(2)} - 4\epsilon_m \right) \right) \\
= \exp \left( n \left( R_1 + R_2 + \epsilon_1 + \epsilon_2 
- I \left( X ; \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)} \right) - I \left( \hat{Y}^{(1)} ; \hat{Y}^{(2)} - \delta_1 \right) \right) \right)
\end{align*}

(9)

Next, we tackle $\mathbb{E} [K^2]$:

$$
\mathbb{E} [K^2] = \mathbb{E} \left[ \sum_{u_1, w_2} \mathbb{E} \left[ 1 \left( F \left( w_1, w_2 \right) \right) \right] \sum_{v_1, v_2} 1 \left( F \left( v_1, v_2 \right) \right) \right]
= \sum_{u_1, w_2} \sum_{v_1, v_2} \mathbb{E} \left[ 1 \left( F \left( w_1, w_2 \right) \right) \right] \mathbb{E} \left[ 1 \left( F \left( v_1, v_2 \right) \right) \right]
= \sum_{u_1, w_2} \sum_{v_1, v_2} \Pr \left( F \left( w_1, w_2 \right) \cap F \left( v_1, v_2 \right) \right)
= \sum_{\Theta \subset \{1, 2\}} \Pr \left( \sum_{u_1, w_2, v_1, v_2, \text{with overlap } \Theta} \Pr \left( F \left( w_1, w_2 \right) \cap F \left( v_1, v_2 \right) \right) \right)
(11)
$$

where in the last step we distinguish four cases of whether $w_1 = v_1$ or not. These cases are described by the four possible subsets of $\{1, 2\}$: $\Theta = \{1, 2\}, \Theta = \{1\}, \Theta = \{2\}, \Theta = \emptyset$. For $i \in \Theta$, we have $w_i = v_i$, whereas for the remaining indices $i \in \Theta^c$ we have $w_i \neq v_i$. Next, we go through these four cases.

**Case $\Theta = \{1, 2\}$:** We have $w_1 = v_1$ and $w_2 = v_2$. Using a derivation similar to the one that gave (4)-(10) we get

$$
\sum_{u_1, w_2, v_1, v_2, \text{with overlap } \Theta = \{1, 2\}} \Pr \left( F \left( w_1, w_2 \right) \cap F \left( v_1, v_2 \right) \right)
= \sum_{u_1, w_2} \Pr \left( F \left( w_1, w_2 \right) \right)
< \sum_{u_1, w_2} \exp \left( n \left( I \left( X ; \hat{Y}^{(1)}, \hat{Y}^{(2)} \right) \right)
+ I \left( \hat{Y}^{(1)} ; \hat{Y}^{(2)} - 4\epsilon_m \right) \right)
\leq \exp \left( n \left( R_1 + R_2 + \epsilon_1 + \epsilon_2 
- I \left( X ; \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)} \right) - I \left( \hat{Y}^{(1)} ; \hat{Y}^{(2)} \right) 
+ \delta_2 \right) \right)
\leq \sum_{\hat{g}^{(1)} \in A^{(n)}_e (\cdot | x)} \Pr \left( \hat{Y}^{(1)} \left( w_1 \right) = \hat{g}^{(1)} \left( x \right) \right)
\cdot \Pr \left( \hat{Y}^{(2)} \left( w_2 \right), \hat{Y}^{(12)} \left( w_1, w_2 \right) \in A^{(n)}_e (\cdot | x, \hat{g}^{(1)}) \right)
\leq \sum_{\hat{g}^{(1)} \in A^{(n)}_e (\cdot | x)} \Pr \left( \hat{Y}^{(1)} \left( w_1 \right) = \hat{g}^{(1)} \left( x \right) \right)
= 1
$$

(12)

where $\delta_2$ accounts for the $4\epsilon_m$-term and the offset from the flooring operations.

