Exploring ISAC: Information Theoretic Insights

Mehrasa Ahmadipour*, Michèle Wigger†, Shlomo Shamai‡

*UMPA, ENS de Lyon, Email: mehrasa.ahmadipour@ens-lyon.fr †LTCI Telecom Paris, IP Paris, 91120 Palaiseau, France, Email: michele.wigger@telecom-paris.fr ‡Technion, Haifa, Israel, Email: sshlomo@ee.technion.ac.il

I. ABSTRACT

This article reviews results from the literature regarding the bottlenecks and tradeoffs of integrated sensing and communication (ISAC) through the lens of information theory, offering a distinct perspective compared to recent works that focus on signal processing, wireless communications, or other related overviews. Different models and scenarios are considered and compared. For example, scenarios where radar sensing is performed at the communication and radar transmitter (mono-static ISAC) and scenarios where the radar receiver differs from the radar transmitter (called bi-static radar). Similarly, we discuss ISAC bottlenecks and tradeoffs both in slowly-varying environments where the main sensing target is described by a single parameter and accordingly sensing performance is described by detection error probabilities, as well as in fast-varying environments where the sensing targets are described by vectors and thus vector-valued performance measures such as average distortions like mean-squared errors are used to determine sensing performances. This overview article further also considers limitations and opponities in network ISAC environments, such as collaborative or interactive sensing, and the influence of secrecy and privacy requirements on ISAC systems, a line of research that has obtained growing interest over the last few years. For all these scenarios we provide and discuss precise models and their limitations and provide either bounds or full characterizations of the fundamental information-theoretic performance limits of these systems. Further extensions as well as important open research directions are also discussed.

II. INTRODUCTION

Integrated Sensing and Communication (ISAC) represents a transformative paradigm that unifies sensing and communication functionalities into a single system, leveraging shared spectral, hardware, and computational resources. This integration is increasingly crucial in modern technological ecosystems, where efficient resource utilization and enhanced system performance are paramount. ISAC offers notable benefits in applications such as autonomous vehicles, industrial automation, smart cities, and wireless networks.

By reducing latency, improving spectrum efficiency, and enhancing situational awareness, ISAC is foundational to emerging technologies like 6G communication and the Internet of Things (IoT). Its dual-purpose design minimizes infrastructure costs while enabling seamless interaction between sensing and communication, paving the way for adaptive and intelligent systems.

As an example of a practical ISAC systems, we can consider the Wi-Fi technology. Based on the IEEE 802.11 standards, it has delivered significant social and economic benefits. Recently, attention has turned to WLAN sensing—also known as Wi-Fi sensing—which leverages the widespread Wi-Fi infrastructure and ubiquitous signals in our environment to perform various sensing tasks. By employing advanced signal processing techniques, received Wi-Fi signals can be used to detect obstructions, monitor environmental changes, and interpret target movement. Despite these innovations and progress, several challenges remain in current standardization efforts, as evidenced by ongoing work towards IEEE 802.11bf for WLAN sensing and ISAC [1].

In general, designing effective ISAC systems involves achieving simultaneous high-performance sensing and communication. Advanced configurations, such as bi-static radar or multi-terminal ISAC systems,

require efficient exchange of sensing information between terminals to enhance collective sensing capabilities rather than relying solely on local data. Furthermore, modern ISAC applications impose additional constraints, such as privacy and security. Balancing these competing requirements necessitates an understanding of the tradeoffs and fundamental performance limits across various system criteria.

Information theory has rich history and provides a robust framework to analyze such complex multi-purpose systems and reveal the inherent tensions and tradeoffs in the fundamental limits of the various performances of such systems. Fundamental results from information-theory literature for communication (e.g., data rate, capacity) [2]–[7], detection and hypothesis testing [8]–[20], estimation [21]–[24], and compression [25]–[31] indeed can form the foundation for analyzing ISAC systems. Prior studies have explored tradeoffs between these performance measures in both distributed and non-distributed setups. Some ISAC scenarios, though not explicitly named as such, have been studied within the information-theoretic community under related contexts like "simultaneous data communication and state estimation." Other scenarios remain unexplored but can benefit from information-theoretic insights derived from analogous setups.

This article aims to synthesize and present both established and emerging information-theoretic results relevant to ISAC systems. Compared to many existing overview ISAC articles that tackle the problem more from a communication and signal-processing angle [32]–[41], here we focus on information-theoretic results that are more closely aligned with [42], [43]. Specifically, we focus on the inherent tradeoffs and fundamental performance limits of ISAC systems, emphasizing coding techniques and proof strategies that enable optimal sensing-communication tradeoffs. We would like to point out that the information-theoretic models studied in this survey are very general, and can be specialized to different practical scenarios of interest. Keeping the model and results general allows us to derive broadly based conclusions and also cover a larger range of application scenarios.

Specifically, in this article, we start with a a brief historical perspective of ISAC (Section III), followed by a first technical section (Section IV) that considers a canonical ISAC point-to-point setup with a single Tx wishing to communicate to a single Rx and where the sensing task is to estimate a state-sequence (such as the accelerations of an obstacle) up to a desired distortion. We will start discussing results on simple memoryless channel models, and then move on to very general models with memory. All the results discussed in this Section IV illustrate the inherent tradeoff between the sensing and communication tasks encountered in such ISAC systems. While Section IV considers a mono-static radar setup, i.e., the sensing task is performed at the Tx of the communication system, the subsequent Section V considers bi-static radar models where sensing is performed at the Rx. Section VI further generalizes the setup to multiple Txs or Rxs and sensing at multiple terminals. Not only the communication problem is of formidable difficulty in these setups but also the sensing task is significantly more involved, as now collaborative and interactive sensing strategies can be applied to provide remote terminals with sensing information gathered at other terminals. As we shall see, in such scenarios, it does not just suffice to exchange sensing information using standard communication schemes, but instead the code construction previously only used for data communication now need to be adapted to also enable the collaborative sensing tasks. Moreover, given the distributed sensing information that has to be conveyed from certain terminals to others, network joint source-channel coding schemes become essential to attain good sensing performances. The subsequent Section VII, then describes how the above ISAC schemes and performance limits need to be adapted so as to ensure secrecy of only messages or of states and sensing targets/states. The last technical section (Section VIII) of this overview then takes a different approach to the sensing task, assuming that the sensing task consists in detecting a single parameter (and not estimating a state-vector as in the previous chapters) which determines the behaviour of the sensing target. This problem seems to be slightly more challenging but first instructive results are presented, in particular when the Tx restricts to non-adaptive coding schemes where the backscattered signals can only be used for the sensing task but not to produce the subsequent inputs. The overview article is then concluded with conclusive

remarks.

There exists a large body of studies also on other aspects of ISAC systems, for example, the works in [44]–[57] have studied ISAC from a more communication-theoretic perspective in environments where Txs and Rxs are equipped with multiple antennas, in particular also in so called massive multi-input and multi-output (MIMO) systems. Interesting research directions result in these MIMO systems regarding whether smart selection of beamformers allow to reduce the tradeoff between communication and sensing performances and how the tradeoff is influenced by the choice of the antenna distance. Recent initiatives to improve understanding of ISAC systems of course also include learning-based studies. The tutorial in [41] provides a comprehensive overview of works using deep-learning based techniques and the summary of the reviewed results are provided in Table I.

A. Notation

In this survey we shall use class notation for our mathematical expressions. For example, upper-case letters like X denote random quantities, while lower-case letters like x represent their deterministic realizations. Sets are represented using calligraphic font (e.g., $\mathcal{X}$). The n-tuples $(X_1, \ldots, X_n)$ and $(x_1, \ldots, x_n)$ are abbreviated as X^n and x^n, respectively. Similarly, the $n - t$-tuples $(X_{t+1}, \ldots, X_n)$ and $(x_{t+1}, \ldots, x_n)$ are written as X_{t+1}^n and x_{t+1}^n. *Independent and identically distributed* is abbreviated as *i.i.d.*, and *probability mass function* as *pmf*. The conditional probability is written as $P_{XY|UV}(x, y|u, v)$, and $P_X(\cdot)$ represents the pmf of a finite random variable X. The expectation of a random variable X is denoted by $\mathbb{E}[X]$. $\mathbb{R}$ and $\mathbb{R}_0^+$ denote the sets of real numbers and nonnegative real numbers, respectively. $\log$ typically represents the base-2 logarithm, accordingly information measures are measured in terms of *nats*. The operator $\oplus$ typically indicates XOR (binary addition modulo 2).

The $\arg\min$ represents the set of minimizers of a function. $\overline{\lim}$ and $\underline{\lim}$ denote the limit superior and limit inferior as $n \to \infty$. The operator $f^{(n)}$ represents a transformation or operation on n-letter sequences. The term ess sup refers to the essential supremum in measure theory. Entropy, conditional entropy, and mutual information are denoted by $H(\cdot)$, $H(\cdot|\cdot)$, and $I(\cdot; \cdot)$, respectively. When the probability mass function (pmf) is not clear from the context, it is included as a subscript, e.g., $H_P(\cdot)$. The Kullback-Leibler divergence between two pmfs is denoted by $D(\cdot\|\cdot)$.

III. PRE-ISAC: SENSING (RADAR) VS. COMMUNICATION

A. Radar Systems

Radar is a system that utilizes radio waves to learn about positions, motions, or the mere presence of target objects in an environment through the analysis of backscattered signals. In fact, a radar terminal radiates a waveform that propagates through space until it reaches a target, where it is reflected in a way that depends on the properties of the target. The radar terminal collects and analyzes the backscattered signals so as to gain information about these properties. In the radar system, if the presence and position of a target are already known, the transmitter tries to steer all the energy of the transmitted waveform towards the target, so as to obtain more information through the backscattered waveform. Radar thus uses Line-of-Sight (LoS) techniques. Traditional radar systems mainly operate within the 24–79 GHz frequency band.

Sensing tasks can be roughly classified into three categories, detection, estimation, and recognition, which are all based on collecting signals/data concerning the sensed objects. Detection refers to making decisions on an object's *state* given some observations, such as the presence/absence of the target or other events related to the target. The detection problem can be modeled as a binary or multi hypothesis testing problem. In the binary hypothesis testing problem as an example, one selects from two hypotheses; the alternative hypothesis $\mathcal{H}_1$ and the null hypothesis $\mathcal{H}_0$. Detection metrics are the probability that $\mathcal{H}_1$ holds but the detector chooses $\mathcal{H}_0$ (often denoted miss-detection probability), and the probability that $\mathcal{H}_0$ holds but the detector chooses $\mathcal{H}_1$ (often denoted false-alarm probability).

TABLE I
A COMPREHENSIVE OVERVIEW OF THIS SURVEY

Category	Result Description	Reference(s)
Sensing as Monostatic Radar	Lemma 1: Optimal estimator for P2P and BC	[58]
	Theorem 1: Exact Capacity-Distortion for Memoryless P2P, asymptotic analysis	[58]
	Strong converse Remark 1	[59]
	Log-Loss distortion Theorem 2	[60]
	Nonasymptotic P2P, Theorem 3	[61]
	Channel with memory, RL approach 3	[62]
Sensing as Bistatic Radar (P2P)	C-D with No CSI at TX 4	[63]
	C-D with Strictly causal CSI at TX 5	[64]
	C-D non-causal CSI, Gaussian channel at Tx 6	[65]
Network-ISAC	General BC outer 7 and inner 1 bounds	[66], [67]
	Optimal symbolwise estimator	–
	Outerbounds for MAC Theorem 8	[68] [69]
	Innerbound MAC Theorem 9	[68]–[71]
	Innerbound D2D 10	[71]
Secrecy-ISAC	Secrecy-capacity-distortion Inner 11 and outer 12 Bounds	[72]
	Secrecy of the mesage and the state 13	[73]
ISAC with Detection-Error Exponents	Non-adaptive rate-detection-exponents Theorem 14	[59], [74]–[77]
	Adaptive rate-detection-exponents Theorem 15	[76]
	Sequential (Variable-Length) rate-detection-exponents Theorem 16	[76]
	Sequential (Variable-Length) ISAC with Change-Point Detection Theorem 17	[78]

Estimation refers to extracting valuable parameters, typically with continuous alphabets, of the sensed object from observations. For example, distance/velocity/angle/quantity/size of targets are possible parameters a radar system desires to estimate. Various interesting performance metrics exist for estimation, whose suitability depends on the application. Prominent examples are the mean squared error (MSE) metric, which measures the expected squared-error of the estimated parameter to the ground truth parameter. Notice that in the case of *unbiased estimators*, i.e., estimators $\hat{S}$ whose conditional expectation is always equal to the true parameter S, $\mathbb{E}[\hat{S}|S] = S$, the Cramér–Rao bound (CRB) expresses an interesting lower bound on the MSE that can be attained by any unbiased estimator.

B. Wireless Communication Systems

In a communication system, a transmitter (Tx) aims to transfer either data bits or source samples (such as audio or video file samples) to a distant receiver (Rx). The data or source information is encoded onto a transmitted waveform, which the receiver then collects and analyzes to estimate the transmitted information. Performance metrics commonly considered for communication systems include energy or spectral efficiency, which measure how many bits of information are communicated using a given energy budget or bandwidth, respectively.

For data transmission, robustness of communication is typically measured by the bit-error rate (BER), symbol-error rate (SER), or frame-error rate (FER), which indicate the likelihood of errors in the received data due to channel disturbances. In source communication, robustness is either measured by the bit-error rate or more often by distortion metrics such as the average mean-squared error. These performance metrics are especially pertinent in traditional wireless communication systems, which predominantly operate in the 2.4 GHz band. To outline the main differences between radar and communication systems, glance at Table III-B.

Communication	Sensing
2.4 GHz	24-79 GHz
Data/Source Transmission	Estimation/Detection
Bit/Signal/Frame Error Rate	Minimum Mean Squared Error (MMSE), Cramer-Rao Bound (CRB)
Distortion	Detection/False Alarm Probability
All Propagation Paths	Line of Sight (LoS)

TABLE II
COMPARISON BETWEEN COMMUNICATION AND SENSING SYSTEMS.

C. Coexisting Communication and Radar Systems

The early work [79] modulates the communication bits on the missile range radar pulse interval. Interference rejection and robustness in multipath fading environments, inherent properties of spread spectrum systems, also make chirp signaling (used in radar application) very active for the expanding wireless communications market.

