An Information-Theoretic Approach to Joint Sensing and Communication

Mehrasa Ahmadipour, Mari Kobayashi, Michèle Wigger, Giuseppe Caire

Abstract—**A communication setup is considered where a single transmitter wishes to convey messages to one or two receivers and simultaneously estimate the states of the receivers through the backscattered signals of the emitted waveform. The scenario at hand is motivated by joint radar and communication, which aims to co-design radar sensing and communication over shared spectrum and hardware. In this paper, we model the communication channel as a simple memoryless channel with independent and identically distributed (i.i.d.) time-varying state sequences and we model the backscattered signals by (strictly causal) generalized feedback. For single-receiver systems of this form, we fully characterize the capacity-distortion tradeoff, defined as the largest rate at which a message can reliably be conveyed to the receiver while simultaneously allowing the transmitter to sense the state sequence with a given allowed distortion. Our results show a tradeoff between the achievable rates and distortions, and that this tradeoff only stems from a common choice of the input distribution (the waveform) but not from other properties of the utilized codes. To better illustrate the capacity-distortion tradeoff, we propose a numerical method to compute the optimal inputs (waveforms) that achieve the desired tradeoff. For two-receiver systems with two states, we characterize the capacity-distortion tradeoff region of physically degraded broadcast channels (BC) as a rather straightforward extension of the single receiver case. Here, a tradeoff not only arises between sensing and communication performances but also between the various rates and the distortions of the different states. Similarly to the single-receiver case, the optimal co-design scheme exploits the generalized feedback signals only for sensing but not for improving communication performance. This is different for general two-receiver BCs, where optimal co-design schemes exploit generalized feedback also to improve capacity. However, as we show, also for BCs the optimal sensing performance only depends on the chosen input distribution (waveform) but not on the code construction used to accomplish the communication task. For general BCs, we provide inner and outer bounds on the capacity-distortion region, as well as a sufficient condition when this capacity-distortion region is equal to the product of the capacity region and the set of achievable distortions, in which case no tradeoff between sensing and communication occurs. A number of illustrative examples demonstrate that the optimal co-design schemes outperform conventional schemes that split the resources between sensing and communication, both for single-receiver and BC systems.**

Part of this material was presented at IEEE Int. Symp. Inf. Theory (ISIT) 2018 [1] and at IEEE Inf. Theory Workshop (ITW) 2020 [2].

M. Ahmadipour and M. Wigger are with LTCI, Telecom Paris, IP Paris, F-91120 Palaiseau, France, {mehrasa.ahmadipour,michele.wigger}@telecom-paris.fr.

M. Kobayashi is with Technical University of Munich, Munich, Germany, mari.kobayashi@tum.de.

G. Caire is with Technical University of Berlin, Germany, caire@tu-berlin.de.

The works of M. Ahmadipour and M. Wigger were supported by the European Research Council (ERC) under the European Union's Horizon 2020 programme, grant agreement number 715111. The works of M. Kobayashi and G. Caire were supported by the DFG, Grant agreement numbers KR 3517/11-1 and CA 1340/11-1, respectively.

I. Introduction

Future generation wireless networks are expected to support several autonomous and intelligent applications that strongly rely on accurate sensing and localization techniques [3]. An example are intelligent transportation systems where vehicles interact in a cooperative radar sensor network with the goal to provide unique safety features and intelligent traffic routing. The key enabler of such applications is the ability to sense the dynamically changing environment continuously, hereafter called the *state*, and to react accordingly by exchanging information. The standard assumption of such a joint radar sensing and communication system is a transmitter equipped with a co-located radar receiver that wishes to convey a message to a (already detected) receiver and simultaneously estimate the state parameters of that receiver.

A common but naive approach to address sensing and communication is to separate the two tasks in independent systems and accordingly split the available resources such as bandwidth and power between the two systems. In our information-theoretic model that we present shortly, such a system corresponds to time-sharing between communication and sensing; we shall call this *basic time-sharing (TS)*. The high cost of spectrum and hardware however encourages integrating the sensing and communications tasks via a single waveform and a single hardware platform (see, e.g., [4], [5] and references therein). First attempts towards such integrated systems use a standard communication system and exploit the backscattered signal from this waveform for sensing purposes, where the employed transmit waveform is either optimized for sensing or for communication; we shall call these *the sensing and communication modes of improved TS.*

The scenario at hand has been extensively studied in the literature (see e.g. [6], [7] and references therein). In particular, several *joint sensing and communication* schemes, or co-design schemes, have been proposed to optimize performance metrics capturing some tension between two performances [8]–[12]. Despite of these works providing system guidelines or proposing waveforms suitable to some specific scenarios, none has addressed the fundamental performance limits above which a joint sensing and communication system cannot operate irrespectively of computational complexities, choices of state parameters, or further assumptions. This observation inspires us to study the fundamental limit of joint sensing and communication from an information-theoretic perspective.

Our work is the first information-theoretic work on joint sensing and communication. We emphasize the difference to the information-theoretic works in [13]–[17] where sensing (state-estimation) is performed at the receiver and not at the transmitter, which models different real-world applications. In [13], [15]–[17], the transmitter even knows the state a priori.

In this paper, we build on a simple single-transmitter communication model with a discrete memoryless channel and independent and identically distributed (i.i.d.) state-sequences. The transmitter observes strictly causal *generalized feedback signals*, used for state sensing, while each receiver is assumed to perfectly know its corresponding channel state. The generalized feedback model captures two underlying assumptions used in radar signal processing. On the one hand, it captures the inherently passive nature of the backscattered signal observed at the transmitter, which cannot be controlled but is determined by its surrounding environment. On the other hand, it models the fact that the backscattered signal depends on the waveform employed by the transmitter. It is thus clear, that the employed waveform affects both the communication and sensing performances of the system and should be designed in a synergistic manner. Our goal is to characterize the fundamental tradeoff between the communication and sensing performance of such systems and the improvements an optimally designed scheme achieves over the separation scheme (i.e., the described basic TS) and over integrated systems that either prioritize sensing or communication (i.e., above described improved TS). To this purpose, we consider the *capacity-distortion tradeoff* as a performance measure since it suitably balances between two ultimate objectives: maximizing communication rate and minimizing state estimation error or *distortion*. The presented model was introduced in our conference publications [1], [2] and was also extended to the two-user multiple-access channel in [18], [19].

In this work we consider the single-transmitter single-receiver point-to-point (P2P) channel and the single-transmitter two-receiver BC. For the P2P channel we exactly characterize the capacity-distortion-cost tradeoff, which allows to quantify the merit of an optimal co-design scheme over the described basic and improved TS schemes. Not surprisingly, our results show that without loss in optimality the communication scheme can ignore the generalized feedback signals, which are only used for state sensing, and communication and sensing performances only depend on each other through the choice of the common waveform. Our results further show that in most situations a tradeoff between the simultaneously achievable sensing and communication performances arises. Based on our results we further identify “matched” situations where the same waveform simultaneously achieves capacity and minimum distortion. A Blahut-Arimoto type algorithm is presented that evaluates the capacity-distortion-cost tradeoff numerically.

While feedback does not increase capacity of memoryless P2P channels, it can significantly increase capacity of memoryless BCs [20]–[22] because it enables the transmitter to send some common information that is useful to both receivers at the same time (see e.g., [23, Section 17]). In our joint sensing and communication-over-BC setup, the generalized feedback thus improves both sensing and communication performances. Nevertheless, like in the P2P setup, the two performances only depend on each other through the common choice of the waveform. In other words, we show that the optimal state-sensing is independent of the employed BC-feedback-code and only depends on the chosen waveform but not on other details of the code construction. This allows to base joint coding and sensing systems on known BC-feedback code constructions such as [20]–[22]. Based on the scheme in [20], we provide a general inner bound on the capacity-distortion region for general memoryless BCs with generalized feedback. We also provide a general outer bound by extending a known converse technique that reveals the outputs at one of the receivers to the other receiver. Inner and outer bounds coincide only in special cases. Completely characterizing the capacity-distortion tradeoff region for a general memoryless state-dependent BC seems extremely challenging since even the capacity region (without sensing) is unknown both in the case without and with feedback (see e.g., [20]–[22], [24], [25]). Instead, we characterize the capacity-distortion region for the special case of physically degraded BCs. Analogously to the single-user case, feedback does not enlarge capacity for physically degraded BCs and is useful only for sensing but not for communication. Through various numerical examples we illustrate the merit of optimal co-design schemes the basic and improved TS for physically degraded and general BCs.

A. Contributions

The paper provides the following technical contributions:

1) It characterizes the capacity-distortion-cost tradeoff of state-dependent memoryless channels in Theorem 1 and states the optimal estimator (a deterministic symbol-by-symbol estimator) in Lemma 1. A modified Blahut-Arimoto algorithm [26], [27] is proposed to calculate the tradeoff region numerically. To this end, the optimality of an alternating optimization approach is proved in Theorem 4.
2) As a rather straightforward extension of Theorem 1, we characterize the capacity-distortion tradeoff region of physically degraded state-dependent memoryless broadcast channels in Theorem 2.
3) For general state-dependent BCs, we provide an outer bound on the capacity-distortion region in Theorem 3 and an inner bound in Proposition 1. The inner bound is based on [20] and can be achieved using a block-Markov strategy that combines Marton coding with a lossy version of Gray-Wyner coding with side-information.
4) Corollary 1 (for single-user channels) and Proposition 2 (for broadcast channels) identify sufficient conditions for channels where no capacity-distortion tradeoff arises.
5) Many illustrative examples are provided to demonstrate the benefits of the optimal co-design scheme compared to the aforementioned baseline schemes. These include a binary channel with a multiplicative Bernoulli state in Corollary 2, a real Gaussian channel, a binary BC with multiplicative Bernoulli states in Corollaries 3 and 4, as well as the state-dependent Dueck BC in Corollaries 7 and 8.

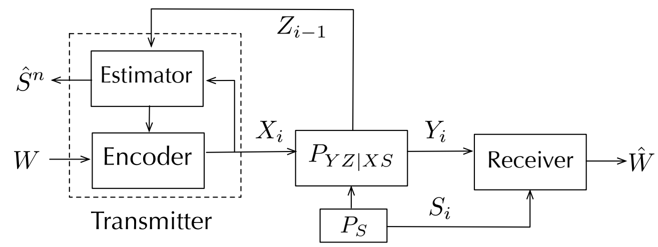

Fig. 1. Joint sensing and communication model.

B. Organization

The rest of this paper is organized as follows. The following Section II formulates the joint sensing and communication problem in a single-receiver channel and provides the corresponding capacity-distortion-cost tradeoff. Section III extends the obtained results to two-user broadcast channels. Finally, Section IV concludes the paper.

C. Notation

We use calligraphic letters to denote sets, e.g., $\mathcal{X}$. The sets of real and nonnegative real numbers, however, are denoted by $\mathbb{R}$ and $\mathbb{R}_0^+$. Random variables are denoted by uppercase letters, e.g., X, and their realizations by lowercase letters, e.g., x. For vectors, we use boldface notation, i.e., lower case boldface letters such as $\boldsymbol{x}$ for deterministic vectors. We use $[1:X]$ to denote the set $\{1,\cdots,X\}$. We use X^n for the tuple of random variables $(X_1,\cdots,X_n)$. We abbreviate *independent and identically distributed* as *i.i.d.* and *probability mass function* as *pmf*. Logarithms are taken with respect to base 2. We use $\perp$ to indicate independence between random variables.

II. A Single Receiver

A. System Model

Consider the point-to-point communication scenario depicted in Fig. 1, where a transmitter wishes to communicate a message to a receiver over a memoryless state-dependent channel and simultaneously estimate the state from generalized feedback. In order to formulate the joint sensing and communication problem, we consider a state-dependent memoryless channel such that the channel output at the receiver Y_i and the feedback signal Z_i at a given time i are generated according to its stationary channel law $P_{YZ|XS}(\cdot,\cdot|x_i,s_i)$ given the time-i channel input $X_i=x_i$ and state realization $S_i=s_i$, irrespective of the past inputs, outputs and state signals. Except for some Gaussian examples, we assume that the channel states S_i, inputs X_i, outputs Y_i, and feedback signals Z_i take value in finite sets $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. The state sequence $\{S_i\}_{i\geq 1}$ is assumed i.i.d. according to a given state distribution $P_S(\cdot)$ and perfectly known to the receiver.

A $(2^{n\mathsf{R}}, n)$ code for the state-dependent memoryless channel (SDMC) consists of

1) a discrete message set $\mathcal{W}$ of size $|\mathcal{W}| \geq 2^{n\mathsf{R}}$;
2) a sequence of encoding functions $\phi_i \colon \mathcal{W}\times\mathcal{Z}^{i-1} \to \mathcal{X}$, for $i = 1, 2, \ldots, n$;
3) a decoding function $g\colon \mathcal{S}^n \times \mathcal{Y}^n \to \mathcal{W}$;
4) a state estimator $h\colon \mathcal{X}^n \times \mathcal{Z}^n \to \hat{\mathcal{S}}^n$, where $\hat{\mathcal{S}}$ denotes a given finite reconstruction alphabet.

For a given code, the random message W is uniformly distributed over the message set $\mathcal{W}$ and the inputs are obtained as $X_i = \phi_i(W, Z^{i-1})$, for $i = 1, \ldots, n$. The corresponding channel outputs Y_i and Z_i at time i are obtained from the state S_i and the input X_i according to the SDMC transition law $P_{YZ|SX}$. Let $\hat{S}^n := (\hat{S}_1, \cdots, \hat{S}_n) = h(X^n, Z^n)$ denote the state estimate at the transmitter and $\hat{W} = g(S^n, Y^n)$ the decoded message at the receiver.

The quality of the state estimates is measured by the expected average per-block distortion

$$\Delta^{(n)} := \mathbb{E}[d(S^n, \hat{S}^n)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(S_i, \hat{S}_i)] \tag{1}$$

where $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}_0^+$ is a given bounded *distortion function*:

$$\max_{(s,\hat{s})\in\mathcal{S}\times\hat{\mathcal{S}}} d(s, \hat{s}) < \infty. \tag{2}$$

In practical communication systems, we typically impose an expected cost constraint on the channel inputs such as an average or peak power constraint. These cost constraints can often be expressed as

$$\mathbb{E}[b(X^n)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[b(X_i)] \tag{3}$$

for some given cost functions $b\colon \mathcal{X} \mapsto \mathbb{R}_0^+$.

Definition 1. *A rate-distortion-cost tuple* $(\mathsf{R}, \mathsf{D}, \mathsf{B})$ *is said achievable if there exists a sequence (in* n*) of* $(2^{n\mathsf{R}}, n)$ *codes that simultaneously satisfy*

$$\lim_{n\to\infty} P_e^{(n)} = 0, \tag{4a}$$

$$\overline{\lim_{n\to\infty}}\, \Delta^{(n)} \leq \mathsf{D}, \tag{4b}$$

$$\overline{\lim_{n\to\infty}}\, \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[b(X_i)] \leq \mathsf{B} \tag{4c}$$

for $P_e^{(n)} := \Pr\left(\hat{W} \neq W\right)$.

The capacity-distortion-cost tradeoff $\mathsf{C}(\mathsf{D}, \mathsf{B})$ *is the largest rate* R *such that the rate-distortion-cost tuple* $(\mathsf{R}, \mathsf{D}, \mathsf{B})$ *is achievable.*

The main result of this section is an exact characterization of $\mathsf{C}(\mathsf{D}, \mathsf{B})$. We begin by describing the optimal estimator h, which is independent of the choice of encoding and decoding functions, and operates on a symbol-by-symbol basis, i.e., it computes estimate $\hat{S}_i$ only in function of X_i and Z_i but not of the other inputs and feedback signals.

Lemma 1. *Define the function*

$$\hat{s}^*(x, z) := \arg\min_{s'\in\hat{\mathcal{S}}} \sum_{s\in\mathcal{S}} P_{S|XZ}(s|x, z) d(s, s'), \tag{5}$$

where ties can be broken arbitrarily and

$$P_{S|XZ}(s|x, z) = \frac{P_S(s)P_{Z|SX}(z|s, x)}{\sum_{\tilde{s}\in\mathcal{S}} P_S(\tilde{s})P_{Z|SX}(z|\tilde{s}, x)}. \tag{6}$$

Irrespective of the choice of encoding and decoding functions, distortion $\Delta^{(n)}$ *in* (4b) *is minimized by the estimator*

$$h^*(x^n, z^n) := (\hat{s}^*(x_1, z_1), \hat{s}^*(x_2, z_2), \ldots, \hat{s}^*(x_n, z_n)). \tag{7}$$

Notice that the function $\hat{s}(\cdot, \cdot)$ *only depends on the SDMC channel law* $P_{YZ|SX}$ *and the state distribution* P_S*.*

Proof: See Appendix A. ■

The optimal state estimator is thus a symbolwise estimator directly applied to the sequences observed at the transmitter. As we shall see later in this article, this optimality of the symbolwise estimator extends also to the broadcast scenario.

Lemma 1 implies that we can focus without loss in optimality on a symbol-by-symbol deterministic estimator. Based on (5), we define the estimation cost $c(x)$ of the optimal estimator as

$$c(x) := \mathsf{E}\left[d(S, \hat{s}^*(X, Z))|X = x\right]. \tag{8}$$

We are ready to present the capacity-distortion-cost tradeoff.

B. Capacity-Distortion-Cost Tradeoff

In order to characterize some useful properties of the capacity-distortion-cost function, we define the following sets:

$$\mathcal{P}_{\mathsf{B}} = \left\{ P_X \;\middle|\; \sum_{x\in\mathcal{X}} P_X(x)b(x) \leq \mathsf{B} \right\}, \tag{9a}$$

$$\mathcal{P}_{\mathsf{D}} = \left\{ P_X \;\middle|\; \sum_{x\in\mathcal{X}} P_X(x)c(x) \leq \mathsf{D} \right\}. \tag{9b}$$

Then, the minimum distortion for a given cost B is given by

$$\mathsf{D}_{\min}(\mathsf{B}) := \min_{P_X \in \mathcal{P}_{\mathsf{B}}} \sum_{x\in\mathcal{X}} P_X(x)c(x). \tag{10}$$

Definition 2. *Define the* information-theoretic tradeoff function $\mathsf{C}_{\mathsf{inf}} : [\mathsf{D}_{\min}(\mathsf{B}), \infty) \times [0,\infty) \to \mathbb{R}_0^+$ *as*

$$\mathsf{C}_{\mathsf{inf}}(\mathsf{D}, \mathsf{B}) := \max_{P_X \in \mathcal{P}_{\mathsf{D}} \cap \mathcal{P}_{\mathsf{B}}} I(X;Y \mid S) \tag{11}$$

where $(X,S,Y,Z) \sim P_X P_S P_{YZ|SX}$ and the maximum is over all P_X satisfying both the distortion and cost constraints (9b) *and* (9a).

