



HAL
open science

Causal Vector-Valued Witsenhausen Counterexamples with Feedback

Mengyuan Zhao, Maël Le Treust, Tobias Oechtering

► **To cite this version:**

Mengyuan Zhao, Maël Le Treust, Tobias Oechtering. Causal Vector-Valued Witsenhausen Counterexamples with Feedback. ITW 2024 - IEEE Information Theory Workshop, Nov 2024, Shenzhen, China. pp.687-692, 10.1109/ITW61385.2024.10806960 . hal-04863320

HAL Id: hal-04863320

<https://hal.science/hal-04863320v1>

Submitted on 3 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Causal Vector-valued Witsenhausen Counterexamples with Feedback

Mengyuan Zhao

Division of Information Science
and Engineering
KTH Royal Institute of Technology
100 44 Stockholm, Sweden
Email: mzhao@kth.se

Maël Le Treust

Univ. Rennes, CNRS, Inria
IRISA UMR 6074
F-35000 Rennes, France
Email: mael.le-treust@cnrs.fr

Tobias J. Oechtering

Division of Information Science
and Engineering
KTH Royal Institute of Technology
100 44 Stockholm, Sweden
Email: oech@kth.se

Abstract—We study the continuous vector-valued Witsenhausen counterexample through the lens of empirical coordination coding. We characterize the region of achievable pairs of costs in three scenarios: (i) causal encoding and causal decoding, (ii) causal encoding and causal decoding with channel feedback, and (iii) causal encoding and noncausal decoding with channel feedback. In these vector-valued versions of the problem, the optimal coding schemes must rely on a time-sharing strategy, since the region of achievable pairs of costs might not be convex in the scalar version of the problem. We examine the role of the channel feedback when the encoder is causal and the decoder is either causal or non-causal, and we show that feedback improves the performance, only when the decoder is non-causal.

I. INTRODUCTION

In 1968, Witsenhausen introduced his famous counterexample demonstrating the suboptimality of affine strategies in the Linear Quadratic Gaussian (LQG) settings featuring non-classical information patterns [1]. Since then, it has become a cornerstone in the field of distributed decision-making [2]–[5] and information-theoretic control [6]–[13].

Due to the vector-valued formulation of the problem [14], many information-theoretic approaches have been adopted for analyzing this open problem [15]–[18]. The concept of empirical coordination, introduced in [19]–[21], plays an important role in establishing cooperative behavior among all agents in the network, providing single-letter solutions for problems such as characterizing the optimal cost, capacity region, and utility functions [22]–[24].

Feedback, as an intrinsic component of communication systems, offers the potential to enhance performance by enabling decision makers (DMs) to refine their actions based on previous outcomes. In many multi-terminal setups, feedback has proven beneficial for increasing the capacity region and assisting communication of the multiple-access channel [25, 26] as well as broadcast channel [27, 28]. Additionally, when considering the empirical coordination coding problem in a point-to-point scenario, [29, 30] showed that channel feedback could enable the DMs to directly coordinate their outputs with the system state, by simplifying the information constraint and by reducing the number of auxiliary random variables, and

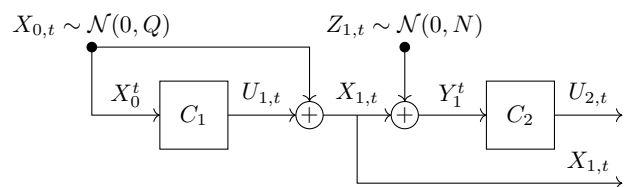


Figure 1. Vector-valued Witsenhausen counterexample with causal encoder and causal decoder.

therefore, enlarge the set of distributions. However, whether feedback facilitates coordination in the vector-valued Witsenhausen settings with causal controllers [31, 32] still remains an open question. Recently, it has been claimed in [33], that when constraint to Gaussian case, feedback does not contribute to performance improvements when the first DM is causal and the second DM is noncausal. In this article, we complete this study and we provide the necessary proof to the claimed information constraint.

In this paper, we aim to deepen the understanding of the problem by incorporating channel feedback for a comparative study, focusing on three novel setups: (i) causal encoding and causal decoding, (ii) causal encoding and causal decoding with channel feedback, and (iii) causal encoding and non-causal decoding with channel feedback. Our analysis shows that time-sharing is crucial in the first setup, in order to randomize between operating points and convexify the region of achievable pairs of costs. The minimum Gaussian cost identified in [33] can also be achieved by time-sharing between two affine policies. Time-sharing corresponds to randomized decision-making process for the original one-shot Witsenhausen counterexample which enables operating points to be achieved by randomization. The characterization of the region of achievable pairs of costs for the second model is derived using the genie-aided argument by focusing on its outer-bound. Comparing the results of the second and third models, we conclude that feedback can enlarge the region of achievable pairs of costs when the decoder is non-causal but not when it is causal.