Another approach in [80] proposed as early as 1962 is based on chirp signals proposed for both analog and digital communication [81] but are also commonly used in radar applications. These works can be categorized as the first steps towards *Integrated Sensing and Communication (ISAC)*.

Since then, significant evolution has lead to an entire set of pre-ISAC systems, see [32] where a category of solutions is revisited. Some straightforward solutions are called *Non-Overlapped Resource Allocation*. In subsequent information-theoretic models we will see later, such a system corresponds to time- or resource-sharing between communication and sensing; we shall call this *Basic time-sharing (TS)*, and with a minor modification, we will introduce *Improved time-sharing*.

A common but naive approach to address sensing and communication is to separate the two tasks into independent systems and split the available resources, such as bandwidth and power, between them so that they do not interfere.

Time-division ISAC can be conveniently implemented into the existing commercial systems by splitting the transmission duration into radar and radio cycles, for example [82]. For radar sensing, frequency-modulated continuous waveform (FMCW) with up-and-down-chirp modulations is used, while various different modulation schemes (e.g., BPSK, PPM) can be used for communication.

In an orthogonal frequency division multiplexing (OFDM) system, frequency-division ISAC can be implemented by allocating different communication and sensing tasks to specific subcarriers, depending on the channel conditions and power budget of the Tx [83].

Similarly, the 3GPP/5G-NR standards were originally designed primarily for communication but have evolved to accommodate additional functionalities, such as positioning and sensing. The 5G-NR standard

primarily uses OFDM due to its flexibility and efficiency, and its inherent structure can also be exploited for sensing tasks such as radar-like functions and localization. In practice, enhancements (e.g., pilot designs and advanced signal processing techniques) are introduced to extract sensing information from these communication signals. Meanwhile, Orthogonal Time Frequency Space (OTFS) modulation [84], [85] has attracted increasing interest as an alternative to OFDM, placing symbols in the delay-Doppler domain to handle high-mobility channels more robustly. By directly leveraging delay and Doppler features, OTFS can inherently support sensing-like operations, making it appealing for ISAC in future releases of 5G and beyond. In information-theoretic studies, these models are often incorporated into resource-splitting approaches. The ideal goal of ISAC is to further serve both tasks, as discussed in this work. For details on the evolution of 3GPP, see [86].

ISAC with non-overlapped resources can also be implemented over orthogonal spatial resources e.g., different antenna groups [87]. Thus, non-overlapping resource allocation can be performed in time, frequency, or spatial domains, as illustrated also in Figure 1.

(a) Time-sharing   (b) Frequency-sharing   (c) Spatial-sharing
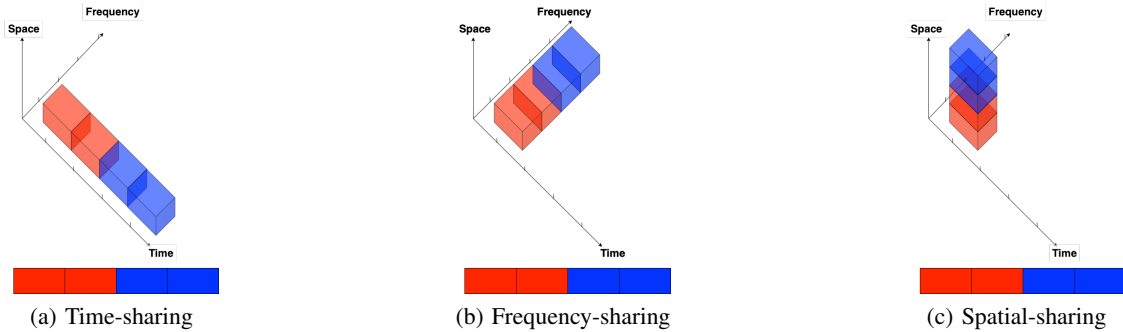
Fig. 1. The red cubes demonstrate the communication waveform, and the black cubes demonstrate the sensing waveform.

From an information-theoretic perspective, we examine pre-ISAC and non-overlapping schemes through two baseline approaches: the *Basic Time-Sharing (TS)* scheme and the *Improved Time-Sharing (TS)* scheme. The Basic TS scheme represents the non-overlapping resource allocation strategy, which divides its resources (time, bandwidth, or spatial dimensions) between the following two modes:

- *Sensing mode*
  The system aims to design a suitable waveform to attain the minimum possible distortion. In this model, the waveform is translated into an input distribution; thus, the input probability mass function (pmf) P_X is chosen to minimize the distortion, and hence the minimum distortion is achieved. The communication rate is zero.

- *Communication mode*
  The system is designed to transfer as much reliable data as possible. Therefore, the input distribution is chosen to maximize the rate and communicates rate equals channel capacity. The estimator is set to a constant value regardless of the feedback and the input signals. The mode thus suffers from a large distortion.

The improved time-sharing *Improved TS* scheme still performs a sort of non-overlapped resource allocation, but resources are not exclusively dedicated to only sensing or only communication. It is simply that one of these tasks is prioritized. 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*
  The input pmf P_X is chosen to achieve minimum distortion. At the same time, the transmitter is

also equipped with a communication encoder. It uses this input pmf to simultaneously transmit data at the rate given by the input-output mutual information of the system.

- *Communication mode with sensing*
  The input distribution is chosen to maximize the communication rate. i.e. achieve the capacity of the channel. The transmitter is however also equipped with a radar estimation device that optimally guesses the state-sequence based on the transmitted and backscattered signals.

D. Integrated Sensing and Communication (ISAC)

The concept originated from observations in communication systems where backscattered signals were typically ignored and not utilized. Subsequent studies revealed that these backscattered signals, though initially overlooked, could provide valuable information that can help the transmitter(s) (Tx) to improve communication performance or simplify coding schemes. In fact, backscattered signals can help any transmitting terminal to better estimate current and future channel conditions at the intended receivers (Rx) or to identify the receivers' uncertainty about the transmitted data. Accordingly, they can improve transmission performance by adapting future transmissions to the uncertainties to resolve or to future channel conditions. Such strategies allow to decrease error probabilities, simplify coding schemes, and when channels vary only slowly in time, with high dependencies between channel conditions at different time, adaptive scheme can even achieve higher reliable rates, i.e., improve capacity.

It is not hard to see that in typical ISAC scenarios, all non-extreme operating points of the Basic and Improved TS schemes are highly suboptimal compared to optimal integrated schemes.

IV. Mono-Static ISAC with Sensing Distortion

In this section, we introduce a first simple information-theoretic model of ISAC that allows to obtain a convenient expression for the information-theoretic limits and tradeoffs for ISAC point-to-point channels with a single Tx and a single Rx. Sensing performance is measured by an arbitrary distortion function, as is typically used by information theorists in rate-distortion theory to formalize lossy compression systems or joint source-channel coding. The model is powerful in the sense that it allows to include desired properties both from a sensing and a communication perspective. For example, any arbitrary number of sensing targets can be modeled, as well as a wide range of sensing metrics. On the other hand, depending on the specific situation one wishes to analyze, this model can describe arbitrary communication and radar channels, which may or may not depend on the sensing targets. Moreover, the model allows to include arbitrary (perfect or imperfect) channel state-information (CSI) at the Rx and arbitrary instantaneous but causal CSI at the Tx, which typically are obtained by transmission of independent pilot signals. To state a note on the down-side of the proposed model, it can only model channels and targets that evolve in a memoryless fashion. Moreover, certain sensing performances like detection probabilities cannot be described using distortion conditions. To remedy this latter drawback, in Section VIII ahead we introduce a related problem where sensing performance is measured in terms of detection error probabilities.

Notice that a very related model was also considered in [55] for a multi-antenna Gaussian fading channel. In this related work, sensing performance is however not measured in terms of distortion but by an averaged inverse Fisher information, which is motivated by the well-known Cramér-Rao bound. Particularly, in [55], I-MMSE is introduced as a unifying relation between distortion-based sensing (MMSE) and communications (mutual information) . It seems more difficult to determine the fundamental limits of this basic ISAC model under this related sensing criterion; however [55] was able to determine some extreme points of this tradeoff: the points of optimum communication performance or optimum sensing performance.
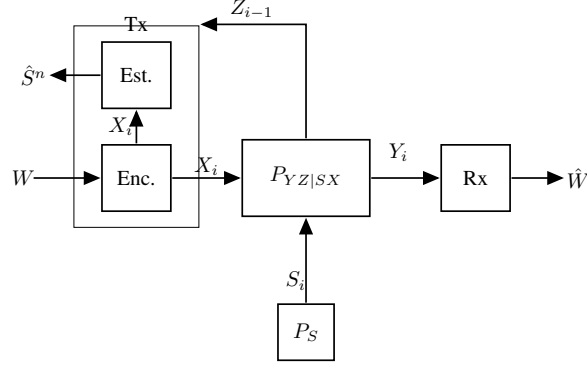
Fig. 2. The first information-theoretic ISAC model.

A. The Memoryless Model

The first information-theoretic model for a single-Tx single-Rx ISAC system was introduced by Kobayashi, Caire, and Kramer in [58] and is depicted in Fig. 2. In this model, a single Tx wishes to communicate a message W of nR independent and uniform data bits to the single Rx by communicating over n uses of a state-dependent discrete memoryless channel. The state-sequence $S_1, \ldots, S_n$ models the parameters one wishes to estimate (e.g., the accelerations of a given target) and in this model are assumed i.i.d. according to a given and known distribution P_S. At the same time, the Tx also aims to estimate the state-sequence $S_1^n = (S_1, \ldots, S_n)$ from generalized feedback signals, which here model the backscatterers observed at the Tx. The communication channel and the radar channel (i.e., the generation of the backscatterers) are jointly modeled by a DMSC with stationary channel law $P_{YZ|XS}(y, z|x, s)$. That means, if at a given time $i \in \{1, \ldots, n\}$ (more precisely for a given channel use i) the Tx feeds input $X_i = x_i$ to the channel and the state realization is $S_i = s_i$, then the Rx's time-i observed channel output Y_i, and the generalized feedback signal Z_i backscattered to the Tx are generated according to the conditional pmf $P_{YZ|XS}(\cdot, \cdot|x_i, s_i)$, irrespective of the past inputs, outputs and state realizations.

The transmitter produces its (potentially random) channel inputs $X_1, \ldots, X_n$ as a function of the message W and the backscattered signals. So, it produces the time-i input as $X_i = \phi_i(W, Z_1, \ldots, Z_{i-1})$, for $i = 1, \ldots, n$. The receiver observes the channel outputs $Y_1, \ldots, Y_n$ corresponding to these inputs and based on the entire sequence produces a guess of the message $\hat{W} = g(Y_1, \ldots, Y_n)$.

Based on the backscattered sequence $Z_1, \ldots, Z_n$ and its produced inputs $X_1, \ldots, X_n$, the transmitter also produces the state estimates $(\hat{S}_1, \ldots, \hat{S}_n) = h(X_1, \ldots, X_n, Z_1, \ldots, Z_n)$. The quality of these 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*:

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

Examples of commonly used distortion functions are the Hamming distortion $d(s, \hat{s}) = \mathbb{1}\{s \neq \hat{s}\}$, which measures the fraction of wrongly reconstructed symbols, or the mean-squared error distortion $d(s, \hat{s}) = (s - \hat{s})^2$, which measures the average squared distance between the reconstruction and source sequences.

A last feature of this model is a cost constraint (such as a power constraint) on the input sequence, as

imposed by many practical communication systems. 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 $b \colon \mathcal{X} \mapsto \mathbb{R}_0^+$. In case of an input power constraint, the cost function $b(x) = x^2$ is implied for radio channels where the power is proportional to the square of the emitted signal (which represents the electromagnetic field).

B. The Capacity-Distortion-Cost Tradeoff

The goal of information-theoretic studies is to identify the optimal performance that can be achieved by choosing the best system implementation under given modeling assumptions. In the present case, we consider the memoryless model introduced in the previous section IV-A and wish to determine the performance of the optimal data encoders and decoders, and state estimators. We will limit the study to encoders/decoders that have arbitrarily small error probabilities when the blocklength n grows without bounds.

Definition 1. *A rate-distortion-cost tuple* $(\mathsf{R}, \mathsf{D}, \mathsf{B})$ *is said achievable if there exists a sequence (in n) of encoding, decoding, and estimation functions* $(\phi_1, \ldots, \phi_n, g, h)$ *that simultaneously satisfy*

$$\lim_{n\to\infty}\Pr[\hat{W} \neq W] = 0, \tag{4a}$$

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

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

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

The main result of Kobayashi et al. in [58], see Theorem 1 ahead,[1] provides an exact characterization of $\mathsf{C}(\mathsf{D}, \mathsf{B})$. A first step to obtain this result is to describe the optimal estimator, which in the present memoryless model is pleasingly simple, because it operates on a symbol-by-symbol basis. That means, estimate $\hat{S}_i$ of the i-th state symbol S_i is solely based on the i-th input X_i and feedback signal Z_i.

Lemma 1 (Lemma 1, [58]). *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

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

---

[1]The work in [58] assumes that the receiver has perfect state information at the receiver. Here, we give the more general result that also includes the case without receiver state-information. Notice that this model is more general because any kind (even imperfect) of receiver side-information can be provided to the receiver as part of the output Y, in which case the original results of [58] are recovered. For a more detailed discussions on this, see [66].