Lemma 2. *Given a SDMC $P_{YZ|SX}$ with state-distribution P_S, the capacity-distortion-cost tradeoff function $\mathsf{C}_{\mathsf{inf}}(\mathsf{D},\mathsf{B})$ has the following properties.*

i) $\mathsf{C}_{\mathsf{inf}}(\mathsf{D},\mathsf{B})$ *is non-decreasing and concave in $\mathsf{D} \geq \mathsf{D}_{\min}(\mathsf{B})$ and $\mathsf{B} \geq 0$.*

ii) $\mathsf{C}_{\mathsf{inf}}(\mathsf{D},\mathsf{B})$ *saturates at the channel capacity:*

$$\mathsf{C}_{\mathsf{inf}}(\mathsf{D},\mathsf{B}) = \mathsf{C}_{\mathsf{NoEst}}(\mathsf{B}), \quad \forall \mathsf{D} \geq \mathsf{D}_{\max}(\mathsf{B}), \tag{12}$$

where $\mathsf{C}_{\mathsf{NoEst}}(\mathsf{B}) := \max_{P_X \in \mathcal{P}_{\mathsf{B}}} I(X;Y|S)$ denotes the classical channel capacity of the SDMC for a given cost B, and $\mathsf{D}_{\max}(\mathsf{B})$ denotes the corresponding distortion

$$\mathsf{D}_{\max}(\mathsf{B}) := \sum_{x\in\mathcal{X}} P_{X_{\max}}(x)c(x). \tag{13}$$

for $P_{X_{\max}} := \operatorname{argmax}_{P_X \in \mathcal{P}_{\mathsf{B}}} I(X;Y|S)$.

Proof: The proof is a straightforward extension of [14, Corollary 1] to the case of two cost functions and the state dependent channel. The nondecreasing property follows immediately from the definition in (11) because we have $\mathcal{P}_{\mathsf{D}_1} \subseteq \mathcal{P}_{\mathsf{D}_2}$ and $\mathcal{P}_{\mathsf{B}_1} \subseteq \mathcal{P}_{\mathsf{B}_2}$ for any $\mathsf{D}_1 \leq \mathsf{D}_2$ and $\mathsf{B}_1 \leq \mathsf{B}_2$.

In order to verify the concavity of $\mathsf{C}_{\mathsf{inf}}(\mathsf{D},\mathsf{B})$ with respect to (D,B), we consider time-sharing between two input distributions, denoted by $P_X^{(1)}$ and $P_X^{(2)}$, that achieve $\mathsf{C}_{\mathsf{inf}}(\mathsf{D}_1,\mathsf{B}_1)$ and $\mathsf{C}_{\mathsf{inf}}(\mathsf{D}_2,\mathsf{B}_2)$, respectively. To make the dependency of the mutual information with respect to the input distribution more explicit, we adapt the following notation: for any pmf P_X over the input alphabet $\mathcal{X}$, let $\mathcal{I}(P_X, P_{Y|XS} \mid P_S) := I(X;Y \mid S)$ for $(S,X,Y) \sim P_S P_X P_{Y|XS}$.

For any $\theta \in (0,1)$, we have:

$$\begin{aligned}
&\theta \mathsf{C}_{\mathsf{inf}}(\mathsf{D}_1,\mathsf{B}_1) + (1-\theta)\mathsf{C}_{\mathsf{inf}}(\mathsf{D}_2,\mathsf{B}_2) \\
&\overset{(a)}{=} \theta \mathcal{I}\left(P_X^{(1)}, P_{Y|XS} \,\middle|\, P_S\right) \\
&\qquad\qquad + (1-\theta)\mathcal{I}\left(P_X^{(2)}, P_{Y|XS} \,\middle|\, P_S\right) \\
&\overset{(b)}{\leq} \mathcal{I}\left(\theta P_X^{(1)} + (1-\theta)P_X^{(2)}, P_{Y|XS} \,\middle|\, P_S\right) \\
&\overset{(c)}{\leq} \mathsf{C}_{\mathsf{inf}}\left(\theta \mathsf{D}_1 + (1-\theta)\mathsf{D}_2, \theta \mathsf{B}_1 + (1-\theta)\mathsf{B}_2\right).
\end{aligned} \tag{14}$$

where (a) follows by definition, (b) follows from the concavity of the mutual information functional with respect to the input distribution, (c) follows by the linearity of the constraints and because for any $k=1,2$ the pmf $P_X^{(k)}$ has expected cost no larger than B_k and expected distortion no larger than D_k. This establishes the concavity of $\mathsf{C}_{\mathsf{inf}}(\mathsf{D},\mathsf{B})$. ■

We now state the main result of this section.

Theorem 1. *The capacity-distortion-cost tradeoff of a SDMC $P_{YZ|SX}$ with state-distribution P_S is:*

$$\mathsf{C}(\mathsf{D},\mathsf{B}) = \mathsf{C}_{\mathsf{inf}}(\mathsf{D},\mathsf{B}), \quad \mathsf{D} \geq \mathsf{D}_{\min}(\mathsf{B}), \;\; \mathsf{B} \geq 0. \tag{15}$$

Proof: See Appendix B. ■

The proof of Theorem 1 is similar to the proof of the classic capacity-cost function [28], except that one also has to account for the sensing performance. Both in the converse proof and the achievability proof, this can be accomplished by evaluating the performance of the optimal (per-symbol) estimator $\hat{s}^*(\cdot,\cdot)$ in Lemma 1. In particular, a standard random coding argument can be used to prove achievability of Theorem 1.

On a different note, capacity of a memoryless channel is known to be achieved with i.i.d. inputs. Also because of the memoryless nature of the optimal estimator $h(\cdot,\cdot)$ in Lemma 1, this observation extends to our joint sensing and communication setup.

Appendix C presents a Blahut-Arimoto type algorithm that can be used to solve the optimization problem (11), which characterizes the capacity-distortion-cost tradeoff $\mathsf{C}_{\mathsf{inf}}(\mathsf{D},\mathsf{B})$. It is used to evaluate the capacity-distortion-cost tradeoff for the Gaussian example in Subsection II-C3 ahead.

Combining Lemma 2 and Theorem 1, we can conclude that the rate-distortion tradeoff function $\mathsf{C}(\mathsf{D},\mathsf{B})$ is non-decreasing and concave in $\mathsf{D} \geq \mathsf{D}_{\min}$ and $\mathsf{B} \geq 0$, and for any $\mathsf{B} \geq 0$ it saturates at the channel capacity $\mathsf{C}_{\mathsf{NoEst}}(\mathsf{B})$. For many channels, given $\mathsf{B} \geq 0$, the tradeoff $\mathsf{C}(\mathsf{D},\mathsf{B})$ is strictly increasing in D until it reaches $\mathsf{C}_{\mathsf{NoEst}}(\mathsf{B})$. However, for SDMBCs and costs $\mathsf{B} \geq 0$ where the capacity-achieving input distribution $P_{X_{\max}} := \operatorname{argmax}_{P_X \in \mathcal{P}_{\mathsf{B}}} I(X;Y \mid S)$ also achieves minimum distortion $\mathsf{D}_{\min}(\mathsf{B})$ in (10), the capacity-distortion tradeoff is constant $\mathsf{C}(\mathsf{D},\mathsf{B}) = \mathsf{C}_{\mathsf{NoEst}}(\mathsf{B})$, irrespective of the allowed distortion D. This is in particular the case, when the expected distortion $\mathsf{E}[d(S,\hat{s}^*(X,Z))]$ does not depend on the input distribution P_X. The following corollary identifies a set of SDMCs $P_{YZ|SX}$ and state distributions P_S where this holds for all costs $\mathsf{B} \geq 0$.

Corollary 1. *Assume that there exists a function $\psi(\cdot)$ with domain $\mathcal{X} \times \mathcal{Z}$ so that irrespective of the input distribution P_X the following two conditions hold:*

$$(S, \psi(X,Z)) \perp X, \tag{16}$$

$$S \multimap \psi(X,Z) \multimap (X,Z), \tag{17}$$

for $(S,X,Z) \sim P_S P_X P_{Z|SX}$. In this case, for any given B, the rate-distortion tradeoff function $\mathsf{C}(\mathsf{D},\mathsf{B})$ is constant over $\mathsf{D} \geq \mathsf{D}_{\min}$ and equal to the channel capacity of the SDMC:

$$\mathsf{C}(\mathsf{D},\mathsf{B}) = \mathsf{C}_{\text{NoEst}}(\mathsf{B}), \qquad \forall \mathsf{D} \geq \mathsf{D}_{\min}(\mathsf{B}), \;\; \mathsf{B} \geq 0. \tag{18}$$

Proof: See Appendix D. ■

The following state-dependent erasure channel satisfies the conditions in above corollary. Let S be Bernoulli-p and Y equal to the erasure symbol "?" when $S=1$ and $Y=X$ when $S=0$. Moreover, assume perfect output feedback, i.e., $Y=Z$. For the choice $\psi(X,Z) = \mathbb{1}\{Z = \text{“?”}\} = S$ both Markov chains in Corollary 1 are trivially satisfied because S and X are independent.

Remark 1. *Theorem 1 is easily adapted to the more general case of imperfect channel state information (CSI), i.e., to a scenario where the receiver does not observe the state-sequence S^n but a related sequence S_R^n, where (S^n, S_R^n) are i.i.d. according to an arbitrary distribution P_{SS_R}. In this case, Theorem 1 remains valid if in Definition* (11) *the state S is replaced by S_R, i.e.,*

$$\mathsf{C}^{\text{imp}}(\mathsf{D},\mathsf{B}) = \max_{P_X \in \mathcal{P}_{\mathsf{D}} \cap \mathcal{P}_{\mathsf{B}}} I(X;Y \mid S_R), \tag{19}$$
$$\mathsf{D} \geq \mathsf{D}_{\min}(\mathsf{B}), \;\; \mathsf{B} \geq 0,$$

where $(X, S_R, Y, Z) \sim P_X P_{SS_R} P_{YZ|SS_RX}$ and the definitions of the sets $\mathcal{P}_{\mathsf{B}}$ and $\mathcal{P}_{\mathsf{D}}$ are kept as in (9a) *and* (9b)*, same as the definition of the function $c(x)$ in* (8)*. Notice that the symbolwise estimator in* (7) *remains optimal also in this related setup.*

Proof. See Appendix E. □

C. Examples

Before presenting our examples, we present two baseline schemes.

1) Baseline Schemes: We consider two baseline schemes that time share (TS) between two operating modes. The first baseline scheme, termed *Basic TS scheme*, is unable to simultaneously perform the sensing and communication tasks and splits its resources (time or bandwidth) between the following two modes:

- *Sensing mode without communication (achieves rate-distortion pair $(0, \mathsf{D}_{\min}(\mathsf{B}))$)*
  The input pmf P_X is chosen to minimize the distortion:
  $$P_{X_{\min}} := \operatorname*{argmin}_{P_X \in \mathcal{P}_{\mathcal{B}}} \sum_x P_X(x) c(x), \tag{20}$$
  and thus the minimum distortion $\mathsf{D}_{\min}(\mathsf{B})$ defined in (10) is achieved. Due to the lack of communication capability, the communication rate is zero.

- *Communication mode without sensing (achieves $(\mathsf{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\text{trivial}}(\mathsf{B}))$)*
  The input pmf P_X is chosen to maximize the rate:
  $$P_{X_{\max}} = \operatorname*{argmax}_{P_X \in \mathcal{P}_{\mathcal{B}}} I(X;Y \mid S), \tag{21}$$
  and this mode thus communicates at a rate equal to the channel capacity $\mathsf{C}_{\text{NoEst}}(\mathsf{B})$. Due to the lack of proper sensing capabilities, the estimator is set to a constant value regardless of the feedback and the input signals. The mode thus achieves distortion
  $$\mathsf{D}_{\text{trivial}}(\mathsf{B}) := \min_{s' \in \hat{\mathcal{S}}} \sum_{s \in \mathcal{S}} P_S(s) d(s, s'). \tag{22}$$

The second baseline scheme is called *Improved TS scheme* and can simultaneously perform the communication and sensing tasks. This scheme time-shares between the following modes.

- *Sensing mode with communication (achieves $(\mathsf{R}_{\min}(\mathsf{B}), \mathsf{D}_{\min}(\mathsf{B}))$)*
  The input pmf P_X is choosen according to (20) to achieve the minimum distortion. The chosen pmf $P_{X_{\min}}$ can achieve the following communication rate:
  $$\mathsf{R}_{\min} := I(X_{\min}; Y \mid S), \qquad \text{for } X_{\min} \sim P_{X_{\min}}. \tag{23}$$

- *Communication mode with sensing (achieves $(\mathsf{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\max}(\mathsf{B}))$)*
  The input pmf $P_{X_{\max}}$ is chosen as in (21) to maximize the communication rate. The mode thus communicates at the capacity $\mathsf{C}_{\text{NoEst}}(\mathsf{B})$ of the channel. Sensing is performed by means of the optimal estimator in (5). The mode thus achieves distortion
  $$\mathsf{D}_{\max} := \sum_{x \in \mathcal{X}} P_{X_{\max}}(x) c(x), \quad \text{for } X_{\max} \sim P_{X_{\max}}. \tag{24}$$

It is worth noticing that for any cost $\mathsf{B} \geq 0$, the two operating points of the two modes in the Improved TS scheme, $(\mathsf{R}_{\min}(\mathsf{B}), \mathsf{D}_{\min}(\mathsf{B}))$ and $(\mathsf{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\max}(\mathsf{B}))$, also lie on the capacity-distortion-cost tradeoff curve $\mathsf{C}(\mathsf{D},\mathsf{B})$ presented in Theorem 1. These two points are thus also operating points of any optimal co-design scheme. As we will see at hand of the following examples, all other operating points of the Improved TS scheme are typically suboptimal compared to an optimal co-design scheme.

2) Example 1: Binary Channel with Multiplicative Bernoulli State: Consider a channel $Y = SX$ with binary alphabets $\mathcal{X} = \mathcal{S} = \mathcal{Y} = \{0, 1\}$ and where the state S is Bernoulli-q, for $q \in (0, 1)$. We assume perfect output feedback to the transmitter $Y = Z$, and consider the Hamming distortion measure $d(s, \hat{s}) = s \oplus \hat{s}$. No cost constraint is imposed.

The following corollary specializes Theorem 1 to this example.

Corollary 2. *The capacity-distortion tradeoff of a binary channel with multiplicative Bernoulli state is given by*

$$\mathsf{C}(\mathsf{D}) = qH_{\mathsf{b}}\left(\frac{\mathsf{D}}{\min\{q, 1-q\}}\right), \tag{25}$$

where $H_{\mathsf{b}}(p)$ denotes the binary entropy function. In other words, the curve $\mathsf{C}(\mathsf{D})$ is parameterized as

$$\{(\mathsf{C} = qH_{\mathsf{b}}(p),\ \mathsf{D} = p\min\{q, 1-q\}) \colon p \in [0, 1/2]\}. \tag{26}$$

Proof: Since Y is deterministic given (S, X), and it equals 0 whenever $S = 0$, we have:

$$\begin{aligned} I(X;Y \mid S) &= P_S(0)H(Y \mid S=0) \\ &\quad + P_S(1)H(Y \mid S=1) \\ &= P_S(1)H(X). \end{aligned} \tag{27}$$

Setting $p := P_X(0)$, we obtain

$$I(X;Y \mid S) = qH_{\mathsf{b}}(p). \tag{28}$$

To calculate the distortion, we notice that the optimal estimator $\hat{s}^*(\cdot,\cdot)$ in Lemma 1 sets

$$\hat{s}^*(x, z) = \begin{cases} z, & \text{if } x = 1 \\ \operatorname{argmax}_{s \in \{0,1\}} P_S(s), & \text{if } x = 0. \end{cases} \tag{29}$$

In fact, whenever $x = 1$ the transmitter acquires full state knowledge because $z = y = s$. In this case $c(x = 1) = 0$. For $x = 0$, the transmitter does not receive any useful information about the state and hence uses the best constant estimator, irrespective of the feedback z. In this case,

$$\begin{aligned} c(x=0) &= \mathsf{E}\left[d\Big(S, \operatorname*{argmax}_{s \in \{0,1\}} P_S(s)\Big)\Big| X = 0\right] \\ &= \min_{s \in \{0,1\}} P_S(s) = \min\{q, 1-q\}, \end{aligned} \tag{30}$$

where we used the independence of S and X. The expected distortion of the optimal estimator thus evaluates to:

$$\mathsf{D} = \sum_x P_X(x)c(x) = P_X(0)c(0) = p\min\{q, 1-q\}. \tag{31}$$

■

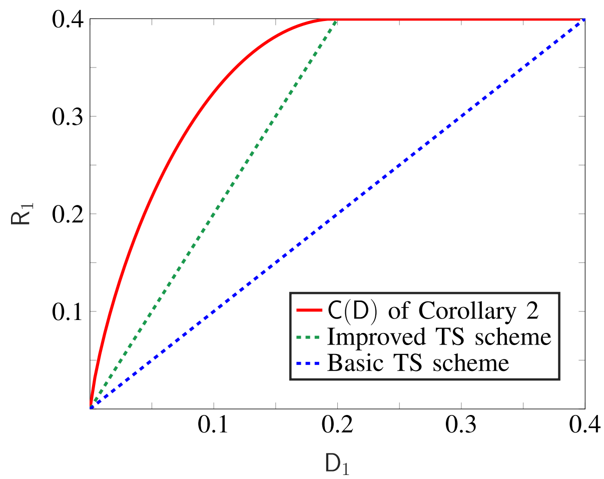

Fig. 2. Capacity-distortion tradeoff of the binary channel with multiplicative Bernoulli state of parameter $q = 0.4$.

The capacity-distortion tradeoff of Corollary 2 is illustrated in Fig. 2 for state parameter $q = 0.4$. The figure also compares the performances of the two baseline TS schemes. We observe a significant gain of an optimal co-design scheme over the two TS baseline schemes. We conclude this example with a derivation of the parameters of the TS schemes.

The capacity-achieving input distribution is easily found as $P_{X_{\max}}(0) = P_{X_{\max}}(1) = 1/2$, and by (28) and (31) we find $\mathsf{C}_{\text{NoEst}} = q$ and $\mathsf{D}_{\max} = \min\{q, 1-q\}/2$. Minimum distortion $\mathsf{D}_{\min} = 0$ is achieved by always sending $X = 1$, i.e., $P_{X_{\min}}(1) = 1$ and $P_{X_{\min}}(0) = 0$, in which case $D_{\min} = 0$ and $R_{\min} = 0$, see also (28) and (31). The Improved TS scheme thus achieves all pairs on the line connecting the two points $(0, 0)$ with $(q, \min\{q, 1-q\}/2)$. To determine the performance of the basic TS scheme, we recall that the best constant estimator (that does not consider the feedback) is $\hat{s}_{\text{const}} = \operatorname{argmax}_{s \in \{0,1\}} P_S(s)$, which allows to conclude that $\mathsf{D}_{\text{trivial}} = \min\{q, 1-q\}$. The basic TS scheme thus achieves all rate-distortion pairs on the line connecting the points $(0, 0)$ and $(q, \min\{q, 1-q\})$.

