

Channels with Action Dependent States and Additional Private Messages

Viswanathan Ramachandran¹, Sibi Raj B Pillai², and Vinod M Prabhakaran³

¹Department of Electrical Engineering, Indian Institute of Technology Bombay

²Department of Electrical Engineering, Indian Institute of Technology Bombay

³School of Technology and Computer Science, Tata Institute of Fundamental Research

Abstract

In channels with action dependent states, a message is conveyed using two encoders operating sequentially, *viz.* an action encoder and a channel encoder. The actions drive the output of a discrete-memoryless channel (DMC), which in turn forms the state process for the DMC between the channel encoder and receiver. Assuming non-causal knowledge of the state-process at the channel encoder, a single letter characterization of the capacity is known in the discrete memoryless case.

We consider the action dependent state channel where an additional private message needs to be communicated by the channel encoder. In addition, we consider a common reconstruction (CR) of the state at the channel encoder and decoder. Capacity characterizations for the discrete memoryless and Gaussian versions are presented. As a consequence, we settle the capacity characterization of the Gaussian action dependent channel with only a common message and CR. We further show that the availability of strictly causal state feedback to the action encoder, even with randomization allowed, does not improve the capacity of the action dependent state channel.

I. INTRODUCTION

The problem of coding for channels with state was introduced in the seminal paper by Shannon [1], wherein causal state information was assumed at the encoder. The capacity for the non-causal setting was established by Gelfand and Pinsker [2]. In these models, the state process is assumed to be given by nature. Motivated by applications involving multi-stage encoding (for instance, two-stage recording on a magnetic storage device), Weissman [3] introduced the notion of a *channel with action dependent states (ADSC)*. In this setting, the transmitter can take *actions* that influence the formation of channel states in the first stage, and the encoding in the second stage is based upon the channel state sequence so generated and the message. Notice that the actions play a dual role of message communication as well as controlling the channel states. The capacity of this model was derived in [3].

Following [3], the action-dependent framework has been extended in several directions. Permuter *et al.* [4] studied the source coding dual in which the decoder can take actions based on the observed compression index that influence the formation of side information. Asnani *et al.* [5] considered a setting in which the encoder as well as the decoder can take probing actions to learn the channel state, with a cost constraint associated with each. Choudhuri *et al.* [6] considered causal state communication over an action-dependent channel and characterized the trade-off between message communication and state estimation distortion. Ahmadi *et al.* [7] studied action-dependent channels with the additional constraint of common reconstructions (CR) [8] of the state at the encoder and decoder. Recently, Kittichokechai *et al.* [9] studied source and channel coding settings with action-dependent states and CR constraints, wherein the actions control the partial channel state information available at the encoder and decoder. The action dependent model has been extended to multi-user channels as well— for instance see Steinberg *et al.* [10] and Steinberg [11].

Notice that single letter capacity characterizations for the discrete-memoryless action dependent state channel (DMADSC) as well as the DMADSC with CR constraints were presented in [3] and [7] respectively. However, the Gaussian counterparts are listed therein as open problems. Another interesting question listed in [3] is whether strictly causal state-information at the action encoder increases the capacity. In this work, we consider a more general model where the channel encoder has an additional private message stream along with CR constraints. Our contributions are summarized below.

- The capacity region for the ADSC with additional private message and CR is derived for both the discrete-memoryless and Gaussian versions.
- We prove the optimality of Gaussian auxiliaries for the Gaussian action dependent model with CR constraints [7].
- We show that the capacity of an ADSC is unchanged by strictly causal state feedback to the action encoder, even if the encoders are allowed to randomize, thereby settling a question left open in [3].

We note that our problem framework without common reconstruction constraints is somewhat similar to the co-operative multiple access model studied by Zaidi *et al.* [12] [13]. The setting in [13] consists of a state dependent multiple access channel with degraded message sets, wherein the encoder sending only the common message (non-cognizant encoder) observes the state strictly causally, while the other encoder observes the state non-causally. The capacity region was derived for both the discrete memoryless as well as Gaussian versions, with the observation that strictly causal state information can enlarge the capacity region compared to the case of no state observation at the non-cognizant encoder.

Though the action encoder in our setting can be viewed to be playing the role of the non-cognizant encoder, the fundamental difference between the two settings is that the state S^n in [13] is generated by nature IID, while it is the output of a DMC fed with the message dependent action sequence in our case. An equivalence can be established between the two settings if and only if $p(s|a) = p(s)$, i.e. the state sequence is independent of the action sequence. As such, our conclusion that strictly causal state feedback does not increase the capacity for the common message case is consistent with the observation that the common message capacity in [13] is independent of strictly causal state observation at the non-cognizant encoder- see Corollary 2 and Remark 5 in [13]. Nevertheless, the capacity region in the Gaussian case without common reconstructions is the same as that established in Theorem 4 and Corollary 3 of [13]. The novelty in our model is that we are interested in common reconstruction of the state process in the action-dependent setting with additional messages, which also settles the optimality of Gaussian auxiliaries for the model studied in [7].

Organization: We introduce the system model and main results in Section II. Section III contains the proof of our main result. The extension to the Gaussian case is given in Section IV, which also deals with the Gaussian action dependent channel with common reconstruction constraints as well as reversible input constraints. The generalization of strictly causal state information with randomization at the action encoder is discussed in Section V. Concluding remarks are given in Section VI.