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

To be used later, define the following sets of input distributions:

$$\mathcal{P}_{\mathsf{B}} = \left\{ P_X \;\middle|\; \sum_{x \in \mathcal{X}} P_X(x) b(x) \leq \mathsf{B} \right\}, \tag{8a}$$

$$\mathcal{P}_{\mathsf{D}} = \left\{ P_X \;\middle|\; \sum_{x \in \mathcal{X}} P_X(x) \mathsf{E}[d(S, \hat{s}^*(X, Z))|X = x] \leq \mathsf{D} \right\}. \tag{8b}$$

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

$$\mathsf{D}_{\min}(\mathsf{B}) := \min_{P_X \in \mathcal{P}_{\mathsf{B}}} \sum_{x \in \mathcal{X}} P_X(x) \mathsf{E}[d(S, \hat{s}^*(X, Z))|X = x]. \tag{9}$$

The main result in [58], [66] is the following theorem:

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

$$\mathsf{C}(\mathsf{D}, \mathsf{B}) = \max_{P_X \in (\mathcal{P}_{\mathsf{B}} \cap \mathcal{P}_{\mathsf{D}})} I(X; Y), \qquad \mathsf{D} \geq \mathsf{D}_{\min} \;\; \mathsf{B} \geq 0. \tag{10}$$

Remark 1. *Notice that above result also remains valid if the expected distortion constraint (4b) is replaced by an* excess distortion constraint *where the probability that the sequences exceed average distortion D is required to vanish asymptotically in the blocklength n. The work in [59] considered such an excess distortion criterion but imposed weaker constraints where both the excess distortion probability and the decoding error probability do not necessarily need to vanish asymptotically but simply be bounded by given positive constants $\delta, \epsilon \in (0, 1)$. The analysis in [59] showed that the fundamental limits remain unchanged when the sum of both allowed error probabilities $\delta + \epsilon < 1$. If maximum error probabilities are considered instead of average error probabilities (over messages), then the results remain for all $\epsilon, \delta \in (0, 1)$, and the so called* strong converse *holds. Notice that the similar non-zero error probability behaviours apply also to with only communication but no sensing, because communication takes place over a compound channel, see [20].*

It has been shown in [67] 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 $\mathsf{B} \geq 0$ saturates at the channel capacity without distortion constraints $\mathsf{C}_{\mathrm{NoDist}}(\mathsf{B})$. For many channels, given $\mathsf{B} \geq 0$, the tradeoff $\mathsf{C}(\mathsf{D}, \mathsf{B})$ is strictly increasing in D until it reaches $\mathsf{C}_{\mathrm{NoDist}}(\mathsf{B})$. However, for SDMBCs and costs $\mathsf{B} \geq 0$ where the capacity-achieving input distribution $P_{X_{\max}} := \mathrm{argmax}_{P_X \in \mathcal{P}_{\mathsf{B}}} I(X; Y \mid S)$ also achieves minimum distortion $\mathsf{D}_{\min}(\mathsf{B})$ in (9), the capacity-distortion tradeoff is constant, 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.

To understand the result in Theorem 1, consider the example of a real Gaussian channel with Rayleigh fading and noisy feedback. (For more examples, see [58], [67].) The channel output is thus given by:

$$Y_i = S_i X_i + N_i, \tag{11}$$

where X_i is the channel input satisfying $\overline{\lim}_{n \to \infty} \frac{1}{n} \sum_i \mathsf{E}\left[|X_i|^2\right] \leq \mathsf{B} = 10\mathrm{dB}$, 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 Tx observes the noisy feedback

$$Z_i = Y_i + N_{\mathrm{fb}, i}, \tag{12}$$

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

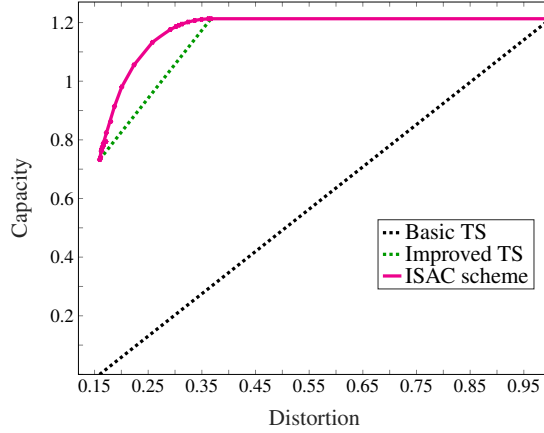The capacity of this channel is achieved with a Gaussian input $X_{\max} \sim \mathcal{N}(0, \mathsf{B})$, and thus the

Fig. 3. Capacity-distortion tradeoff of fading AWGN channel with $\mathsf{B} = 10$ dB and $\sigma_{\mathrm{fb}}^2 = 1$. (Rate is measured in *nats*.)

communication mode with sensing achieves the rate-distortion pair

$$\mathsf{C}_{\mathrm{NoEst}}(\mathsf{B}) = \frac{1}{2}\mathsf{E}\left[\ln(1 + |S|^2\mathsf{B})\right] = 1.213, \tag{13}$$

$$\mathsf{D}_{\mathrm{max}}(\mathsf{B}) = \mathsf{E}\left[\frac{(1 + \sigma_{\mathrm{fb}}^2)}{1 + |X_{\mathrm{max}}|^2 + \sigma_{\mathrm{fb}}^2}\right] = 0.367, \tag{14}$$

where the numerical values correspond to $\sigma_{\mathrm{fb}}^2 = 1$ and $P = 10\mathrm{dB}$ and the logarithm here is with respect to the natural unit and thus messured in *nats*.

Minimum distortion $\mathsf{D}_{\mathrm{min}}$ is achieved by 2-ary pulse amplitude modulation (PAM), and thus the sensing mode with communication achieves rate-distortion pair $(\mathsf{R}_{\mathrm{min}}(\mathsf{B}), \mathsf{D}_{\mathrm{min}}(\mathsf{B})) = (0.733, \frac{1 + \sigma_{\mathrm{fb}}^2}{1 + P + \sigma_{\mathrm{fb}}^2} = 0.166)$ where the numerical value again corresponds to $\sigma_{\mathrm{fb}} = 1$ and $B = 10$ dB. Next, they 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 $(\mathsf{C}_{\mathrm{NoDist}}(\mathsf{B}), \mathsf{D}_{\mathrm{trivial}}(\mathsf{B}) = 1)$. The sensing mode without communication achieves rate-distortion pair $(0, \mathsf{D}_{\mathrm{min}}(\mathsf{B}))$.

In Fig. 3, the rate-distortion tradeoff achieved by these two TS baseline schemes is compared 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 $(\mathsf{R}_{\mathrm{min}}(\mathsf{B}), \mathsf{D}_{\mathrm{min}}(\mathsf{B}))$ and $(\mathsf{C}_{\mathrm{NoEst}}(\mathsf{B}), \mathsf{D}_{\mathrm{max}}(\mathsf{B}))$ of the Improved TS scheme. To obtain a numerical approximation of the points on $\mathsf{C}(\mathsf{D}, \mathsf{B})$ in between these two operating points an alternating optimization method similar to the Blahut-Arimoto algorithm is used in [58].

C. Log-Loss Distortion

The work in [60] considered a related setup, where distortion is measured in terms of log-loss distortion. The goal of sensing is thus rather to obtain a soft estimate, i.e., probability distribution $Q_{\hat{S}^n|X^nZ^n}(\cdot|x^n, z^n)$ for the state estimate, instead of a state-sequence $\hat{S}^n$. The model described in the previous section can be adapted to account for a log-loss distortion constraint simply, where in particular the achievability criterion (4b) has to be replaced by the following requirement:

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\log \frac{1}{Q_{\hat{S}^n|X^nZ^n}(S^n|X^n, Z^n)}\right] \leq \mathsf{D}. \tag{15}$$

In analogy to Lemma 1, it can be argued that the optimal log-loss estimator $Q_{\hat{S}^n|X^nZ^n}$ is in product form $Q_{\hat{S}|XZ}^{\otimes n}$ and given by the posterior probability in (6)

$$Q_{\hat{S}|XZ}^*(s|x,z) = P_{S|XZ}(s|x,z) = \frac{P_S(s)P_{Z|SX}(z|s,x)}{\sum_{\tilde{s}\in\mathcal{S}} P_S(\tilde{s})P_{Z|SX}(z|\tilde{s},x)}, \tag{16}$$

which is solely determined by the channel law and the state distribution but not by the utilized coding scheme.

We can thus conclude that the equivalent of the capacity-distortion-cost tradeoff for log-loss distortion is:

Theorem 2.

$$\mathsf{C}_{\text{LogLoss}}(\mathsf{D},\mathsf{B}) = \max_{\substack{P_X\in\mathcal{P}_\mathcal{B}:\\ H(\hat{S}|XZ)\leq\mathsf{D}}} I(X;Y), \tag{17}$$

where the entropy and the mutual information are calculated according to the joint law $P_{SXYZ} = P_S P_X P_{YZ|XS}$.

Notice again, that any kind of receiver side-information can be incorporated in the received signal Y, and thus can be treated within the exposed framework. Notice further that [60] considers a slightly different approach where the distortion constraint needs to be satisfied for any realization of the message $W = w$. Interestingly, the two models lead to the same capacity-distortion-cost tradeoff and under both models the optimal estimator is the posterior estimator.

D. Finite Block Length Results

A similar model as introduced in [58] is considered in [61]. Definition 1 needs to be adapted for the nonasymptotic regime as follows:

Definition 2. *Given a blocklength n, the rate-distortion-error triple (R,D,ϵ) is said to be achievable if there exist encoding, decoding, and estimation functions $\{f^{(n)}, g^{(n)}, h^{(n)}\}$ satisfying*

$$\frac{1}{n}\log_2(W) \geq R, \tag{18}$$

$$\epsilon^{(n)} \leq \epsilon, \tag{19}$$

$$\Delta^{(n)} \leq \mathsf{D}. \tag{20}$$

The main results of [61] include the derivation of achievability and converse bounds on the rate-distortion-error tradeoff in the finite blocklength regime.

Theorem 3. *Given a blocklength n, the rate-distortion-error tradeoff (R,D,ϵ) is achievable if there exists a P_X and a constant $K > 0$ such that the following conditions are satisfied:[2]*

$$R \leq I(X;Y) - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon - \beta_u) - \frac{K\log(n)}{n}, \tag{21}$$

$$\mathsf{D} \geq \sum_{x\in\mathcal{X}}\sum_{s\in\mathcal{S}}\sum_{z\in\mathcal{Z}} d(s,\hat{s}^*(x,z))P_X(x)P_S(s)P_{Z|XS}(z|x,s), \tag{22}$$

where

$$\beta_u := \frac{1}{nK} + \frac{0.7975T}{\sqrt{nV^3}},$$

---

[2]Notice that here distortion is measured as an expected distortion over all messages. For small number of messages, i.e., small blocklengths n, the encoder and decoder might need additional randomness to construct the desired distribution. A different approach was taken in [88], where sensing performance is measured with an excess distortion criteria.
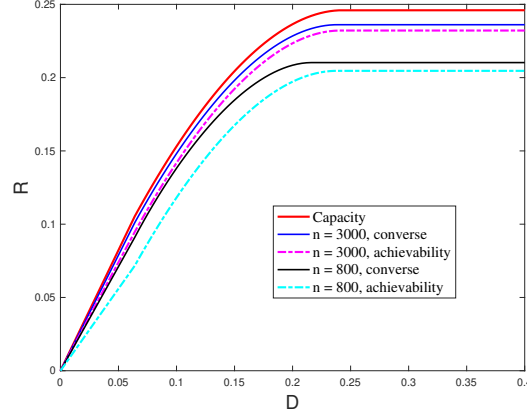
Fig. 4. Achievability and converse bounds on the rate-distortion-error trade-off for $\epsilon = 10^{-3}$, $q = 0.4$, $\mathsf{K} = 0.5$ and different values of the blocklength n.

and the mutual information $I(X;Y)$ and the central moments V and T are defined based on the joint pmf $P_{XY}(x,y) = P_X(x)P_{Y|X}(y|x)$. *On the other hand any rate-distortion-error triple* (R, D, ϵ) *is not achievable if for all* $\delta > 0$ *and pmfs* P_X *satisfying the distortion condition the following lower bound holds:*

$$R \geq I(X;Y) - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon + \beta_l) + \frac{\log(n)}{2n} - \frac{\log \delta}{n},$$

where

$$\beta_l := \frac{0.7975T}{\sqrt{nV^3}} + \frac{\delta}{\sqrt{n}}.$$

Example 1. *Consider the binary channel with multiplicative Bernoulli state:*

$$Y = SX, \tag{23}$$

where all alphabets are binary $\mathcal{X} = \mathcal{S} = \mathcal{Y} \in \{0,1\}$, *the state is Bernoulli-q with* $q \in (0,1)$ *and the feedback is perfect, i.e.,* $Z = Y$. *We consider the Hamming distortion measure* $d(s,\hat{s}) = s \oplus \hat{s}$.

Fig. 4 illustrates the achievability and converse bounds in above theorem for $\epsilon = 10^{-3}$, $q = 0.4$, $\mathsf{K} = 0.5$. *As can be seen from this figure the bounds are tight for large values of* n. *Notice that for* $q = 0.4$ *the capacity of the channel is* $\mathsf{C} = 0.246$ *and the achieved distortion is* $\mathsf{D}_{\mathrm{comm}} = 0.2432$.

E. Channels with Memory

The previous sections assumed a memoryless stationary model both for the channel and the distribution of the target/state that the transmitter wishes to estimate. The work [62] relaxed both assumptions and considered a general model, where the state process $\{S_t\}$ follows an arbitrary joint distribution and the channel is characterized by a general sequence of transition laws $P_{Y_i Z_i | X^i S^i Z^{i-1} Y^{i-1}}$, for $i = 1, 2, \ldots$. The distortion constraint is also generalized beyond average block distortion constraints, by requiring that

$$p - \overline{\lim_{n \to \infty}} \frac{1}{n} d(S^n, \hat{S}^n) \leq \mathsf{D}, \tag{24}$$

for a general non-negative distortion function $d(\cdot, \cdot)$.

Using Han and Verdu's information-spectrum method, [62] characterized the capacity-distortion tradeoff[3] for this setup with memory. It is given by [62]

$$\mathsf{C}(\mathsf{D}) := \sup_{\{P_{X^n}\}_n} p - \varliminf_{n \to \infty} \frac{1}{n} i(X^n; Y^n) \tag{25}$$

where the supremum is over all input distributions $\{P_{X^n}\}$ and estimators $\{\hat{S}^n(Z^n, X^n)\}$ satisfying $\varlimsup_{n\to\infty} \frac{1}{n} d(S^n, \hat{S}^n(X^n, Z^n)) \leq$ D. Here, $i(X^n; Y^n)$ is the previously defined information density between sequences X^n and Y^n.

A slightly different model was considered in [89], where the transmitter has to estimate the state-sequence in an online manner, i.e., state-estimate $\hat{S}_i$ has to be produced after having produced the time-i channel input X_i and having observed the time-i feedback signal Z_i. The capacity-distortion tradeoff was derived for this related model, but limiting to the class of ergodic channels where the sequences of information-densities are sure to converge.