3) Example 2: Real Gaussian Channel with Rayleigh Fading: We consider the real Gaussian channel with Rayleigh fading:

$$Y_i = S_i X_i + N_i, \tag{32}$$

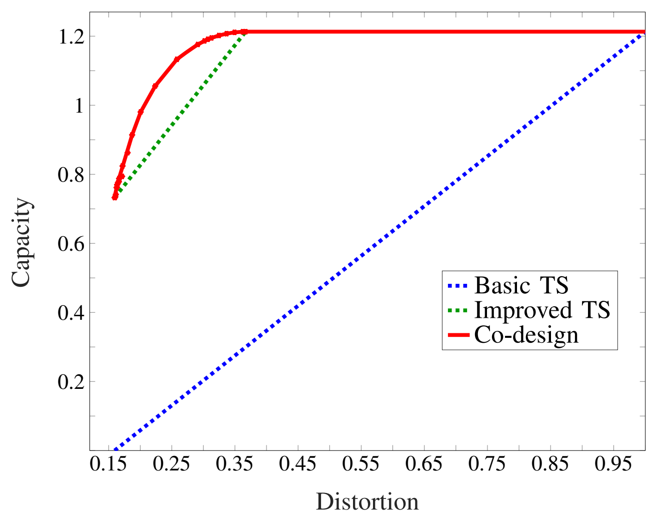

Fig. 3. Capacity-distortion tradeoff of fading AWGN channel $\mathsf{B} = 10$ dB and $\sigma_{\text{fb}}^2 = 1$.

where X_i is the channel input satisfying $\overline{\lim}_{n\to\infty} \frac{1}{n} \sum_i \mathsf{E}\left[|X_i|^2\right] \leq \mathsf{B} = 10$dB, and both sequences $\{N_i\}$ and $\{S_i\}$ are independent of each other and i.i.d. Gaussian with zero mean and unit variance. The transmitter observes the noisy feedback

$$Z_i = Y_i + N_{\text{fb},i}, \tag{33}$$

where $\{N_{\text{fb},i}\}$ are i.i.d. zero-mean Gaussian of variance $\sigma_{\text{fb}}^2 \geq 0$. We consider the quadratic distortion measure $d(s, \hat{s}) = (s - \hat{s})^2$.

First, we characterize the two operating points achieved by the Improved TS baseline scheme. The capacity of this channel is achieved with a Gaussian input $X_{\max} \sim \mathcal{N}(0, \mathsf{B})$, and thus the communication mode with sensing achieves the rate-distortion pair

$$\mathsf{C}_{\text{NoEst}}(\mathsf{B}) = \frac{1}{2}\mathsf{E}\left[\log(1 + |S|^2\mathsf{B})\right] = 1.213, \tag{34}$$

$$\mathsf{D}_{\max}(\mathsf{B}) = \mathsf{E}\left[\frac{(1+\sigma_{\text{fb}}^2)}{1 + |X_{\max}|^2 + \sigma_{\text{fb}}^2}\right] = 0.367, \tag{35}$$

where we have set $\sigma_{\text{fb}}^2 = 1$ and $P = 10$dB to obtain the numerical values. Minimum distortion $\mathsf{D}_{\min}$ is achieved by 2-ary pulse amplitude modulation (PAM), and thus the sensing mode with communication achieves rate-distortion pair

$$\mathsf{R}_{\min}(\mathsf{B}) = 0.733, \quad \mathsf{D}_{\min}(\mathsf{B}) = \frac{1 + \sigma_{\text{fb}}^2}{1 + P + \sigma_{\text{fb}}^2} = 0.166, \tag{36}$$

where the numerical value again corresponds to $\sigma_{\text{fb}} = 1$ and $B = 10$dB. Next, we characterize the performance of the basic TS baseline scheme. The best constant estimator for this channel is $\hat{s} = 0$, and the communication mode without sensing achieves rate-distortion pair $(\mathsf{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\text{trivial}}(\mathsf{B}) = 1)$. The sensing mode without communication achieves rate-distortion pair $(0, \mathsf{D}_{\min}(\mathsf{B}))$.

In Fig. 3, we compare the rate-distortion tradeoff achieved by these two TS baseline schemes with a numerical approximation of the capacity-distortion-cost tradeoff $\mathsf{C}(\mathsf{D}, \mathsf{B})$ of this channel. As previously explained, $\mathsf{C}(\mathsf{D}, \mathsf{B})$ also passes through the two end points $(\mathsf{R}_{\min}(\mathsf{B}), \mathsf{D}_{\min}(\mathsf{B}))$ and $(\mathsf{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\max}(\mathsf{B}))$ of the Improved TS scheme. We use the Blahut-Arimoto type Algorithm 1 to obtain a numerical approximation of the points on $\mathsf{C}(\mathsf{D}, \mathsf{B})$ in between these two operating points. Specifically, the input alphabet is quantized to a $M = 16$-ary PAM constellation

$$\mathcal{X}_{\text{q}} := \{(2m - 1 - M)\kappa, m = 1, \cdots, M\}, \tag{37}$$

where $\kappa := \sqrt{3P/(M^2 - 1)}$. The Gaussian noise N is quantized with a centered equally-spaced 50-points alphabet, and the state S is quantized by applying an equally-spaced 8000-points quantizer on the Chi-square distributed random variable S^2. Denoting the quantized input, noise, and state by X_{q}, N_{q}, and S_{q}, we keep our multiplicative-state, additive-noise channel model to generate the channel outputs used to run Algorithm 1 to obtain the numerical approximations:

$$Y_{\text{q}} = S_{\text{q}} X_{\text{q}} + N_{\text{q}}. \tag{38}$$

III. Multiple Receivers

In this section, we consider joint sensing and communication over two-receiver broadcast channels.

A. System Model

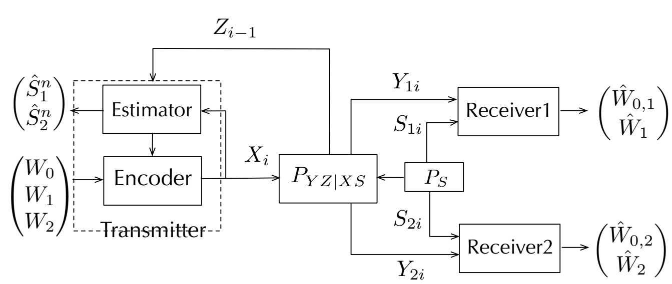

Fig. 4. State-dependent broadcast channel with generalized feedback and state-estimator at the transmitter.

Consider the two-receiver broadcast channel scenario depicted in Fig. 4. The model comprises a two-dimensional memoryless state sequence $\{(S_{1,i}, S_{2,i})\}_{i\geq 1}$

whose samples at any given time i are distributed according to a given joint law $P_{S_1S_2}$ over the state alphabets $\mathcal{S}_1 \times \mathcal{S}_2$. Receiver 1 observes state sequence $\{S_{1,i}\}$ and Receiver 2 observes state sequence $\{S_{2,i}\}$. The transmitter communicates with both receivers over a state-dependent memoryless broadcast channel (SDMBC), where given time-i input $X_i = x$ and state realizations $S_{1,i} = s_1$ and $S_{2,i} = s_2$, the time-i outputs $Y_{1,i}$ and $Y_{2,i}$ observed at the receivers and the transmitter's feedback signal Z_i are distributed according to the stationary channel transition law $P_{Y_1Y_2Z|S_1S_2X}(\cdot,\cdot,\cdot|s_1,s_2,x)$. We again assume that all alphabets $\mathcal{X},\mathcal{Y}_1,\mathcal{Y}_2,\mathcal{Z},\mathcal{S}_1,\mathcal{S}_2$ are finite.

The goal of the transmitter is to convey a common message W_0 to both receivers and individual messages W_1 and W_2 to Receivers 1 and 2, respectively, while estimating the states sequences $\{S_{1,i}\}$ and $\{S_{2,i}\}$ within some target distortions. For simplicity, the input cost constraint is omitted.

A $(2^{n\mathsf{R}_0}, 2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ code for an SDMBC thus consists of

1) three message sets $\mathcal{W}_0 = [1 : 2^{n\mathsf{R}_0}]$, $\mathcal{W}_1 = [1 : 2^{n\mathsf{R}_1}]$, and $\mathcal{W}_2 = [1 : 2^{n\mathsf{R}_2}]$;
2) a sequence of encoding functions $\phi_i \colon \mathcal{W}_0 \times \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{Z}^{i-1} \to \mathcal{X}$, for $i = 1, 2, \ldots, n$;
3) for each $k = 1, 2$ a decoding function $g_k \colon \mathcal{S}_k^n \times \mathcal{Y}_k^n \to \mathcal{W}_0 \times \mathcal{W}_k$;
4) for each $k = 1, 2$ a state estimator $h_k \colon \mathcal{X}^n \times \mathcal{Z}^n \to \hat{\mathcal{S}}_k^n$, where $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$ are given reconstruction alphabets.

For a given code, we let the random messages W_0, W_1, and W_2 be uniform over the message sets $\mathcal{W}_0$, $\mathcal{W}_1$, and $\mathcal{W}_2$ and the inputs $X_i = \phi_i(W_0, W_1, W_2, Z^{i-1})$, for $i = 1, \ldots, n$. The corresponding outputs $Y_{1,i}Y_{2,i}, Z_i$ at time i are obtained from the states $S_{1,i}$ and $S_{2,i}$ and the input X_i according to the SDMBC transition law $P_{Y_1Y_2Z|S_1S_2X}$. Further, for $k = 1, 2$ let $\hat{S}_k^n := (\hat{S}_{k,1}, \cdots, \hat{S}_{k,n}) = h_k(X^n, Z^n)$ be the transmitter's estimates for state S_k^n and $(\hat{W}_{0,k}, \hat{W}_k) = g_k(S_k^n, Y_k^n)$ the messages decoded by Receiver k. The quality of the state estimates $\hat{S}_k^n$ is again measured by bounded per-symbol distortion functions $d_k \colon \mathcal{S}_k \times \hat{\mathcal{S}}_k \mapsto [0, \infty)$, i.e., we assume

$$\max_{s_k \in \mathcal{S}_k, \hat{s}_k \in \hat{\mathcal{S}}_k} d_k(s_k, \hat{s}_k) < \infty, \quad k = 1, 2. \tag{39}$$

Our interest is in the two *expected average per-block distortions*

$$\Delta_k^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d_k(S_{k,i}, \hat{S}_{k,i})], \quad k = 1, 2, \tag{40}$$

and the joint probability of error

$$P_e^{(n)} := \Pr\Big((\hat{W}_{0,k}, \hat{W}_1) \neq (W_0, W_1) \quad \text{or} \quad (\hat{W}_{0,k}, \hat{W}_2) \neq (W_0, W_2)\Big). \tag{41}$$

Definition 3. *A rate-distortion tuple* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *is achievable if there exists a sequence (in* n*) of* $(2^{n\mathsf{R}_0}, 2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ *codes that simultaneously satisfy*

$$\lim_{n\to\infty} P_e^{(n)} = 0 \tag{42a}$$

$$\overline{\lim_{n\to\infty}} \, \Delta_k^{(n)} \leq \mathsf{D}_k, \quad \text{for } k = 1, 2. \tag{42b}$$

Definition 4. *The capacity-distortion region* $\mathcal{CD}$ *is given by the closure of the union of all achievable rate-distortion tuples* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$.

In the remainder of the section, we present bounds on the capacity-distortion region $\mathcal{CD}$. As in the single-receiver case, one can easily determine the optimal estimator functions h_1 and h_2, which are independent of the encoding and decoding functions and operate on a symbol-by-symbol basis.

Lemma 3. *For each* $k = 1, 2$*, define the function*

$$\hat{s}_k^*(x, z) := \arg\min_{s' \in \hat{\mathcal{S}}_k} \sum_{s_k \in \mathcal{S}_k} P_{S_k|XZ}(s_k|x, z) d(s_k, s'), \tag{43}$$

where ties can be broken arbitrarily.

Irrespective of the choice of encoding and decoding functions, distortions $\Delta_1^{(n)}$ *and* $\Delta_2^{(n)}$ *are minimized by the estimators for* $k = 1, 2$

$$h_k^*(x^n, z^n) = (\hat{s}_k^*(x_1, z_1), \hat{s}_k^*(x_2, z_2), \ldots, \hat{s}_k^*(x_n, z_n)). \tag{44}$$

Proof: See Appendix A. ■

Analogously to the definition in Equation (8) we can then define the optimal estimation cost for each input symbol $x \in \mathcal{X}$:

$$c_k(x) := \mathsf{E}[d_k(S_k, \hat{s}_k^*(X, Z))|X = x], \qquad k = 1, 2. \tag{45}$$

Characterizing the capacity-distortion region is very challenging in general, because even the capacity regions of the SDMBC with and without feedback are unknown to date. We first present the exact capacity-distortion region for the class of physically degraded SDMBCs and then provide bounds for general SDMBCs. We shall also compare our results on the capacity-distortion regions to the performances achieved by simple TS baseline schemes, in analogy to the single-receiver setup.

Specifically, we again have a *basic TS baseline scheme* that performs either sensing or communication at a time, and an *improved TS baseline scheme* that is able to perform both functions simultaneously via a common waveform by prioritizing either sensing or communication. Analogously to the single-receiver setup, each of the two baseline schemes time-shares between a sensing mode and a communication mode. However, since we now have two distortions and three rates, the choice of the "optimal" pmf P_X for each mode is not necessarily unique, but rather a continuum, depending on which function of the two distortions or the three rates one wishes to optimize. For fixed input pmf, the difference between the communication mode *without sensing* (employed by the basic TS scheme) and the communication mode *with sensing* (employed by the improved TS scheme) lies in the choice of the estimators. In the former mode, the transmitter applies the best *constant estimators* for the two state-sequences, irrespective of its inputs and feedback outputs. In the latter mode, it applies the optimal estimators in Lemma 3, which depend on the input and the feedback output. Similarly, the difference between the communication modes *without and with sensing* is that in the former all rates are zero and in the latter the chosen input pmf P_X can be used for communication at positive rates.

B. Capacity-Distortion Region for Physically Degraded SDMBCs

This section characterizes the capacity-distortion region for *physically degraded SDMBCs* and evaluates it for two binary examples.

Definition 5. *An SDMBC $P_{Y_1Y_2Z|S_1S_2X}$ with state pmf $P_{S_1S_2}$ is called* physically degraded *if there are conditional laws $P_{Y_1|XS_1}$ and $P_{S_2Y_2|S_1Y_1}$ such that*

$$P_{Y_1Y_2|S_1S_2X}P_{S_1S_2} = P_{S_1}P_{Y_1|S_1X}P_{S_2Y_2|S_1Y_1}. \tag{46}$$

That means for any arbitrary input P_X, the tuple $(X, S_1, S_2, Y_1, Y_2) \sim P_XP_{S_1S_2}P_{Y_1Y_2|S_1S_2X}$ satisfies the Markov chain

$$X \multimap\!- (S_1, Y_1) \multimap\!- (S_2, Y_2). \tag{47}$$

Theorem 2. *The capacity-distortion region $\mathcal{CD}$ of a physically degraded SDMBC is given by the closure of the set of all tuples $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ for which there exists a joint law P_{UX} so that the tuple $(U, X, S_1, S_2, Y_1, Y_2, Z) \sim P_{UX}P_{S_1S_2}P_{Y_1Y_2Z|S_1S_2X}$ satisfies the two rate constraints*

$$\mathsf{R}_1 \leq I(X; Y_1 \mid U, S_1) \tag{48}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \leq I(U; Y_2 \mid S_2), \tag{49}$$

and the distortion constraints

$$\mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z))] \leq \mathsf{D}_k, \quad k = 1, 2. \tag{50}$$

Proof: The achievability can be proved by standard superposition coding and using the optimal estimators in Lemma 3. The converse also follows from standard steps and the details are provided in Appendix F. ■

As mentioned in the proof, data communication is performed by simple superposition coding that ignores the feedback. Thus, also for physically degraded BCs feedback only facilitates state sensing but is useless for communications.

Remark 2. *Similarly to the single-receiver case, an input cost-constraint as in* (4c) *can be added to our model. Theorem 2 remains valid in this case, if the choice of the input distribution P_X is limited to satisfy the cost constraint*

$$\sum_{x \in \mathcal{X}} P_X(x)b(x) \leq B. \tag{51}$$

The analogous remark also applies to the non-physically degraded BC ahead and the presented inner and outer bounds.

Remark 3. *Similarly to what we described in Remark 1, the result in Theorem 2 can be extended to the case with imperfect receiver state-informations $S_{R,1}^n$ and $S_{R,2}^n$. For $(S^n, S_{R,1}^n, S_{R,2}^n)$ i.i.d. $\sim P_{SS_{R,1},S_{R,2}}$ it suffices to replace in the rate-constraints* (48) *and* (49) *of Theorem 2 the state S_1 by $S_{R,1}$ and the state S_2 by $S_{R,2}$. The analogous remark also applies to the non-physically degraded BC ahead and the presented inner and outer bounds.*

In what follows, we evaluate above Theorem 2 for two examples.

1) Example 3: Binary BC with Multiplicative Bernoulli States: Consider the physically degraded SDMBC with binary input and output alphabets $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$ and binary state alphabets $\mathcal{S}_1 = \mathcal{S}_2 = \{0, 1\}$. The channel input-output relation is described by

$$Y_k = S_kX, \qquad k = 1, 2, \tag{52}$$

with the joint state pmf

$$P_{S_1S_2}(s_1, s_2) = \begin{cases} 1-q, & \text{if } (s_1, s_2) = (0, 0) \\ 0, & \text{if } (s_1, s_2) = (0, 1) \\ q\gamma, & \text{if } (s_1, s_2) = (1, 1) \\ q(1-\gamma) & \text{if } (s_1, s_2) = (1, 0), \end{cases} \tag{53}$$

for $\gamma, q \in [0, 1]$. Notice that S_2 is a degraded version of S_1, which together with the transition law (52) ensures the Markov chain $X \multimap\!- (S_1, Y_1) \multimap\!- (S_2, Y_2)$ and the physically degradedness of the SDMBC. We consider output feedback

$$Z = (Y_1, Y_2), \tag{54}$$

and set the common rate $\mathsf{R}_0 = 0$ for simplicity.

In this SDMBC, zero distortions $\mathsf{D}_1 = \mathsf{D}_2 = 0$ can be achieved by deterministically choosing $X = 1$ exactly as for the single-receiver case. This choice however cannot achieve any positive communication rates, i.e., $\mathsf{R}_1 = \mathsf{R}_2 = 0$. In the sensing mode with and without communication, we thus have:

$$(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) = (0, 0, 0, 0). \tag{55}$$

The optimal input distribution for communication is $X_{\max} \sim \mathcal{B}(1/2)$, in which case all rate-pairs $(\mathsf{R}_1, \mathsf{R}_2)$ satisfying

$$\mathsf{R}_k \leq P_{S_k}(1), \qquad k = 1, 2, \tag{56}$$

are achievable. The input $X_{\max} \sim \mathcal{B}(1/2)$ simultaneously maximizes both communication rates $\mathsf{R}_1, \mathsf{R}_2$.

In the communication mode *without* sensing, the transmitter applies the optimal constant estimator for each state, namely

$$\hat{s}_{\text{const},k} := \underset{\hat{s} \in \{0,1\}}{\operatorname{argmax}}\, P_{S_k}(\hat{s}), \qquad k = 1, 2, \tag{57}$$

and thus achieves all tuples

$$(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) = (qr, \gamma q(1-r), \mathsf{D}_{1,\max}, \mathsf{D}_{2,\max}) \tag{58}$$

where $\mathsf{D}_{1,\max} := \min\{q, 1-q\}$ and $\mathsf{D}_{2,\max} := \min\{\gamma q, 1-\gamma q\}$, and $r \in [0, 1]$ denotes the time-sharing parameter between the two communication rates.

In the communication mode *with* sensing, the same input $X_{\max}$ is used. The transmitter however applies the optimal estimator for $k = 1, 2$:

$$\hat{s}_k^*(x, y_1, y_2) = \begin{cases} y_k, & \text{if } x = 1 \\ \hat{s}_{\text{const},k}, & \text{if } x = 0, \end{cases} \tag{59}$$

and achieves the tuple

$$(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) = \left(qr, \gamma q(1-r), \frac{\mathsf{D}_{1,\max}}{2}, \frac{\mathsf{D}_{2,\max}}{2}\right), \tag{60}$$

where r again denotes the time-sharing parameter between the two communication rates.