**Case $\Theta = \{1\}$:** We have $w_1 = v_1$ and $w_2 \neq v_2$. Using the shorthands $A^{(n)}_e (\cdot | x)$ for $A^{(n)}_e (p_{X, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)} | X})$ or $A^{(n)}_e (p_{X, \hat{Y}^{(1)} | X})$, respectively (it should be clear from the context, which distribution needs to be plugged in), and $A^{(n)}_e (\cdot | x, \hat{g}^{(1)})$ for $A^{(n)}_e (p_{X, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)} | X, \hat{g}^{(1)}}, \hat{g}^{(1)}$, we get

$$
\Pr \left( F \left( w_1, w_2 \right) \cap F \left( v_1, v_2 \right) \right)
= \Pr \left( \left\{ \left( \hat{Y}^{(1)} \left( w_1 \right), \hat{Y}^{(2)} \left( w_2 \right), \hat{Y}^{(12)} \left( w_1, w_2 \right) \right) \in A^{(n)}_e (\cdot | x) \right\} 
\cap \left\{ \left( \hat{Y}^{(1)} \left( v_1 \right), \hat{Y}^{(2)} \left( v_2 \right), \hat{Y}^{(12)} \left( v_1, v_2 \right) \right) \in A^{(n)}_e (\cdot | x) \right\} \right)
= \Pr \left( \left\{ \left( \hat{Y}^{(1)} \left( w_1 \right) \in A^{(n)}_e (\cdot | x) \right\} 
\cap \left\{ \left( \hat{Y}^{(2)} \left( w_2 \right), \hat{Y}^{(12)} \left( w_1, w_2 \right) \right) \in A^{(n)}_e (\cdot | x, \hat{Y}^{(1)} \left( w_1 \right)) \right\} 
\cap \left\{ \left( \hat{Y}^{(2)} \left( v_2 \right), \hat{Y}^{(12)} \left( v_1, v_2 \right) \right) \in A^{(n)}_e (\cdot | x, \hat{Y}^{(1)} \left( v_1 \right)) \right\} \right) \right)
= \Pr \left( \left\{ \left( \hat{Y}^{(1)} \left( w_1 \right) \right) \cap A^{(n)}_e (\cdot | x, \hat{Y}^{(1)} \left( w_1 \right)) \right\} \right)
= \Pr \left( \left\{ \left( \hat{Y}^{(2)} \left( w_2 \right), \hat{Y}^{(12)} \left( w_1, w_2 \right) \right) \in A^{(n)}_e (\cdot | x, \hat{Y}^{(1)} \left( w_1 \right)) \right\} \right)
= \Pr \left( \left\{ \left( \hat{Y}^{(1)} \left( w_1 \right) \right) \right\} \right)
= \Pr \left( \left\{ \left( \hat{Y}^{(2)} \left( w_2 \right), \hat{Y}^{(12)} \left( w_1, w_2 \right) \right) \right\} \right)
\leq \sum_{\hat{g}^{(1)} \in A^{(n)}_e (\cdot | x)} \Pr \left( \hat{Y}^{(1)} \left( w_1 \right) = \hat{g}^{(1)} \left( x \right) \right)
= 1
$$

(13)

where in (13) we have used Lemma 1 (in Appendix); (14) must be understood as the definitions of the events $A_a, A_b, A_c$; and where (15) follows from the chain rule. Note that conditionally on $A_a$, the events $A_b$ and $A_c$ are independent of each other, i.e., in (15) we have the two terms $\Pr \left( A_b | A_a \right)$ and $\Pr \left( A_c | A_a \right)$ that are basically the same. Let us investigate them more closely. We have

$$
\Pr \left( A_b | A_a \right)
= \sum_{\hat{g}^{(1)} \in A^{(n)}_e (\cdot | x)} \Pr \left( \hat{Y}^{(1)} \left( w_1 \right) = \hat{g}^{(1)} \left( x \right) \right)
\cdot \Pr \left( \left\{ \left( \hat{Y}^{(2)} \left( w_2 \right), \hat{Y}^{(12)} \left( w_1, w_2 \right) \right) \in A^{(n)}_e (\cdot | x, \hat{g}^{(1)}) \right\} \right)
\leq \sum_{\hat{g}^{(1)} \in A^{(n)}_e (\cdot | x)} \Pr \left( \hat{Y}^{(1)} \left( w_1 \right) = \hat{g}^{(1)} \left( x \right) \right)
= 1
$$