While the generality of the presented "arbitrary/ergodic non-i.i.d." models is appealing, the complexity of the expressions (both from an analytical perspective as well as in view of numerical evaluations) limits the utility of the results. An interesting approach is to consider larger (not only i.i.d.) classes of channels and source sequences for which the capacity-distortion tradeoff still has a relatively simple form.

In this spirit, the work in [89] characterized the capacity-distortion tradeoff of a class of channels that have previously been introduced and studied in the context of pure capacity calculations. This class of channels is also particularly interesting because the numerical calculation of capacity [90] as well as of the capacity-distortion tradeoff can be cast into the framework of Markov decision processes and thus solved using reinforcement learning (RL) as well as its many more advanced alternatives that have been introduced in recent years such as Q-learning etc.

An RL approach has been followed in [89] to evaluate the capacity-distortion tradeoff for a specific class of binary channels. Interestingly, the authors in [89] also analyzed the influence of the size of the state-space considered in the RL approach, which corresponds to the memory in the coding strategy employed at the transmitter. Figure 5 plots a weighted sum between the information rate and the distortion in function of the weight factor β. It shows the performances achieved by four versions of the RL approach allowing for different sizes of the state-spaces: a full state-space; a highly-reduced state-space that only allows to implement memoryless policies; and intermediate state-spaces with sizes equal to 10% or 40% of the full state-space and thus allowing to implement coding strategies with a limited amount of memory.
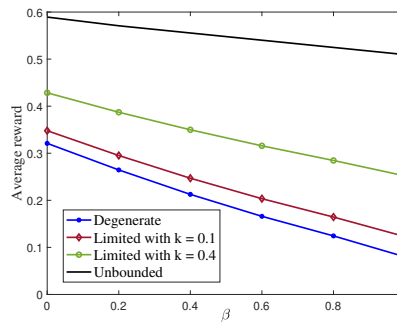

Fig. 5. RL-Average Reward composed of β times the information rate and $(1 - \beta)$ times the negative sensing distortion.

V. SENSING AT THE RX (RX-ISAC) WITH SENSING DISTORTION

In certain practical systems, sensing is performed at a device that differs from the radar-emitting device. Such situations are often referred to as bi-static radar. Information-theory literature has considered various

---

[3]Additional cost constraints can be included in the model and the results in a standard way.

bi-static ISAC scenarios [63]–[65], [91], [92]. In this section, we focus on bi-static ISAC where the sensing task is performed at the communication Rx, and the sensing task is to estimate the target (state) up to a given distortion constraint.
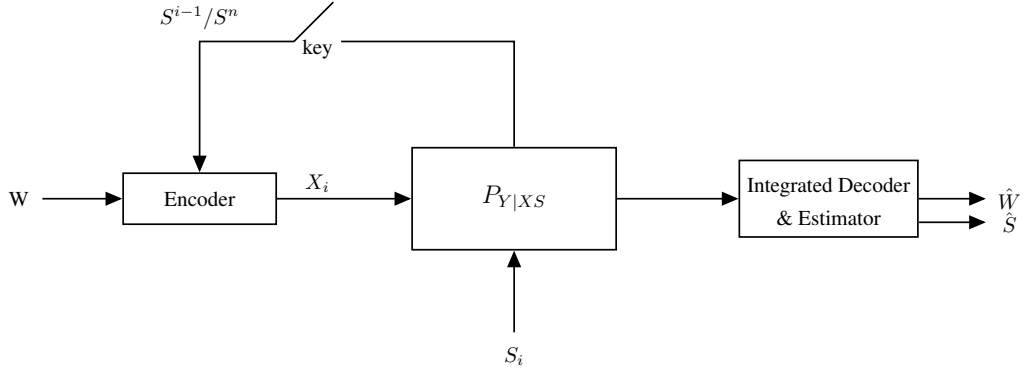

Fig. 6. ISAC with Rx sensing

The presented model captures one of the major challenges in bi-static ISAC, which is that the sensing terminal is à priori not aware of the channel input sequence and thus due to the memory in the channel input sequence symbol-by-symbol estimators based solely on the observations are suboptimal. We shall see that when the sensing is performed at the communication Rx this difficulty is easily solved by first decoding the data and reconstructing the input sequence which contains all the memory in the system. In this case, a symbol-wise estimator based on this reconstructed input sequence and the observed sequence of channel outputs achieves minimum distortion. (As we shall see in the section on Network ISAC, it is more complicated to characterize the optimal estimation strategy when the sensing terminal is not à priori required to decode all the transmitted data and codewords.)

A. A Memoryless Model

We consider a similar memoryless model as in the previous section. A single Tx wishes to communicate a message W to a single-receiver over a state-dependent channel and the Rx aims to decode this message and at the same time also estimate the channel state-sequence up to the allowed distortion, see Figure 6. In other words, the Rx applies a decoding function $g(\cdot)$ to its outputs to produce a message guess $\hat{W} = g(Y^n)$ and also a estimation function $h(\cdot)$ to produce the estimates $\hat{S}^n = h(Y^n)$. To allow for a general model, we also include the models where the Tx knows the state-sequence S^n in one of the following ways:

- the Tx has no information about S^n;
- the Tx knows the entire sequence S^n non-causally, i.e., already before the entire transmission starts;
- the Tx knows S^n in a strictly causal way, i.e., it learns S_i only after channel use i and prior to channel use $i + 1$;
- the Tx knows S^n in a causal way, i.e., it learns S_i just before channel use i.

Depending on the available state-information, the Tx produces its time-i channel input either as a function of only the message W and the previous generalized feedback Z^{i-1}, or also in function of the entire state-sequence S^n (for the non-causal case), of the previous and the current state S^i (for the causal case), or of the previous states S^{i-1} only (for the strictly causal case).

The definition of the capacity-distortion function is analogous to Definition 1, but where encoding, decoding, and state-estimation functions are as described above. Moreover, here we do not consider cost constraints (which however could easily be included).

B. Capacity-Distortion Tradeoffs

We start with the model without state-information at the Tx. In this case, the optimal estimator at the Rx is a symbol-by-symbol estimator based on the observed sequence of outputs and the decoded codeword, and the capacity-distortion tradeoff was characterized in [63].

Theorem 4. *When the Tx has no knowledge about the state-sequence S^n, the capacity-distortion function is given by*

$$\mathcal{C}_{\text{No-CSI}} := \max_{P_X \in \mathcal{P}_{\text{D}}} I(X; Y), \tag{26}$$

where

$$\mathcal{P}_{\text{D}} = \left\{ P_X \ \Big| \ \sum_{x \in \mathcal{X}} P_X(x) E[d(S, \hat{s}^*(X, Y))] \le \text{D} \right\}, \tag{27}$$

and $\hat{s}^*(\cdot, \cdot)$ is the optimal estimator introduced in (51) *with the feedback output Z replaced by the decoder output Y.*

This capacity-distortion tradeoff for the setup where the Tx is not informed about the state-sequence was also extended to a multi-access setup with multiple transmitter, see [91] and to a two-hop setup [93]. In the latter work it is shown that a decode-(indirectly)-compress-and-forward strategy achieves the capacity-distortion function.

Consider next the scenarios where the Tx does learn the state-sequence S^n either causally or strictly causally, see [64], [92]. In these cases, the Tx wishes to assist the Rx in the sensing task by conveying information about the state-sequence to the receiver, in the same spirit as it sends data. In other words, the Tx will compress the observed state-sequence and send the compression information to the Rx which then reconstructs the compressed version of the state. The Rx finally applies an optimal symbol-by-symbol estimator to this compressed sequence as well as to the decoded input codewords and the observed channel outputs.

In case the Tx observes the state-sequence only causally or even strictly-causally, it has to employ a block-Markov coding scheme, where in each block it sends compression information about the state-sequence from the previous block. Transmission of this compression information and of the data is performed using an optimal data-communication scheme. Specifically, in the setup with strictly-causal state-information, a standard channel code is used that ignores the state-information completely. For the setup with causal state information the Tx has to resort to Shannon strategies which have been shown to achieve capacity in these setups. Notice that under Shannon strategies the channel inputs are generated symbolwise from an auxiliary codeword and the state-sequence. The Rx thus does not have access to the channel inputs even when it decodes the codewords correctly. Nevertheless, it can be shown that the symbolwise estimator based on the decoded codewords and the observed channel output sequence achieves the optimal Rx sensing performance.

Theorem 5 (Theorem 2, [64])**.** *The capacity–distortion function for strictly causal state communication is*

$$\mathsf{C}_{\text{Str-caus.}}(\mathsf{D}) = \max_{P_X P_{U|XS}} \left(I(U, X; Y) - I(U, X; S) \right), \tag{28}$$

where the maximum is over all laws $P_X P_{U|XS}$ such that $\mathbb{E}[d(S, \hat{s}^*(U, X, Y))] \le \mathsf{D}$, where (U, S, X, Y) are distributed according to $P_S P_X P_{U|XS} P_{Y|XS}$ and

$$\hat{s}^*(u, x, y) = \arg\min_{s' \in \hat{\mathcal{S}}} \sum_{s \in \mathcal{S}} P_{S|UXY}(s|u, x, y) d(s, s'). \tag{29}$$

The capacity-distortion function for causal state-communication is

$$\mathsf{C}_{\text{Caus.}}(\mathsf{D}) = \max \big(I(U, V; Y) - I(U, V; S) \big) \tag{30}$$

where the maximum is over all laws $P_V P_{U|V,S}$ and functions $x(v, s)$ such that $\mathbb{E}[d(S, \hat{s}^*(U, V, Y))] \leq \mathsf{D}$, for (U, V, S, X, Y) distributed according to $P_V P_S P_{U|VS} \mathbb{1}\{X = x(V, S)\} P_{Y|XS}$ and here

$$\hat{s}^*(u, v, y) = \arg\min_{s' \in \hat{\mathcal{S}}} \sum_{s \in \mathcal{S}} P_{S|UVY}(s|u, x, y) d(s, s'). \tag{31}$$

In above expressions, the U- and V-auxliaries stand for the auxiliary codewords. The subtracted mutual information terms arise because the Tx transmits compression information together with the data, and thus the rate of the compression information needs to be deduced from the total rate of communication that can be sustained from the Tx to the Rx.

When the Tx observes the state-sequence non-causally, no block-Markov strategies are necessary. Gel'fand-Pinsker (GP) coding [94], which achieves capacity for channels with non-causal state-information at the Tx, is used to transmit the data and the compression information to the Rx. In GP coding, again the channel inputs are obtained as a function of auxiliary codewords and the state-sequence. The Rx thus again cannot reconstruct the sequence of channel inputs, even after decoding the messages correctly. However, again, a symbol-by-symbol estimator based on the decoded codewords and the observed sequences achieves the optimal Rx sensing perfomance.

For this setup, with non-causal state-information at the Tx, the exact capacity-distortion tradeoff is generally still an open problem, only upper and lower bounds are known [92]. The work in [65] has caracterized the exact capacity-distortion tradeoff in case of a Gaussian model with mean squared-error (MSE) distortion. This is, for a scenario where the time-t channel output is given by $Y_t = X_t + S_t + N_t$ for X_t the channel input, S_t a Gaussian state of variance Q; and N_t a Gaussian noise of variance N, and $d(s, \hat{s}) = (s - \hat{s})^2$. In this case, the capacity-distortion tradeoff was derived in [65].

Theorem 6. *The capacity-distortion tradeoff with non-causal state-information at the Tx in the Gaussian case is given by:*

$$\mathsf{C}_{\text{Gaus.}}(\mathsf{D}) = \max_{r \in [0,1]} \frac{1}{2} \log\left(1 + \frac{rP}{N}\right), \tag{32}$$

where the maximum is over all values of r satisfying

$$\mathsf{D} \geq Q \frac{rP + N}{(\sqrt{Q} + \sqrt{(1-r)P})^2 + rP + N}. \tag{33}$$

In above theorem, the parameter $r \in [0, 1]$ indicates the fraction of the transmit power that the Tx uses for data transmission, i.e., to encode the message. The rest of the power, i.e., a fraction $1 - r$ of the total power, is used to send channel state-information (in an uncoded manner) to the Rx.

VI. NETWORK ISAC WITH SENSING DISTORTION

Modern communication systems are often multi-user and network-oriented, meaning that multiple Txs wish to simultaneously transmit data to multiple Rxs and some of these terminals have to accomplish sensing tasks. Characterizing the information-theoretic fundamental limits of multi-user network systems has been an active area of research for decades [95], and a vast majority of the systems still lack complete and computable characterizations of the fundamental performance limits, even when only data has to be transmitted, i.e., for systems without sensing tasks. Nevertheless, different interesting and insightful code constructions have been proposed for network communication systems and it has been shown that they perform reasonably close to the fundamental limits. In recent studies, people have introduced sensing tasks

into these code constructions, so as to obtain information-theoretic network ISAC schemes. Information-theoretic converse (infeasibility) results have also been derived for certain network ISAC scenarios. In this section, we shall review both network ISAC coding schemes and converse results. Notice that network ISAC has also received significant attention in the signal-processing and communication theory literature. We refer to [33], [96] for these results.

We will start by reviewing a broadcast ISAC where communication is from a single Tx to multiple Rxs and the sensing is performed at the Tx. As we shall see, the sensing problem is the same as in the point-to-point communication scenario, and thus the simple symbol-wise estimator in (5) is optimal, so that the sensing problem and the communication problem "decouple" similarly to the point-to-point case. The second scenario that we consider in this section is the multi-access ISAC problem with Tx sensing. Since the sensing task is accomplished at multiple distributed terminals with heterogeneous sensing information, this sensing problem is fundamentally different and allows for more complicated strategies, e.g. collaborative sensing strategies and interactive exchange of sensing information between the different terminals. We shall present different ISAC coding schemes that perform the required communication tasks, and at the same time also exchange sensing information, thus allowing to implement collaborative sensing strategies. Similar strategies have been proposed also for device-to-device (D2D) communication (the two-way channel) and the interference channel (IC).