The basic and improved TS baseline schemes achieve the time-sharing lines between points (55) and (58) and points (55) and (60), respectively. The following corollary evaluates Theorem 2 to obtain the performance of the optimal co-design scheme.

Corollary 3. *The capacity-distortions region $\mathcal{CD}$ of the binary physically degraded SDMBC in* (52)–(54) *is the set of all tuples* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *satisfying*

$$\mathsf{R}_0 + \mathsf{R}_1 \leq qH_\mathrm{b}(p)r, \tag{61a}$$
$$\mathsf{R}_0 + \mathsf{R}_2 \leq \gamma qH_\mathrm{b}(p)(1-r), \tag{61b}$$
$$\mathsf{D}_1 \geq p\min\{q, 1-q\}, \tag{61c}$$
$$\mathsf{D}_2 \geq p\min\{\gamma q, 1-\gamma q\}, \tag{61d}$$

for some choice of the parameters $r, p \in [0, 1]$.

Proof: We start by noticing that for this example $I(X;Y_1 \mid U, S_1) = qH(X|U)$ and $I(U;Y_2 \mid S_2) = q\gamma(H(X) - H(X \mid U))$. Setting $p := P_X(0)$ and $r := \frac{H(X|U)}{H(X)}$, directly leads to the desired rate constraints. The distortion constraints are obtained from the optimal estimators in (59). Following the same steps as in the single-receiver case, i.e. (30) and (31), we obtain

$$\mathsf{D}_k \geq p\min\{P_{S_k}(0), P_{S_k}(1)\}, \tag{62}$$

which concludes the proof. ■

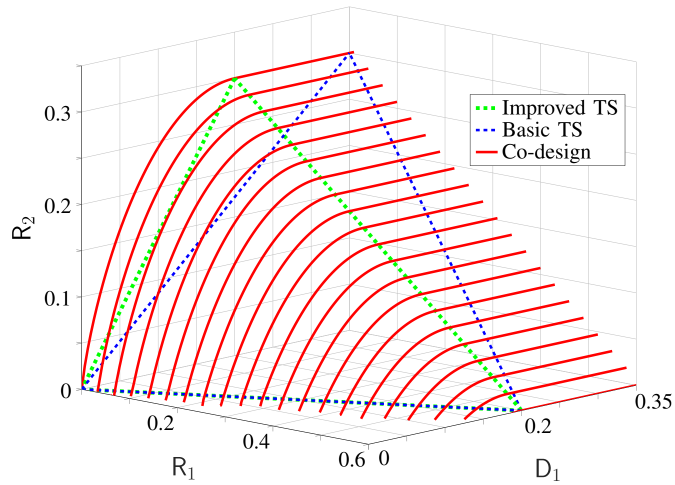

Fig. 5. Boundary of the capacity-distortion region $\mathcal{CD}$ for Example 3 in Subsection III-B1.

Notice that above Corollary 3 reduces to Corollary 2 in the special case of $\mathsf{R}_0 = \mathsf{R}_2 = 0$ and $\mathsf{D}_2 = \infty$, i.e., when we ignore Receiver 2.

Fig. 5 shows in red colour the boundary of the projection of the tradeoff region $\mathcal{CD}$ of this example onto the 3-dimensional plane $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1)$, for parameters $\gamma = 0.5$ and $q = 0.6$. The tradeoff with D_2 is omitted for simplicity and because D_2 is a scaled version of D_1. The figure also shows the boundaries of the basic and improved TS baseline schemes. We again notice a significant gain for an optimal co-design scheme compared to the TS baseline schemes.

So far, there was no tradeoff between the two distortion constraints D_1 and D_2. This is different in the next example, which otherwise is very similar.

2) Example 4: Binary BC with Multiplicative Bernoulli States and Flipping Inputs: Reconsider the same state pmf $P_{S_1S_2}$ as in the previous example, but now an SDMBC with a transition law that flips the input for receiver 2:

$$Y_1 = S_1 X, \qquad Y_2 = S_2(1 - X). \tag{63}$$

As in the previous example we consider output feedback $Z = (Y_1, Y_2)$.

Corollary 4. *The capacity-distortion region $\mathcal{CD}$ of the binary SDMBC with flipping inputs in* (63) *and output feedback is the set of all tuples* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *satisfying*

$$\mathsf{R}_1 \leq qH_\text{b}(p)r, \tag{64a}$$
$$\mathsf{R}_0 + \mathsf{R}_2 \leq \gamma q H_\text{b}(p)(1-r), \tag{64b}$$
$$\mathsf{D}_1 \geq p\min\{q(1-\gamma), (1-q)\}, \tag{64c}$$
$$\mathsf{D}_2 \geq (1-p)q\min\{\gamma, 1-\gamma\}, \tag{64d}$$

for some choice of the parameters $r, p \in [0,1]$.

The capacity-distortion region expression above captures the tradeoffs between the two rates through the parameter r, between the rates and the distortions through the parameter p, and between the two distortions through the parameter p.

Comparing above Corollary 4 to the previous Corollary 3, we remark the identical rate constraints and the relaxed distortion contraints for both D_1 and D_2 in Corollary 4. The reason is that the flipping input allows the transmitter to perfectly estimate S_1 from (X, Y_1, Y_2) not only when $X = 1$ but also when $X = 0$ and $Y_2 = 1$ because they imply that $S_2 = 1$ and by (53) also $S_1 = 1$.

Proof: The proof is similar to the proof of Corollary 3, except for the description of the optimal estimators. To determine these optimal estimators, we remark that only four input-output relations are possible: $(x, y_1, y_2) \in \{(0,0,0), (0,0,1), (1,0,0), (1,1,0)\}$. Moreover, when $X = 1$, then $Y_1 = S_1$, and when $X = 0$, then $Y_2 = S_2$. In particular, when $X = 0$ and $Y_2 = 1$, then $S_2 = 1$ and also $S_1 = 1$, see (53). The optimal estimator for state S_1 thus is:

$$\hat{s}_1^*(x, y_1, y_2) = \begin{cases} y_1, & \text{if } x = 1 \\ 1, & \text{if } (x, y_2) = (0,1) \\ \arg\min_s P_{S_1|S_2}(s|0), & \text{else,} \end{cases} \tag{65}$$

and $\hat{s}_1^*(X, Y_1, Y_2) = S_1$ unless $X = 0$, $Y_2 = 0$, and $S_1 \neq \arg\min_s P_{S_1|S_2}(s|0)$, which is equivalent to $(X = 0, S_2 = 0)$ and $S_1 \neq \arg\min_s P_{S_1|S_2}(s|0)$. This yields $c_1(1) = 0$ and because S_2 is independent of X:

$$c_1(0) = P_{S_2}(0)\min_s P_{S_1|S_2}(s|0). \tag{66}$$

Recalling $p = P_X(0)$, we readily obtain the distortion for state S_1:

$$\mathsf{D}_1 = p\min_s P_{S_1,S_2}(s, 0) = p\min\{q(1-\gamma), 1-q\}. \tag{67}$$

The optimal estimator and the corresponding distortion for state S_2 can be obtained in a similar way. ■

C. Capacity-Distortion Region for General SDMBCs

In the remainder of this section, we reconsider general SDMBCs, for which we present bounds on $\mathcal{CD}$. We start with a simple outer bound.

Theorem 3 (Outer Bound on $\mathcal{CD}$)**.** *If* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *lies in $\mathcal{CD}$ for a given SDMBC $P_{Y_1Y_2Z|S_1S_2X}$ with state pmf $P_{S_1S_2}$, then there exist pmfs $P_X, P_{U_1|X}, P_{U_2|X}$ such that the random tuple $(U_k, X, S_1, S_2, Y_1, Y_2, Z) \sim P_{U_k|X}P_XP_{S_1S_2}P_{Y_1Y_2Z|S_1S_2X}$ satisfies the rate constraints*

$$\mathsf{R}_0 + \mathsf{R}_k \leq I(U_k; Y_k \mid S_k), \quad k = 1, 2, \tag{68a}$$
$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq I(X; Y_1, Y_2 \mid S_1, S_2), \tag{68b}$$

and the average distortion constraint

$$\mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z))] \leq \mathsf{D}_k, \quad k = 1, 2, \tag{69}$$

where the function $\hat{s}_k^*(\cdot,\cdot)$ is defined in (43).

Proof: See Appendix F. ■

Achievability results are easily obtained by combining existing achievability results for SDMBCs with generalized feedback with the optimal estimator in Lemma 3. We consider the block-Markov coding scheme in [20], which in each block applies Marton coding to transmit fresh data to the receivers as well as compression information describing the inputs and outputs of the previous block. The receivers decode the Marton codewords backwards, starting from the last block, and using both their channel outputs as well as the previously decoded compression information pertaining to the block. Combining this scheme with the optimal estimator in Lemma 3 yields the following proposition.

Proposition 1 (Inner Bound on $\mathcal{CD}$)**.** *Consider an SDMBC $P_{Y_1Y_2Z|S_1S_2X}$ with state pmf $P_{S_1S_2}$. The capacity-distortion region $\mathcal{CD}$ includes all tuples* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *that satisfy inequalities* (70) *on top of this page and the distortion constraints* (69)*. where $(U_0, U_1, U_2, X, S_1, S_2, Y_1, Y_2, Z, V_0, V_1, V_2) \sim P_{U_0U_1U_2X}P_{S_1S_2}P_{Y_1Y_2Z|S_1S_2X}P_{V_0V_1V_2|U_0U_1U_2Z}$, for some choice of (conditional) pmfs $P_{U_0U_1U_2X}$ and $P_{V_0V_1V_2|U_0U_1U_2Z}$.*

$$\mathsf{R}_0 + \mathsf{R}_1 \leq I(U_0, U_1; Y_1, V_1 \mid S_1) - I(U_0, U_1, U_2, Z; V_0, V_1 \mid S_1, Y_1) \tag{70a}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \leq I(U_0, U_2; Y_2, V_2 \mid S_2) - I(U_0, U_1, U_2, Z; V_0, V_2 \mid S_2, Y_2) \tag{70b}$$

$$\begin{aligned}\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq\ & I(U_1; Y_1, V_1 | U_0, S_1) + I(U_2; Y_2, V_2 \mid U_0, S_2) + \min_{k\in\{1,2\}} I(U_0; Y_k, V_k \mid S_k) - I(U_1; U_2 \mid U_0) \\ & - I(U_0, U_1, U_2, Z; V_1 \mid V_0, S_1, Y_1) - I(U_0, U_1, U_2, Z; V_2 | V_0, S_2, Y_2) \\ & - \max_{k\in\{1,2\}} I(U_0, U_1, U_2, Z; V_0 \mid S_k, Y_k)\end{aligned} \tag{70c}$$

$$\begin{aligned}2\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq\ & I(U_0, U_1; Y_1, V_1 \mid S_1) + I(U_0, U_2; Y_2, V_2 \mid S_2) - I(U_1; U_2 \mid U_0) \\ & - I(U_0, U_1, U_2, Z; V_0, V_1 \mid S_1, Y_1) - I(U_0, U_1, U_2, Z; V_0, V_2 \mid S_2, Y_2)\end{aligned} \tag{70d}$$

Proof: Similar to [20] and omitted. ■

In analogy to Corollary 1 for the single-receiver case, for some SDMBCs there is no tradeoff between the achievable distortions and communication rates. In this case, for the BC, the capacity-distortion region is given by the Cartesian product between the SDMBC's capacity region:

$$\mathcal{C} := \{(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2) \colon \mathsf{D}_1 \geq 0,\ \mathsf{D}_2 \geq 0 \text{ s.t. } (\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) \in \mathcal{CD}\}, \tag{71}$$

and its distortion region:

$$\mathcal{D} := \{(\mathsf{D}_1, \mathsf{D}_2) \colon \mathsf{R}_0 \geq 0,\ \mathsf{R}_1 \geq 0,\ \mathsf{R}_2 \geq 0 \text{ s.t. } (\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) \in \mathcal{CD}\}. \tag{72}$$

Proposition 2 (No Rate-Distortion Tradeoff)**.** *Consider an SDMBC $P_{Y_1Y_2Z|S_1S_2X}$ with state pmf $P_{S_1S_2}$ for which there exist functions ψ_1 and ψ_2 with domain $\mathcal{X} \times \mathcal{Z}$ so that irrespective of the input distribution P_X the relations*

$$(S_k, \psi_k(Z, X)) \perp X, \tag{73}$$

$$S_k \multimap\!\!- \psi_k(Z, X) \multimap\!\!- (Z, X), \quad k = 1, 2, \tag{74}$$

hold for $(S_1, S_2, X_2, Z) \sim P_{S_1}P_{S_2}P_X P_{Z|XS_1,X_2}$. The capacity-distortion region of this SDMBC is the product of the capacity region and the distortion region:

$$\mathcal{CD} = \mathcal{C} \times \mathcal{D}. \tag{75}$$

Proof: Analogous to the proof of Corollary 1. Specifically, the proof is obtained from Appendix D by replacing $(S, \hat{S}, \psi, Y, T)$ with $(S_k, \hat{S}_k, \psi_k, Y_k, T_k)$, for $k = 1, 2$. ■

D. Example 5: Erasure BC with Noisy Feedback

Our first example satisfies Conditions (73) and (74) in Proposition 2 for an appropriate choice of ψ_1 and ψ_2, and its capacity-distortion region is thus given by the product of the capacity region and the distortion region.

Let $(E_1, S_1, E_2, S_2) \sim P_{E_1S_1E_2S_2}$ over $\{0,1\}^4$ be given but arbitrary. Consider the state-dependent erasure BC

$$Y_k = \begin{cases} X & \text{if } S_k = 0, \\ ? & \text{if } S_k = 1, \end{cases}, \qquad k = 1, 2, \tag{76}$$

where the feedback signal $Z = (Z_1, Z_2)$ is given by

$$Z_k = \begin{cases} Y_k & \text{if } E_k = 0, \\ ? & \text{if } E_k = 1, \end{cases}, \qquad k = 1, 2. \tag{77}$$

Further consider Hamming distortion measures $d_k(s, \hat{s}) = s \oplus \hat{s}$, for $k = 1, 2$. For the choice

$$\psi_k(Z_k) = \begin{cases} 1, & \text{if } \quad Z_k = ?, \\ 0, & \text{else}, \end{cases} \tag{78}$$

the described SDMBC satisfies the conditions in Proposition 2, thus yielding the following corollary.

Corollary 5. *The capacity-distortion region of the state-dependent erasure BC with noisy feedback in* (76)–(77) *is the Cartesian product between the capacity region of the SDMBC and its distortion region:*

$$\mathcal{CD} = \mathcal{C} \times \mathcal{D}. \tag{79}$$

When $P_{E_1S_1E_2S_2} = P_{E_1S_1}P_{E_2S_2}$, then the distortion region is given by:

$$\mathcal{D} = \{(\mathsf{D}_1, \mathsf{D}_2) \colon \mathsf{D}_k \geq P_{E_kS_k}(1, 0)\}. \tag{80}$$

Proof. The state can perfectly be estimated ($S_k = 0$) with zero distortion if $(S_k, E_k) = (0, 0)$. Otherwise, the feedback is $Z_k = ?$ and provides no information. The optimal estimator is then given by the best constant estimator, which in this example is:

$$\hat{s}_{\text{const},k} = \mathbf{1}\{P_{S_k}(1) \geq P_{S_kE_k}(0, 1)\}. \tag{81}$$

This immediately yields the distortion constraint in (80). □

Notice that the capacity region $\mathcal{C}$ of the SDMBC (76)

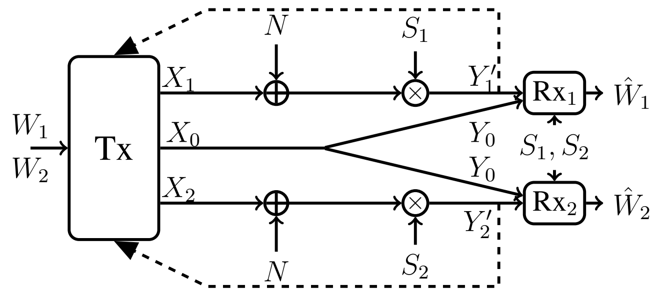

Fig. 6. State-dependent Dueck Broadcast Channel.

is unknown even with perfect feedback. In [29], [30], the capacity region of this SDMBC with perfect feedback was characterized when each receiver is informed about the state realizations at *both* receivers.

E. Example 6: State-Dependent Dueck's BC with Multiplicative Bernoulli States

We consider a state-dependent version of Dueck's example in [31], which first served to show that feedback can increase capacity of a memoryless BC. Interestingly, despite its simplicity, the state-dependent extension of this example allows observing various kinds of tradeoffs between communication and sensing performances and also between performances at the various receivers. For example, for specific choices of parameters, the problems of sensing and communication decompose (Corollary 6), and it is possible to simultaneously achieve the optimal sensing and communication performances. For other parameters a tradeoff arises. The present example also shows nicely that our presented co-design scheme can significantly outperform the two TS methods.

Consider the state-dependent BC in Figure 6 with input $X = (X_0, X_1, X_2) \in \{0,1\}^3$, i.i.d. Bernoulli states $S_1, S_2 \sim \mathcal{B}(q)$, for $q \in [0,1]$, and outputs

$$Y_k = (X_0, Y_k', S_1, S_2), \qquad k = 1,2, \tag{82}$$

where

$$Y_k' = S_k(X_k \oplus N), \qquad k = 1,2, \tag{83a}$$

and the noise $N \sim \mathcal{B}(1/2)$ is independent of the inputs and the states. The feedback signal is

$$Z = (Y_1', Y_2'), \tag{84}$$

and for simplicity we again ignore the common rate R_0.

We notice that only X_1 and X_2 are corrupted by the state and the noise. Since X_0 is received without any state or noise, it is thus completely useless for sensing. In fact, the optimal estimator of Lemma 3 for $k = 1,2$ is (see Appendix H-A)

$$\hat{s}_k^*(x_1, x_2, y_1', y_2') = \begin{cases} \mathbb{1}\{q \geq (1-q)\} & y_k' = 0, y_{\bar{k}}' = 1, x_1 \neq x_2 \\ 0 & y_k' = 0, y_{\bar{k}}' = 1, x_1 = x_2 \\ 1 & y_k' = 1 \\ 0 & y_1 = y_2 = 0, x_1 \neq x_2 \\ \mathbb{1}\{q \geq (1-q)(2-q)\} & y_1' = y_2' = 0, x_1 = x_2 \end{cases} \tag{85}$$

where we slightly abuse notation by omitting the argument x_0 for the estimator $\hat{s}_k^*$ because this latter does not depend on x_0.