II. SYSTEM MODEL

The vector-valued Witsenhausen counterexample setup consists of source states and channel noises that are drawn independently according to the i.i.d. Gaussian distributions $X_0^n \sim \mathcal{N}(0, Q\mathbb{I})$ and $Z_1^n \sim \mathcal{N}(0, N\mathbb{I})$, for some block length $n > 0$, and $Q, N > 0$. We denote by X_1 the memoryless interim state and Y_1 and output of the memoryless additive channel, generated by

$$X_1 = X_0 + U_1 \quad \text{with } X_0 \sim \mathcal{N}(0, Q), \quad (1)$$

$$Y_1 = X_1 + Z_1 = X_0 + U_1 + Z_1 \quad \text{with } Z_1 \sim \mathcal{N}(0, N). \quad (2)$$

We denote by $\mathcal{P}_{X_0} = \mathcal{N}(0, Q)$ the generative Gaussian probability distribution of the state, and by $\mathcal{P}_{X_1, Y_1 | X_0, U_1}$ the channel probability distribution according to (1) and (2).

We now present the three models and their corresponding results. The proofs of these results are given in Appendix.

A. Causal encoding and Causal decoding

Let's first consider the model with two causal DMs, as illustrated in Fig. 1.

Definition II.1. For $n \in \mathbb{N}$, a ‘‘control design’’ with causal encoder and causal decoder is a tuple of stochastic functions $c = (\{f_{U_1, t | X_0^t}^{(t)}\}_{t=1}^n, \{g_{U_2, t | Y_1^t}^{(t)}\}_{t=1}^n)$ defined by

$$f_{U_1, t | X_0^t}^{(t)} : \mathcal{X}_0^t \longrightarrow \mathcal{U}_{1, t}, \quad g_{U_2, t | Y_1^t}^{(t)} : \mathcal{Y}_1^t \longrightarrow \mathcal{U}_{2, t}, \quad (3)$$

which induces a distribution over sequences of symbols:

$$\prod_{t=1}^n \mathcal{P}_{X_0, t} \prod_{t=1}^n f_{U_1, t | X_0^t}^{(t)} \prod_{t=1}^n \mathcal{P}_{X_1, t, Y_1, t | X_0, t, U_1, t} \prod_{t=1}^n g_{U_2, t | Y_1^t}^{(t)}. \quad (4)$$

We denote by $\mathcal{C}_{ed}(n)$ the set of control designs with causal encoder and causal decoder.

Definition II.2. We define the two long-run cost functions $c_P(u_1^n) = \frac{1}{n} \sum_{t=1}^n (u_{1, t})^2$ and $c_S(x_1^n, u_2^n) = \frac{1}{n} \sum_{t=1}^n (x_{1, t} - u_{2, t})^2$. The pair of costs $(P, S) \in \mathbb{R}^2$ is said to be achievable if for all $\varepsilon > 0$, there exists $\bar{n} \in \mathbb{N}$ such that for all $n \geq \bar{n}$, there exists a control design $c \in \mathcal{C}_{ed}(n)$ such that

$$\mathbb{E} \left[|P - c_P(U_1^n)| + |S - c_S(X_1^n, U_2^n)| \right] \leq \varepsilon. \quad (5)$$

We denote by \mathcal{R}_{ed} the set of achievable pairs of costs for control designs in \mathcal{C}_{ed} .

Next, we characterize the costs region \mathcal{R}_{ed} .

Theorem II.3. The pair of Witsenhausen costs (P, S) is achievable if and only if there exists a joint distribution over the random variables $(X_0, T, U_1, X_1, Y_1, U_2)$ that decomposes according to

$$\mathcal{P}_{X_0} \mathcal{P}_T \mathcal{P}_{U_1 | X_0, T} \mathcal{P}_{X_1, Y_1 | X_0, U_1} \mathcal{P}_{U_2 | T, Y_1}, \quad (6)$$

such that

$$P = \mathbb{E} \left[U_1^2 \right], \quad S = \mathbb{E} \left[(X_1 - U_2)^2 \right], \quad (7)$$

where \mathcal{P}_{X_0} and $\mathcal{P}_{X_1, Y_1 | X_0, U_1}$ are the given Gaussian distributions, and T is the time-sharing auxiliary random variable with cardinality bound $|\mathcal{T}| \leq 2$.