(15)
Hence, we get the following bound:

\[
\max_{\hat{y}^{(1)} \in A_2^*(n) \mid |x|} \Pr \left[ \left( \hat{Y}^{(2)}(w_2), \hat{Y}^{(12)}(w_1, w_2) \right) \in A_2^*(n) \left( |x, \hat{y}^{(1)} \right) \right] \leq \exp \left( n \left( R_1 + 2R_2 + \epsilon_1 + 2\epsilon_2 \right) - 2I(X; \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) - 2I(\hat{Y}^{(1)}, \hat{Y}^{(2)}) + I(X; \hat{Y}^{(1)} + \delta_3) \right),
\]

(21)

where \( \delta_3 \) accounts for \( 8\epsilon_m \) and the rounding offsets.

Case \( \Theta = \{2\} \): This is the case as the case \( \Theta = \{1\} \), but with exchanged roles of \( Y^{(1)} \) and \( Y^{(2)} \):

\[
\sum_{w_1, w_2, v_1 \neq w_1} \Pr \left( F(w_1, w_2) \cap F(v_1, v_2) \right)
< \exp \left( n \left( 2R_1 + R_2 + 2\epsilon_1 + \epsilon_2 \right) - 2I(X; \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) - 2I(\hat{Y}^{(1)}, \hat{Y}^{(2)}) + I(X; \hat{Y}^{(2)} + \delta_4) \right).
\]

(22)

Case \( \Theta = \emptyset \): In this case we have both \( w_1 \neq v_1 \) and \( w_2 \neq v_2 \), i.e., the two events \( F(w_1, w_2) \) and \( F(v_1, v_2) \) are independent. Hence, we have

\[
\sum_{w_1, w_2, v_1 \neq w_1} \Pr \left( F(w_1, w_2) \cap F(v_1, v_2) \right) = \sum_{w_1, w_2} \Pr \left( F(w_1, w_2) \right) \Pr \left( F(v_1, v_2) \right)
= \sum_{w_1, w_2} \Pr \left( F(w_1, w_2) \right) \sum_{v_1 \neq w_1} \Pr \left( F(v_1, v_2) \right)
\leq \sum_{w_1, w_2} \Pr \left( F(w_1, w_2) \right) \sum_{v_1 \neq w_1} \Pr \left( F(v_1, v_2) \right)
= \left( \sum_{w_1, w_2} \Pr \left( F(w_1, w_2) \right) \right)^2 = (\mathbb{E}[K])^2.
\]

(23)

(24)

Here, in (23) we increase the number of terms in the second sum; and in (24) we use (4). Hence, by plugging (12), (21), (22), (24) in (11), and using (10), we get from (3) the following bound for \( \Pr(\cdot) \):

\[
\Pr(\text{Case } b) < 4\exp \left( n \left( -R_1 - R_2 - \epsilon_1 - \epsilon_2 \right) + I(X; \hat{Y}^{(1)} + \hat{Y}^{(2)}, \hat{Y}^{(12)}) \right) + 4\exp \left( n \left( -R_1 + \epsilon_1 + I(X; \hat{Y}^{(1)}) \right) \right) + \delta_4 + 2\delta_1) \]

\[
\leq \delta_5,
\]

where \( \delta_5 \) is arbitrarily small if \( n \) is large enough and \( \epsilon \) small enough such that \( \delta_2 + 2\delta_1 < \epsilon_1 + \epsilon_2 \).
B. Proof of Theorem 2

Assume that there exists a third encoder whose index always safely arrives at the decoder, i.e., the third encoder sees a noise-free channel. We assign the rate $R_0$ to this third encoder.