For a given input pmf with probability $t := \Pr[X_1 \neq X_2]$, the expected distortion achieved by the optimal estimators in (85) is (see Appendix H-B):

$$\begin{aligned} &\mathsf{E}\left[d_k(S_k, \hat{s}_k^*(X_1, X_2, Y_1', Y_2'))\right] \\ &\quad = \frac{1}{2} tq \left(\min\{q, 1-q\} + (1-q)\right) \\ &\quad + \frac{1}{2}(1-t)\min\{q, (1-q)(2-q)\} \end{aligned} \tag{86}$$

We observe different cases: i) for $q \in [0, 1/2]$, both minima are achieved by q; ii) for $q \in \left(1/2, 2-\sqrt{2}\right]$, the first and second minima are achieved by $1-q$ and q, respectively; iii) for $q \in \left(2-\sqrt{2}, 1\right]$, the first and second minimum are achieved by $(1-q)$ and $(1-q)(2-q)$, respectively. The distortion constraint (69) thus evaluates to:

$$\mathsf{D}_k \geq \begin{cases} q/2 & q \in [0, 1/2] \\ q(1-t(2q-1))/2 & q \in \left(1/2, 2-\sqrt{2}\right] \\ (1-q)(2-q+t(3q-2))/2 & q \in \left(2-\sqrt{2}, 1\right]. \end{cases} \tag{87}$$

We notice that for $q \in [0, 1/2]$, the distortion constraint is independent of t and thus of P_X, and the minimum expected distortions are $\mathsf{D}_{\min,1} = \mathsf{D}_{\min,2} = \frac{1}{2}q$. For $q \in \left(1/2, 2-\sqrt{2}\right]$, the minimum expected distortions are achieved for $t = 1$ and the same holds also for $q \in \left(2-\sqrt{2}, 2/3\right]$. For $q \in [2/3, 1]$, the distortions are minimized for $t = 0$. We thus have $\mathsf{D}_{\min,1} = \mathsf{D}_{\min,2} = \mathsf{D}_{\min}$, where

$$\mathsf{D}_{\min} := \begin{cases} q/2, & q \in [0, 1/2] \\ q(1-q), & q \in [1/2, 2/3] \\ (1-q)(2-q)/2, & q \in [2/3, 1]. \end{cases} \tag{88}$$

We obtain a characterization of the distortion region:

$$\mathcal{D} = \{(\mathsf{D}_1, \mathsf{D}_2) \colon \mathsf{D}_1 \geq \mathsf{D}_{\min},\ \mathsf{D}_2 \geq \mathsf{D}_{\min}\}. \tag{89}$$

The private-messages capacity region is:

$$\mathcal{C} = \{(\mathsf{R}_1, \mathsf{R}_2) \colon \mathsf{R}_1 \leq 1,\ \mathsf{R}_2 \leq 1,$$

$$\text{and} \quad \mathsf{R}_1 + \mathsf{R}_2 \leq 1 + q^2\}. \quad (90)$$

The converse and achievability proofs are provided in Appendices H-C and H-D, respectively.

Reconsider now the case where $q \in [0, 1/2]$. As previously explained, the distortion is independent of the input distribution, and the capacity-distortion region $\mathcal{CD}$ degenerates to the product of the capacity and distortion regions:

Corollary 6. *[No Rate-Distortion Tradeoff] For above state-dependent Dueck example with $q \in [0, 1/2]$:*

$$\mathcal{CD} = \mathcal{C} \times \mathcal{D}. \quad (91)$$

For the general case, we only have bounds on the capacity-distortion region $\mathcal{CD}$. We first present our outer bound, which is based on Theorem 3 and proved in Appendix H-C.

Corollary 7 (Outer Bound)**.** *The capacity-distortion region $\mathcal{CD}$ (without common message) of Dueck's state-dependent BC is included in the set of tuples* $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *that for some choice of the parameters $t \in [0, 1]$ satisfy the rate-constraints*

$$\mathsf{R}_k \leq 1, \qquad k = 1, 2, \quad (92)$$

$$\mathsf{R}_1 + \mathsf{R}_2 \leq 1 + q^2 H_\text{b}(t) \quad (93)$$

and the distortion constraints in (87)*.*

The inner bound is based on Proposition 1, see Appendix H-D. Together with the outer bound in Corollary 7 it characterizes both the distortion region $\mathcal{D}$ and the capacity region $\mathcal{C}$ in (89) and (90).

Corollary 8 (Inner bound)**.** *The capacity-distortion region $\mathcal{CD}$ of the state-dependent Dueck BC includes all rate-distortion tuples* $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *that for some choice of $t \in [0, 1]$ satisfy* (87) *and*

$$\mathsf{R}_k \leq 1, \quad k = 1, 2, \quad (94)$$

$$\mathsf{R}_1 + \mathsf{R}_2 \leq 1 + qH_\text{b}(t) - q(1-q), \quad (95)$$

as well as the convex hull of all these tuples.

Fig. 7 shows our outer and inner bounds in Corollaries 7 and 8 for $q = 3/4$, where in the inner bound we consider the convex hull operation through convex combinations between values of $t > 0$ and $t = 0$. The figure also shows the performances of the basic and improved TS baseline schemes, whose modes we explain next. (Recall that the basic TS scheme time-shares the sensing mode without communication and the communication mode without sensing, and the improved TS scheme time-shares the sensing mode with communication and the communication mode with sensing.)

Sensing mode with and without communication:

In the sensing mode with communication, one can choose an arbitrary pmf for X_0, e.g., X_0 Bernoulli-$1/2$ because this input does not affect the sensing. From (88), the minimum distortions of $\mathsf{D}_{\min,1} = \mathsf{D}_{\min,2} = 5/32$ are achieved by setting $X_1 = X_2$ with probability 1. For $X_1 = X_2$ the sum-rate cannot exceed $\mathsf{R}_1 + \mathsf{R}_2 \leq 1$, because $I(X_0, X_1, X_2; Y_1, Y_2) = I(X_0, X_2; Y_1, Y_2) \leq H(X_0) + I(X_2; Y_1', Y_2' | X_0) \leq 1$ as Y_1' and Y_2' are corrupted by the Bernoulli-$1/2$ noise N. On the other hand, any rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ of sum-rate $\mathsf{R}_1 + \mathsf{R}_2 = 1$ is trivially achievable by communicating only with the noiseless X_0-input.

We conclude that the sensing mode with communication achieves the rate-distortion tuple $(\mathsf{R}_1, \mathsf{R}_1, \mathsf{D}_1, \mathsf{D}_2)$ satisfying

$$\mathsf{R}_1 + \mathsf{R}_2 \leq 1 \quad \text{and} \quad \mathsf{D}_k \geq 5/32, \quad k = 1, 2. \quad (96)$$

If the transmitter cannot perform communication and sensing tasks simultaneously, the same minimum distortions are achieved but the rates are trivially zero.

$$\mathsf{R}_1 + \mathsf{R}_2 = 0 \quad \text{and} \quad \mathsf{D}_k \geq 5/32, \quad k = 1, 2. \quad (97)$$

Communication mode with and without sensing:

The optimal pmf P_X achieving the capacity region in (90) corresponds to i.i.d. Bernoulli-$1/2$ distributed X_0, X_1, X_2 (Appendix H-D). The corresponding sum rate is $\mathsf{R}_1 + \mathsf{R}_2 = 1 + q^2 = 25/16$. The minimum achievable distortions are thus obtained from (87) by setting $t = \Pr[X_1 \neq X_2] = 1/2$, i.e., $\mathsf{D}_{\max,1} = \mathsf{D}_{\max,2} = 11/64$. The best constant estimator is $\hat{S}_1 = \hat{S}_2 = 1$ because $3/4 = P_{S_k}(1) > P_{S_k}(0) = 1/4$, which achieves distortions $\mathsf{D}_{\text{trivial},1} = \mathsf{D}_{\text{trivial},2} = 1/4$. We can conclude that the communication mode with sensing achieves all rate-distortion tuples $(\mathsf{R}_1, \mathsf{R}_1, \mathsf{D}_1, \mathsf{D}_2)$ satisfying

$$\mathsf{R}_1 + \mathsf{R}_2 \leq 25/16, \\ \mathsf{R}_k \leq 1 \quad \text{and} \quad \mathsf{D}_k \geq 11/64 \quad k = 1, 2 \quad (98)$$

and the communication mode without sensing achieves all rate-distortion tuples $(\mathsf{R}_1, \mathsf{R}_1, \mathsf{D}_1, \mathsf{D}_2)$ satisfying

$$\mathsf{R}_1 + \mathsf{R}_2 \leq 25/16 \\ \mathsf{R}_k \leq 1 \quad \text{and} \quad \mathsf{D}_k \geq 1/4, \quad k = 1, 2. \quad (99)$$

IV. Conclusion

Motivated by the paradigm of integrated sensing and communication systems, we studied joint sensing and communication in memoryless state-dependent channels. We fully characterized the capacity-distortion tradeoff

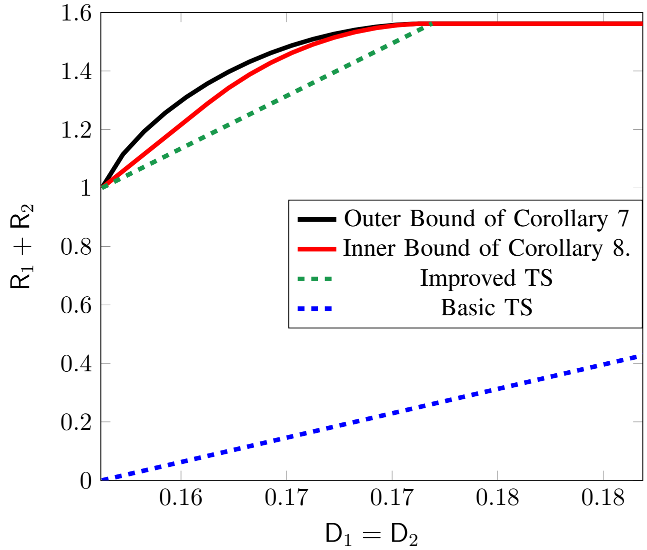

Fig. 7. Sum-rate $R_1 + R_2$ vs. symmetric distortion $D_1 = D_2$ for the state-dependent Dueck BC with $q = 3/4$.

for the single-user channels as well as physically-degraded broadcast channels. For general broadcast channels, we presented inner and outer bounds on the capacity-distortion region. Through a number of illustrative examples, we demonstrated that the optimal co-design scheme offers non-negligible gain compared to the basic time-sharing scheme that performs either sensing or communication, as well as compared to the improved time-sharing scheme that integrates both tasks into a single system but chooses the common waveform to *prioritize* one of the tasks. Interestingly, there are ideal situations where the capacity is achieved without compromising the sensing performance. Our results also showed that for the single-transmitter systems studied in this paper the optimal sensing depends only on the employed waveform, but not on the underlying coding scheme. This holds also for broadcast channels where the two tasks are not only connected through the employed waveform but also through the generalized feedback, which in this case should be exploited to improve the set of achievable rates. Notice that the situation is different in multi-transmitter situations [19], such as multiple-access channels, where coding can be used to improve the sensing performance a the multiple transmitters (by conveying information from one transmitter to the other through the generalized feedback links) and thus the code construction used for data communication should be adapted to integrate also coding for sensing.

An interesting line of future research is the characterization of the capacity-distortion tradeoff for channels with memory. In this case, feedback increases capacity even on the point-to-point channel. For channels with memory, obtaining good state estimation (sensing) and communication performances seem less contradicting goals, because a good state estimation is also useful to improve communication.

Acknowledgements

The authors would like to thank Gerhard Kramer for his support on an early version of this paper.

Appendix A
Proof of Lemma 1

Recall that $\hat{S}^n = h(X^n, Z^n)$, and write for each $i = 1, \cdots, n$:

$$\begin{aligned}
&\mathbb{E}\left[d(S_i, \hat{S}_i)\right] = \\
&\mathbb{E}_{X^n,Z^n}\left[\mathbb{E}[d(S_i, \hat{S}_i)|X^n, Z^n]\right] \\
&\overset{(a)}{=} \sum_{x^n,z^n} P_{X^nZ^n}(x^n, z^n) \\
&\qquad \sum_{\hat{s}\in\mathcal{S}} P_{\hat{S}_i|X^nZ^n}(\hat{s}|x^n, z^n) \\
&\qquad\qquad \sum_{s} P_{S_i|X_iZ_i}(s|x_i, z_i) d(s, \hat{s}) \\
&\geq \sum_{x^n,z^n} P_{X^nZ^n}(x^n, z^n) \\
&\qquad\qquad \min_{\hat{s}\in\mathcal{S}} \sum_{s} P_{S_i|X_iZ_i}(s|x_i, z_i) d(s, \hat{s}) \\
&= \mathbb{E}[d(S_i, \hat{s}^*(X_i, Z_i))], \qquad (100)
\end{aligned}$$

where (a) holds by the Markov chain

$$\left(X^{i-1}, X_{i+1}^n, Z^{i-1}, Z_{i+1}^n, \hat{S}_i\right) -\!\!\circ\!\!- (X_i, Z_i) -\!\!\circ\!\!- S_i. \qquad (101)$$

Summing over all $i = 1, \ldots, n$, we thus obtain:

$$\Delta^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[d(S_i, \hat{S}_i)\right] \qquad (102)$$

$$\geq \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(S_i, \hat{s}^*(X_i, Z_i))], \qquad (103)$$

which yields the desired conclusion.

Appendix B
Proof of Theorem 1

1) Converse: Fix a sequence (in n) of $(2^{n\mathsf{R}}, n)$ codes such that Limits (4) hold. By Fano's inequality there exists a sequence $\epsilon_n \to 0$ as $n \to \infty$ so that:

$$\begin{aligned}
n\mathsf{R} &\leq I(W; Y^n, S^n) + n\epsilon_n \\
&= I(W; Y^n \mid S^n) + n\epsilon_n
\end{aligned}$$

$$= \sum_{i=1}^{n} H(Y_i \mid Y^{i-1}, S^n) - H(Y_i \mid W, Y^{i-1}, S^n) + n\epsilon_n$$
$$\overset{(a)}{\leq} \sum_{i=1}^{n} H(Y_i \mid S_i) - H(Y_i \mid X_i, Y^{i-1}, W, S^n) + n\epsilon_n$$
$$\overset{(b)}{=} \sum_{i=1}^{n} H(Y_i \mid S_i) - H(Y_i \mid X_i, S_i) + n\epsilon_n$$
$$= \sum_{i=1}^{n} I(X_i; Y_i \mid S_i) + n\epsilon_n \tag{104}$$

where (a) holds because conditioning can only reduce entropy; and (b) holds because $(W, Y^{i-1}, S^{i-1}, S_{i+1}^n) - (S_i, X_i) - Y_i$ form a Markov chain. We continue as:

$$\begin{aligned}
\mathsf{R} &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i \mid S_i) + \epsilon_n \\
&\overset{(c)}{\leq} \frac{1}{n}\sum_{i=1}^{n} \mathsf{C}_{\text{inf}}\left(\sum_x P_{X_i}(x)c(x), \sum_x P_{X_i}(x)b(x)\right) + \epsilon_n \\
&\overset{(d)}{\leq} \mathsf{C}_{\text{inf}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_x P_{X_i}(x)c(x), \frac{1}{n}\sum_{i=1}^{n}\sum_x P_{X_i}(x)b(x)\right) + \epsilon_n \\
&\overset{(e)}{\leq} \mathsf{C}_{\text{inf}}(\mathsf{D}, \mathsf{B})
\end{aligned} \tag{105}$$

where (c) holds by the definition of $\mathsf{C}_{\text{inf}}(\mathsf{D}, \mathsf{B})$, and (d) and (e) hold by Lemma 2.

2) Achievability: Fix $P_X(\cdot)$ and functions $\hat{h}(x,z)$ that achieve $C(\mathsf{D}/(1+\epsilon), \mathsf{B})$, where D is the desired distortion and B is the target cost, for a small positive number $\epsilon > 0$. We define the joint pmf $P_{SXY} := P_S P_X P_{Y|SX}$.

a) Codebook generation: Generate $2^{n\mathsf{R}}$ sequences $\{x^n(w)\}_{w=1}^{2^{n\mathsf{R}}}$ by randomly and independently drawing each entry according to P_X. This defines the codebook $\mathcal{C} = \{x^n(w)\}_{w=1}^{2^{n\mathsf{R}}}$, which is revealed to the encoder and the decoder.

b) Encoding: To send a message $w \in \mathcal{W}$, the encoder transmits $x^n(w)$.

c) Decoding: Upon observing outputs $Y^n = y^n$ and state sequence $S^n = s^n$, the decoder looks for an index $\hat{w}$ such that

$$(s^n, x^n(\hat{w}), y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{SXY}). \tag{106}$$

If exactly one such index exists, it declares $\hat{W} = \hat{w}$. Otherwise, it declares an error.

d) Estimation: Assuming that it sent the input sequence $X^n = x^n$ and observed the feedback signal $Z^n = z^n$, the encoder computes the reconstruction sequence as:

$$\hat{S}^n = (\hat{s}^*(x_1, z_1), \hat{s}^*(x_2, z_2), \ldots, \hat{s}^*(x_n, z_n)). \tag{107}$$

e) Analysis: We start by analyzing the probability of error and the distortion averaged over the random code construction. Given the symmetry of the code construction, we can condition on the event $W = 1$.

We then notice that the decoder makes an error, i.e., declares nothing or $\hat{W} \neq 1$ if, and only if, one or both of the following events occur:

$$\mathcal{E}_1 = \left\{(S^n, X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}(P_{SXY})\right\} \tag{108}$$
$$\mathcal{E}_2 = \left\{(S^n, X^n(w'), Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{SXY}) \text{ for some } w' \neq 1\right\}. \tag{109}$$

where we defined $P_{SXY} := P_S P_X P_{Y|SX}$. Thus, by the union bound:

$$P_e^{(n)} = P(\mathcal{E}_1 \cup \mathcal{E}_2) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2). \tag{110}$$

The first term goes to zero as $n \to \infty$ by the weak law of large numbers. The second term also tends to zero as $n \to \infty$ if $\mathsf{R} < I(X;Y|S)$ by the independence of the codewords and the packing lemma [23, Lemma 3.1]. Therefore, $P_e^{(n)}$ tends to zero as $n \to \infty$ whenever $R < I(X;Y|S)$.

The expected distortion (averaged over the random codebook, state and channel noise) can be upper bounded as

$$\Delta^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left[d(S_i, \hat{S}_i)\right] \tag{111}$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left[d(S_i, \hat{S}_i) \middle| \hat{W} \neq 1\right] \Pr(\hat{W} \neq 1) + \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left[d(S_i, \hat{S}_i) \middle| \hat{W} = 1\right] \Pr(\hat{W} = 1) \tag{112}$$
$$\leq \mathsf{D}_{\max} P_e + \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left[d(S_i, \hat{S}_i) \middle| \hat{W} = 1\right] \cdot (1 - P_e). \tag{113}$$

In the event of correct decoding, i.e., $\hat{W} = 1$,

$$(S^n, X^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_S P_X P_{Y|SX}), \tag{114}$$

and since $\hat{S}_i = \hat{s}^*(X_i, Z_i)$, also

$$(S^n, X^n(1), \hat{S}^n) \in \mathcal{T}_\epsilon^{(n)}\left(P_{SX\hat{S}}\right), \tag{115}$$

where $P_{SX\hat{S}}$ denotes the joint marginal pmf of $P_{SXZ\hat{S}}(s,x,z,\hat{s}) := P_S(s)P_X(x)P_{Z|SX}(z|s,x)\mathbb{1}\{\hat{s} = \hat{s}^*(x,z)\}$. Then,

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left[d(S_i,\hat{S}_i)|\hat{W}=1\right] \leq (1+\epsilon)\mathsf{E}\left[d(S,\hat{S})\right], \tag{116}$$

for $(S,\hat{S})$ following the marginal of the pmf $P_{SXZ\hat{S}}$ defined above. Assuming that $\mathsf{R} < I(X;Y|S)$, and thus $P_e \to 0$ as $n \to \infty$, we obtain from (113) and (116):

$$\overline{\lim_{n\to\infty}} \Delta^{(n)} = (1+\epsilon)\mathsf{E}\left[d(S,\hat{S})\right]. \tag{117}$$

Taking finally $\epsilon \downarrow 0$, we can conclude that the error probability and distortion constraint (4a), (4b) hold (averaged over the random code constructions, the random states, and the noise in the channel) whenever

$$\mathsf{R} < I(X;Y \mid S), \tag{118}$$

$$\mathsf{E}\left[d(S,\hat{S})\right] < \mathsf{D}. \tag{119}$$

Notice that the cost constraint (4c) is fullfilled by construction. By standard arguments it can then be shown that there must exist at least one sequence of deterministic code books $\mathcal{C}_n$ so that constraints (4) hold.