Remark II.4. The probability distribution (6) satisfies

$$\begin{cases} X_0 \text{ is independent of } T, \\ (X_1, Y_1) \text{---} \ominus (X_0, U_1) \text{---} \ominus T, \\ U_2 \text{---} \ominus (T, Y_1) \text{---} \ominus (X_0, U_1, X_1). \end{cases} \quad (8)$$

The first property comes from the fact that the time-sharing random variable is independent of the source. The second Markov chain is due to the memoryless property of the channel. The last Markov chain comes from the causal decoding and the symbol-wise reconstruction.

Remark II.5. The cost region

$$\mathcal{R}_{ed} = \{(P, S) : P = \mathbb{E}[U_1^2], S = \mathbb{E}[(X_1 - U_2)^2], \text{ for } \mathcal{P} \text{ of the form of (6)}\} \quad (9)$$

characterized in Theorem II.3 is convex. This is because time-sharing synthesizes DMs to agree on the operating points. Therefore, we can construct a control scheme that achieves any point on the cord between two pairs of Witsenhausen costs.

Let's look at the following example delving into the effect of convexification using time-sharing:

Example. We consider a binary source X_0 with $\mathbb{P}(X_0 = 0) = \mathbb{P}(X_0 = 1) = \frac{1}{2}$, binary symmetric stochastic encoder $\sim \text{Bern}(\alpha)$ and decoder $\sim \text{Bern}(\beta)$ and a perfect channel with the set of two symbols $\mathcal{X}_0 = \mathcal{U}_1 = \mathcal{U}_2 = \{0, 1\}$, as represented in Fig. 2.

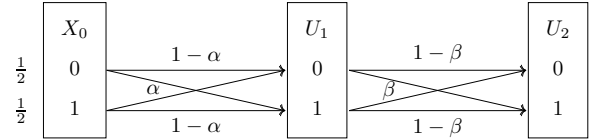


Figure 2. Binary information source and binary symmetric stochastic encoder and decoder.

The joint distribution of this system is presented in Fig. 3. The goal of the two DMs is to design their strategies through pairs (α, β) to achieve a desired empirical distribution.

	$U_2 = 0$	$U_2 = 1$		$U_2 = 0$	$U_2 = 1$
$U_1 = 0$	$\frac{1}{2}(1-\alpha)(1-\beta)$	$\frac{1}{2}(1-\alpha)\beta$	$U_1 = 0$	$\frac{1}{2}\alpha(1-\beta)$	$\frac{1}{2}\alpha\beta$
$U_1 = 1$	$\frac{1}{2}\alpha\beta$	$\frac{1}{2}\alpha(1-\beta)$	$U_1 = 1$	$\frac{1}{2}(1-\alpha)\beta$	$\frac{1}{2}(1-\alpha)(1-\beta)$
	$X_0 = 0$			$X_0 = 1$	

Figure 3. The joint distribution induced by a binary system depicted in Fig. 2

In this example, we initially don't apply the time-sharing technique. Fig. 4 shows an empirical distribution (third row) that is generated by directly combining the distributions given

by $(\alpha = 0, \beta = 0)$ (first row) and $(\alpha = 1, \beta = 1)$ (second row). However, one can easily show that this combined empirical distribution can not be achieved by any single choice of pair $(\alpha, \beta) \in [0, 1]^2$, thus it is achievable only through time-sharing.

		$U_2 = 0$	$U_2 = 1$			$U_2 = 0$	$U_2 = 1$
$U_1 = 0$	$U_1 = 1$	$\frac{1}{2}$	0		$U_1 = 0$	0	0
$U_1 = 1$	$U_1 = 1$	0	0		$U_1 = 1$	0	$\frac{1}{2}$
		$X_0 = 0$	$X_0 = 1$			$X_0 = 0$	$X_0 = 1$
		$U_2 = 0$	$U_2 = 1$			$U_2 = 0$	$U_2 = 1$
$U_1 = 0$	$U_1 = 1$	0	0		$U_1 = 0$	0	$\frac{1}{2}$
$U_1 = 1$	$U_1 = 1$	$\frac{1}{2}$	0		$U_1 = 1$	0	0
		$X_0 = 1$	$X_0 = 1$			$X_0 = 1$	$X_0 = 1$
		$U_2 = 0$	$U_2 = 1$			$U_2 = 0$	$U_2 = 1$
$U_1 = 0$	$U_1 = 1$	$\frac{1}{4}$	0		$U_1 = 0$	0	$\frac{1}{4}$
$U_1 = 1$	$U_1 = 1$	$\frac{1}{4}$	0		$U_1 = 1$	0	$\frac{1}{4}$
		$X_0 = 0$	$X_0 = 1$			$X_0 = 0$	$X_0 = 1$

Figure 4. First row: joint distribution given by $(\alpha = 0, \beta = 0)$, second row: joint distribution given by $(\alpha = 1, \beta = 1)$, third row: a convex combination of the above two cases, but is not achievable by any single pair of (α, β) .