- **Setup:** Choose $\epsilon_0, \epsilon_1, \epsilon_2 > 0$, a PMF $p_{Y(0)}(Y(0), Y(1), Y(2)|X)$, compute the marginals $p_{Y(0)}(Y(0))$, $p_{Y(1)}(Y(1)|Y(0))$, and $p_{Y(2)}(Y(2)|Y(0), Y(1))$ as marginal distributions of $p_{X}p_{Y(0)}(Y(0), Y(1), Y(2)|X)$. Then, we define all cases back into (2), now gives us the characterization of the Theorem. This completes the proof.

- **Encoder Design:** For a given sequence $x$, the encoders try to find a triple $(w_0, w_1, w_2)$ such that
  \[
  \Pr\left( F(w_0, w_1, w_2) \cap F(v_0, v_1, v_2) \right) \leq \left( \sum_{w_0, w_1, w_2} \Pr(F(w_0, w_1, w_2)) \right)^2.
  \]
  If $w_0 \neq v_0$, then, the two events $F(w_0, w_1, w_2)$ and $F(v_0, v_1, v_2)$ are disjoint because $w_0$ is a counter that was used in the generation of all codewords simultaneously so the cases $\Theta = \{1, 2\}, \Theta = \{1\}, \Theta = \{2\}$ and $\Theta = \emptyset$ can be treated jointly.

- **Case $\Theta = \{1, 2\}, \Theta = \{1\}, \Theta = \{2\}$ and $\Theta = \emptyset$: 
  \[
  \sum_{w_0, w_1, w_2} \Pr\left( F(w_0, w_1, w_2) \cap F(v_0, v_1, v_2) \right)
  \]
  \[
  = \sum_{w_0, w_1, w_2} \Pr(F(w_0, w_1, w_2))
  \]
  \[
  \leq \left( \sum_{w_0, w_1, w_2} \Pr(F(w_0, w_1, w_2)) \right)^2
  \]
  \[
  = \left( \mathbb{E}[K] \right)^2.
  \]

\section*{Performance Analysis}

We again distinguish three different cases: (a) $x \notin \mathcal{K}_c^{(n)}(p_X)$ (b) $x \in \mathcal{K}_c^{(n)}(p_X)$, and for every pair $(w_0, w_1, w_2)$, \eqref{case-c} is not satisfied (c) $x \in \mathcal{K}_c^{(n)}(p_X)$ and exists a pair $(w_0, w_1, w_2)$ such that \eqref{case-c} is satisfied. The analysis of (a) and (c) is similar to what we did in Theorem 1 thus, we omit it. We only study $\Pr(C = \text{Case b})$. \begin{align}
F(w_0, w_1, w_2) &\subseteq \left\{ (Y(0), Y(1), Y(2)) | (w_0, Y(0), Y(1), Y(2)) \right\}
\end{align}
\[ \leq \exp \left( n \left( R_0 + R_1 + 2R_2 + \epsilon_0 + \epsilon_1 + 2\epsilon_2 \right) \right) \]

\[ + I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) \]
\[ - 2I (\hat{Y}^{(1)}, \hat{Y}^{(2)}|\hat{Y}^{(0)}) + \delta_3 \right) \right) . \]

\[ \sum_{w_0,w_1,w_2,v_0,v_1,v_2} \Pr \left( F (w_0, w_1, w_2) \cap F (v_1, v_1, v_2) \right) \]

\[ \leq \exp \left( n \left( R_0 + 2R_1 + 2R_2 + \epsilon_0 + \epsilon_1 + 2\epsilon_2 \right) \right) \]
\[ + I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) - 2I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) \]
\[ - 2I (\hat{Y}^{(1)}, \hat{Y}^{(2)}|\hat{Y}^{(0)}) + \delta_4 \right) . \]

\[ \sum_{w_0,w_1,w_2,v_0,v_1,v_2} \Pr \left( F (w_0, w_1, w_2) \cap F (v_0, v_1, v_2) \right) \]

\[ \leq \exp \left( n \left( R_0 + 2R_1 + 2R_2 + \epsilon_0 + \epsilon_1 + 2\epsilon_2 \right) \right) \]
\[ + I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) - 2I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) \]
\[ - 2I (\hat{Y}^{(1)}, \hat{Y}^{(2)}|\hat{Y}^{(0)}) + \delta_5 \right) . \]