II. SYSTEM MODEL

Consider the model shown in Fig. 1. There are two encoders, namely an action encoder $\mathcal{E}_{\text{act}}$ and a channel encoder $\mathcal{E}_{\text{chan}}$. A common message W_c is observed by both $\mathcal{E}_{\text{act}}$ and $\mathcal{E}_{\text{chan}}$. The action encoder chooses the output symbol $a \in \mathcal{A}$ which is fed to a DMC $p(s|a)$. The output of this DMC forms the state sequence of a channel with input $x \in \mathcal{X}$, output $y \in \mathcal{Y}$ and transition probability $p(y|x, s)$. The encoder $\mathcal{E}_{\text{chan}}$ has non-causal access to the state sequence S^n . $\mathcal{E}_{\text{chan}}$ also needs to convey a private message W_1 . We term this model as *DMADSC with additional private message*. More generally, one can consider common reconstruction (CR) of the state process [8], wherein $\mathcal{E}_{\text{chan}}$ and the decoder must agree on a reconstruction of the state viz. $\hat{S}^n$. We call this model as *DMADSC with CR*. The inputs have to satisfy an average cost constraint defined by a vector function $\gamma : \mathcal{A} \times \mathcal{X} \rightarrow [0, \infty)^2$, where the cost function for sequences is

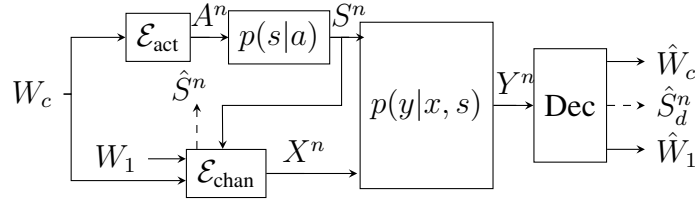


Fig. 1. Action-Dependent Channel with additional private message, and with/without common reconstructions

defined as $\gamma(a^n, x^n) = \frac{1}{n} \sum_{i=1}^n \gamma(a_i, x_i)$. Define a single letter distortion measure $d : \mathcal{S} \times \hat{\mathcal{S}} \rightarrow [0, \infty)$ for state reconstruction, where the distortion between sequences is defined as $d(s^n, \hat{s}^n) = \frac{1}{n} \sum_{i=1}^n d(s_i, \hat{s}_i)$. We assume that the distortion measure is bounded and let $D_{max} = \max_{s \in \mathcal{S}, \hat{s} \in \hat{\mathcal{S}}} d(s, \hat{s})$. We also assume that all the alphabets $|\mathcal{A}|, |\mathcal{S}|, |\mathcal{X}|, |\mathcal{Y}|, |\hat{\mathcal{S}}|$ are finite.

Definition 1. Let Γ be a vector in $\mathbb{R}_+^2$. For the DMADSC with additional private message, an $(n, R_1, R_c, \Gamma, \epsilon)$ scheme consists of two encoder maps:

$$A^n(W_c) : [1 : 2^{nR_c}] \rightarrow \mathcal{A}^n,$$

$$X^n(W_c, W_1, S^n) : [1 : 2^{nR_c}] \times [1 : 2^{nR_1}] \times \mathcal{S}^n \rightarrow \mathcal{X}^n,$$

and a decoding map $\psi(Y^n) : \mathcal{Y}^n \rightarrow [1 : 2^{nR_1}] \times [1 : 2^{nR_c}]$ such that for independent and uniformly distributed choices of (W_c, W_1) , we have

$$\mathbb{P}(\psi(Y^n) \neq (W_1, W_c)) \leq \epsilon, \quad (1)$$

$$\sum_{i=1}^n \mathbb{E}[\gamma_k(A_i, X_i)] \leq n(\Gamma_k + \epsilon), k = 1, 2. \quad (2)$$

(where γ_k and Γ_k represent the k th coordinates of γ and Γ .)

For the DMADSC with CR, we additionally define a sender quantization map $\phi : \mathcal{S}^n \times \{1, \dots, 2^{nR_c}\} \times \{1, \dots, 2^{nR_1}\} \rightarrow \hat{\mathcal{S}}^n$ and a decoder reconstruction map $\psi_2(Y^n) : \mathcal{Y}^n \rightarrow \hat{\mathcal{S}}^n$ s.t.

$$\mathbb{P}(\psi_2(Y^n) \neq \phi(S^n, W_c, W_1)) \leq \epsilon, \quad (3)$$

$$\mathbb{E}[d(S^n, \psi_2(Y^n))] \leq n(D + \epsilon). \quad (4)$$

We say that a tuple (R_1, R_c, D, Γ) is achievable if an $(n, R_1, R_c, D, \Gamma, \epsilon)$ coding scheme exists for every $\epsilon > 0$ for sufficiently large n . Let $\mathcal{C}_{CR}^{dmadsc}$ be the collection of all achievable (R_1, R_c, D, Γ) tuples. Our main result is stated now.

Theorem 2. For the DMADSC $(p(s|a), p(y|x, s))$ with CR, $\mathcal{C}_{CR}^{dmadsc}$ is the closure of the union of all (R_1, R_c, D, Γ) tuples such that

$$R_1 \leq I(U; Y|A) - I(U; S|A), \quad (5)$$

$$R_1 + R_c \leq I(A, U; Y) - I(U; S|A), \quad (6)$$

where the union is over distributions of the form $p(a)p(s|a)p(u|a, s)p(x|u, s)p(y|x, s)$ for which there exists a map $\phi : \mathcal{U} \rightarrow \hat{\mathcal{S}}$ such that $\mathbb{E}[d(S, \phi(U))] \leq D$ and $\mathbb{E}[\gamma(A, X)] \leq \Gamma$ (where the cost constraint is to be understood componentwise), with $|\mathcal{U}| \leq |\mathcal{A}| \cdot |\mathcal{X}| \cdot |\mathcal{S}| + 2$. Notice that $U \rightarrow (X, S) \rightarrow Y$ and $A \rightarrow (U, S) \rightarrow X$ are Markov chains.