B. Causal-causal with Channel Feedback

Now suppose the causal encoder has a delayed access to the sequence of channel output Y_1^{t-1} at each stage t to help its decision, see Fig. 5.

Definition II.6. For $n \in \mathbb{N}$, a “control design” with causal encoder and causal decoder with channel feedback is a tuple of stochastic functions $c = (\{f_{U_{1,t}|X_0^t, Y_1^{t-1}}^{(f,t)}\}_{t=1}^n, \{g_{U_{2,t}|Y_1^t}\}_{t=1}^n)$ defined by

$$f_{U_{1,t}|X_0^t, Y_1^{t-1}}^{(f,t)} : \mathcal{X}_0^t \times \mathcal{Y}_1^{t-1} \longrightarrow \mathcal{U}_{1,t}, \quad g_{U_{2,t}|Y_1^t}^{(t)} : \mathcal{Y}_1^t \longrightarrow \mathcal{U}_{2,t}, \quad (10)$$

which induces a distribution over sequences of symbols:

$$\prod_{t=1}^n \mathcal{P}_{X_{0,t}} \prod_{t=1}^n f_{U_{1,t}|X_0^t, Y_1^{t-1}}^{(f,t)} \prod_{t=1}^n \mathcal{P}_{X_{1,t}, Y_{1,t}|X_{0,t}, U_{1,t}} \prod_{t=1}^n g_{U_{2,t}|Y_1^t}^{(t)}, \quad (11)$$

where $Y_1^0 = \emptyset$. We denote by $\mathcal{C}_{ed,f}(n)$ the set of control designs with causal encoder and causal decoder with channel feedback.

We define the achievable pairs of cost (P, S) similarly as in Definition II.2 and we denote by $\mathcal{R}_{ed,f}$ the region of achievable pairs of costs.

Theorem II.7. $\mathcal{R}_{ed,f} = \mathcal{R}_{ed}$.

The proof of Theorem II.7 is stated in Appendix B, where a genie-aided method is used to prove the converse result.

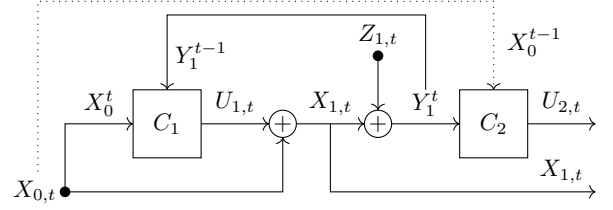


Figure 5. Witsenhausen counterexample for causal-encoding and causal-decoding with channel feedback Y_1^{t-1} to the encoder. The dotted line describes the source feed-forward X_0^{t-1} to the decoder, that is used for the genie-aided argument in the converse proof of Theorem II.7 in Appendix B.

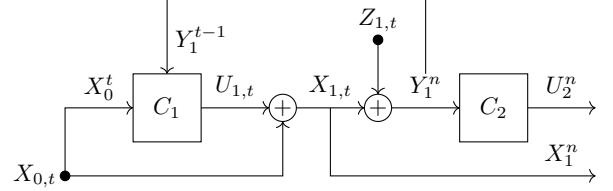


Figure 6. Witsenhausen counterexample for causal-encoding and noncausal-decoding with channel feedback Y_1^{t-1} to the encoder.

Adding channel feedback does not help for enlarging the region of achievable pairs of costs when both DMs are causal. This result is a consequence of the two causalities such that the two DMs only deal with new information: Even though at time $t \in [1 : n]$ the encoder has the opportunity to provide more insights for its previous decisions based on the past channel output Y_1^{t-1} , the causal decoder has already responded to its past actions and therefore does not allow further modifications for this time.

C. Causal-noncausal with Channel Feedback

In this section, we consider the model of causal-encoding and noncausal-decoding with channel feedback as illustrated in Fig. 6. The main result in this chapter provides the information constraint that completes the proof of Corollary III.6 in [33].