\[ \{ \Pr (\text{Case } b) \leq \left( \frac{E[K^2] - (E[K])^2}{(E[K])^2} \right) \right) \]
\[ < 4 \exp \left( n \left( - R_0 - R_1 - R_2 - \epsilon_0 - \epsilon_1 - 2\epsilon_2 \right) \right) \]
\[ + \delta_2 + 2\delta_1 + I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) \]
\[ + I (\hat{Y}^{(1)}, \hat{Y}^{(2)}|\hat{Y}^{(0)}) \right) + 4 \exp \left( n \left( - R_0 \right) \right) \]
\[ - R_1 - \epsilon_0 - \epsilon_1 + I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)} + \delta_3 + 2\delta_1) \right) \]
\[ + 4 \exp \left( n \left( - R_0 - R_2 - \epsilon_0 - \epsilon_2 \right) \right) \]
\[ + I (X; \hat{Y}^{(0)}, \hat{Y}^{(2)} + \delta_4 + 2\delta_1) \right) \]
\[ + 4 \exp \left( n \left( - R_0 - \epsilon_0 + I (X; \hat{Y}^{(0)}) \right) \right) \]
\[ + \delta_5 + 2\delta_1 \right) \right) \equiv \delta_6 , \]

where \( \delta_6 \) is arbitrarily small if \( n \) is large enough and \( \epsilon \) small enough such that \( \delta_2 + 2\delta_1 < \epsilon_0 + \epsilon_1 + \epsilon_2 \), \( \delta_3 + 2\delta_1 < \epsilon_0 + \epsilon_1 + \epsilon_2 \), and \( \delta_5 + 2\delta_1 < \epsilon_0 \), and if \( R_0 \geq I (X; \hat{Y}^{(0)}) \), \( R_0 + R_1 > I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)}) \), \( R_0 + R_2 > I (X; \hat{Y}^{(0)}, \hat{Y}^{(2)}) \), \( R_0 + R_1 + R_2 \geq I (X; \hat{Y}^{(0)}, \hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(12)}) \) \( I (\hat{Y}^{(1)}, \hat{Y}^{(2)}|\hat{Y}^{(0)}) \).

In fact, we do not have access to such a third guaranteed channel. However, we can simulate it by adding \( nR_0 \) nats to both of the other channels. Then, in all interesting three cases (only \( w_1 \) arrives, only \( w_2 \) arrives, and both arrive) we have these nats and they act like they had come over the virtual third channel. This now means that we adapt our rates \( R_1 \) and \( R_2 \), \( R_1 \equiv R_1 + R_0 \) and \( R_2 \equiv R_2 + R_0 \). Plugging this into our relationships, using that \( R_0 \geq I (X; \hat{Y}^{(0)}) \) and rename \( \hat{Y}^{(0)} \) as \( U \), in combination will the other two cases, give us the characterization of the Theorem. This completes the proof.

V. APPENDIX

Definition 7 (Strongly \( \epsilon \)-typical set [5]): Fix an \( \epsilon > 0 \), a PMF \( p_{X,Y} (x,y) \), and a blocklength \( n \). The strongly typical set \( A^{(n)}_\epsilon (p_{X,Y}) \) with respect to the PMF \( p_{X,Y} (x,y) \) is defined as

\[ A^{(n)}_\epsilon (p_{X,Y}) \triangleq \left\{ (x,y) \in X^n \times Y^n : \left| p_{X,Y} (a,b) - p_{X,Y} (a,b) \right| < \epsilon \frac{\sqrt{\|X\|Y}}{2^n}, \forall (a,b) \in X \times Y, \right\} \]

Definition 8 (Conditionally strongly typical set [5]): For some joint PMF \( p_{X,Y} \) with marginal \( p_X \) and for some given sequence \( x \in A^{(n)}_\epsilon (p_X) \), we define the conditionally strongly typical set with respect to \( p_{X,Y} \) as

\[ A^{(n)}_\epsilon (p_{X,Y} | x) \triangleq \left\{ y \in Y^n : (x,y) \in A^{(n)}_\epsilon (p_{X,Y}) \right\} \]