Remark 3. For the special case of no CR (i.e. $D \geq D_{max}$) and $R_c = 0$ ($W_c = \phi$), the private message capacity is

$$R_1 \leq \max_{a \in \mathcal{A}, p(u|a, s), p(x|u, s)} [I(U; Y|A = a) - I(U; S|A = a)],$$

which corresponds to selecting the action sequence that leads to the maximum Gelfand-Pinsker rate. On the other hand, when $D \geq D_{max}$ and $R_1 = 0$ ($W_1 = \phi$), the common message capacity is

$$R_c \leq \max_{p(a), p(u|a,s), p(x|u,s)} [I(A, U; Y) - I(U; S|A)],$$

which is nothing but the characterization in [3, Theorem 1].

III. PROOF OF THEOREM 2

Proof of Achievability: The achievability is proven using a combination of Gelfand-Pinsker (GP) coding and superposition coding. Here, an action codebook is built first based on the common message. Then for each a^n (action) sequence, a conditional GP codebook U shall be generated according to the private message. The details are given in Appendix A.

Proof of Converse: By Fano's inequality, we can write:

$$H(W_1, W_c, \hat{S}^n | Y^n) \leq n\epsilon_n, \text{ where } \epsilon_n \xrightarrow{n \rightarrow \infty} 0. \quad (7)$$

We suppress the $n\epsilon_n$ terms in the sequel. We can write the following chain of inequalities for the private rate:

$$\begin{aligned} nR_1 &= H(W_1) \stackrel{(a)}{=} H(W_1 | W_c) \\ &= H(\hat{S}^n, W_1 | W_c) - H(\hat{S}^n | W_1, W_c) \\ &\stackrel{(b)}{=} H(\hat{S}^n, W_1 | W_c) - H(\hat{S}^n | W_1, W_c) + H(\hat{S}^n | W_1, W_c, S^n) \\ &\stackrel{(c)}{\approx} I(\hat{S}^n, W_1; Y^n | W_c, A^n) - I(\hat{S}^n; S^n | W_1, W_c, A^n) \\ &\stackrel{(d)}{=} I(\hat{S}^n, W_1; Y^n | W_c, A^n) - I(\hat{S}^n, W_1; S^n | W_c, A^n) \\ &= \sum_{i=1}^n \left[I(\hat{S}^n, W_1; Y_i | W_c, Y^{i-1}, A^n) - I(\hat{S}^n, W_1; S_i | W_c, A^n, S_{i+1}^n) \right] \\ &\stackrel{(e)}{\leq} \sum_{i=1}^n \left[I(\hat{S}^n, W_1, W_c, Y^{i-1}, A^{i-1}, A_{i+1}^n; Y_i | A_i) - I(\hat{S}^n, W_1, W_c, S_{i+1}^n, A^{i-1}, A_{i+1}^n; S_i | A_i) \right] \\ &\stackrel{(f)}{=} \sum_{i=1}^n \left[I(\hat{S}^n, W_1, W_c, Y^{i-1}, S_{i+1}^n, A^n; Y_i | A_i) - I(\hat{S}^n, W_1, W_c, Y^{i-1}, S_{i+1}^n, A^n; S_i | A_i) \right] \\ &\stackrel{(g)}{=} \sum_{i=1}^n [I(U_i; Y_i | A_i) - I(U_i; S_i | A_i)], \end{aligned} \quad (8)$$

where (a) follows since $W_1 \perp W_c$, (b) follows since $\hat{S}^n$ is a deterministic function of (W_1, W_c, S^n) , (c) follows from Fano's inequality and since W_c determines A^n , (d) follows since $(W_1, W_c) \rightarrow A^n \rightarrow S^n$, (e) follows since $S_i \rightarrow A_i \rightarrow (W_c, S_{i+1}^n, A^{i-1}, A_{i+1}^n)$, (f) follows from the Csiszár sum lemma and (g) follows with a choice of $U_i = (\hat{S}^n, W_1, W_c, Y^{i-1}, S_{i+1}^n, A^n)$ that satisfies the Markov conditions $U_i \rightarrow (X_i, S_i) \rightarrow Y_i$ and $A_i \rightarrow (U_i, S_i) \rightarrow X_i$. Now by introducing a time-sharing random variable Q , independent of everything else and uniformly drawn on $[1 : n]$, we have

$$\begin{aligned} nR_1 &\leq n [I(U_Q; Y_Q | A_Q, Q) - I(U_Q; S_Q | A_Q, Q)] \\ &= n [I(U_Q; Y_Q | A_Q, Q) - I(U_Q, Q; S_Q | A_Q)] \\ &\leq n [I(U_Q, Q; Y_Q | A_Q) - I(U_Q, Q; S_Q | A_Q)]. \end{aligned} \quad (9)$$

Now the proof is completed by replacing $(Q, U_Q) = U$, $Y_Q = Y$, $S_Q = S$, $X_Q = X$ and $A_Q = A$ and noting that the Markov conditions $U \rightarrow (X, S) \rightarrow Y$ and $A \rightarrow (U, S) \rightarrow X$ hold. For the sum rate, consider the

following chain of inequalities:

$$\begin{aligned}
n(R_1 + R_c) &= H(W_1, W_c) \\
&= H(\hat{S}^n, W_1, W_c) - H(\hat{S}^n | W_1, W_c) \\
&\stackrel{(a)}{=} H(\hat{S}^n, W_1, W_c, A^n) - I(\hat{S}^n; S^n | W_1, W_c, A^n) \\
&\stackrel{(b)}{\approx} I(\hat{S}^n, W_1, W_c, A^n; Y^n) - I(\hat{S}^n; S^n | W_1, W_c, A^n) \\
&\stackrel{(c)}{=} I(\hat{S}^n, W_1, W_c, A^n; Y^n) - I(\hat{S}^n, W_1, W_c; S^n | A^n) \\
&\stackrel{(d)}{\leq} \sum_{i=1}^n \left[I(\hat{S}^n, W_1, W_c, Y^{i-1}, A^n; Y_i) - I(\hat{S}^n, W_1, W_c, S_{i+1}^n, A^{i-1}, A_{i+1}^n; S_i | A_i) \right] \\
&\stackrel{(e)}{=} \sum_{i=1}^n \left[I(\hat{S}^n, W_1, W_c, Y^{i-1}, S_{i+1}^n, A^n; Y_i) - I(\hat{S}^n, W_1, W_c, Y^{i-1}, S_{i+1}^n, A^n; S_i | A_i) \right] \\
&\stackrel{(f)}{=} \sum_{i=1}^n [I(U_i, A_i; Y_i) - I(U_i; S_i | A_i)], \tag{10}
\end{aligned}$$

where (a) follows since $\hat{S}^n$ is a deterministic function of (W_1, W_c, S^n) and since W_c determines A^n , (b) follows from Fano's inequality, (c) follows since $(W_1, W_c) \rightarrow A^n \rightarrow S^n$, (d) follows since $S_i \rightarrow A_i \rightarrow (S_{i+1}^n, A^{i-1}, A_{i+1}^n)$, (e) follows from the Csiszár sum lemma and (f) follows since $U_i = (\hat{S}^n, W_1, W_c, Y^{i-1}, S_{i+1}^n, A^n)$. Now by introducing a time-sharing RV Q , as in (9), we have

$$\begin{aligned}
n(R_1 + R_c) &\leq \sum_{i=1}^n [I(U_i, A_i; Y_i | Q = i) - I(U_i; S_i | A_i, Q = i)] \\
&= n [I(U_Q, A_Q; Y_Q | Q) - I(U_Q; S_Q | A_Q, Q)] \\
&= n [I(U_Q, A_Q; Y_Q | Q) - I(U_Q, Q; S_Q | A_Q)] \\
&\leq n [I(U_Q, Q, A_Q; Y_Q) - I(U_Q, Q; S_Q | A_Q)]. \tag{11}
\end{aligned}$$

For the bound on input costs, we proceed as follows:

$$\begin{aligned}
\mathbb{E}[\gamma_k(A, X)] &= \sum_{a,x} \mathbb{P}(A = a, X = x) \gamma_k(a, x) \\
&= \sum_{a,x} \mathbb{P}(A_Q = a, X_Q = x) \gamma_k(a, x) \\
&= \sum_{a,x} \sum_{i=1}^n \mathbb{P}(A_Q = a, X_Q = x | Q = i) \mathbb{P}(Q = i) \gamma_k(a, x) \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{a,x} \mathbb{P}(A_i = a, X_i = x) \gamma_k(a, x) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\gamma_k(A_i, X_i)] \leq (\Gamma_k + \epsilon), k = 1, 2, \tag{12}
\end{aligned}$$

where the last step follows from the fact that for any successful $(n, R_1, R_c, D, \Gamma, \epsilon)$ scheme, we have from (2)

$$n(\Gamma_k + \epsilon) \geq \sum_{i=1}^n \mathbb{E}[\gamma_k(A_i, X_i)], k = 1, 2. \tag{13}$$

We next consider the bound on the distortion. Let $\hat{S}_d^n$ be the receiver's reconstruction of the state sequence.

Define the event $\mathcal{B} := \{\hat{S}^n = \hat{S}_d^n\}$. Notice that $P(\mathcal{B}) \geq 1 - \epsilon, \forall \epsilon > 0$. Hence

$$\begin{aligned}
n(D + \epsilon) &\geq \sum_{i=1}^n \mathbb{E}[d(S_i, \hat{S}_{di})] \\
&= \mathbb{E}_{\mathcal{B}} \left[\sum_{i=1}^n \mathbb{E}[d(S_i, \hat{S}_{di})] \middle| \mathcal{B} \right] \\
&\geq P(\mathcal{B}) \sum_{i=1}^n \mathbb{E}[d(S_i, \hat{S}_i)] \\
&\stackrel{(a)}{\geq} (1 - \epsilon) \sum_{i=1}^n \mathbb{E}[d(S_i, \phi_i(U_i))] \\
&= n(1 - \epsilon) \mathbb{E}_Q[\mathbb{E}[d(S_Q, \phi_Q(U_Q)) | Q]] \\
&= n(1 - \epsilon) \mathbb{E}[d(S_Q, \phi_Q(U_Q))] = n(1 - \epsilon) \mathbb{E}[d(S, \phi_Q(U_Q))] \\
&\stackrel{(b)}{=} n(1 - \epsilon) \mathbb{E}[d(S, \phi(U_Q, Q))] \\
&\stackrel{(c)}{=} n(1 - \epsilon) \mathbb{E}[d(S, \phi(U))],
\end{aligned}$$

where (a) follows by taking $\phi_i(U_i) = \hat{S}_i$, (b) follows by defining $\phi : (Q, U_Q) \mapsto \phi_Q(U_Q)$ and (c) follows since $(U_Q, Q) = U$. Finally, the bound on $|\mathcal{U}|$ follows by an application of the support lemma [14, Appendix C].

IV. GAUSSIAN SETTING

The Gaussian action-dependent state channel [3] (GADSC) is given by

$$Y = X + A + W + Z, \quad (14)$$

with $W \sim \mathcal{N}(0, \sigma_W^2)$ and $Z \sim \mathcal{N}(0, \sigma_Z^2)$, $W \perp Z$, and A and X being constrained in average power to P_A and P_X respectively. The state itself is given by $S = A + W$, with A being independent of W . We have the following theorem.