Definition II.8. For $n \in \mathbb{N}$, a “control design” with causal encoder and noncausal decoder with channel feedback is a tuple of stochastic functions $c = (\{f_{U_{1,t}|X_0^t, Y_1^{t-1}}^{(f,t)}\}_{t=1}^n, g_{U_2^n|Y_1^n})$ defined by

$$f_{U_{1,t}|X_0^t, Y_1^{t-1}}^{(f,t)} : \mathcal{X}_0^t \times \mathcal{Y}_1^{t-1} \longrightarrow \mathcal{U}_{1,t}, \quad g_{U_2^n|Y_1^n} : \mathcal{Y}_1^n \longrightarrow \mathcal{U}_2^n, \quad (12)$$

which induces a distribution over sequences of symbols:

$$\prod_{t=1}^n \mathcal{P}_{X_{0,t}} \prod_{t=1}^n f_{U_{1,t}|X_0^t, Y_1^{t-1}}^{(f,t)} \prod_{t=1}^n \mathcal{P}_{X_{1,t}, Y_{1,t}|X_{0,t}, U_{1,t}} g_{U_2^n|Y_1^n}, \quad (13)$$

where $Y_1^0 = \emptyset$. We denote by $\mathcal{C}_{e,f}(n)$ the set of control designs with causal encoder and noncausal decoder with channel feedback.

Again, we denote by $\mathcal{R}_{e,f}$ the region of achievable pairs of rates for control designs in $\mathcal{C}_{e,f}(n)$.

Theorem II.9. *The pair of Witsenhausen costs is achievable (P, S) if and only if there exists a joint distribution over the random variables $(X_0, W_1, U_1, X_1, Y_1, U_2)$ that decomposes according to*

$$\mathcal{P}_{X_0} \mathcal{P}_{W_1} \mathcal{P}_{U_1|X_0, W_1} \mathcal{P}_{X_1, Y_1|X_0, U_1} \mathcal{P}_{U_2|X_0, W_1, Y_1}, \quad (14)$$

such that

$$I(W_1; Y_1) - I(U_2; X_0|W_1, Y_1) \geq 0, \quad (15)$$

$$P = \mathbb{E}[U_1^2], \quad S = \mathbb{E}[(X_1 - U_2)^2], \quad (16)$$

where \mathcal{P}_{X_0} and $\mathcal{P}_{X_1, Y_1|X_0, U_1}$ are given Gaussian distributions, and W_1 is an auxiliary random variables.

Remark II.10. *Distribution (14) satisfies the follows*

$$\begin{cases} X_0 \text{ is independent of } W_1, \\ (X_1, Y_1) \text{ --- } (X_0, U_1) \text{ --- } W_1, \\ U_2 \text{ --- } (X_0, Y_1, W_1) \text{ --- } (U_1, X_1). \end{cases} \quad (17)$$

Remark II.11. *Compared to the single-letter result in [32, Theorem II.3], the presence of channel feedback enables the decoder to directly coordinate with the source state X_0 instead of its noisy representation (W_2 in the reference paper). Therefore, feedback enlarges the set of achievable pairs. A similar observation has been also pointed out in [29, 30, 34]*

Sketch of the proof for Theorem II.9. The converse proof for Theorem II.9 can be adapted from the converse proof of Theorem III.2 in [29], which dealt with a state-independent channel. We can modify the arguments from this reference to apply to the state-dependent channel in our setting, without affecting the outcome of the result.

Additionally, the coding scheme for the achievability proof for Theorem II.9 also extends the approach provided in [29] with changing the channel from state-independent to state-dependent. The cost analysis is directly derived from the arguments presented in [32]. \square

APPENDIX A: PROOF OF THEOREM II.3

Converse. For a pair (P, S) , assume that we have a control design $c \in \mathcal{C}_{ed}(n)$ of block length $n \in \mathbb{N}$ and small $\varepsilon > 0$ which induces a joint p.m.f. \mathcal{P}^n of the form (4) such that

$$|P - \mathbb{E}[c_P(U_1^n)]| + |S - \mathbb{E}[c_S(X_1^n, U_2^n)]| < \varepsilon. \quad (18)$$

This is implied by condition (5).