Lemma 1 (5): The event \( \{ (X,Y) \in A^{(n)}_\epsilon (p_{X,Y}) \} \) is equivalent to the event

\[ \{ X \in A^{(n)}_\epsilon (p_X) \} \cap \{ Y \in A^{(n)}_\epsilon (p_{X,Y} | X) \} . \]
Definition 9 ($\epsilon_m, \delta_l [5]$): We name one particular $\epsilon$ and one particular $\delta$ that we meet often:

$$
\epsilon_m(p_{X,Y}(x,y)) \triangleq -\epsilon \log(p_{\min}^{X,Y}(x,y)),
$$

$$
\delta_l(n, \epsilon, \mathbb{X} \times \mathbb{Y}) \triangleq (n+1)|\mathbb{X} \times \mathbb{Y}| e^{-n \frac{\epsilon^2}{2(n+1)|\mathbb{X} \times \mathbb{Y}|^2} \log \epsilon},
$$

where $p_{\min}^{X,Y}$ denotes the smallest positive value of $p_{X,Y}(x,y)$.

Lemma 2 ([5]): Let $(x,y) \in A^*_e(n)(p_{X,Y})$. Then,

$$
e^{-n(H(X,Y) + \epsilon_m(p_{X,Y}(x,y)))} < p_{X,Y}^n(x,y),
$$

$$
e^{-n(H(X,Y) - \epsilon_m(p_{X,Y}(x,y)))} < e^{n(H(Y|X)+\epsilon_m(p_{X,Y}(x,y)))},
$$

Moreover,

$$
1 - \delta_l(n, \epsilon, \mathbb{X} \times \mathbb{Y}) \leq \Pr[(x,y) \in A^*_e(n)(p_{X,Y})] \leq 1.
$$

Lemma 3 ([5]): For every $y \in A^*_e(n)(p_{X,Y}|x)$ we obtain

$$
e^{-n(H(Y|X)+\epsilon_m(p_{X,Y}(x,y)))} < p_{X|Y}^n(y|x),
$$

$$
e^{-n(H(Y|X) - \epsilon_m(p_{X,Y}(x,y)))} < e^{n(H(Y|X)+\epsilon_m(p_{X,Y}(x,y)))},
$$

The size of the conditionally strongly typical set is bounded as

$$
|A^*_e(n)(p_{X,Y}|x)| < e^{n(H(Y|X)+\epsilon_m(p_{X,Y}(x,y)))},
$$

If also $x \in A^*_e(n)(p_X)$, then, we can obtain

$$
|A^*_e(n)(p_{X,Y}|x)| > (1 - \delta_l(n, \epsilon, \mathbb{X} \times \mathbb{Y})) e^{n(H(Y|X)-\epsilon_m(p_{X,Y}(x,y)))},
$$

Lemma 4 ([5]): Let $p_{X,Y}(x,y)$ be a joint PMF with marginals $p_X(x), p_Y(y)$. Let $(x,y)$ be generated:$\{(x_k,y_k)\}_k=1$ IID $\sim p_X(x)p_Y(y)$. Then,

$$
(1 - \delta_l(n, \epsilon, \mathbb{X} \times \mathbb{Y})) e^{-n(I(X;Y)+\epsilon_2)} < \Pr[(x,y) \in A^*_e(n)(p_{X,Y})] < e^{-n(I(X;Y)-\epsilon_2)},
$$

where $\epsilon_2 \triangleq \epsilon_m(p_{X,Y}(x,y)) + \epsilon_m(p_X(x)) + \epsilon_m(p_Y(y)) \leq 3\epsilon_m(p_{X,Y}(x,y)).$