A. One-to-Many Communication (Broadcast Channels) with Tx-Sensing

1) The Memoryless Model: Consider the single-Tx two-Rx broadcast ISAC system, which is depicted in Figure 7. Extensions to multiple Rxs follow standard techniques. The setup is similar
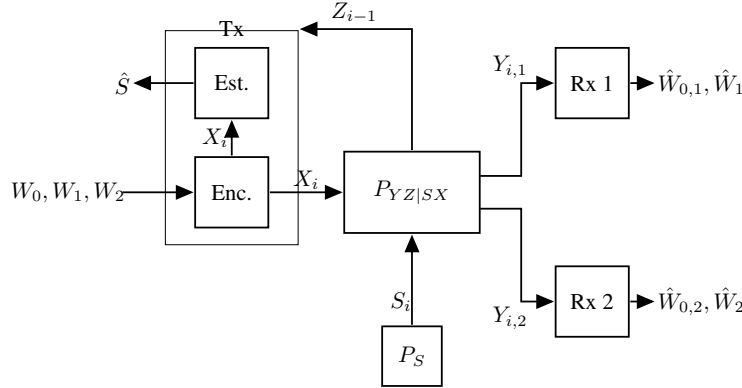

Fig. 7. State-dependent broadcast channel with generalized feedback and state-estimator at the Tx.

to the single-user setup in Subsection IV-A, however communication is to two distinct Rxs 1 and 2. Specifically, the Tx wishes to communicate the rate-R_0 message W_0 to both Rxs, the rate-R_1 message W_1 to Rx 1, and the rate-R_2 message W_2 to Rx 2. The Tx thus produces inputs of the form $X_i = \phi_i(W_0, W_1, W_2, Z_1, \ldots, Z_{i-1})$, for $i = 1, \ldots, n$. The communication channel and the generalized feedback channel are governed by a state-sequence S^n that is i.i.d. according to P_S and is jointly modelled by a stationary memoryless channel of transition probabilities $P_{Y_1 Y_2 Z | XS}(y_1, y_2, z | x, s)$ determining the outputs $\{Y_{1,i}\}$ at Rx 1, the outputs $\{Y_{2,i}\}$ at Rx 2, and the generalized feedback signals $\{Z_i\}$ at the Tx. Based on its observed channel outputs $Y_{k,1}, \ldots, Y_{k,n}$, each Rx $k \in \{1, 2\}$ produces the guesses $\hat{W}_{0,k}$ and $\hat{W}_k$ of the messages W_0 and W_k using appropriate decoding functions $(\hat{W}_{0,k}, \hat{W}_k) = g_k(Y_{k,1}, \ldots, Y_{k,n})$, and the Tx estimates the state-sequences as $(\hat{S}_1, \ldots, \hat{S}_n) = h(X_1, \ldots, X_n, Z_1, \ldots, Z_n)$. As before, communication performance is measured in terms of decoding error probabilities and sensing performance in terms of expected average per-block distortion.

Accordingly, we have the following achievability definition, for a given bounded and non-negative distortion function $d(\cdot, \cdot)$.

Definition 3. *A rate-distortion tuple* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D})$ *is achievable if there exists a sequence (in n) of encoding, decoding, and state estimation functions such that*

$$\lim_{n \to \infty} \Pr\left(\hat{W}_k \neq W_k \quad \text{or} \quad \hat{W}_{0,k} \neq W_0\right) = 0, \qquad k \in \{1, 2\}, \tag{34a}$$

$$\overline{\lim_{n \to \infty}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d(S_i, \hat{S}_i)] \leq \mathsf{D} \tag{34b}$$

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

Remark 2. *Above model was considered in [66]. The authors in [67] considered a slightly different model where the state is composed of two components $S = (S_1, S_2)$, where each S_k is revealed to the corresponding Rx and has to be estimated at the Tx up to a maximum allowed distortion D_k. As already mentioned for the point-to-point setup, receiver state-information is included in above model as a special case, by including the state informations as part of the channel outputs. Similarly, since the state-alphabet can be arbitrary but finite, in above model S can also be a pair of finite states (S_1, S_2). To be able to fully capture the setup and the results in [67] as special cases, it thus suffices to extend above model from [66] to multiple distortion constraints, which can easily be done. The advantage of the model in [66] is that it is more general and allows to model all kinds perfect or imperfect state-information at the Rxs.*

2) Results: The optimal estimator is again given by Lemma 1. That means, the optimal estimator is (irrespective of the choice of the encoding and decoding functions)

$$h_k^*(x^n, z^n) = (\hat{s}_k^*(x_1, z_1), \hat{s}_k^*(x_2, z_2), \ldots, \hat{s}_k^*(x_n, z_n)). \tag{35}$$

where $\hat{s}^*(x, z) = \arg\min_{s' \in \hat{\mathcal{S}}} \sum_{s \in \mathcal{S}} P_{S|XZ}(s|x, z) d(s, s')$ was defined in (5) and the conditional probability distribution $P_{S|XZ}(s|x, z) = \frac{P_S(s) P_{Z|SX}(z|s, x)}{\sum_{\tilde{s} \in \mathcal{S}} P_S(\tilde{s}) P_{Z|SX}(z|\tilde{s}, x)}$ again only depends on the channel and the state distribution.

Identification of this optimal estimator immediately allows to reduce the problem of characterizing the capacity-distortion tradeoff region $\mathcal{CD}$ of the broadcast channel (BC) to the problem of identifying the set of communication rates that are achievable under a given constraint on the statistics of the channel input symbols. In this sense, we again notice a decoupling of the sensing problem and the communication problem for the BC as for the point-to-point channel. The communication problem needs to be solved under a constraint on the channel input statistics, but otherwise the sensing part does not interfere.

The pure communication problem over a memoryless BC with feedback is still open, and only inner and outer bounds are known for general channels. Notable exceptions are the classes of physically degraded BCs [97] (where feedback does not increase capacity) and other classes of BCs with states [7]. For these classes, with help of the optimal estimator in (35) one can immediately characterize the capacity distortion region $\mathcal{CD}$, see [67]. For all other classes, the optimal estimator can be combined with the proposed coding schemes for BCs with feedback [2]–[6] the known infeasibility proofs (converses) to obtain inner and outer bounds on $\mathcal{CD}$ for general ISAC BCs.

The following inner and outer bounds on $\mathcal{CD}$ were reported in [66], see also [67].

Theorem 7 (Outer Bound). *If* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D})$ *lies in* $\mathcal{CD}$, *then there exist pmfs* $P_X, P_{U_1|X}, P_{U_2|X}$ *such that the random tuple* $(U_k, X, S, Y_1, Y_2, Z) \sim P_{U_k|X} P_X P_S P_{Y_1 Y_2 Z|SX}$ *satisfies the rate constraints*

$$\mathsf{R}_0 + \mathsf{R}_k \leq I(U_k; Y_k), \quad k = 1, 2, \tag{36a}$$

$$\mathsf{R}_0 + \mathsf{R}_1 \le I(U_0, U_1; Y_1, V_1) - I(U_0, U_1, U_2, Z; V_0, V_1 \mid Y_1) \tag{38a}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \le I(U_0, U_2; Y_2, V_2) - I(U_0, U_1, U_2, Z; V_0, V_2 \mid Y_2) \tag{38b}$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \le I(U_1; Y_1, V_1 | U_0) + I(U_2; Y_2, V_2 \mid U_0) + \min_{k \in \{1,2\}} I(U_0; Y_k, V_k) - I(U_1; U_2 \mid U_0)$$

$$- I(U_0, U_1, U_2, Z; V_1 \mid V_0, Y_1) - I(U_0, U_1, U_2, Z; V_2 | V_0, Y_2)$$

$$- \max_{k \in \{1,2\}} I(U_0, U_1, U_2, Z; V_0 \mid Y_k) \tag{38c}$$

$$2\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \le I(U_0, U_1; Y_1, V_1) + I(U_0, U_2; Y_2, V_2) - I(U_1; U_2 \mid U_0)$$

$$- I(U_0, U_1, U_2, Z; V_0, V_1 \mid Y_1) - I(U_0, U_1, U_2, Z; V_0, V_2 \mid Y_2) \tag{38d}$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \le I(X; Y_1, Y_2), \tag{36b}$$

and the average distortion constraints

$$\mathbb{E}[d(S, \hat{s}^*(X, Z))] \le \mathsf{D}. \tag{37}$$

Proposition 1 (Inner Bound)**.** *The capacity-distortion region $\mathcal{CD}$ includes all tuples $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D})$ that for some choice of the auxiliaries $(U_0, U_1, U_2, X, S, Y_1, Y_2, Z, V_0, V_1, V_2) \sim P_{U_0 U_1 U_2 X} P_S P_{Y_1 Y_2 Z | SX} P_{V_0 V_1 V_2 | U_0 U_1 U_2 Z}$ satisfy inequalities (38) on top of this page and the distortion constraint (37).*
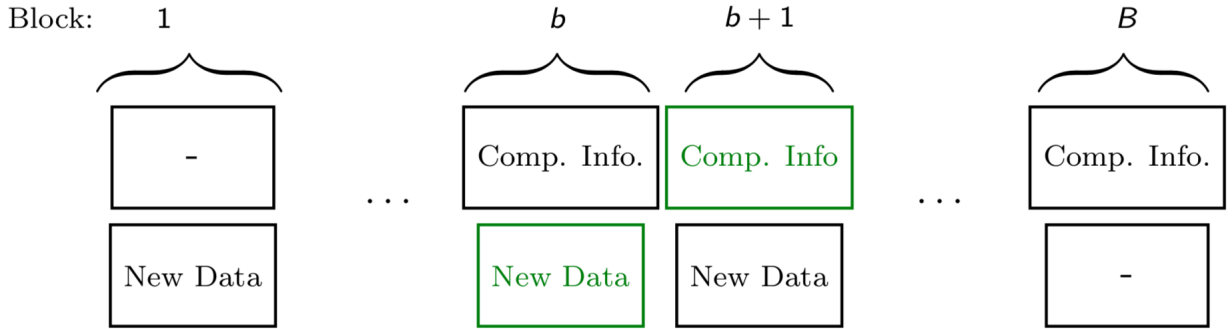

Fig. 8. Block Markov coding structure.

The outer bound is obtained by considering a genie-aided system where Rx 2 observes not only the Y_2 outputs but also the Y_1 outputs, and by using the optimal estimator in (35). The inner bound is obtained by combining again this optimal estimator with the scheme in [5] for broadcast communication with generalized feedback. The scheme in [5] is based on a block-Markov strategy (see Figure 8) where the Tx uses the generalized feedback signals in a block and its own transmitted signal in the same block to identify correlated compression information for both Rxs to improve their decodings. It then sends this update (compression) information in the following block, where the correlation allows the Tx to send part of the compression information as common information that is simultaneously useful for both Rxs, which is more efficient than sending individual information to the Rxs and thus improves over no-feedback communication. Technically speaking, the common information is created using distributed compression techniques à la Gray-Wyner [29]. Decoding is performed backward, starts from the last block, where first the refinement information is decoded and then used to facilitate decoding of the previous block.

Both the inner bounds and the outer bounds are expressed with the help of *auxiliary random variables*. Examining the details of the proof of the inner bound in [67], the auxiliary random variables U_0, U_1, U_2 are easily identified with the different types of codewords used in the code construction. The auxiliary random variables V_0, V_1, V_2 are identified with codewords compressing the feedback signals and the auxiliary codewords corresponding to U_0, U_1, U_2. The U_k auxiliary random variables however again point to a superposition structure given the Markov chains $U_k - X - Y_k$.

More from a technical perspective, the auxiliary random variables allow to obtain inner and outer bounds that can be expressed as single-letter optimization problems. On the negative side, these optimization problems often still have high computational complexities.

3) Example: Consider the physically degraded broadcast channel with binary input and output alphabets $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$ and two-bit state alphabet $\mathcal{S} = \{0, 1\}^2$, i.e., the state S can be written as $S = (S_1, S_2)$ with binary S_1 and S_2. To describe the channel, let $Y_k' = S_k \cdot X$ for each Rx $k \in \{1, 2\}$, where the joint state pmf is:

$$P_{S_1 S_2}(s_1, s_2) = \begin{cases} 1 - q, & \text{if } (s_1, s_2) = (0, 0) \\ 0, & \text{if } (s_1, s_2) = (0, 1) \\ q\gamma, & \text{if } (s_1, s_2) = (1, 1) \\ q(1 - \gamma) & \text{if } (s_1, s_2) = (1, 0), \end{cases} \tag{39}$$

for a real number $\gamma, q \in [0, 1]$. The generalized feedback signals are $Z = (Y_1', Y_2')$ and the Rx outputs $Y_k = (Y_k', S_k)$, which means that each Rx is informed of its corresponding state. Distortion is measured in terms of Hamming distortion between S_1 and an optimal estimator of S_1 based on (X, Z).

Notice that S_2 is a degraded version of S_1, which together with the transition law ensures the Markov chain $X \multimap (S_1, Y_1') \multimap (S_2, Y_2')$ and the physically degradedness of the BC. For physically degraded BCs the presented inner and outer bounds coincide [67] and thus we can obtain the exact characterization of the capacity-distortion tradeoff of this example, which is shown numerically in Figure 9.
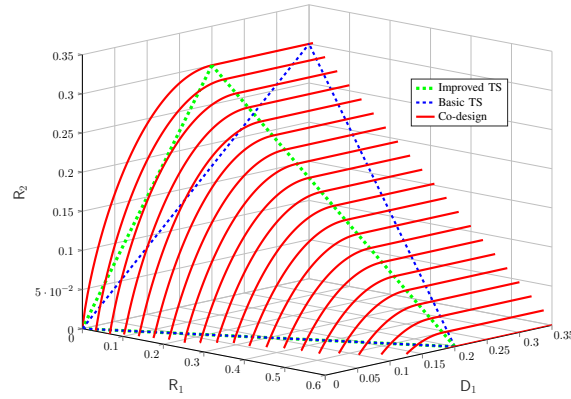

Fig. 9. Capacity-distortion region for proposed example.

We observe a tradeoff between the two rates R_1 and R_2 and the permissible distortion D. Moreover, resource-sharing strategies are highly suboptimal as for the point-to-point case.

B. Multi-Access ISAC: Collaborative Sensing and Suboptimality of Symbol-Wise Estimators

This section reviews information-theoretic models for ISAC over multi-access channels (MAC). The first information-theoretic ISAC MAC scheme was proposed in [68] based on Willems' coding scheme

for data communication [98]. Willem's scheme is again based on a block-Markov strategy where in each block the Txs not only send fresh data but also update information pertaining to the previous block. Again, the update information can be sent in a collaborative way, which renders the communication more efficient.