let Q be an independent time random variable uniformly distributed over $\{1, \dots, n\}$. Define auxiliary random variables $X_0 = X_{0,Q}, U_1 = U_{1,Q}, X_1 = X_{1,Q}, Y_1 = Y_{1,Q}, U_2 = U_{2,Q}$ with distribution

$$\mathbb{P}((X_0, U_1, X_1, Y_1, U_2) = (x_0, u_1, x_1, y_1, u_2)) \quad (19)$$

$$= \frac{1}{n} \sum_{q=1}^n \mathbb{P}((X_{0,q}, U_{1,q}, X_{1,q}, Y_{1,q}, U_{2,q}) = (x_0, u_1, x_1, y_1, u_2))$$

$$\forall (x_0, u_1, x_1, y_1, u_2) \in \mathcal{X}_0 \times \mathcal{U}_1 \times \mathcal{X}_1 \times \mathcal{Y}_1 \times \mathcal{U}_2.$$

Then, the expected long-run costs could be reformulated as

$$\mathbb{E}[c_P(U_1^n)] = \mathbb{E}\left[\frac{1}{n} \sum_{q=1}^n U_{1,q}^2\right] = \mathbb{E}[U_1^2], \quad (20)$$

$$\mathbb{E}[c_S(X_1^n, U_2^n)] = \mathbb{E}\left[\frac{1}{n} \sum_{q=1}^n (X_{1,q} - U_{2,q})^2\right] \quad (21)$$

$$= \mathbb{E}[(X_1 - U_2)^2]. \quad (22)$$

Therefore, given (18), we obtain that

$$|P - \mathbb{E}[U_1^2]| + |S - \mathbb{E}[(X_1 - U_2)^2]| < \varepsilon, \quad (23)$$

which is valid for all $\varepsilon \geq 0$. Hence, we have (7).

Now, we define the new auxiliary random variables $W_q = Y_1^{q-1}$ for $q = 1, \dots, n$ and $T = (W_Q, Q)$. These auxiliary random variables satisfy the following Markov chains

- $X_0 \perp\!\!\!\perp T$: This is because the source is i.i.d. generated, independent of the time stage Q as well as the past channel output sequence Y_1^{Q-1} due to causal encoding.
- $(X_1, Y_1) \text{ --- } (X_0, U_1) \text{ --- } T$: This comes from the discrete memoryless channel, the characterization of the output (X_1, Y_1) depends only on the input (X_0, U_1) .
- $U_2 \text{ --- } (T, Y_1) \text{ --- } (X_0, U_1, X_1)$: This is the consequence of the causal decoding that the reconstruction is fully characterized by sequence Y_1^Q up to stage $Q = 1, \dots, n$.

Therefore, the distribution of all the introduced auxiliary random variables decomposes as (6). \square

Achievability. Consider a joint distribution \mathcal{P} of the form of (6) with $\mathbb{E}[U_1^2] = P$ and $\mathbb{E}[(X_1 - U_2)^2] = S$. Fix a small $\varepsilon > 0$ and a blocklength $n \in \mathbb{N}$.

The encoder and decoder simply conduct symbol-by-symbol approaches based on the given distribution \mathcal{P} : Before the transmission, the encoder selects a typical sequence $T^n \in \mathcal{T}_\varepsilon(\mathcal{P}_T)$ and shares it to the decoder. Then, at each time $t \in \{1, \dots, n\}$, the encoder observes $X_{0,t}$, and outputs $U_{1,t} \sim \mathcal{P}_{U_1|X_0, T}(\cdot|X_{0,t}, T_t)$. The channel generates $X_{1,t}, Y_{1,t} \sim \mathcal{P}_{X_1, Y_1|X_0, U_1}(\cdot|X_{0,t}, U_{1,t})$. Then, the decoder draws $U_{2,t} \sim \mathcal{P}_{U_2|T, Y_1}(\cdot|T_t, Y_{1,t})$.

In such case, since each symbol is generated i.i.d. according to its distribution, from the law of large numbers (LLN), we have

$$c_P(U_1^n) = \frac{1}{n} \sum_{t=1}^n U_{1,t}^2 \xrightarrow{n \rightarrow \infty} P. \quad (24)$$

$$c_S(X_1^n, U_2^n) = \frac{1}{n} \sum_{t=1}^n (X_{1,t} - U_{2,t})^2 \xrightarrow{n \rightarrow \infty} S \quad (25)$$

Since convergence in probability implies convergence in \mathcal{L}^1 measure, given the existence of finite second moment, we have

$$\mathbb{E}[|c_P(U_1^n) - P|] < \frac{1}{2}\varepsilon, \quad (26)$$

$$\mathbb{E}[|c_S(X_1^n, U_2^n) - S|] < \frac{1}{2}\varepsilon. \quad (27)$$

for sufficiently large n . \square

Proof of the Cardinality Bound of T . Consider the set \mathcal{P} of all p.m.f.s on $\mathcal{U}_1 \times \mathcal{X}_1 \times \mathcal{U}_2$ satisfying the form of (6) and the following two continuous functions

$$f_1(\mathcal{P}, t) = \mathbb{E} \left[U_1^2 \mid T = t \right], \quad (28)$$

$$f_2(\mathcal{P}, t) = \mathbb{E} \left[(X_1 - U_2)^2 \mid T = t \right]. \quad (29)$$