Appendix C
Blahut-Arimoto Type Algorithm to evaluate Theorem 1

Through simple time-sharing arguments, it can be shown that for given feasible B, the set of achievable (R,D) pairs over the single-receiver channel is convex. Its boundary is thus characterized by solving the following parameterized optimization problem for each $\mu \geq 0$:

$$L_\mu(\mathsf{B}) := \max_{P_X\in\mathcal{P}_\mathcal{B}} \Bigg[\mathcal{I}(P_X, P_{Y|XS} \mid P_S) - \mu\sum_{x\in\mathcal{X}} P_X(x)c(x)\Bigg]. \tag{120}$$

Notice that the conditional mutual information functional can explicitly be written as:

$$\mathcal{I}(P_X, P_{Y|XS} \mid P_S) = \sum_{x\in\mathcal{X}}\sum_{s\in\mathcal{S}}\sum_{y\in\mathcal{Y}} P_S(s)P_X(x)P_{Y|XS}(y|xs) \log\frac{P_{Y|XS}(y|xs)}{P_{Y|S}(y|s)}. \tag{121}$$

for the state pmf P_S and the SDMB transition law $P_{YZ|XS}$.

For $\mu = 0$, the optimization in (120) yields the capacity of the SDMC under the input cost constraint (disgarding the distortion constraint), while for $\mu \to \infty$, it yields the minimum possible distortion subject to the same input cost constraint. We remark that the parameterized optimization problem above differs from the standard Blahut-Arimoto algorithm with cost constraints [27, Section IV] only in that 1) the objective function (120) includes an additional penalty term $-\mu\sum_{x\in\mathcal{X}} P_X(x)c(x)$ and 2) the mutual information functional is $I(X;Y \mid S)$ instead of $I(X;Y)$, which reflects the state-dependent channel and the state knowledge at the receiver. Since the penalty term is additive and linear in P_X, all concavity properties desired for a Blahut-Arimoto type algorithm remain valid. The following Theorem 4 can then be proved by standard alternating optimization techniques, in analogy to the proof of the Blahut-Arimoto algorithm [26], [27].

For any conditional pmf $Q_{X|YS}$ on X given (Y,S), define the function

$$J_\mu(P_X, P_{Y|XS}, P_S, Q_{X|YS}) := \sum_{x\in\mathcal{X}}\sum_{s\in\mathcal{S}}\sum_{y\in\mathcal{Y}} P_X(x)P_S(s)P_{Y|XS}(y|x,s) \cdot\log\frac{Q_{X|YS}(x|y,s)}{P_X(x)} - \mu\sum_{x\in\mathcal{X}} P_X(x)c(x). \tag{122}$$

Theorem 4. *Let the state pmf P_S and the SDMC transition law $P_{YZ|XS}$ be given. The following statements hold:*
a) For any $\mu, \mathsf{B} \geq 0$:

$$L_\mu(\mathsf{B}) = \max_{P_X\in\mathcal{P}(\mathsf{B})}\max_{Q_{X|YS}} J_\mu(P_X, P_S, P_{YZ|XS}, Q_{X|YS}). \tag{123}$$

b) Fix $P_X \in \mathcal{P}(\mathsf{B})$. Then, $J_\mu(P_X, P_S, P_{YZ|XS}, Q_{X|YS})$ is maximized by choosing $Q_{X|YS}$ as

$$Q^\star_{X|YS}(x|y,s) = \frac{P_X(x)P_{Y|XS}(y|xs)}{\sum_{x'} P_X(x')P_{Y|XS}(y|x's)}, \quad (x,y,s) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{S}, \tag{124}$$

c) Fix $Q_{X|YS}$. Then, $J_\mu(P_X, P_S, P_{YZ|XS}, Q_{X|YS})$ is maximized by choosing $P_X \in \mathcal{P}(B)$ as

$$P^\star_X(x) = \frac{2^{g(x)}}{\sum_{x'} 2^{g(x')}}, \quad x \in \mathcal{X}, \tag{125}$$

where

$$g(x) = \sum_s \sum_y P_S(s) P_{Y|XS}(y|xs) \log Q_{X|YS}(x|ys) - \lambda b(x) - \mu c(x) \quad (126)$$

and $\lambda \geq 0$ *is chosen so that* $\sum_{x\in\mathcal{X}} P_X^\star(x) b(x) = B$ *when evaluated for* $P_X^\star$ *in* (125)*, or if no such* λ *exists, then it is set to* $\lambda = 0$. □

Proof: We give the proofs for the three results $a)$–$c)$

$a)$ Fix pmfs $P_S, P_X, P_{Y|XS}$ and define $P_{SXY}(s,x,y) := P_S(s)P_X(x)P_{Y|XS}(y|x,s)$. Notice that

$$\begin{aligned}
&J_\mu(P_X, P_{Y|XS}, P_S, Q_{X|YS}) \\
&= \sum_{x\in\mathcal{X}} \sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_{SXY}(s,x,y) \cdot \log \frac{Q_{X|YS}(x|ys)}{P_X(x)} \\
&\quad -\mu \sum_{x\in\mathcal{X}} P_X(x)c(x) &(127)\\
&= \sum_{x\in\mathcal{X}} \sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_{SXY}(s,x,y) \cdot \log \frac{Q_{X|YS}(x|ys)P_{SY}(s,y)}{P_X(x)P_{SY}(s,y)} \\
&\quad -\mu \sum_{x\in\mathcal{X}} P_X(x)c(x) &(128)\\
&= \sum_{x\in\mathcal{X}} \sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_{SXY}(s,x,y) \left[\log \frac{Q_{X|YS}(x|ys)P_{SY}(s,y)}{P_{SXY}(s,x,y)} + \log \frac{P_{SXY}(s,x,y)}{P_X(x)P_{SY}(s,y)} \right] \\
&\quad -\mu \sum_{x\in\mathcal{X}} P_X(x)c(x) &(129)\\
&= -\underbrace{D(P_{SXY} \| Q_{X|YS}P_{SY})}_{\leq 0} + \sum_{x\in\mathcal{X}} \sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_{SXY}(s,x,y) \cdot \log \frac{P_{SXY}(s,x,y)}{P_X(x)P_{SY}(s,y)} \\
&\quad -\mu \sum_{x\in\mathcal{X}} P_X(x)c(x) &(130)\\
&\leq \sum_{x\in\mathcal{X}} \sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_{SXY}(s,x,y) \cdot \log \frac{P_{SXY}(s,x,y)}{P_X(x)P_{SY}(s,y)} \\
&\quad -\mu \sum_{x\in\mathcal{X}} P_X(x)c(x) &(131)\\
&= \sum_{x\in\mathcal{X}} \sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_{SXY}(s,x,y) \cdot \log \frac{P_{Y|XS}(y|x,s)}{P_{Y|S}(y|s)} \\
&\quad -\mu \sum_{x\in\mathcal{X}} P_X(x)c(x) &(132)\\
&= \mathcal{I}(P_X, P_{Y|XS} \mid P_S) - \mu \sum_{x\in\mathcal{X}} P_X(x)c(x), &(133)
\end{aligned}$$

where $D(\cdot\|\cdot)$ denotes the Kullback-Leibler Divergence [32]. Above inequality holds with equality when $Q_{X|YS} = P_{X|YS}$, where the latter stands for the conditional marginal pmf of P_{SXY}. Therefore, $\max_{Q_{X|YS}} J_\mu(P_X, P_{Y|XS}, P_S, Q_{X|YS})$ equals the right-hand side of (133), which directly implies (123).

$b)$ For fixed P_X, according to (127)–(133), $J_\mu(P_X, P_{Y|XS}, P_S, Q_{X|YS})$ is maximized by the choice

$$\begin{aligned}
Q^\star_{X|YS}(x|y,s) &= \frac{P_{SXY}(s,x,y)}{P_{SY}(s,y)} \\
&= \frac{P_S(s)P_X(x)P_{Y|XS}(y|x,s)}{\sum_{x'} P_S(s)P_X(x')P_{Y|XS}(y|x',s)}. \quad (134)
\end{aligned}$$

$c)$ The function $J_\mu(P_X, P_{Y|XS}, P_S, Q_{X|YS})$ is concave in P_X and we can thus use the KKT conditions to find the maximum value $\max_{P_X} J_\mu(P_X, P_{Y|XS}, P_S, Q_{X|YS})$ over all pmfs P_X satisfying $\sum_{x\in\mathcal{X}} P_X(x)b(x) \leq B$. In this case, the KKT conditions are summarized by the two constraints

$$\begin{aligned}
&\sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_S(s)P_{Y|XS}(y|x,s) \log \frac{Q_{X|YS}(x|y,s)}{P_X(x)} \\
&- \sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_S(s)P_{Y|XS}(y|x,s) \ln(2)^{-1} \\
&-\mu c(x) - \lambda b(x) \\
&= \xi, &(135)
\end{aligned}$$

and

$$\sum_{x\in\mathcal{X}} P_X(x)b(x) \leq B, \quad (136)$$

and $\lambda = 0$ if above inequality is strict, and the Lagrange multiplier ξ ensures that the pmf P_X sums to 1. Since $\sum_{s\in\mathcal{S}} \sum_{y\in\mathcal{Y}} P_S(s)P_{Y|XS}(y|x,s) = 1$, Equation (135) is equivalent to

$$\log P_X(x)$$

$$= \sum_{s\in\mathcal{S}}\sum_{y\in\mathcal{Y}} P_S(s)P_{Y|XS}(y|x,s)\log Q_{X|YS}(x|y,s) - \mu c(x) - \lambda b(x) - \xi - \ln(2)^{-1}, \quad (137)$$

and thus to

$$P_X(x) = 2^{\sum_{s\in\mathcal{S}}\sum_{y\in\mathcal{Y}} P_S(s)P_{Y|XS}(y|x,s)\log Q_{X|YS}(x|y,s) - \mu c(x) - \lambda b(x)} \cdot 2^{-\xi - \ln(2)^{-1}}. \quad (138)$$

Choosing finally the Lagrange multipliers ξ and λ so that the pmf P_X sums to 1 and the cost constraint $\sum_{x\in\mathcal{X}} P_X(x)b(x) \le B$ holds with equality, we obtain the result in (125). If no such λ exists, then we set $\lambda = 0$.

■

Each of the two maximizations in (123) is a convex optimization problem. The solution $L_\mu(B)$ can thus be obtained by an alternating maximization procedure. For our problem at hand, this alternating maximization procedure is described in Algorithm 1. The algorithm produces an optimal convergent input distribution $P_{X,\mu}^{\infty}$, which can be used to compute a pair of capacity-distortion values $(\mathsf{C}_\mu(\mathsf{B}), \mathsf{D}_\mu(\mathsf{B}))$ on the boundary of the capacity-distortion tradeoff for given input cost B:

$$\mathsf{C}_\mu(\mathsf{B}) = \mathcal{I}\left(P_{X,\mu}^{(\infty)}, P_{Y|XS}\middle| P_S\right) \quad (139a)$$

$$\mathsf{D}_\mu(\mathsf{B}) = \sum_x c(x)P_{X,\mu}^{(\infty)}(x). \quad (139b)$$

Varying μ, the entire capacity-distortion tradeoff is obtained for fixed input cost B. Moreover, by varying the input cost B, the whole boundary of the achievable capacity-distortion-cost tradeoff region is obtained.

Algorithm 1 Blahut-Arimoto Type Algorithm for SDMCs

Fix $\mu \ge 0$.

procedure TRADEOFF $(\mathsf{C}_\mu(\mathsf{B}), \mathsf{D}_\mu(\mathsf{B}))$

Initialize $P_X^{(0)}(x) = \frac{1}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$.

for $k = 1, 2, 3, \ldots$ **do**

$$Q_{X|YS}^{(k)}(x|y,s) = \frac{P_X^{(k-1)}(x)P_{Y|XS}(y|x,s)}{\sum_{x'} P_X^{(k-1)}(x')P_{Y|XS}(y|x',s)}. \quad (140)$$

Choose $\lambda^{(0)} > 0$.

for $\ell = 1, 2, \ldots$ **do**

Compute $p^{(\ell)}(x) = \frac{e^{g^{(\ell)}(x)}}{\sum_{x'} e^{g^{(\ell)}(x')}}$ with

$$g^{(\ell)}(x) = \sum_{s,y} P_S(s)P_{Y|XS}(y|x,s)\log Q_{X|YS}^{(k)}(x|y,s) - \lambda^{(\ell-1)}b(x) - \mu \sum_{(x,s,z)\in\mathcal{X}\times\mathcal{S}\times\mathcal{Z}} P_X(x)P_S(s)P_{Z|XS}(z|x,s) \, d(s, s^*(x,z)) \quad (141)$$

Update dual variables:

$$\lambda^{(\ell)} = \left[\lambda^{(\ell-1)} + \alpha_\ell\left(\sum_x b(x)p^{(\ell)}(x) - B\right)\right]_+ \quad (142)$$

▷ where α_ℓ is the gradient adaptation step

Let $P_X^{(k)}(x) = \lim_{\ell\to\infty} p^{(\ell)}(x)$.

APPENDIX D
PROOF OF COROLLARY 1

It suffices to show that under the described conditions, the distortion constraint (4b) does not depend on P_X. To this end, we define $T = \psi(X,Z)$ and rewrite the expected distortion as:

$$\mathbb{E}[d(S,\hat{S})] = \sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}} P_{XZ}(x,z)\sum_{s\in\mathcal{S}} P_{S|XZ}(s|x,z) \cdot d(s, \hat{s}^*(x,z)) \quad (143)$$

$$\stackrel{(a)}{=} \sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}} P_{XZ}(x,z) \cdot \min_{s'\in\hat{\mathcal{S}}} \sum_{(s,t)\in\mathcal{S}\times\mathcal{T}} P_{ST|XZ}(s,t|x,z)d(s,s') \quad (144)$$

$$\stackrel{(b)}{=} \sum_{(x,z,t)\in\mathcal{X}\times\mathcal{Z}\times\mathcal{T}} P_{XZ}(x,z)\mathbb{1}\{t = \psi(x,z)\} \cdot \min_{s'\in\hat{\mathcal{S}}}\sum_{s\in\mathcal{S}} P_{S|T}(s|t)d(s,s') \quad (145)$$

$$= \sum_{t\in\mathcal{T}} P_T(t)\min_{s'\in\hat{\mathcal{S}}}\sum_{s\in\mathcal{S}} P_{S|T}(s|t)d(s,s') \quad (146)$$

where (a) holds by the definition of $\hat{s}^*(x,z)$ and the law of total probability; and (b) by the Markov chain S —◦— T —◦— (X,Z), see (17), and because T is a function of X,Z. The independence of the pair (T,S) with X from (16), together with the above expression implies that the expected distortion does not depend on the choice of the input distribution P_X. Hence, we can conclude that for any given $\mathsf{B} \ge 0$, the rate-distortion tradeoff function $\mathsf{C}(\mathsf{D},\mathsf{B})$ is constant over all $\mathsf{D} \ge \mathsf{D}_{\min}$ and coincides with the capacity of the SDMC $\mathsf{C}_{\text{NoEst}}(\mathsf{B})$.

Appendix E
Proof of Remark 1

1) Converse: Fix a sequence (in n) of $(2^{n\mathsf{R}}, n)$ codes such that Limits (4) hold. By Fano's inequality there exists a sequence $\epsilon_n \to 0$ as $n \to \infty$ so that:

$$\begin{aligned} n\mathsf{R} &\le I(W; Y^n, S_R^n) + n\epsilon_n \\ &= I(W; Y^n \mid S_R^n) + n\epsilon_n \\ &= \sum_{i=1}^n H(Y_i \mid Y^{i-1}, S_R^n) \\ &\qquad - H(Y_i \mid W, Y^{i-1}, S_R^n) + n\epsilon_n \\ &\overset{(a)}{\le} \sum_{i=1}^n H(Y_i \mid S_{R,i}) \\ &\qquad - H(Y_i \mid X_i, Y^{i-1}, W, S_R^n) + n\epsilon_n \\ &\overset{(b)}{=} \sum_{i=1}^n H(Y_i \mid S_{R,i}) \\ &\qquad - H(Y_i \mid X_i, S_{R,i}) + n\epsilon_n \\ &= \sum_{i=1}^n I(X_i; Y_i \mid S_{R,i}) + n\epsilon_n \end{aligned} \tag{147}$$

where (a) holds because conditioning can only reduce entropy; and (b) holds because $(W, Y^{i-1}, S_R^{i-1}, S_{R,i+1}^n) - (S_{R,i}, X_i) - Y_i$ form a Markov chain.

Define

$$\mathsf{C}_{\text{inf}}^{\text{imp}}(\mathsf{D}, \mathsf{B}) := \max_{P_X \in \mathcal{P}_\mathsf{D} \cap \mathcal{P}_\mathsf{B}} I(X; Y \mid S_R). \tag{148}$$

Then, we have

$$\begin{aligned} \mathsf{R} &\le \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i \mid S_{R,i}) + \epsilon_n \\ &\overset{(c)}{\le} \frac{1}{n} \sum_{i=1}^n \mathsf{C}_{\text{inf}}^{\text{Imp}} \Bigg(\sum_x P_{X_i}(x) c(x), \\ &\qquad \sum_x P_{X_i}(x) b(x) \Bigg) + \epsilon_n \\ &\overset{(d)}{\le} \mathsf{C}_{\text{inf}}^{\text{Imp}} \Bigg(\frac{1}{n} \sum_{i=1}^n \sum_x P_{X_i}(x) c(x), \\ &\qquad \frac{1}{n} \sum_{i=1}^n \sum_x P_{X_i}(x) b(x) \Bigg) + \epsilon_n \\ &\overset{(e)}{\le} \mathsf{C}_{\text{inf}}^{\text{Imp}}(\mathsf{D}, \mathsf{B}) \end{aligned} \tag{149}$$

where (c) holds by the definition of $\mathsf{C}_{\text{inf}}^{\text{Imp}}(\mathsf{D}, \mathsf{B})$ in (148), and (d) and (e) hold by similar monotonicity and concavity properties as stated in Lemma 2.