In [70], [71] an improved collaborative ISAC scheme has been proposed where the two Txs not only cooperate for the purpose of data transmission but also for the purpose of exchanging sensing information from one Tx to the other, so as to allow it to improve its sensing performance. More specifically, in [70], [71] the two Txs exchange sensing information and data in each block, where the exchanged data is then retransmitted in the next block to improve the decoding performance at the Rx. Recently, in [69], proposed a further improvement where the common update information sent by the two Txs not only consists of data but also includes sensing information, allowing the Rx to obtain better state-information and thus improve its decoding performance. (In the discussed models, the Rx has no sensing task, which however could easily be included.)

1) The memoryless model: The model is similar to before, however now we have two Txs and a single Rx, see Fig. 10. Each Tx $k \in \{1, 2\}$ wishes to send a rate-R_k message to the Rx and estimate a memoryless state-sequence $\{S_{k,i}\}$. The sequence of pairs $\{(S_{1,i}, S_{2,i})\}_{i \geq 1}$ are i.i.d. according to a given joint pmf $P_{S_1 S_2}$. The channel input-output relation is specified by the memoryless and stationary channel transition law $P_{Y Z_1 Z_2 | S_1 S_2 X_1 X_2}$. Based on the two messages W_1 and W_2 and the past generalized feedback signals $Z_{k,1}, \ldots, Z_{k,i-1}$, each Tx $k \in \{1, 2\}$ generates its time-i channel input as $X_{k,i} = \phi_{k,i}(W_k, Z_{k,1}, \ldots, Z_{k,i-1})$ and at the end of the communication it estimates the state-sequence as $\hat{S}_k^n = h_k(X_k^n, Z_k^n)$. The estimated sequence $\hat{S}_k^n$ should match the state-sequence S_k^n up to distortion level D_k when measured by a given per-symbol distortion function $d_k(\cdot, \cdot)$. The receiver decodes both messages W_1 and W_2 based on its observed channel outputs as $(\hat{W}_1, \hat{W}_2) = g(Y^n)$.
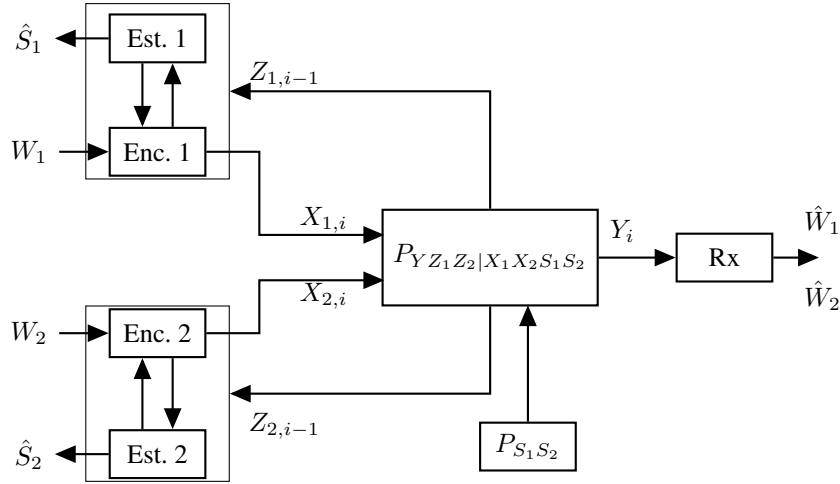

Fig. 10. State-dependent discrete memoryless multiaccess channel with sensing at the transmitters.

Definition 4. *A rate-distortion tuple* (R_1, R_2, D_1, D_2) *is called achievable in the setup above if there exists a sequence (in n) of encoding, decoding, and estimation functions such that*

$$\lim_{n \to \infty} \Pr\left(\hat{W}_1 \neq W_1 \quad \text{or} \quad \hat{W}_2 \neq W_2\right) = 0 \tag{40a}$$

$$\varlimsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d_k(S_{k,i}, \hat{S}_{k,i})] \leq D_k, \quad \text{for } k \in \{1, 2\}. \tag{40b}$$

In this multi-access ISAC setup, the closure of the set of all achievable tuples $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *is called the ISAC-MAC capacity-distortion region* $\mathcal{CD}$.

Like the previous point-to-point and broadcast ISAC models, also above MAC model includes scenarios with (perfect or imperfect) Rx channel state information as special cases. (The state-information can simply be added as part of the output.) Notice further that above model also includes scenarios where the channel is governed by an internal i.i.d. state-sequence S^n of pmf P_S and the states S_1^n, S_2^n are obtained from S^n over an independent memoryless channel $P_{S_1 S_2 | S}$.

2) Results: Determining the set of all achievable rates for the MAC with feedback is even open for only data communication, without the sensing task. Only a non-computable multi-letter expression is known in the general case [99]. Exceptions are the Gaussian MAC with perfect feedback [100] and a class of semi-deterministic MACs with one- or two-sided perfect feedback [101]. Various coding schemes [98], [102]–[104] have been proposed, as well as an outer bound on the feedback capacity based on the dependence balance bound [105]. A recurrent theme in the presented coding schemes is that the feedback links to the two Txs allow to build up cooperation between the Txs. In fact, the feedback links establish a communication path from one Tx to the other, and the two Txs can thus (either implicitly as in [100] or explicitly as in [98], [102]–[104]) align future channel inputs through cooperation, which amplifies the signals compared to the noise and allows for a better decoding performance at the Rx. As we shall see, the same idea is also key for proposing good MAC ISAC schemes.

There is a second fundamental idea that is required to achieve good ISAC MAC schemes, as we shall see in the following. It is inspired from and closely related to the works on multi-access communication over state-dependent channels where the Txs both have state-information, see for example, [106]–[110]

We first present infeasibility results for the ISAC multi-access problem. A first outer bound on the capacity distortion region $\mathcal{CD}$ was established in [68] and then improved in [69]. The outer bound in [69] is:

Theorem 8 (Outer Bound)**.** *The capacity-distortion region of the ISAC MAC* $\mathcal{CD}$ *is included in the set of all tuples* (R_1, R_2, D_1, D_2) *that for some pomf* $P_{QQ_Z} P_{X_1 X_2 | QQ_Z}$ *satisfy:*

$$R_1 \leq I(X_1; YZ_1Z_2 | X_2 QQ_Z), \tag{41}$$

$$R_2 \leq I(X_2; YZ_1Z_2 | X_1 QQ_Z), \tag{42}$$

$$R_1 + R_2 \leq I(X_1 X_2; YZ_1Z_2 | QQ_Z), \tag{43}$$

$$R_1 + R_2 \leq I(X_1 X_2; Y), \tag{44}$$

with the dependence balance constraint:

$$I(X_1; X_2 | QQ_Z) \leq I(X_1; X_2 | Z_1 Z_2 QQ_Z), \tag{45}$$

and the sensing constraints:

$$D_k \geq \mathsf{E}[d(S_k, \hat{s}^*(Z_1, Z_2, X_1, X_2))], \quad k \in \{1, 2\}, \tag{46}$$

$$f_{k,R-D}(D_k) \leq I(S_k X_{k'}; Z_k | X_k Q), \quad k', k \in \{1, 2\}, \ k' \neq k, \tag{47}$$

$$f_{k,R-D}(D_k) \leq I(S_k; Z_1 Z_2 | X_1 X_2 Q), \quad k \in \{1, 2\}, \tag{48}$$

where $f_{k,R-D}(D_k)$ *is the standard rate-distortion function of source* S_k.

It suffices to consider Q *and* Q_Z *whose alphabets* $\mathcal{Q}$ *and* $\mathcal{Q}_Z$ *have cardinalities satisfying* $|\mathcal{Q}| \cdot |\mathcal{Q}_Z| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 3$.

The outer bound is obtained by combining standard information-theoretic bounding steps with the following three key ideas: 1) providing Tx k also with Tx k' inputs and outputs $(X_{k'}^n, Z_{k'}^n)$, for $k, k' \in \{1, 2\}$ and $k \neq k'$, during the sensing task can only improve sensing performance and leads to constraint

(46); 2) applying dependence balance considerations à la Hekstra and Willems [105] based on the pair of generalized feedback outputs (Z_1, Z_2) yields a valid constraint, see (45); and 3) the sensing distortion at Tx k cannot be smaller than the minimum sensing distortion in a joint-source channel coding problem where the source S_k^n is transmitted from Tx k' to k, see constraints (47) and (48). The former two key ideas were already exploited to derive the outer bound in [68]. Idea 3) was proposed in [69] and allows to obtain a strictly improved bound.

A first coding scheme (and thus achievability result) for the ISAC MAC was proposed in [68] based on Willems' scheme for multi-access communication with feedback in [101]. The scheme is again based on a block-Markov strategy where in each block the two Txs send new independent data as well as common update information that will decoded at the Rx. More specifically, the scheme is illustrated in Figure 11 and each block consists of three layers, where the top-most layer is most difficult to decode and the lowest-layer easiest. Both Txs send the same lowest-layer which thus can be transmitted in a
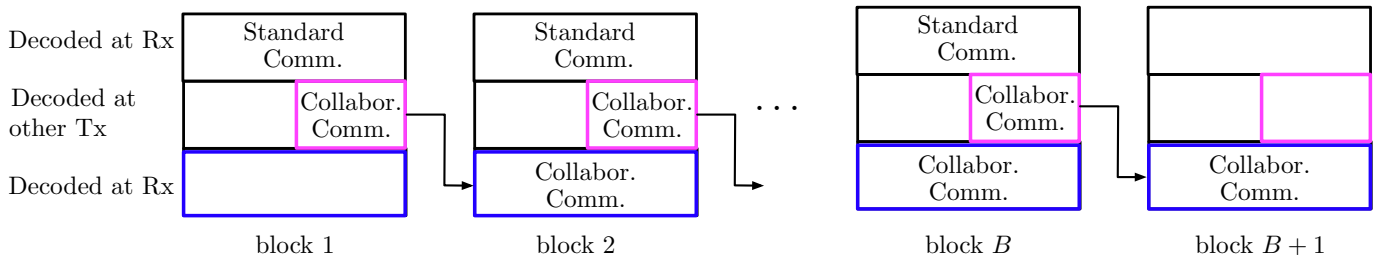

Fig. 11. Block-Markov strategy of Willems' multi-access scheme with generalized feedback.

cooperative manner, while the upper two layers are independent across the two transmitters. The details of the three layers are as follows:

- In the top layer, each Tx independently sends new data in each block. This data is decoded at the Rx only, following the backward decoding algorithm described later.
- In the medium layer, each Tx independently sends new data in each block. This data is decoded at the other Tx at the end of the block and at the Rx following the backward decoding algorithm described later.
- In the lowest layer, the two Txs cooperate and jointly resend the data sent by the two Txs in the medium layer of the previous block. (Recall that the medium layer data of the previous block has been decoded by the other Tx at the end of the previous block.) This data is decoded at the Rx following the backward decoding algorithm described next.

The receiver decodes all transmitted data using a backward decoding procedure, starting from the last block. Specifically, for each block it decodes the data in the top and lowest layer, while it already is informed of the data sent in the middle layer, because it has decoded it in the previous step..

Each Tx k produces its state estimates $\hat{S}_k^n$ by using an optimal symbolwise estimator based on its own inputs X_k^n, its own observed generalized feedback signals Z_k^n, and also the middle-layer codeword symbols $U_{k'}^n$ decoded from the other Tx $k' \neq k$. It is clearly suboptimal for Tx k to estimate its state-sequence S_k^n simply based on its inputs X_k^n and its feedback signals Z_k^n, and an improved performance can be obtained by attempting to decode also the codewords transmitted by the other Tx k'.

In [68] the sensing tasks and data communication tasks are thus still considered individually. A first joint approach was considered in [70], [71], where sensing information was introduced to the coding scheme so as to allow for collaborative sensing, in other words to allow each Tx to exploit sensing information available at the other Tx. On a technical level, this was enabled by having each Tx k compress the signals $U_{k'}^N$, X_k^N, and Z_k^N of a given block and send the compression information (described in bits) as additional information in the medium-layer of the next codeword, see Figure 12. This way, any of the two Txs can

convey sensing information to the other Tx over the Tx-to-Tx path, because the information in the medium layer is decoded at the other Tx. The compression information is also decoded at the Rx, and used to improve decoding of the transmitted data. For the compression of the sensing information, the scheme in [70], [71] uses implicit binning, i.e., the Tx as well as the Rx use their side-information about the compressed sequences from their own inputs and observations to reconstruct the sensing information. This allows the scheme to occupy less rate in the medium-layer codewords and thus improve communication efficiency.

The other encoding and decoding steps are as in the scheme in [68] and described previously. Each Tx k now performs the sensing task by producing symbol-wise estimates based on the triples $U_{k'}^N$, X_k^N, and Z_k^N and of the compression information obtained from the other Tx k'.
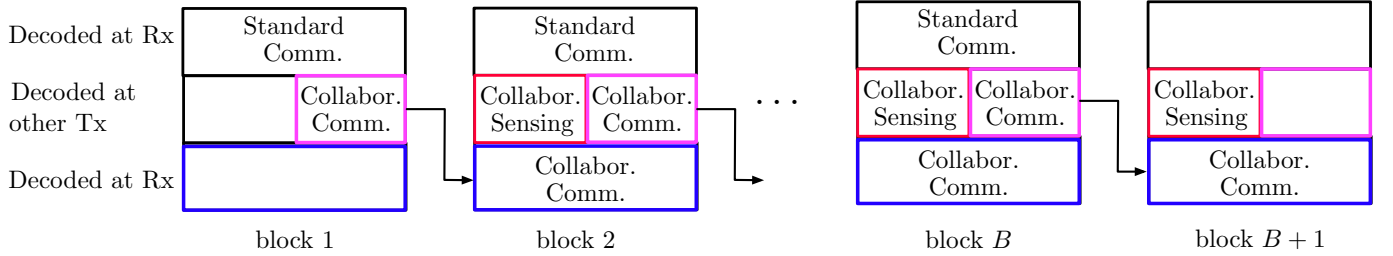

Fig. 12. Block-Markov strategy of the ISAC multi-access scheme in [71].

Above coding schemes establish the following inner bound to the capacity-distortion tradeoff region $\mathcal{CD}$ [70], [71].