Theorem 4. *For the GADSC with CR constraints, the capacity region is achieved by appropriate jointly Gaussian choices of $p(a)$, $p(u, x|a, s)$ in Theorem 2. Specifically, the capacity region $\mathcal{C}_{CR}^{gads}$ is given by*

$$\mathcal{C}_{CR}^{gads} = \bigcup \{(R_1, R_c) | R_1 \leq R_{priv}, R_1 + R_c \leq R_{sum}\}, \quad (15)$$

where the union is over $\rho_1, \rho_2 \in [-1, 1]$ satisfying $\rho_1^2 + \rho_2^2 \leq 1$ and $D \geq \frac{(1-\rho_2^2)P_X\sigma_W^2}{P_X + \alpha^2\sigma_W^2 + 2\alpha\rho_2\sqrt{P_X\sigma_W^2}}$, for the rate functions (R_{priv}, R_{sum}) as in (19), (20) and α as in (23).

Proof. We will prove this theorem using the single-letter expression in Theorem 2. While Theorem 2 was shown for the discrete memoryless case, it does hold for the Gaussian case as well. The achievability of Theorem 2 for the Gaussian case follows from an application of the discretization procedure [14, Sec. 3.4.1]. But note that the converse for the discrete memoryless case assumed finite $|\hat{\mathcal{S}}|$ (see (7), i.e. Fano's inequality for $\hat{S}^n$). However, without loss of generality, we can restrict attention to an exponential number of agreed reconstructions, i.e. $|\hat{\mathcal{S}}^n| = \mathcal{O}(2^{nc})$, where c is a constant, as explained next.

Remark 5. *It can be shown that if there exists a scheme without any cardinality bounds on the reproductions which achieves (R_1, R_c, D) , then for any $\delta > 0$, $(R_1, R_c, D + \delta)$ can be achieved using a scheme where the reproductions are confined to lie in an alphabet of size which scales as $2^{n \cdot c(\delta)}$. This can be proven using, for instance, a scalar quantizer whose average distortion is δ and whose rate scales with n as $2^{n \cdot c(\delta)}$. Thus, the converse of Theorem 2 also applies for the Gaussian case.*

$$R_{priv}(\rho_1, \rho_2) = \frac{1}{2} \log \left[\left(\frac{\sigma_Z^2 + (1 - \rho_1^2 - \rho_2^2)P_X}{\sigma_Z^2} \right) \right], \quad (19)$$

$$R_{sum}(\rho_1, \rho_2) = \frac{1}{2} \log \left[\left(\frac{\sigma_Z^2 + (1 - \rho_1^2 - \rho_2^2)P_X}{\sigma_Z^2} \right) \left(\frac{P_X + P_A + \sigma_Z^2 + \sigma_W^2 + 2\rho_1\sqrt{P_X P_A} + 2\rho_2\sqrt{P_X \sigma_W^2}}{(1 - \rho_1^2)P_X + \sigma_Z^2 + \sigma_W^2 + 2\rho_2\sqrt{P_X \sigma_W^2}} \right) \right]. \quad (20)$$

■

We now prove the rest of the converse starting with expressions (5) and (6) from Section II:

$$\begin{aligned} R_1 &\leq h(Y|A) - h(Y|A, U) - h(S|A) + h(S|A, U) \\ &= h(Y|A) - h(Y|A, U, S) - h(S|A) + h(S|Y, A, U) \\ &\leq h(Y|A) - h(Y|A, U, S, X) - h(S|A) + h(S|Y, A, U) \\ &= h(Y|A) - h(Z) - h(S|A) + h(S|Y, A, U) \\ &\leq \frac{1}{2} \log \left[\left(\frac{\sigma_{Y|A}^2}{\sigma_Z^2} \right) \left(\frac{\sigma_{S|Y, A, U}^2}{\sigma_W^2} \right) \right]. \end{aligned} \quad (16)$$

Similarly, for the sum rate

$$\begin{aligned} R_1 + R_c &\leq h(Y) - h(Y|U, A) - h(S|A) + h(S|A, U) \\ &= h(Y) - h(Y|A, U) - h(S|A) + h(S|A, U) \\ &= h(Y) - h(Y|A, U, S) - h(S|A) + h(S|Y, A, U) \\ &\leq h(Y) - h(Y|A, U, S, X) - h(S|A) + h(S|Y, A, U) \\ &\leq h(Y) - h(Z) - h(S|A) + h(S|Y, A, U) \\ &\leq \frac{1}{2} \log \left[\left(\frac{\sigma_Y^2}{\sigma_Z^2} \right) \left(\frac{\sigma_{S|Y, A, U}^2}{\sigma_W^2} \right) \right]. \end{aligned} \quad (17)$$

where the variance terms in (16) and (17) are defined as

$$\sigma_{S|Y, A, U}^2 = \min_{\alpha, \beta, \gamma} \mathbb{E}[S - \alpha Y - \beta A - \gamma U]^2. \quad (18)$$

Thus it can be seen that, via the differential entropy maximizing property of Gaussian random variables for a given variance, the optimal auxiliary U must be jointly Gaussian with (A, S, X, Y) . This completes the proof of converse. For the achievability, we choose $p(a)$, $p(u, x|a, s)$ as follows. The action input is chosen as $A \sim \mathcal{N}(0, P_A)$. We take the channel input to be

$$X = \rho_1 \sqrt{\frac{P_X}{P_A}} A + \rho_2 \sqrt{\frac{P_X}{\sigma_W^2}} W + G, \quad (21)$$

where ρ_1 and ρ_2 satisfy $\rho_1^2 + \rho_2^2 \leq 1$, and $G \sim \mathcal{N}(0, (1 - \rho_1^2 - \rho_2^2)P_X)$ is independent of (A, W, Z) . We choose the auxiliary random variable as follows:

$$U = \delta X + A + \alpha \delta W, \quad (22)$$

where $\delta = -1 / \left(\rho_1 \sqrt{\frac{P_X}{P_A}} \right)$ and the coefficient α is chosen as

$$\alpha = \frac{(1 - \rho_1^2 - \rho_2^2)P_X - \rho_2 \sqrt{\frac{P_X}{\sigma_W^2}} \sigma_Z^2}{(1 - \rho_1^2 - \rho_2^2)P_X + \sigma_Z^2}. \quad (23)$$