From the support lemma in [35], [15, Appendix C], we can establish the time-sharing cardinality bound $|\mathcal{T}| \leq 2$ for the convex-hull operation. \square

APPENDIX B: PROOF OF THEOREM II.7

Converse. Consider a control design illustrated in Fig. 5 together with the dotted line. This induces a distribution of the following form

$$\prod_{t=1}^n \mathcal{P}_{X_0,t} \prod_{t=1}^n f_{U_1,t|X_0^t,Y_1^{t-1}}^{(f,t)} \prod_{t=1}^n \mathcal{P}_{X_1,t,Y_1,t|X_0,t,U_1,t} \prod_{t=1}^n g_{U_2,t|Y_1^t,X_0^{t-1}}^{(f,t)}. \quad (30)$$

This control design is clearly more powerful than that in Definition II.6 because the causal decoder also receives a past sequence of source information X_0^{t-1} at each instant t . Therefore, the achievable cost region induced by this superior system of adding the source feed-forward serves as an outerbound for the region without the source feed-forward. We denote this new region by $\mathcal{R}_{ed,f,f}$ and we have

$$\mathcal{R}_{ed,f,f} \supseteq \mathcal{R}_{ed,f} \supseteq \mathcal{R}_{ed} \quad (31)$$

Now, we show that $\mathcal{R}_{ed,f,f} = \mathcal{R}_{ed}$.

Similar to the converse proof of Theorem II.3 in Appendix A, we introduce a time-sharing random variable $Q \sim \text{Unif}[1, \dots, n]$ and define a sequence of new random variables $X_0 = X_{0,Q}, U_1 = U_{1,Q}, X_1 = X_{1,Q}, Y_1 = Y_{1,Q}, U_2 = U_{2,Q}$ with their joint distribution (19). In this way, the n -stage long-run costs for the control design of (30) also could be reformulated to (20) and (22).

Next, let $W_q = (X_0^{q-1}, Y_1^{q-1})$ for $q = 1, \dots, n$ and $T = (W_Q, Q)$ be new auxiliary random variables. Then, it holds that

- $X_0 \perp\!\!\!\perp T$. This follows from the i.i.d. source and causal encoding.
- $(X_1, Y_1) \ominus (X_0, U_1) \ominus T$: This comes from the discrete memoryless channel.
- $U_2 \ominus (T, Y_1) \ominus (X_0, U_1, X_1)$: This is the consequence of the causal decoding at the presence of Y_1^{Q-1}, X_0^{Q-1} at each time instant Q .

Given the above Markov chains, the single-letter joint distribution that characterizes a desired control scheme of adding a source feed-forward also decomposes as (6).

Therefore, we have shown that $\mathcal{R}_{ed,f,f} = \mathcal{R}_{ed}$. Namely, adding both the channel feedback AND the source feed-forward information does not enlarge the achievable Witsenhausen cost region of \mathcal{R}_{ed} . Since we also have (31), by the genie-aided argument, we conclude $\mathcal{R}_{ed,f} = \mathcal{R}_{ed}$. \square