Theorem 9. *The capacity-distortion region $\mathcal{CD}$ of an ISAC MAC system includes any rate-distortion tuple (R_1, R_2, D_1, D_2) that for some choice of pmfs $P_{U_0}, P_{U_1|U_0}, P_{U_2|U_0}, P_{X_1|U_0U_1}, P_{X_2|U_0U_2}, P_{V_1|U_0U_2X_1Z_1}, P_{V_2|U_0U_1X_2Z_2}$ satisfies Inequalities* (49) *on top of the next page (where $\underline{U} := (U_0, U_1, U_2)$) as well as the distortion constraints*

$$D_k \geq \mathrm{E}[d_k(S_k, \phi_k^*(X_k, Z_k, U_{k'}, V_{k'})], \tag{50}$$

for

$$\phi_k^*(x_k, z_k, u_{k'}, v_{k'}) := \arg\min_{s' \in \hat{\mathcal{S}}_k} \sum_{s_k \in \mathcal{S}_k} P_{S_k|X_k Z_k U_{k'} V_{k'}}(s_k|x_k, z_k, u_{k'}, v_{k'})\, d_k(s_k, s').$$

$$\tag{51}$$

In above theorem, the U_0 random variable stands for the common lowest-layer codeword of both Txs, U_1 and U_2 stand for the medium-layer codewords of the two Txs, and X_1 and X_2 for the top codewords sent by the two Txs. The random variables V_1 and V_2 stand for the compression informations produced at Tx 1 and 2, respectively. Accordingly, the previous achievable region in [68] is obtained as a special case from above theorem by setting $V_1 = V_2 = $ constants.

A further improvement has been obtained in [69] by adding sensing information also to the lowest and the top codewords, see Figure 13. In other words, the two Txs jointly resend the two parts of exchanged compression informations in a block in the next following block as part of the lowest codeword, and in each block they individually add compression information to the top layer, which is not decoded at other Tx but only at the Rx. Indeed, as already mentioned, the Rx can be interested in receiving compression information so as to improve its observations and thus decoding performance of the transmitted data.

It is rather straightforward to identify further ways of obtaining improved multi-access ISAC schemes. For example, one could add additional coding layers as in Marton coding in a way that the Rx is not

$$R_k \le I(U_k; X_{k'}Z_{k'} \mid U_0 U_{k'}) + I(V_k; X_{k'}Z_{k'} \mid \underline{U}) - I(V_k; X_k Z_k \mid \underline{U})$$
$$+ \min\{I(X_k; Y \mid U_0 X_{k'}) + I(V_k; X_1 X_2 Y \mid \underline{U}) + I(V_{k'}; X_1 X_2 Y V_k \mid \underline{U}) - I(V_k; X_k Z_k \mid \underline{U}),$$
$$I(X_1 X_2; Y \mid U_0 U_k) + I(V_k; X_1 X_2 Y \mid \underline{U}) + I(V_{k'}; X_1 X_2 Y V_k \mid \underline{U}) - I(V_{k'}; X_{k'}Z_{k'} \mid \underline{U}),$$
$$I(X_1 X_2; Y \mid U_0) + I(V_k; X_1 X_2 Y \mid \underline{U}) + I(V_{k'}; X_1 X_2 Y V_k \mid \underline{U})$$
$$-I(V_k; X_k Z_k \mid \underline{U}) - I(V_{k'}; X_{k'}Z_{k'} \mid \underline{U}), \ I(X_k; Y V_1 V_2 \mid \underline{U} X_{k'})\}$$
$$k', k \in \{1, 2\}, \ k \ne k', \ (49a)$$

$$R_1 + R_2 \le I(U_2; X_1 Z_1 \mid U_0 U_1) + I(V_2; X_1 Z_1 \mid \underline{U}) - I(V_2; X_2 Z_2 \mid \underline{U})$$
$$+ I(U_1; X_2 Z_2 \mid U_0 U_2) + I(V_1; X_2 Z_2 \mid \underline{U}) - I(V_1; X_1 Z_1 \mid \underline{U})$$
$$+ \min\{I(X_1 X_2; Y \mid U_0 U_2) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) - I(V_1; X_1 Z_1 \mid \underline{U}),$$
$$I(X_1 X_2; Y \mid U_0 U_1) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) - I(V_2; X_2 Z_2 \mid \underline{U}),$$
$$I(X_1 X_2; Y \mid U_0) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U})$$
$$-I(V_1; X_1 Z_1 \mid \underline{U}) - I(V_2; X_2 Z_2 \mid \underline{U}),$$
$$I(X_1 X_2; Y V_1 V_2 \mid \underline{U})\}$$
$$(49b)$$

$$R_1 + R_2 \le I(X_1 X_2; Y) + I(V_1; X_1 X_2 Y \mid \underline{U}) - I(V_1; X_1 Z_1 \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) - I(V_2; X_2 Z_2 \mid \underline{U})$$
$$(49c)$$

and for $k', k \in \{1, 2\}$ and $k' \ne k$, the following satisfies

$$I(U_k; X_{k'}Z_{k'} \mid U_0 U_{k'}) + I(V_k; X_{k'}Z_{k'} \mid \underline{U}) \ge I(V_k; X_k Z_k \mid \underline{U}), \qquad (49d)$$

$$I(X_1 X_2; Y \mid U_0) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) \ge I(V_1; X_1 Z_1 \mid \underline{U}) + I(V_2; X_2 Z_2 \mid \underline{U}) \ (49e)$$

$$I(X_k; Y \mid U_0 X_{k'}) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) \ge I(V_k; X_k Z_k \mid \underline{U}). \qquad (49f)$$
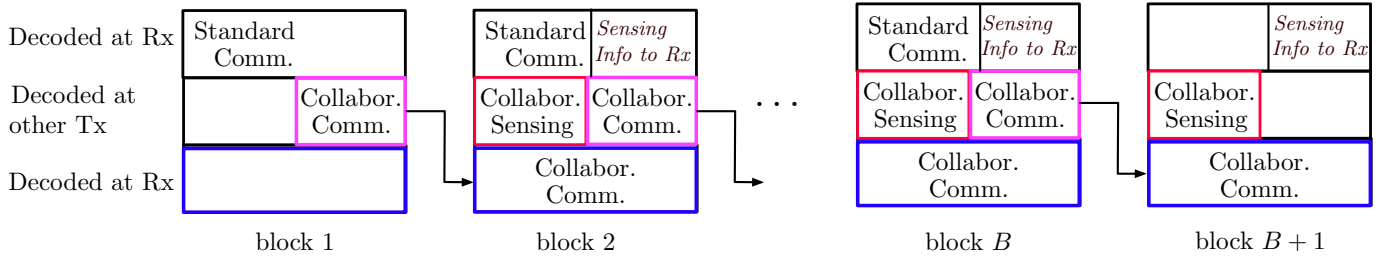

Fig. 13. Block-Markov strategy of the improved ISAC multi-access scheme in [69].

required to decode all sensing information. In fact, in certain scenarios sensing information is useless at the Rx, and moreover the Rx has worse decoding capabilities than the Txs. Moreover, joint source-channel coding methods could be applied for the transmission of sensing information. In fact, the sensing informations sent at the two Txs are correlated and it is well known that in such a scenario a joint source-channel coding approach can achieve improved performances. In an upcoming section, we briefly discuss a joint source-channel coding approach for the two-way channel, i.e., for device-to-device (D2D) communication.

3) Example: The following example shows the improvement of Theorem 9 over the previous scheme in [68]. As mentioned, a further improvement is achieved by the scheme in [69].

Example 2. *Consider binary noise, states and channel inputs $B_0, B_k, S_k, X_k \in \{0, 1\}$. The noise to the Rx B_0 is Bernoulli-t_0, and B_k, the noise on the feedback link to Tx k, is Bernoulli-t_k. All noises are independent and also independent of the states S_1, S_2, which are i.i.d. Bernoulli-p_s. We can then describe the channel as*

$$Y' = S_1 X_1 + S_2 X_2 + B_0, \qquad Y = (Y', S_1, S_2), \qquad (52)$$

$$Z_1 = S_1 X_1 + S_2 X_2 + B_1, \qquad Z_2 = S_1 X_1 + S_2 X_2 + B_2. \qquad (53)$$

In this example the Rx has perfect channel state-information Hamming distance is considered as a distortion measure: $d(s, \hat{s}) = s \oplus \hat{s}$.

Figure 14 shows the maximum sum-rate $R_1 + R_2$ in function of distortion D_2 achieved by Theorem 9 (with collaborative sensing) and the region in [68] without collaborative sensing where $V_1 = V_2 =$ constants. Both curves are strictly concave and thus improve over classic time- and resource sharing strategies. The minimum distortions achieved are $D_{2,\min} = 0.035$ with collaborative sensing and $D_{2,\min} = 0.04$ without.
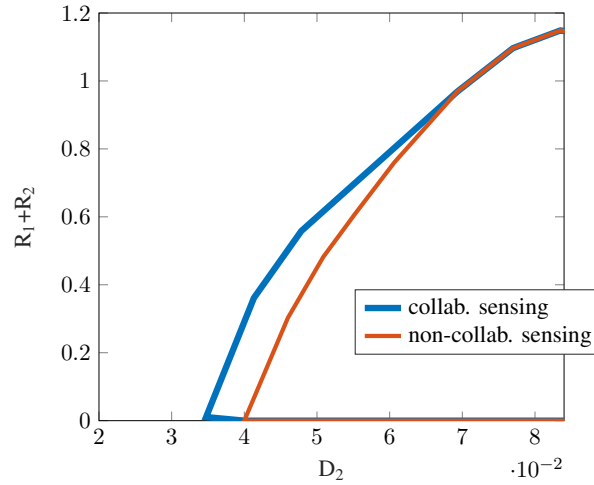

Fig. 14. Sum-rate distortion tradeoff in Example 2 achieved without and with collaborative sensing, for given channel parameters $p_s = 0.9$, $t_0 = 0.3$, $t_1 = 0.1$ and $t_2 = 0.1$.

Similar ISAC coding ideas were also proposed for the interference channel (IC) where two Txs communicate to two Rxs [111]. The idea is to use a block-Markov coding as for the MAC and that the two Txs compress and convey sensing information in addition to the cooperative data communication in previous blocks. The corresponding set of achievable rate-distortion tuples can be found in [111].

C. Device-to-Device (D2D) Communication (Two-Way Channel)

Besides ISAC multi-access systems, [71] also studied the related two-way channel, i.e., device-to-device (D2D) communication. The D2D setup is illustrated in Fig. 15 and is similarly defined to the MAC, except that message W_1 has to be decoded at Tx 2 and message W_2 at Tx 1. There is thus no receiver terminal. The capacity-distortion region $\mathcal{CD}$ is defined in analogy to the MAC setup.

The capacity region for D2D data communication (without sensing task), and thus the optimal coding scheme, is still open in general. Various inner and outer bounds on the capacity region have been proposed. Han [112] and Kramer [113] proposed schemes that correlate the inputs of the two terminals in a block-fashion. While for Han's coding scheme the correlation ensures a stationary distribution of the inputs and outputs across the blocks and thus still allows for single-letter rate-expressions, Kramer has to resort
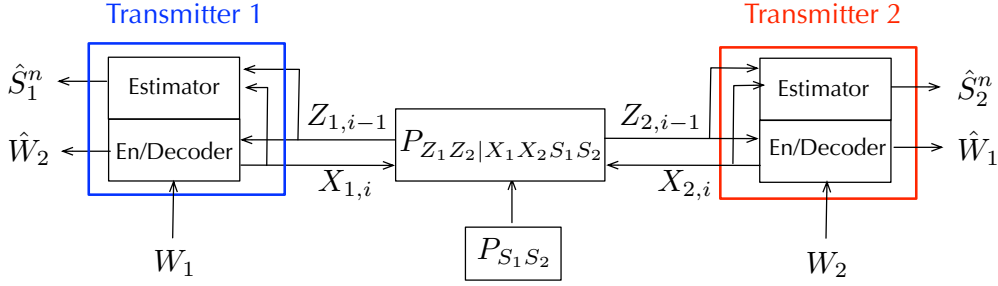
Fig. 15. State-dependent discrete memoryless two-way channel with sensing at the terminals.

$$
P_{U_1' U_2' Z_1 Z_2 X_1 X_2 U_1 U_2}(u_1', u_2', z_1, z_2, x_1, x_2)
$$
$$
= \sum_{\tilde{u}_1, \tilde{u}_2, \tilde{x}_1, \tilde{x}_2, \tilde{z}_1, \tilde{z}_2} P_{U_1'|X_1 Z_1 \tilde{U}_1 \tilde{X}_1 \tilde{Z}_1}(u_1'|u_1, x_1, z_1, \tilde{u}_1, \tilde{x}_1, \tilde{z}_1) P_{U_2'|X_2 Z_2 \tilde{U}_2 \tilde{X}_2 \tilde{Z}_2}(u_2'|x_2, z_2, u_2, \tilde{u}_2, \tilde{x}_2, \tilde{z}_2)
$$
$$
\cdot P_{Z_1 Z_2|X_1 X_2}(z_1, z_2|x_1, x_2) \mathbb{1}\{x_1 = f_1(u_1, \tilde{u}_1, \tilde{x}_1, \tilde{z}_1)\} \mathbb{1}\{x_2 = f_2(u_2, \tilde{u}_2, \tilde{x}_2, \tilde{z}_2)\}
$$
$$
\cdot P_{U_1' U_2' Z_1 Z_2 X_1 X_2 U_1 U_2}(u_1, u_2, \tilde{z}_1, \tilde{z}_2, \tilde{x}_1, \tilde{x}_2, \tilde{u}_1, \tilde{u}_2), \tag{54}
$$

to multi-letter rate-expressions based on directed mutual informations. An interesting outer bound on the capacity region was proposed by Hekstra and Willems [105] again based on the dependence-balance idea, similar to the MAC with feedback.

The work in [71] proposed two coding schemes for the ISAC D2D problem. The idea of the first scheme is to extend Han's D2D coding scheme in a similar way as [71] extended Willems' scheme for the MAC. That means, the two terminals generate compression information, which they convey to the other Tx as part of the indices sent in the data communication scheme. A second, more advanced coding scheme based on joint source-channel coding, was also propose in [71]. In this second scheme, the compression information is not just transmitted by means of indices sent instead of data, but by correlating the channel inputs with the sensing information (i.e., the compression codewords), as is typically done in hybrid coding [28]. This allows the two terminals to directly transfer the correlation of the sensing information to the channel inputs, which often allows for improved decoding performances at the two Txs.