Moreover,

$$
\Pr[Y \in A^*_e(n)(p_{X,Y}|X)] < e^{-n(I(X;Y)-\epsilon_3)},
$$

where $\epsilon_3 \triangleq \epsilon_m(p_{X,Y}(x,y)) + \epsilon_m(p_{X}(x)) \leq 2\epsilon_m(p_{X,Y}(x,y)).$ Further, if $x \in A^*_e(n)(p_X)$, then, we obtain

$$
\Pr[Y \in A^*_e(n)(p_{X,Y}|X)] > (1 - \delta_l(n, \epsilon, \mathbb{X} \times \mathbb{Y})) e^{-n(I(X;Y)+\epsilon_3)}.
$$

References

[1] A. E. Gamal and T. Cover, “Achievable rates for multiple descriptions,” IEEE Trans. Inf. Theory, vol. 28, no. 6, pp. 851–857, November 1982.
[2] Zhen Zhang and T. Berger, “New results in binary multiple descriptions,” IEEE Trans. Inf. Theory, vol. 33, no. 4, pp. 502–521, July 1987.
[3] J. K. Wolf, A. D. Wyner, and J. Ziv, “Source coding for multiple descriptions,” The Bell System Technical Journal, vol. 59, no. 8, pp. 1417–1426, Oct 1980.
[4] M. Mylonakis, P. A. Stavrou, and M. Skoglund, “Empirical coordination subject to a fidelity criterion,” in Proc. IEEE Inf. Theory Workshop, 2019. [Online]. Available: arxiv.org
[5] S. M. Moser, “Lecture notes: Advanced topics in information theory,” 2019. [Online]. Available: http://moser-isi.ethz.ch/docs.
[6] L. Ozarow, “On a source-coding problem with two channels and three receivers,” Bell Labs Tech. J., vol. 59, no. 10, pp. 1909–1921, Dec 1980.
[7] R. Zamir, “Gaussian codes and Shannon bounds for multiple descriptions,” IEEE Trans. Inf. Theory, vol. 45, no. 7, pp. 2629–2636, Nov 1999.
[8] P. L. Dragotti, S. D. Servetto, and M. Vetterli, “Optimal filter banks for multiple description coding: analysis and synthesis,” IEEE Trans. Inf. Theory, vol. 48, no. 7, pp. 2036–2052, July 2002.
[9] R. Venkataramani, G. Kramer, and V. K. Goyal, “Multiple description coding with many channels,” IEEE Trans. Inf. Theory, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.
[10] S. S. Pradhan, R. Puri, and K. Ramchandran, “n-channel symmetric multiple descriptions - part I: (n, k) source-channel erasure codes,” IEEE Trans. Inf. Theory, vol. 50, no. 1, pp. 47–61, Jan 2004.
[11] R. Puri, S. S. Pradhan, and K. Ramchandran, “n-channel symmetric multiple descriptions-part II: an achievable rate-distortion region,” IEEE Trans. Inf. Theory, vol. 51, no. 4, pp. 1377–1392, April 2005.
[12] K. B. Viswanatha, E. Akyol, and K. Rose, “Combinatorial message sharing and a new achievable region for multiple descriptions,” IEEE Trans. Inf. Theory, vol. 62, no. 2, pp. 769–792, Feb 2016.
[13] P. W. Cuff, H. H. Permuter, and T. M. Cover, “Coordination capacity,” IEEE Trans. Inf. Theory, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.
[14] A. Bereyhi, M. Bahrami, M. Miromolaei, and M. R. Aref, “Empirical coordination in a triangular multiterminal network,” in Proc. IEEE Int. Symp. Inf. Theory, July 2013, pp. 2149–2153.
[15] M. Le Treust and T. Tomala, “Information design for strategic coordination of autonomous devices with non-aligned utilities,” in Proc. Annual Allerton Conf. on Comm., Control, and Comput., Sep. 2016, pp. 233–242.
[16] E. Shafieepoorfard and M. Raginsky, “Sequential empirical coordination under an output entropy constraint,” IEEE Trans. Inf. Theory, vol. 64, no. 10, pp. 6830–6841, Oct 2018.
[17] G. Kramer and S. A. Savari, “Communicating probability distributions,” IEEE Trans. Inf. Theory, vol. 53, no. 2, pp. 518–525, Feb 2007.