REFERENCES

- [1] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, 1968.
- [2] R. Bansal and T. Başar, "Stochastic teams with nonclassical information revisited: When is an affine law optimal?," in *1986 American Control Conference*, pp. 45–50, IEEE, 1986.
- [3] E. I. Silva, G. C. Goodwin, and D. E. Quevedo, "Control system design subject to SNR constraints," *Automatica*, vol. 46, no. 2, pp. 428–436, 2010.
- [4] S. Yüksel and T. Başar, *Stochastic networked control systems: Stabilization and optimization under information constraints*. Springer Science & Business Media, 2013.
- [5] A. Gupta, S. Yüksel, T. Başar, and C. Langbort, "On the existence of optimal policies for a class of static and sequential dynamic teams," *SIAM Journal on Control and Optimization*, vol. 53, no. 3, pp. 1681–1712, 2015.
- [6] N. C. Martins and M. A. Dahleh, "Fundamental limitations of performance in the presence of finite capacity feedback," in *Proceedings of the 2005, American Control Conference, 2005.*, pp. 79–86, IEEE, 2005.
- [7] J. S. Freudenberg and R. H. Middleton, "Feedback control performance over a noisy communication channel," in *2008 IEEE Information Theory Workshop*, pp. 232–236, IEEE, 2008.
- [8] M. S. Derpich and J. Ostergaard, "Improved upper bounds to the causal quadratic rate-distortion function for gaussian stationary sources," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3131–3152, 2012.
- [9] A. Agrawal, F. Danard, B. Larrousse, and S. Lasaulce, "Implicit coordination in two-agent team problems with continuous action sets. application to the Witsenhausen cost function," in *European Control Conference, ECC 2015, Linz, Austria, July 15-17, 2015*, pp. 1854–1859, IEEE, 2015.
- [10] E. Akyol, C. Langbort, and T. Başar, "Information-theoretic approach to strategic communication as a hierarchical game," *Proceedings of the IEEE*, vol. 105, no. 2, pp. 205–218, 2017.
- [11] C. D. Charalambous, C. Kourtellis, and I. Tzortzis, "Hierarchical optimality of linear controllers-encoders-decoders operating at control-coding capacity of lqg control systems," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 3682–3687, IEEE, 2017.
- [12] M. Wiese, T. J. Oechtering, K. H. Johansson, P. Papadimitratos, H. Sandberg, and M. Skoglund, "Secure estimation and zero-error secrecy capacity," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1047–1062, 2018.
- [13] P. A. Stavrou, M. Skoglund, and T. Tanaka, "Sequential source coding for stochastic systems subject to finite rate constraints," *IEEE Transactions on Automatic Control*, vol. 67, no. 8, pp. 3822–3835, 2022.
- [14] P. Grover and A. Sahai, "Witsenhausen's counterexample as assisted interference suppression," *International Journal of Systems, Control and Communications*, vol. 2, pp. 197–237, 2010.
- [15] A. El Gamal and Y. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [16] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1850–1859, 2008.
- [17] O. Sumszyk and Y. Steinberg, "Information embedding with reversible stegotext," in *2009 IEEE International Symposium on Information Theory*, pp. 2728–2732, IEEE, 2009.
- [18] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Causal state communication," *IEEE Transactions on Information Theory*, vol. 59, no. 6, pp. 3709–3719, 2013.
- [19] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.
- [20] P. Cuff and C. Schieler, "Hybrid codes needed for coordination over the point-to-point channel," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 235–239, IEEE, 2011.
- [21] P. Cuff and L. Zhao, "Coordination using implicit communication," in *2011 IEEE Information Theory Workshop*, pp. 467–471, IEEE, 2011.
- [22] M. Le Treust, "Joint empirical coordination of source and channel," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5087–5114, 2017.

- [23] B. Laroousse, S. Lasaulce, and M. Wigger, "Coordination in state-dependent distributed networks: The two-agent case," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 979–983, IEEE, 2015.
- [24] B. Laroousse, S. Lasaulce, and M. R. Bloch, "Coordination in distributed networks via coded actions with application to power control," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3633–3654, 2018.
- [25] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback (corresp.)," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 100–102, 1975.
- [26] L. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 623–629, 1984.
- [27] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the gaussian broadcast channel with feedback (corresp.)," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 667–671, 1984.
- [28] G. Dueck, "The capacity region of the two-way channel can exceed the inner bound," *Information and Control*, vol. 40, no. 3, pp. 258–266, 1979.
- [29] M. Le Treust, "Empirical coordination with channel feedback and strictly causal or causal encoding," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015.
- [30] M. Le Treust and M. R. Bloch, "State leakage and coordination with causal state knowledge at the encoder," *IEEE Transactions on Information Theory*, vol. 67, no. 2, pp. 805–823, 2021.
- [31] M. Le Treust and T. J. Oechtering, "Power-estimation trade-off of vector-valued Witsenhausen counterexample with causal decoder," *IEEE Transactions on Information Theory*, vol. 70, no. 3, pp. 1588–1609, 2024.
- [32] M. Zhao, M. Le Treust, and T. J. Oechtering, "Coordination coding with causal encoder for vector-valued Witsenhausen counterexample," *arXiv preprint arXiv:2401.16926*, 2024.
- [33] M. Zhao, T. J. Oechtering, and M. Le Treust, "Optimal gaussian strategies for vector-valued witsenhausen counterexample with non-causal state estimator." <https://people.kth.se/~oech/CDC24.pdf>, 2024.
- [34] S. I. Bross and A. Lapidoth, "The rate-and-state capacity with feedback," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1893–1918, 2017.
- [35] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.