2) Achievability: Fix $P_X(\cdot)$ and a function $\hat{h}(x, z)$ that achieve $C(\mathsf{D}/(1+\epsilon), \mathsf{B})$, where D is the desired distortion and B is the target cost, for a small positive number $\epsilon > 0$. We define the joint pmf $P_{SS_RXY} := P_{SS_R} P_X P_{Y|SS_RX}$. Codebook generation, encoding, and estimation are as described in the proof of Theorem 1; the only difference is in the decoding at the receiver, where the state S^n has to be replaced by S_R. In more details:

a) Decoding: Upon observing outputs $Y^n = y^n$ and state sequence $S_R^n = s_R^n$, the decoder looks for an index $\hat{w}$ such that

$$(s_R^n, x^n(\hat{w}), y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{S_RXY}) \tag{150}$$

where $P_{S_RXY} = \sum_\mathcal{S} P_{SS_RXY}$. If exactly one such index exists, it declares $\hat{W} = \hat{w}$. Otherwise, it declares an error.

b) Analysis: We start by analyzing the probability of error and the distortion averaged over the random code construction. Given the symmetry of the code construction, we can condition on the event $W = 1$. We then notice that the decoder makes an error, i.e., declares nothing or $\hat{W} \neq 1$ if, and only if, one or both of the following events occur:

$$\mathcal{E}_1 = \left\{ (S_R^n, X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}(P_{XS_RY}) \right\} \tag{151}$$

or

$$\mathcal{E}_2 = \left\{ (S_R^n, X^n(w'), Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XS_RY}) \text{ for some } w' \neq 1 \right\}. \tag{152}$$

Thus, by the union bound:

$$P_e^{(n)} = P(\mathcal{E}_1 \cup \mathcal{E}_2) \le P(\mathcal{E}_1) + P(\mathcal{E}_2), \tag{153}$$

where we consider the average probability of error not only over the random channel noise and states but also over the random codeconstruction. The first term goes to zero as $n \to \infty$ by the weak law of large numbers. By the independence of the codewords and the packing lemma [23, Lemma 3.1], the second term also tends to zero as $n \to \infty$

$$R < I(X; Y | S_R). \tag{154}$$

Following similar steps as in the analysis in Appendix B, and using the fact that by the weak law of large numbers with probability tending to 1 as $n \to \infty$:

$$(S^n, S_R^n, X^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_X P_S P_{S_R} P_{Y|SS_RX}), \tag{155}$$

it can be shown that

$$\overline{\lim_{n\to\infty}} \Delta^{(n)} = (1+\epsilon) \mathsf{E}\left[d(S, \hat{s}^*(X, Z)) \right]. \tag{156}$$

Thus when $\epsilon \downarrow 0$, the distortion constraint (4b) holds (averaged over the random code constructions, the random states, and the noise in the channel) whenever

$$\mathsf{E}[d(S, \hat{s}^*(X, Z))] < \mathsf{D}. \tag{157}$$

Notice that the cost constraint (4c) is fullfilled by construction.

By standard arguments it can then be shown that there must exist at least one sequence of deterministic code books $\mathcal{C}_n$ so that constraints (4) are satisfied under conditions (154) and (157).

Appendix F
Converse Proof of Theorem 2

Fix a sequence (in n) of $(2^{n\mathsf{R}_0}, 2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ codes satisfying (42). Fix a blocklength n and start with Fano's inequality:

$$\begin{aligned}
\mathsf{R}_0 + \mathsf{R}_2 &= \frac{1}{n} H(W_0, W_2) \\
&\leq \frac{1}{n}\sum_{i=1}^{n} I(W_0, W_2; Y_{2i}, S_{2,i} \mid Y_2^{i-1}, S_2^{i-1}) + \epsilon_n \\
&\leq \frac{1}{n}\sum_{i=1}^{n} I(W_0, W_2, Y_2^{i-1}, S_2^{i-1}; Y_{2,i}, S_{2,i}) + \epsilon_n \\
&= I(W_0, W_2, Y_2^{T-1}, S_2^{T-1}; Y_{2,T}, S_{2,T} \mid T) + \epsilon_n \\
&\leq I(W_0, W_2, Y_2^{T-1}, S_2^{T-1}, T; Y_{2,T}, S_{2,T}) + \epsilon_n \\
&\overset{(a)}{=} I(U; Y_2 \mid S_2) + \epsilon_n,
\end{aligned} \tag{158}$$

where T is chosen uniformly over $\{1, \cdots, n\}$ and independent of $X^n, Y_1^n, Y_2^n, W_0, W_1, W_2, S_1^n, S_2^n$; ϵ_n is a sequence that tends to 0 as $n \to \infty$; and $U := (W_0, W_2, Y_2^{T-1}, S_2^{T-1}, T)$, $Y_2 := Y_{2,T}$ and $S_2 := S_{2,T}$. Here (a) holds because $S_2 \sim P_{S_2}$ independent of (U, X), where we define $X := X_T$.

Following similar steps, we obtain:

$$\begin{aligned}
R_1 &= \frac{1}{n} H(W_1 \mid W_0, W_2) \\
&\leq \frac{1}{n} I(W_1; Y_1^n, S_1^n \mid W_0, W_2) + \epsilon_n \\
&\leq \frac{1}{n} I(W_1; Y_1^n, S_1^n, Y_2^n, S_2^n \mid W_0, W_2) + \epsilon_n \\
&= \frac{1}{n}\sum_{i=1}^{n} I(W_1; Y_{1,i}, Y_{2,i}, S_{1,i}, S_{2,i} \\
&\qquad \mid Y_1^{i-1}, Y_2^{i-1}, S_1^{i-1}, S_2^{i-1}, W_0, W_2) + \epsilon_n \\
&\leq \frac{1}{n}\sum_{i=1}^{n} I(X_i, W_1, Y_1^{i-1}, S_1^{i-1}; Y_{1,i}, Y_{2,i}, S_{1,i}, S_{2,i} \\
&\qquad \mid Y_2^{i-1}, S_2^{i-1}, W_0, W_2) + \epsilon_n \\
&\overset{(b)}{\leq} \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_{1,i}, S_{1,i} \\
&\qquad \mid Y_2^{i-1}, S_2^{i-1}, W_0, W_2) + \epsilon_n
\end{aligned} \tag{159}$$

$$\begin{aligned}
&= I(X_T; Y_{1T}, S_{1,T} \mid Y_2^{T-1}, S_2^{T-1}, W_0, W_2, T) + \epsilon_n \\
&\overset{(c)}{=} I(X; Y_1 \mid S_1, U) + \epsilon_n,
\end{aligned} \tag{160}$$

where we defined $Y_1 := Y_{1,T}$ and $S_1 := S_{1,T}$; and where (b) holds by the physically degradedness of the SDMBC which implies the Markov chain $(W_0, W_2, W_1, Y_1^{i-1}, S_1^{i-1}, Y_2^{i-1}, S_2^{i-1}) \to X_i \to (S_{1,i}, Y_{1,i}) \to (S_{2,i}, Y_{2,i})$, and (c) holds because $S_1 \sim P_{S_1}$ independent of (U, X).

Recall that we assume the optimal estimators (43) in Lemma 3. Using the definitions of T, X, S_k above and defining $Z := Z_T$, we can write the average expected distortions as:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d_k(S_{k,i}, \hat{s}_k^*(X_i, Z_i)] = \mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z)]. \tag{161}$$

Combining (158), (160), and (161) and letting $n \to \infty$, we obtain that there exists a limiting pmf P_{UX} such that the tuple $(U, X, S_1, S_2, Y_1, Y_2, Z) \sim P_{UX} P_{S_1 S_2} P_{Y_1 Y_2 Z | S_1 S_2 X}$ satisfies the rate-constraints

$$\mathsf{R}_0 + \mathsf{R}_2 \leq I(U; Y_2 \mid S_2) \tag{162}$$

$$\mathsf{R}_1 \leq I(X; Y_1 \mid S_1, U) \tag{163}$$

and the distortion constraints

$$\mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z)] \leq \mathsf{D}_k, \quad k = 1, 2, \tag{164}$$

This completes the proof.

Appendix G
Proof of Theorem 3

Fix a sequence (in n) of $(2^{n\mathsf{R}_0}, 2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ codes satisfying (42). Fix then a blocklength n and consider an enhanced SDMBC where Receiver 1 observes the pair of states $\tilde{S}_1 = (S_1, S_2)$ and the pair of outputs $\tilde{Y}_1 = (Y_1, Y_2)$. The enhanced SDMBC is clearly physically degraded because for any input pmf P_X the Markov chain

$$X \multimap (\tilde{S}_1, \tilde{Y}_1) \multimap (S_2, Y_2) \tag{165}$$

holds.

Following the steps in the previous Appendix F, we can conclude that

$$\mathsf{R}_0 + \mathsf{R}_2 \leq I(U_2; Y_2 \mid S_2) + \epsilon_n \tag{166}$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq I(X; Y_1, Y_2 \mid S_1, S_2) + \epsilon_n \tag{167}$$

and for $k = 1, 2$

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d_k(S_{k,i}, \hat{s}_k^*(X_i, Z_i)] = \mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z)]. \tag{168}$$

Consider next a reversely enhanced SDMBC where Receiver 1 observes only (Y_1, S_1) but Receiver 2 observes both state sequences $\tilde{S}_2 := (S_1, S_2)$ and both outputs $\tilde{Y}_2 := (Y_1, Y_2)$. Following again the steps in the previous Appendix F, but now with exchanged indices 1 and 2, we obtain:

$$\mathsf{R}_0 + \mathsf{R}_1 \leq I(U_1; Y_1 \mid S_1) + \epsilon_n \quad (169)$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq I(X; Y_1, Y_2 \mid S_1, S_2) + \epsilon_n. \quad (170)$$

Combining all these inequalities and letting first $n \to \infty$ and then $\epsilon_n \downarrow 0$, establishes the desired converse result.

Appendix H
Proofs for Dueck's State-Dependent BC

A. Optimal Estimator of Lemma 3

We first derive the optimal estimator $\hat{s}_k^*(x_1, x_2, y_1', y_2')$ of Lemma 3 for this example.

Case $y_1' = y_2' = 1$**:** In this case, $S_1 = S_2 = 1$ deterministically, and thus

$$\hat{s}_k^*(x_1, x_2, 1, 1) = 1, \quad \forall (x_1, x_2), \qquad k = 1, 2. \quad (171)$$

Case $y_1 = 1'$ and $y_2' = 0$: In this case, $S_1 = 1$ deterministically and

$$\hat{s}_1^*(x_1, x_2, 1, 0) = 1, \quad \forall (x_1, x_2). \quad (172)$$

To derive the optimal estimator for state S_2, we notice that $y_1' = 1$ implies $x_1 \oplus N = 1$, i.e., $N = x_1 \oplus 1$. As a consequence,

$$y_2' = (x_2 \oplus x_1 \oplus 1) S_2. \quad (173)$$

So, for $x_2 = x_1$ we have $y_2' = S_2 = 0$ and the optimal estimator sets

$$\hat{s}_2^*(x_1, x_2, 1, 0) = 0, \quad x_1 = x_2. \quad (174)$$

Instead for $x_2 \neq x_1$, the feedback output $y_2' = 0$, irrespective of the state S_2. The optimal estimator then is the constant estimator

$$\hat{s}_2^*(x_1, x_2, 1, 0) = \operatorname*{argmax}_{\hat{s} \in \{0,1\}} P_S(\hat{s}), \quad x_1 \neq x_2. \quad (175)$$

Case $y_1' = 1, y_2' = 0$: Symmetric to the previous case $y_1' = 0, y_2' = 1$. The optimal estimators are as in (174) and (175), but with exchanged indices 1 and 2.

Case $y_1' = y_2' = 0$: To find the optimal estimators, we calculate the conditional probabilities $P_{S_k|X_1X_2Y_1'Y_2'}(\cdot|x_1, x_2, y_1', y_2')$ for $y_1' = y_2' = 0$.

We again distinguish the two cases $x_1 = x_2$ and $x_1 \neq x_2$ and start by considering $x_1 = x_2$. In this case, $x_1 \oplus N = x_2 \oplus N$, and so if $S_k = 1$ then $y_1' = y_2' = 0$ only if $x_1 \oplus N = x_2 \oplus N = 0$, which happens with probability $1/2$ because N is Bernoulli-$1/2$. By the independence of the states and the inputs for $x_1 = x_2$ and $k = 1, 2$:

$$\begin{aligned} &P_{S_k|X_1X_2Y_1'Y_2'}(1|x_1, x_2, 0, 0) \\ &\quad = \frac{P_{S_k}(1) P_{Y_1'Y_2'|X_1X_2S_k}(0, 0|x_1, x_2, 1)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)} \\ &\quad = \frac{P_S(1) 1/2}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)}. \end{aligned} \quad (176a)$$

Let $\bar{k} := 3 - k$ for $k = 1, 2$. If $S_k = 0$, then $y_1' = y_2' = 0$ happens when either $x_1 \oplus N = x_2 \oplus N = 0$ or when $S_{\bar{k}} = 0$ and $x_1 \oplus N = x_2 \oplus N = 1$. Since these are exclusive events and have total probability of $1/2 + P_S(0)1/2$, we obtain for $x_1 = x_2$ and $k \in \{1, 2\}$:

$$\begin{aligned} &P_{S_k|X_1X_2Y_1'Y_2'}(0|x_1, x_2, 0, 0) \\ &\quad = \frac{P_{S_k}(0) P_{Y_1'Y_2'|X_1X_2S_k}(0, 0|x_1, x_2, 0)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)} \\ &\quad = \frac{P_S(0)(1/2 + P_S(0)1/2)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)}. \end{aligned} \quad (176b)$$

We conclude from (176) that for $y_1' = y_2' = 0$ and $x = x_1 = x_2$, the optimal estimators are

$$\begin{aligned} &\hat{s}_k^*(x, x, 0, 0) \\ &= \mathbb{1}\{P_S(0)(1 + P_S(0)) < P_S(1)\}, \ k = 1, 2. \end{aligned} \quad (177)$$

We turn to the case $x_1 \neq x_2$, where $x_1 \oplus N = 1 \oplus (x_2 \oplus N)$. As before, if $S_k = 1$, then $Y_k' = 0$ only if $x_1 \oplus N = 0$, which happens with probability $1/2$. Now this implies $x_2 \oplus N = 1$, and thus $Y_{\bar{k}}' = 0$ only if $S_{\bar{k}} = 0$, which happens with probability $P_S(0)$. We thus obtain for $x_1 \neq x_2$ and $k = 1, 2$:

$$\begin{aligned} &P_{S_k|X_1X_2Y_1'Y_2'}(1|x_1, x_2, 0, 0) \\ &\quad = \frac{P_{S_k}(1) P_{Y_1'Y_2'|X_1X_2S_k}(0, 0|x_1, x_2, 1)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)} \\ &\quad = \frac{P_S(1) P_S(0) 1/2}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)}. \end{aligned} \quad (178a)$$

If $S_k = 0$, then $Y_1' = Y_2' = 0$ happens when $x_{\bar{k}} \oplus N = 0$ or when $x_{\bar{k}} \oplus N = 1$ and $S_{\bar{k}} = 0$. Since these are exclusive events with total probability $1/2 + P_S(0)1/2$, we obtain for $x_1 \neq x_2$ and $k = 1, 2$:

$$\begin{aligned} &P_{S_k|X_1X_2Y_1'Y_2'}(0|x_1, x_2, 0, 0) \\ &\quad = \frac{P_{S_k}(0) P_{Y_1'Y_2'|X_1X_2S_k}(0, 0|x_1, x_2, 0)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)} \\ &\quad = \frac{P_S(0)(1/2 + P_S(0)1/2)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)}. \end{aligned} \quad (178b)$$

Since $P_S(1) < 1 + P_S(0)$, we conclude that for $y_1' =$

$0, y'_2 = 0$ and $x_1 \neq x_2$, the optimal estimator is

$$\hat{s}_k^*(x_1, x_2, 0, 0) = 0, \qquad x_1 \neq x_2,\ k = 1, 2. \tag{179}$$

B. Minimum distortion

We evaluate the expected distortion of the optimal estimators in (85), for a given input pmf $P_{X_0X_1X_2}$. Let $t := \Pr[X_1 \neq X_2]$. We first consider the distortion on state S_2:

$$\begin{aligned}&\mathbb{E}[d(S_2, \hat{s}_2^*(X_1, X_2, Y'_1, Y'_2))]\\ &= \sum_{(x_1,x_2,y'_1,y'_2)\in\{0,1\}^4} P_{X_1X_2Y'_1Y'_2}(x_1, x_2, y'_1, y'_2)\\ &\quad \cdot \Pr\big[S_2 \neq \hat{s}_2^*(x_1, x_2, y'_1, y'_2) \mid X_1 = x_1, X_2 = X_2,\\ &\qquad\qquad Y'_1 = y_1, Y'_2 = y_2\big]\end{aligned} \tag{180}$$

$$\begin{aligned}&\stackrel{(a)}{=} \sum_{(x_1,x_2,y'_1,y'_2)\in\{0,1\}^4} P_{X_1X_2Y'_1Y'_2}(x_1, x_2, y'_1, y'_2)\\ &\qquad \cdot \min_{\hat{s}\in\{0,1\}} P_{S_2|X_1X_2Y'_1Y'_2}(\hat{s}|x_1, x_2, y'_1, y'_2),\end{aligned} \tag{181}$$

where (a) follows by the definition of the function $\hat{s}_2^*$.

In the previous Subsection H-A, we argued that for $y'_2 = 1$ or for $(y'_2 = 0, y'_1 = 1, x_1 = x_2)$, the state S_2 is deterministic ($S_2 = 1$ in the former case and $S_2 = 0$ in the latter) and thus $\min_{\hat{s}\in\{0,1\}} P_{S_2|X_1X_2Y'_1Y'_2}(\hat{s}|x_1, x_2, y'_1, y'_2) = 0$. We further argued that for $(y'_1 = 1, y'_2 = 0, x_1 \neq x_2)$ the transmitter learns nothing about state S_2, which is thus still distributed according to P_S. Based on these observations, we continue from (181) as:

$$\begin{aligned}&\mathbb{E}[d(S_2, \hat{s}_2^*(X_1, X_2, Y'_1, Y'_2)]\\ &= \Pr\big[X_1 \neq X_2, Y'_1 = 1, Y'_2 = 0\big] \min\{P_S(0), P_S(1)\}\\ &\quad + \sum_{(x_1,x_2)\in\{0,1\}^2} P_{X_1X_2Y'_1Y'_2}(x_1, x_2, 0, 0)\\ &\qquad \min\{P_{S_1|X_1X_2Y'_1Y'_2}(0|x_1, x_2, 0, 0),\\ &\qquad\qquad P_{S_1|X_1X_2Y'_1Y'_2}(1|x_1, x_2, 0, 0)\}\\ &\stackrel{(b)}{=} \Pr[X_1 \neq X_2, N = X_1 \oplus 1, S_1 = 1]\\ &\qquad\qquad \min\{P_S(0), P_S(1)\}\\ &\quad + \Pr[X_1 = X_2]\,\frac{1}{2}\min\{P_S(1), P_S(0)(1 + P_S(0))\}\\ &\quad + \Pr[X_1 \neq X_2]\,\frac{1}{2} P_S(0)P_S(1)\end{aligned} \tag{182}$$

$$\begin{aligned}&= \frac{1}{2} tq\Big(\min\{q, (1-q)\} + (1-q)\Big)\\ &\quad + \frac{1}{2}(1-t)q\min\{q, (1-q)(2-q)\}.\end{aligned} \tag{183}$$

where in (b) we used (176)–(179) and the fact that when $X_1 \neq X_2$, then event $\{Y'_1 = 1, Y'_2 = 0\}$ is equivalent to event $\{N = X_1 \oplus 1, S_1 = 1\}$.