The common reconstruction constraint is as follows:

$$D \geq \mathbb{E}[S - \hat{S}]^2 = \text{Var}[S|U, A] = \frac{(1 - \rho_2^2)P_X\sigma_W^2}{P_X + \alpha^2\sigma_W^2 + 2\alpha\rho_2\sqrt{P_X\sigma_W^2}}.$$

Now on evaluating the terms in (16), (17) with the above jointly Gaussian choices, we arrive at the rate constraints in (19), (20).

Remark 6. For the GADSC with only a common message and CR constraints considered in [7], the capacity characterization can be obtained by setting $W_1 = \phi$ i.e. $R_1 = 0$ in Theorem 4, which is given by (24). Note that [7] only gave an achievable region, while its optimality was not established therein.

$$C(D) = \max_{(\rho_1, \rho_2)} R_{\text{sum}}(\rho_1, \rho_2). \quad (24)$$

$$\rho_1^2 + \rho_2^2 \leq 1, D \geq \frac{(1 - \rho_2^2)P_X\sigma_W^2}{P_X + \alpha^2\sigma_W^2 + 2\alpha\rho_2\sqrt{P_X\sigma_W^2}}$$

Remark 7. On further setting $D \geq \sigma_W^2$, i.e. the case of no CR constraints, the characterization simplifies to

$$C = \max_{\substack{(\rho_1, \rho_2) \\ \rho_1^2 + \rho_2^2 \leq 1}} R_{\text{sum}}(\rho_1, \rho_2), \quad (25)$$

which is a complete characterization of the Gaussian action dependent channel. This can also be inferred from the results in Zaidi et al. [12] [13, Theorem 4].

A. Action-Dependent Channels with Reversible inputs

We consider a GADSC with reversible inputs, i.e. with an additional constraint that the input to the channel must be reconstructed reliably at the receiver, in addition to the message. [9] derived the capacity of this model in the discrete case to be

$$C = \max I(A, X; Y) - I(X; S|A), \quad (26)$$

where the joint pmf is of the form $p(a)p(s|a)p(x|a, s)p(y|x, s)$ and the maximization is over all $p(a)$ and $p(x|a, s)$ such that $0 \leq I(X; Y|A) - I(X; S|A)$. In the following, we establish the capacity characterization in the Gaussian setting. The converse can be proved from the single letter characterization due to [9]:

$$\begin{aligned} 0 &\leq h(Y|A) - h(Y|A, X) - h(S|A) + h(S|A, X) \\ &= h(Y|A) - h(Y|A, X, S) - h(S|A) + h(S|Y, A, X) \\ &\leq h(Y|A) - h(Z) - h(W) + h(S|Y, A, X) \\ &\leq \frac{1}{2} \log \left[\left(\frac{\text{Var}[Y|A]}{\sigma_Z^2} \right) \left(\frac{\text{Var}[S|Y, A, X]}{\sigma_W^2} \right) \right], \end{aligned} \quad (27)$$

Similarly, we have

$$\begin{aligned} R &\leq h(Y) - h(Y|A, X) - h(S|A) + h(S|A, X) \\ &= h(Y) - h(Y|A, X) - h(S|A) + h(S|A, X) \\ &= h(Y) - h(Y|A, X, S) - h(S|A) + h(S|Y, A, X) \\ &\leq h(Y) - h(Z) - h(W) + h(S|Y, A, X) \\ &\leq \frac{1}{2} \log \left[\left(\frac{\text{Var}[Y]}{\sigma_Z^2} \right) \left(\frac{\text{Var}[S|Y, A, X]}{\sigma_W^2} \right) \right]. \end{aligned} \quad (28)$$

Thus it can be seen that via the max-entropy principle that (A, S, X, Y) must be jointly Gaussian. This completes the proof of converse. For the achievability, we now choose $A \sim \mathcal{N}(0, P_A)$ and $p(x|a, s)$ as in (21) and compute (27), (28). The final characterization evaluates to (29), (30).

$$C_{RI} = \max_{(\rho_1, \rho_2): \rho_1^2 + \rho_2^2 \leq 1} R_{RI}(\rho_1, \rho_2), \quad (29)$$

$$R_{RI}(\rho_1, \rho_2) = \frac{1}{2} \log \left[\left(\frac{1 - \rho_1^2 - \rho_2^2}{(1 - \rho_1^2 - \rho_2^2)\sigma_W^2 + (1 - \rho_1^2)\sigma_Z^2} \right) \left(P_X + P_A + \sigma_Z^2 + \sigma_W^2 + 2\rho_1 \sqrt{P_X P_A} + 2\rho_2 \sqrt{P_X \sigma_W^2} \right) \right]. \quad (30)$$

V. STRICTLY CAUSAL STATE WITH COMMON RANDOMNESS AT THE ACTION ENCODER

In [3], the question of whether strictly causal state feedback to the action encoder can increase capacity was left open. In other words, the problem can be stated as whether actions of the form $A_i(M, S^{i-1})$ help compared to $A^n(M)$. We additionally assume a more general encoding mechanism in which an unlimited amount of common randomness is allowed between the channel and action encoders, in the form of a real-valued RV Θ uniformly distributed on $[0, 1]$. We see that such state feedback does not help, and the capacity still remains

$$C = \max_{p(a), p(u|a,s), p(x|u,s)} [I(A, U; Y) - I(U; S|A)]. \quad (31)$$

In other words, the optimal strategy is the one which ignores the extra information S^{i-1} . The converse can be proved by a slightly different choice of auxiliary RV in [3], as shown next.