C. Proof of the Outer Bound

The outer bound is based on Theorem 3, as detailed out in the following. The single-rate constraints (68a) specialize to

$$\mathsf{R}_k \leq I(U_k; Y'_k, X_0 \mid S_1, S_2) \tag{184}$$

$$\stackrel{(c)}{=} I(U_k; X_0) \tag{185}$$

$$\leq 1, \tag{186}$$

where the equality holds by the chain rule, because (U_1, X_0) and (S_1, S_2) are independent, and because $I(U_1; Y'_1 \mid X_0, S_1, S_2) = 0$ due to the Bernoulli-$1/2$ noise N.

Defining $t := \Pr[X_1 \neq X_2]$, Bound (68b) specializes to:

$$\begin{aligned}&\mathsf{R}_1 + \mathsf{R}_2\\ &\quad \leq I(X_0, X_1, X_2; Y'_1, Y'_2, X_0 \mid S_1, S_2) &(187)\\ &\quad \stackrel{(d)}{=} H(X_0) + I(X_1, X_2; Y'_2 \mid S_1, S_2, Y'_1, X_0) &(188)\\ &\quad \stackrel{(e)}{=} H(X_0)\\ &\qquad + I(X_1, X_2; Y'_2 \mid S_1 = 1, S_2 = 1, Y'_1, X_0) &(189)\\ &\quad = H(X_0)\\ &\qquad + I(X_1, X_2; Y'_2 \oplus Y'_1 \mid S_1 = 1, S_2 = 1, X_0) &(190)\\ &\quad \stackrel{(f)}{\leq} H(X_0) + P_{S_1S_2}(1,1)H(X_1 \oplus X_2) &(191)\\ &\quad \leq 1 + q^2 H_b(t). &(192)\end{aligned}$$

where (d) holds by the chain rule and because $I(X_1, X_2; Y'_1 \mid X_0, S_1, S_2) = 0$ due to the Bernoulli-$1/2$ noise N; (e) holds because for $(s_1, s_2) \neq (1, 1)$ the mutual information term $I(X_1, X_2; Y'_2 \mid S_1 = s_1, S_2 = s_2, Y'_1, X_0) = 0$ due to the Bernoulli-$1/2$ noise N; and (f) holds because for $S_1 = S_2 = 1$ we have $Y'_2 \oplus Y'_1 = (X_2 \oplus N) \oplus (X_1 \oplus N) = X_2 \oplus X_1$ and because conditioning can only reduce entropy.

The sum-rate constraint (192) is maximized for $t = 1/2$, which combined with (186) establishes the converse to the capacity region in (90).

D. Proof of Achievability Results

We evaluate Proposition 1 for different choices of the involved random variables. Since we ignore the common rate R_0, bound (70d) is not active and can be ignored.

1) First choice:

- X_0, X_1, X_2 Bernoulli-1/2 with X_0 independent of (X_1, X_2) and $X_1 = X_2 = x$ with probability $\frac{1-t}{2}$ for all $x \in \{0,1\}$;
- $U_k = X_k$, for $k = 0, 1, 2$;
- $V_1 = (X_0, X_1)$, $V_2 = (X_0, X_2)$, $V_0 = X_1 \oplus Y_1'$.

We plug this choice into Proposition 1. Constraint (70a) evaluates to:

$$\begin{aligned}\mathsf{R}_1 &\leq I(U_0, U_1; Y_1, V_1 \mid S_1) \\ &\quad -I(U_0, U_1, U_2, Z; V_0, V_1 \mid S_1, Y_1) && (193)\\ &= I(X_0, X_1; X_0, Y_1', S_1, S_2, X_1 \mid S_1) \\ &\quad -I(X_0, X_1, X_2, Y_1', Y_2'; X_1 \oplus Y_1', X_0, X_1 \\ &\qquad \mid S_1, S_2, X_0, Y_1') && (194)\\ &\overset{(e)}{=} H(X_0) + H(X_1) - H(X_1 \mid Y_1') && (195)\\ &= H(X_0) = 1 && (196)\end{aligned}$$

where (e) holds because Y_1' is independent of X_1 due to the Bernoulli-1/2 noise N.

Constraint (70b) evaluates to:

$$\begin{aligned}\mathsf{R}_2 &\leq I(U_0, U_2; Y_2, V_2 \mid S_2) \\ &\qquad -I(U_0, U_1, U_2, Z; V_0, V_2 | S_2, Y_2) && (197)\\ &= I(X_0, X_2; X_0, Y_2', S_1, S_2, X_2 \mid S_1) \\ &\quad -I(X_0, X_1, X_2, Y_1', Y_2'; X_1 \oplus Y_1', X_0, X_2 \\ &\qquad \mid S_1, S_2, X_0, Y_2') && (198)\\ &\overset{(f)}{=} H(X_0) + H(X_2) - H(X_2) \\ &\quad -H(X_1 \oplus Y_1' \mid S_1, S_2, X_0, Y_2', X_2) && (199)\\ &\overset{(g)}{=} 1 - (1-q)(H_\mathsf{b}(t) + q), && (200)\end{aligned}$$

where (f) holds because of the chain rule and the independence of X_2 and Y_2'; and (g) holds because for $S_1 = 0$ the XOR $X_1 \oplus Y_1' = X_1$ and thus $H(X_1 \oplus Y_1' \mid S_1, S_2, X_0, Y_2', X_2) = H(X_1 \mid X_2)$, for $S_1 = S_2 = 1$ the XOR $X_1 \oplus Y_1' = X_2 \oplus Y_2'$, and finally for $S_1 = 1$ and $S_2 = 0$ the XOR $X_1 \oplus Y_1' = N$ independent of $(Y_2' = 0, X_2)$.

Constraint (70b) evaluates to:

$$\begin{aligned}&\mathsf{R}_1 + \mathsf{R}_2 \\ &\leq I(U_1; Y_1, V_1 \mid U_0, S_1) + I(U_2; Y_2, V_2 \mid U_0, S_2) \\ &\quad + \min_{k\in\{1,2\}} I(U_0; Y_k, V_k \mid S_k) - I(U_1; U_2 | U_0) \\ &\quad -I(U_0, U_1, U_2, Z; V_1 \mid V_0, S_1, Y_1) \\ &\quad -I(U_0, U_1, U_2, Z; V_2 \mid V_0, S_2, Y_2) \\ &\quad - \max_{k\in\{1,2\}} I(U_0, U_1, U_2, Z; V_0 \mid S_k, Y_k) && (201)\\ &= \underbrace{I(X_1; X_0, Y_1', S_1, S_2, X_1 \mid X_0, S_1)}_{=H(X_1)} \\ &\quad + \underbrace{I(X_2; X_0, Y_2', S_1, S_2, X_2 \mid X_0, S_2)}_{=H(X_2)} \\ &\quad + \min_{k\in\{1,2\}} \underbrace{I(X_0; X_0, Y_k', S_1, S_2, X_k \mid S_k)}_{=H(X_0)} \\ &\quad - \underbrace{I(X_1; X_2)}_{=H(X_1)-H(X_1|X_2)} \\ &\quad - \underbrace{I(\underline{X}, Y_1', Y_2'; X_0, X_1 \mid X_1 \oplus Y_1', X_0, \underline{S}, Y_1')}_{=0} \\ &\quad - \underbrace{I(\underline{X}, Y_1', Y_2'; X_0, X_2 \mid X_1 \oplus Y_1', X_0, \underline{S}, Y_2')}_{=H(X_2|X_1\oplus Y_1', S_1, S_2, Y_2')} \\ &\quad - \max_{k\in\{1,2\}} \underbrace{I(\underline{X}, Y_1', Y_2'; X_1 \oplus Y_1' \mid \underline{S}, X_0, Y_k')}_{=H(X_1\oplus Y_1'|S_1, S_2, Y_k')} && (202)\\ &\overset{(h)}{=} 2 + H_\mathsf{b}(t) - H(X_2 \mid X_1 \oplus Y_1', S_1, S_2, Y_2') \\ &\quad -H(X_1 \oplus Y_1') && (203)\\ &\overset{(i)}{=} 1 + H_\mathsf{b}(t) - (1-q)H_\mathsf{b}(t) - q(1-q), && (204)\end{aligned}$$

where we used the abbreviations $\underline{X} = (X_0, X_1, X_2)$ and $\underline{S} = (S_1, S_2)$ and (h) holds because $X_1 \oplus Y_1$ is independent of (S_1, S_2, Y_k'), for $k = 1, 2$; and (i) holds because for $S_1 = S_2 = 1$ we have $X_2 = Y_2' \oplus Y_1' \oplus X_1$ and thus $H(X_2 \mid X_1 \oplus Y_1', S_1 = 1, S_2 = 1, Y_2') = 0$, for $S_1 = 0$ the XOR $X_1 \oplus Y_1' = X_1$ and thus $H(X_2 \mid X_1 \oplus Y_1', S_1 = 1, S_2, Y_2') = H(X_2 | X_1) = H_\mathsf{b}(t)$, and finally for $S_1 = 1$ and $S_2 = 0$, we have $X_1 \oplus Y_1' = N$ and $Y_2' = 0$ and thus $H(X_2 \mid X_1 \oplus Y_1', S_1 = 1, S_2 = 0, Y_2') = H(X_2) = 1$.

The presented choice of parameters can thus achieve all rate-distortion tuples $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ satisfying the distortion constraints in (87) (which only depends on the probability $t := \Pr[X_1 \neq X_2]$) and

$$\mathsf{R}_1 \leq 1 \tag{205a}$$
$$\mathsf{R}_2 \leq 1 - (1-q)(H_\mathsf{b}(t) + q) \tag{205b}$$
$$\mathsf{R}_1 + \mathsf{R}_2 \leq 1 + qH_\mathsf{b}(t) - q(1-q). \tag{205c}$$

2) Second choice: Same as the first choice except that $V_0 = X_2 \oplus Y_2'$. Following symmetric arguments as above, we conclude that for this choice the constraints in (70) evaluate to:

$$\mathsf{R}_1 \leq 1 - (1-q)(H_\mathsf{b}(t) + q) \tag{206a}$$
$$\mathsf{R}_2 \leq 1 \tag{206b}$$
$$\mathsf{R}_1 + \mathsf{R}_2 \leq 1 + qH_\mathsf{b}(t) - q(1-q). \tag{206c}$$

3) Combining the Choices and Time-Sharing: From the two previous subsections, we conclude that for any $t \in [0,1]$ the set of rate-distortion tuples $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ is achievable if it satisfies (87) and

$$\mathsf{R}_0 + \mathsf{R}_1 \leq 1 \tag{207a}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \leq 1 \tag{207b}$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq 1 + qH_{\mathrm{b}}(t) - q(1-q). \tag{207c}$$

As previously discussed, for $q \leq 1/2$, the distortion constraints (87) do not depend on t, and thus without loss in optimality in (207) one can set $t = 1/2$, which results in a sum-rate constraint

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq 1 + q^2. \tag{208}$$

Combined with (207a) and (207b), this sum-rate bound establishes the achievability of the capacity region in (90).

For $q > 1/2$ the distortion constraints (87) are either increasing or decreasing in t. The set of achievable rate-distortion tuples is then obtained by varying t either over $[0, 1/2]$ or over $[1/2, 1]$. Numerical results indicate that the so obtained set is not convex and the convex hull is obtained by considering convex combinations between different values of $t > 0$ and $t = 0$ for $q \in [2/3, 1]$ and $t = 1$ for $q \in [1/2, 2/3]$.

References

[1] M. Kobayashi, G. Caire, and G. Kramer, "Joint state sensing and communication: Optimal tradeoff for a memoryless case," in *Proc. IEEE Int. Symp. Info. Theory (ISIT)*, 2018, pp. 111–115.

[2] M. Ahmadipour, M. Wigger, and M. Kobayashi, "Joint sensing and communication over memoryless broadcast channels," in *Proc. IEEE Info. Theory Workshop (ITW)*, 2021, pp. 1–5.

[3] G. Flagship, "6G white paper on localization and sensing," *University of Oulu, Finland*, June 2020.

[4] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang, "Radar and communication coexistence: An overview: A review of recent methods," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 85–99, 2019.

[5] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, 2020.

[6] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," vol. 99, no. 7, pp. 1236–1259, July 2011.

[7] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "On the effectiveness of OTFS for joint radar parameter estimation and communication," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5951–5965, 2020.

[8] D. W. Bliss, "Cooperative radar and communications signaling: The estimation and information theory odd couple," in *Radar Conf., 2014 IEEE*. IEEE, 2014, pp. 0050–0055.

[9] A. R. Chiriyath, B. Paul, G. M. Jacyna, and D. W. Bliss, "Inner bounds on performance of radar and communications co-existence." *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 464–474, 2016.

[10] B. Paul, A. R. Chiriyath, and D. W. Bliss, "Survey of RF communications and sensing convergence research," *IEEE Access*, vol. 5, pp. 252–270, 2017.

[11] P. Kumari, J. Choi, N. González-Prelcic, and R. W. Heath, "IEEE 802.11ad-based radar: An approach to joint vehicular communication-radar system," vol. 67, no. 4, pp. 3012–3027, April 2018.

[12] P. Kumari, D. H. Nguyen, and R. W. Heath, "Performance trade-off in an adaptive ieee 802.11 ad waveform design for a joint automotive radar and communication system," in *IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*. IEEE, 2017, pp. 4281–4285.

[13] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. Info. Theory*, vol. 54, no. 5, pp. 1850–1859, 2008.

[14] W. Zhang, S. Vedantam, and U. Mitra, "Joint transmission and state estimation: A constrained channel coding approach," *IEEE Trans. Info. Theory*, vol. 57, no. 10, pp. 7084–7095, 2011.

[15] C. Choudhuri, Y. Kim, and U. Mitra, "Causal state communication," *IEEE Trans. Info. Theory*, vol. 59, no. 6, pp. 3709–3719, 2013.

[16] S. I. Bross and A. Lapidoth, "The Gaussian source-and-data-streams problem," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5618–5628, 2019.

[17] V. Ramachandran, S. R. B. Pillai, and V. M. Prabhakaran, "Joint state estimation and communication over a state-dependent gaussian multiple access channel," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 6743–6752, 2019.

[18] M. Kobayashi, H. Hamad, G. Kramer, and G. Caire, "Joint state sensing and communication over memoryless multiple access channels," in *Proc. IEEE Int. Symp. Info. Theory (ISIT)*, 2019, pp. 270–274.

[19] M. Ahmadipour, M. Wigger, and M. Kobayashi, "Coding for sensing: An improved scheme for integrated sensing and communication over macs," 2022.

[20] O. Shayevitz and M. Wigger, "On the capacity of the discrete memoryless broadcast channel with feedback," *IEEE Trans. Info. Theory*, vol. 59, no. 3, pp. 1329–1345, 2013.

[21] R. Venkataramanan and S. S. Pradhan, "An achievable rate region for the broadcast channel with feedback," *IEEE Trans. Info. Theory*, vol. 59, no. 10, pp. 6175–6191, 2013.

[22] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. Wigger, "Coding schemes and asymptotic capacity for the gaussian broadcast and interference channels with feedback," *IEEE Trans. Info. Theory*, vol. 60, no. 1, pp. 54–71, 2014.

[23] A.El Gamal and Y.H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[24] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 25, no. 3, pp. 306–311, 1979.

[25] A. Gohari and C. Nair, "New outer bounds for the two-receiver broadcast channel," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 1492–1497.

[26] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Info. Theory*, vol. 18, no. 1, pp. 14–20, 1972.

[27] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Info. Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[28] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 4—5, pp. 265–444, 2008.

[29] C. C. Wang, "On the capacity of 1-to-k broadcast packet erasure channels with channel output feedback," *IEEE Trans. Info. Theory*, vol. 58, no. 2, pp. 931–956, 2012.

[30] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser broadcast erasure channel with feedback: capacity and algorithms," *IEEE Trans. Info. Theory*, vol. 59, no. 9, pp. 5779–5804, 2013.

[31] G. Dueck, "Partial feedback for two-way and broadcast channels," *Information and Control*, vol. 46, no. 1, pp. 1–15, 1980.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory (2nd Edition)*. Wiley-Interscience, 2006.

Mehrasa Ahmadipour received the B.Sc. in Electrical Engineering from Iran University of Science and Technology in 2015, received the M.Sc. degree in electrical engineering in electrical engineering from University of Tehran in 2019. She is currently a Ph.D. student in Telecom Paris, France. Her particular fields of interest include multi-terminal information theory, networks with states, user cooperation, and statistical learning.

Mari Kobayashi (M'06–SM'15) received the B.E. degree in electrical engineering from Keio University, Yokohama, Japan, in 1999, and the M.S. degree in mobile radio and the Ph.D. degree from École Nationale Supérieure des Télécommunications, Paris, France, in 2000 and 2005, respectively. From November 2005 to March 2007, she was a postdoctoral researcher at the Centre Tecnològic de Telecomunicacions de Catalunya, Barcelona, Spain. In May 2007, she joined the Telecommunications department at Centrale Supélec, Gif-sur-Yvette, France, where she is now a professor. She is the recipient of the Newcom++ Best Paper Award in 2010, and IEEE Comsoc/IT Joint Society Paper Award in 2011, and ICC Best Paper Award in 2019. She was an Alexander von Humboldt Experienced Research Fellow (September 2017- April 2019) and an August-Wihelm Scheer Visiting Professor (August 2019-April 2020) at Technical University of Munich (TUM).

Michèle Wigger (S'05–M'09–SM'14) received the M.Sc. degree in electrical engineering, with distinction, and the Ph.D. degree in electrical engineering both from ETH Zurich in 2003 and 2008, respectively. In 2009, she was first a post-doctoral fellow at the University of California, San Diego, USA, and then joined Telecom Paris, France, where she is currently a full professor. She has held visiting professor appointments at the Technion–Israel Institute of Technology and ETH Zurich. Michèle Wigger has previously served as an Associate Editor of the IEEE Communication Letters and as an Associate Editor for Shannon Theory for the IEEE Transactions on Information Theory. During 2016–2019 she also served on the Board of Governors of the IEEE Information Theory Society. Her research interests are in multi-terminal information theory, in particular in distributed source coding and in capacities of networks with states, feedback, user cooperation, or caching.

Giuseppe Caire (S '92 – M '94 – SM '03 – F '05) was born in Torino in 1965. He received the B.Sc. in Electrical Engineering from Politecnico di Torino in 1990, the M.Sc. in Electrical Engineering from Princeton University in 1992, and the Ph.D. from Politecnico di Torino in 1994. He has been a post-doctoral research fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) in 1994-1995, Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, a Professor of Electrical Engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, and he is currently an Alexander von Humboldt Professor with the Faculty of Electrical Engineering and Computer Science at the Technical University of Berlin, Germany.

He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, the IEEE Communications Society and Information Theory Society Joint Paper Award in 2004 and in 2011, the Okawa Research Award in 2006, the Alexander von Humboldt Professorship in 2014, the Vodafone Innovation Prize in 2015, an ERC Advanced Grant in 2018, the Leonard G. Abraham Prize for best IEEE JSAC paper in 2019, the IEEE Communications Society Edwin Howard Armstrong Achievement Award in 2020, and he is a recipient of the 2021 Leibniz Prize of the German National Science Foundation (DFG). Giuseppe Caire is a Fellow of IEEE since 2005. He has served in the Board of Governors of the IEEE Information Theory Society from 2004 to 2007, and as officer from 2008 to 2013. He was President of the IEEE Information Theory Society in 2011. His main research interests are in the field of communications theory, information theory, channel and source coding with particular focus on wireless communications.