$$\begin{aligned} nR &= H(M) \stackrel{(a)}{=} H(M|\Theta) \\ &\stackrel{(b)}{\approx} I(M; Y^n | \Theta) \\ &= \sum_{i=1}^n I(M; Y_i | Y_{i+1}^n, \Theta) \leq \sum_{i=1}^n I(M, Y_{i+1}^n; Y_i | \Theta) \\ &= \sum_{i=1}^n [I(M, Y_{i+1}^n, S^{i-1}; Y_i | \Theta) - I(S^{i-1}; Y_i | M, Y_{i+1}^n, \Theta)] \\ &\stackrel{(c)}{\leq} \sum_{i=1}^n [I(M, Y_{i+1}^n, S^{i-1}, \Theta; Y_i) - I(S_i; Y_{i+1}^n | M, S^{i-1}, \Theta)] \\ &\stackrel{(d)}{=} \sum_{i=1}^n [I(M, Y_{i+1}^n, S^{i-1}, \Theta, A^i; Y_i) - I(S_i; Y_{i+1}^n | M, S^{i-1}, \Theta, A^i)] \\ &\stackrel{(e)}{=} \sum_{i=1}^n [I(M, Y_{i+1}^n, S^{i-1}, \Theta, A^{i-1}, A_i; Y_i) - I(M, Y_{i+1}^n, S^{i-1}, \Theta, A^{i-1}; S_i | A_i)] \\ &\stackrel{(f)}{=} \sum_{i=1}^n [I(U_i, A_i; Y_i) - I(U_i; S_i | A_i)] \\ &\leq n \cdot \max [I(U, A; Y) - I(U; S | A)], \end{aligned} \quad (32)$$

where (a) follows since $M \perp \Theta$, (b) follows from Fano's inequality, (c) follows from the Csiszár sum lemma, (d) follows since (M, S^{i-1}, Θ) determines A^i , (e) follows since $S_i \rightarrow A_i \rightarrow (M, S^{i-1}, \Theta, A^{i-1})$ and (f) follows with a choice of $U_i = (M, S^{i-1}, \Theta, Y_{i+1}^n, A^{i-1})$ that satisfies the Markov conditions $U_i \rightarrow (X_i, S_i) \rightarrow Y_i$ and $A_i \rightarrow (U_i, S_i) \rightarrow X_i$.

VI. CONCLUSION

We studied a generalization of the action-dependent channel with multiple messages and CR constraints, and derived complete characterizations for both discrete memoryless as well as Gaussian models. As a result, we obtained the capacity for a Gaussian action dependent model with only common message and CR. Furthermore, we proved that the capacity of the action-dependent model remains unchanged even if the actions are randomly drawn based on past channel states and the message.

APPENDIX A ACHIEVABILITY PROOF FOR THEOREM 2

Codebook Generation: Fix the p.m.f. $p(a)p(u|a, s)p(x|u, s)$. Randomly and independently generate 2^{nR_c} sequences $a^n(w_c)$, $w_c \in [1 : 2^{nR_c}]$ i.i.d. according to $\prod_{i=1}^n p_A(a_i)$. For each $a^n(w_c)$ sequence, randomly and conditionally independently generate $2^{n(R_1+R')}$ sequences $u^n(w_c, w_1, j)$ for $w_1 \in [1 : 2^{nR_1}]$ and $j \in [1 : 2^{nR'}]$, i.i.d. according to $\prod_{i=1}^n p_{U|A}(u_i|a_i(w_c))$. Let $\mathcal{B}_U(w_1)$ be the set of sequences within the bin indexed by $w_1 \in [1 : 2^{nR_1}]$ in the u -codebook.

Encoding: Fix $\epsilon' > 0$. Given $w_c \in [1 : 2^{nR_c}]$, the action encoder selects the corresponding sequence $a^n(w_c)$ in the A codebook. The state sequence S^n is generated in response to the action sequence via the channel $p(s|a)$. Given $w_1 \in [1 : 2^{nR_1}]$, the channel encoder picks the least index j such that $(a^n(w_c), u^n(w_c, w_1, j), s^n) \in \mathcal{T}_\epsilon^n(U, A, S)$. An error is declared if no such index is found. The channel encoder then draws x^n i.i.d. conditionally given (u^n, s^n) according to $\prod_{i=1}^n p(x_i|u_i, s_i)$, and sends it. The channel encoder also generates its state reconstruction $\hat{S}_i = \phi(u_i)$, $i \in [1 : n]$.

Decoding: Let $\epsilon > \epsilon'$. We use simultaneous decoding. The decoder declares that $(\hat{w}_c, \hat{w}_1)$ is sent if it is the unique message pair such that $(a^n(\hat{w}_c), u^n(\hat{w}_c, \hat{w}_1, \hat{j}), y^n) \in \mathcal{T}_\epsilon^n(U, A, Y)$ for some $\hat{j} \in \mathcal{B}_U(\hat{w}_1)$. An error is declared otherwise.

Error Analysis: Assume without loss of generality that the messages $W_1 = 1$ and $W_c = 1$ were sent, and the index of the chosen U^n sequence is J . The encoding error events are

$$\mathcal{E}_1 = \{(A^n(1), U^n(1, 1, j), S^n) \notin \mathcal{T}_\epsilon^n \forall j \in [1 : 2^{nR'}]\}. \quad (33)$$

The decoding error events at the receiver are as follows:

$$\begin{aligned} \mathcal{E}_2 &= \{(A^n(1), U^n(1, w_1, j), Y^n) \in \mathcal{T}_\epsilon^n \text{ for some } w_1 \neq 1, j \in \mathcal{B}_U(w_1)\}, \\ \mathcal{E}_3 &= \{(A^n(w_c), U^n(w_c, w_1, j), Y^n) \\ &\quad \in \mathcal{T}_\epsilon^n(U, A, Y) \text{ for some } w_c \neq 1, w_1 \neq 1, j \in \mathcal{B}_U(w_1)\}, \\ \mathcal{E}_4 &= \{(A^n(w_c), U^n(w_c, 1, J), Y^n) \in \mathcal{T}_\epsilon^n \text{ for some } w_c \neq 1\}. \end{aligned}$$

By the covering lemma [14], we have,

$$\mathbb{P}(\mathcal{E}_1) \xrightarrow{n \rightarrow \infty} 0 \text{ if } R' \geq I(U; S|A) + \delta(\epsilon'). \quad (34)$$

By the packing lemma [14], we can write

$$\begin{aligned} \mathbb{P}(\mathcal{E}_2) &\xrightarrow{n \rightarrow \infty} 0 \text{ if } R_1 + R' \leq I(U; Y|A) - \delta(\epsilon), \\ \mathbb{P}(\mathcal{E}_3), \mathbb{P}(\mathcal{E}_4) &\xrightarrow{n \rightarrow \infty} 0 \text{ if } R_c + R_1 + R' \leq I(U, A; Y) - \delta(\epsilon). \end{aligned}$$

Thus we arrive at the rate constraints:

$$R_1 \leq I(U; Y|A) - I(U; S|A) - \delta(\epsilon') - \delta(\epsilon), \quad (35)$$

$$R_1 + R_c \leq I(U, A; Y) - I(U; S|A) - \delta(\epsilon') - \delta(\epsilon). \quad (36)$$

Input Cost Analysis: From the error analysis, we have

$$\begin{aligned} \mathbb{P}(\mathcal{E}_1) &= \mathbb{P}((U^n(W_1, W_c, J), S^n, A^n(W_c)) \notin \mathcal{T}_\epsilon^n) \\ &= \mathbb{P}((U^n(W_1, W_c, J), S^n, X^n, A^n(W_c)) \notin \mathcal{T}_\epsilon^n) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

But if $(x^n, a^n) \in \mathcal{T}_{\epsilon'}^n$, then we have $\gamma(x^n, a^n) \leq \Gamma$ by the typical average lemma [14]. Hence we have

$$\begin{aligned} \mathbb{E}[\gamma(X^n, A^n)] &= \mathbb{P}(\mathcal{E}_1) \mathbb{E}[\gamma(X^n, A^n) | \mathcal{E}_1] + \mathbb{P}(\mathcal{E}_1^c) \mathbb{E}[\gamma(X^n, A^n) | \mathcal{E}_1^c] \\ &\leq \mathbb{P}(\mathcal{E}_1) \Gamma_{max} + \mathbb{P}(\mathcal{E}_1^c) \Gamma, \text{ where } \Gamma_{max} = \max_{a \in \mathcal{A}, x \in \mathcal{X}} \gamma(a, x) \\ &\implies \limsup_{n \rightarrow \infty} \mathbb{E}[\gamma(X^n, A^n)] \leq \Gamma, \text{ since } \mathbb{P}(\mathcal{E}_1^c) \xrightarrow{n \rightarrow \infty} 1. \end{aligned}$$

Distortion Analysis: The mentioned distortion can be obtained by making the estimate on a per-letter basis. Since $\phi(U)$ satisfies the distortion constraint, it follows from the random codebook construction that as $n \rightarrow \infty$, we have

$$\begin{aligned} \mathbb{E}[d(S^n, \phi(U^n))] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[d(S_i, \phi(U_i))] \\ &\xrightarrow{n \rightarrow \infty} \mathbb{E}[d(S, \phi(U))] \leq D. \end{aligned} \quad (37)$$

Also, since both the encoder as well as the decoder (assuming correct decoding) can generate $\hat{S}_i = \phi(U_i), i \in [1 : n]$, the common reconstruction constraint is met.

REFERENCES

- [1] C. E. Shannon, “Channels with side information at the transmitter,” *IBM journal of Research and Development*, vol. 2, no. 4, pp. 289–293, 1958.
- [2] S. Gelfand and M. Pinsker, “Coding for channels with random parameters,” *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [3] T. Weissman, “Capacity of channels with action-dependent states,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5396–5411, 2010.
- [4] H. Permuter and T. Weissman, “Source coding with a side information vending machine,” *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4530–4544, 2011.
- [5] H. Asnani, H. Permuter, and T. Weissman, “Probing capacity,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7317–7332, 2011.
- [6] C. Choudhuri and U. Mitra, “Action dependent strictly causal state communication,” in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 3058–3062.
- [7] B. Ahmadi and O. Simeone, “On channels with action-dependent states,” in *Information Theory Workshop (ITW), 2012 IEEE*. IEEE, 2012, pp. 167–171.
- [8] Y. Steinberg, “Coding and common reconstruction,” *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 4995–5010, 2009.
- [9] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, “Coding with action-dependent side information and additional reconstruction requirements,” *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6355–6367, 2015.
- [10] Y. Steinberg and T. Weissman, “The degraded broadcast channel with action-dependent states,” in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 596–600.
- [11] Y. Steinberg, “The degraded broadcast channel with non-causal action-dependent side information,” in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 2965–2969.
- [12] A. Zaidi, P. Piantanida, and S. Shamai, “Wyner-ziv type versus noisy network coding for a state-dependent mac,” in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 1682–1686.
- [13] —, “Capacity region of cooperative multiple-access channel with states,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6153–6174, 2013.
- [14] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.