Theorem 10 (Inner Bound via Joint Source-Channel Coding)**.** *The capacity-distortion region of the D2D ISAC problem* $\mathcal{CD}$ *contains all rate-distortion quadruples* (R_1, R_2, D_1, D_2) *for which there exists a choice of the pmf* $P_{U_1' U_2' Z_1 Z_2 X_1 X_2 U_1 U_2}$ *and functions* f_1 *and* f_2 *satisfying the stationarity condition* (54) *on top of this page and so that the following two rate-constraints*

$$
R_k \leq I(\tilde{U}_k; X_{k'}, Z_{k'}, \tilde{U}_{k'}, \tilde{X}_{k'}, \tilde{Z}_{k'}) - I(U_k; X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k | X_{k'}, Z_{k'}, \tilde{U}_{k'}, \tilde{X}_{k'}, \tilde{Z}_{k'}),
$$
$$
k, k' \in \{1, 2\}, \ k' \neq k, \tag{55}
$$

and the two distortion constraints in

$$
\mathrm{E}\Big[d_k\big(S_k, \ \phi_{2,k}^*(U_{k'}', X_{k'}, Z_{k'}, U_{k'}, X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k, \tilde{U}_{k'}) \big) \Big] \leq D_k, \qquad k, k' \in \{1, 2\}, \ k' \neq k, \tag{56a}
$$

hold.

VII. Secrecy of ISAC Systems

Secrecy is a major concern in ISAC systems, both for the communication and the sensing tasks. Depending on the applications, adversaries should not be able to learn the transmitted data and/or infer information about the sensing targets. The information-theoretic literature has mostly studied the problem of ensuring secrecy of messages [72], [88], [114], [115], but first results also exist to ensure secrecy of sensing information [73]. The model is important because ISAC systems enable the surveillance of the environment, and in many scenarios, it is crucial to prevent unauthorized access to user or channel information. In this section we review both these lines of work.

A. Secrecy of the Message: The Memoryless Model

Based on the memoryless ISAC model in [58], a wiretap equivalent was introduced in [72], see Fig 16. In this model, communication needs to be such that the eavesdropper cannot learn part of the message, which is formalized by the requirement that the equivocation between this specific message part and Eve's observations should vanish for large blocklengths. (This requirement is also known as strong secrecy in the information-theoretic literature.)
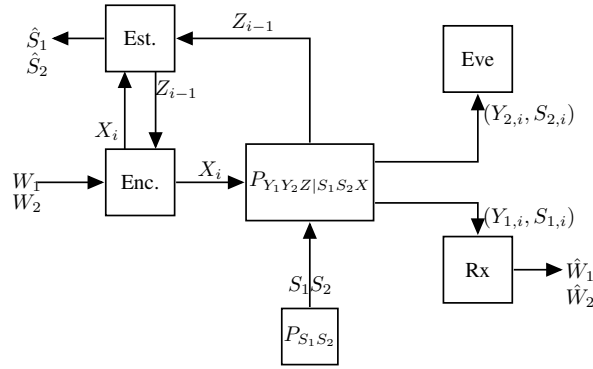

Fig. 16. ISAC model under partial secrecy, where only W_2 should be kept secret from Eve.

Formally, the problem is defined with a Tx, a legitimate Rx, and an eavesdropper (Eve). The Tx aims to communicate a pair of messages (W_1, W_2) of rates R_1 and R_2 to the legitimate receiver, in a way that Eve cannot learn any information about the message W_1. (There is no constraint on how much Eve learns about the other message W_2. Communication is over a memoryless stationary state-dependent channel $P_{Y_1 Y_2 Z|SX}$ where X is the channel input, S_1 and S_2 are the states, Y_1 the outputs at the legitimate receiver, Y_2 the outputs at the eavesdropper and Z The generalized feedback. The state-sequences $\{(S_{1,i}, S_{2,i})\}$ are assumed i.i.d. according to the given law $P_{S_1 S_2}$, and the transmitter creates the time-i channel inputs as $X_i = \phi_i(W_1, W_2, Z_{i-1})$ using some appropriate encoding function ϕ_i. At the end of the communication, the Tx estimates the state-sequence as $(\hat{S}_1, \hat{S}_2^n)^n = h(X^n, Z^n)$. The receiver decodes the two messages as $(\hat{W}_1, \hat{W}_2) = g(Y_1^n, S_1^n)$ using an appropriate encoding function. The goal of the communication is that decoding error probability vanish asymptotically, that the reconstructed state-sequence matches the correct state up to a given distortion constraints D_1 and D_2 under given per-symbol distortion measures $d_1(\cdot, \cdot)$ and $d_2(\cdot, \cdot)$, and that Eve learns nothing about message W_2 from her observations Y_2^n and S_2^n.

Definition 5. *A secrecy-rate-distortion tuple (R_1, R_2, D_1, D_2) is achievable if it is possible to find a sequence (in the blocklength n) of encoding, decoding, and estimation functions satisfying*

$$\lim_{n \to \infty} \Pr[W_k \neq \hat{W}_l] = 0, \quad k \in \{1, 2\}, \tag{57}$$

$$\lim_{n \to \infty} I(W_2; Y_2^n, S_2^n) = 0, \tag{58}$$

$$\varlimsup_{n \to \infty} E[d(S_k^n, \hat{S}_k^n)] \leq D_k, \quad k \in \{1, 2\}. \tag{59}$$

The closure of the set of all achievable secrecy-rate-distortion tuple (R_1, R_2, D_1, D_2) is called the secrecy-capacity-distortion region $\mathcal{SCD}$.

Here we chose to present the slightly restricted model where the Rxs learn the two state-sequences, thus not allowing for no or only imperfect state information. A more general model can however easily be obtained similar to the models presented in the previous sections. The reason for considering this special case is that in the following we will limit to the special case where the Tx observes perfect output feedback, i.e., $Z = (Y_1, Y_2)$ without the two states, which does not allow to incorporate arbitrary channel state-information distributions at the Tx and the Rx/Eve.

B. Secrecy of Messages: Results

Most of the results have been derived under the assumption of perfect feedback from both the Rx and Eve, i.e., $Z = (Y_1, Y_2)$ [72]. Only the outer bounds in [72] apply for a slightly more general scenario where Z is a noisy version of (Y_1, Y_2).

Notice that the optimal estimator at the Tx is the same as in the setup without secrecy constraint, see (5). In the case of perfect output feedback $Z = (Y_1, Y_2)$ and two states, these optimal estimators are:

$$\hat{s}_k^*(x, y_1, y_2) = \arg \min_{s' \in \hat{\mathcal{S}}_k} \sum_{s \in \mathcal{S}_k} P_{S_k|XY_1Y_2}(s|x, y_1, y_2) d_k(s, s'), \quad k \in \{1, 2\}. \tag{60}$$

Combined with these optimal estimators, the output statistics of random binning (OSRB) proof technique [116] allows to achieve the following result [72].

Theorem 11 (Inner Bound). *The secrecy-capacity-distortion region $\mathcal{SCD}$ contains all secrecy-rate-distortion tuples (R_1, R_2, D_1, D_2) that satisfy the following inequalities for some pmf P_{UVX}:*

$$R_1 \leq I(U; Y_1 S_1) \tag{61}$$
$$R_2 \leq \min\{[I(V; Y_1|S_1U) - I(V; Y_2|S_2U)]^+ + H(Y_1S_1|Y_2S_2V), \quad (I(V; Y_1|S_1) - R_1)\} \tag{62}$$
$$D_k \geq \mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Y_1, Y_2))], \quad k \in \{1, 2\}. \tag{63}$$

Theorem 12 (Outer Bound). *The secrecy-capacity-distortion region $\mathcal{SCD}$ is included in the union over all joint distributions $P_{UVX} = P_{UV}P_{X|V}$ of all rate tuples (R_1, R_2, D_1, D_2) satisfying[4]*

$$R_1 + R_2 \leq I(V; Y_1|S_1), \tag{64}$$
$$R_2 \leq I(V; Y_1S_1|Y_2S_2), \tag{65}$$
$$D_k \geq \mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Y_1Y_2))], \quad k \in \{1, 2\}, \tag{66}$$

One can limit $\mathcal{V}$ to $|\mathcal{V}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1||\mathcal{S}_1|, |\mathcal{Y}_2||\mathcal{S}_2|\} + 1$.

The above results assume that only a part of the message (namely W_2) has to be kept secure from Eve. Corresponding results where all messages have to be kept secure are easily obtained by setting $R_1 = 0$ and interpreting R_2 as the total rate of all communicated messages.

Above inner and outer bounds do not coincide in the general case. They do in the case of degraded channels where $P_{Y_2S_2|XS_1Y_1} = P_{Y_2S_2|Y_1S_1}$ and reversely degraded channels where $P_{Y_1S_1|XS_2Y_2} = P_{Y_1S_1|Y_2S_2}$. They have also been specialized to several interesting and practical channels. In particular the results for the Gaussian fading examples are worth being mentioned for binary states [114] as well as for Rayleigh fading states [115].

---

[4]The bounds are slightly simpler and stronger than the bounds in [72] and can be proved using similar steps.

Finally, notice that a finite blocklength analysis for ISAC with security constraints has been performed in [88].

C. Secrecy of Data and Sensing Information

In [73] not only the message (data) has to be kept secure from an external eavesdropper, but also the channel state-sequence S^n. Depending on the channel, Eve will always learn about the channel state, however it is required that this knowledge stays beyond a given threshold. In other words, the Tx has to choose transmission strategies in a way that not too much information is leaked about the sensing target. In a practical application this could mean that the Tx has to restrict to beamforming strategies where it points its beam towards a given direction.
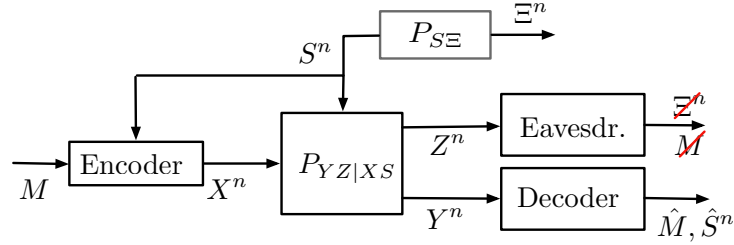

Fig. 17. ISAC model with state-information at the Tx and secrecy constraints on messages and states .

To make the problem more interesting from a technical point of view, the authors in [73] focused on the setup in Figure 17 where the Tx learns the state-sequence S^n in a non-causal manner (i.e., before transmission starts) and the sensing is performed at the Rx. An external eavesdropper is not allowed to learn any information about the message nor the sequence Ξ^n which is obtained by passing the state-sequence S^n through a memoryless channel $P_{\Xi|S}$ independent of the message and the communication channel. Notice that the channel $P_{\Xi|S}$ needs to be carefully chosen in the model to reflect the desired security constraint. For example, it could select a part of the state if S is bipartite $S = (S_1, S_2)$ and only one of these states needs to be kept secret, or it could implement a function $\Xi = \nu(S)$ when only certain characteristics of the target have to be kept secret. In general, the setup in [73] allows to model any stochastic relationship between the state S and and the part that needs to be kept secret Ξ.

In the setup of this section there is only one message W of rate R and the Tx produces its channel inputs as $X_i = \phi_i(W, S^n)$, for S^n the i.i.d. state-sequence following a given pmf P_S. The channel outputs Y^n observed at the legitimate Rx and Z^n observed at the eavesdropper are produced from inputs and states according to a given stationary and memoryless channel law $P_{YZ|XS}$. Based on the observed outputs Y^n, the Rx decodes the message as $\hat{W} = g(Y^n)$ and produces an estimate of the state-sequence $\hat{S}^n = h(Y^n)$. The goal of the Tx is to find an encoding strategy for which the Rx can decode with arbitrary small probabilities of error and reconstruct the state-sequence with desired distortion D but such that the eavesdropper does not learn about the related sequence Ξ^n nor the message W. This leads to the following definition of achievability.

Definition 6. *A rate-distortion pair* (R, D) *is called securely-achievable if there exists a sequence (in n) of encoding, decoding, and estimation functions such that*

$$\lim_{n \to \infty} \Pr[\hat{W} \neq W] = 0 \tag{67a}$$

$$\lim_{n \to \infty} I(W, \Xi^n; Z^n). = 0 \tag{67b}$$

$$\varlimsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d(S_i, \hat{S}_i)] \leq D. \tag{67c}$$

The following inner bound was proved in [73].

Theorem 13. *For any pmf $P_{UVX|S}$ so that for the associated tuple $(S, \Xi, U, V, X, YZ) \sim P_S P_{\Xi|S} P_{UVX|S} P_{YZ|XS}$, the random variable Ξ is independent of the pair (U, Z) and any function $g(\cdot)$ on appropriate domains, all pairs (R_M, D) satisfying the following inequalities*

$$R \leq I(UV; Y) - I(UV; S) \tag{68}$$

$$R \leq I(V; Y \mid U) - I(V; \Xi Z \mid U) + \min\{0, I(U; Y) - I(U; S)\} \tag{69}$$

and

$$D \leq E[d(S, g(U, V, Y))] \tag{70}$$

are securely achievable.

Above achievability result is based on the following coding scheme. A two-level superposition code with cloud-center codewords U^n and satellite codewords V^n is considered. The Tx uses the U^n-codewords to describe information about the state-sequence S^n to the receiver, where this cloud-center codeword can also be decoded by the eavesdropper. It further uses the V^n-codewords to send more refined information about S^n as well as the message W to the Rx. The Rx decodes both the U^n and V^n codewords so as to recover the transmitted message W. It also reconstructs the state-sequence based on the two decoded codewords and its own observed sequence of channel outputs. Security of the scheme against the external eavesdropper is obtained by choosing the U^n-codewords so that the decoded does not reveal information about the Ξ^n-sequence (because the U^n-codeword is also decoded by the eavesdropper). In fact, in the construction, only the V^n-codeword can contain information about Ξ^n and W, and they are chosen of sufficiently high rate so that the eavesdropper cannot decode them.

The theorem includes several interesting special cases. When Z is independent of the input-state pair (X, S), the setup reduces to the setup without secrecy constraint studied in [117], in which case Theorem 13 can be simplified by choosing $U =$const. On a different note, when the entire state S has to be kept secret, $\Xi = S$, then U has to be chosen independently of S and thus $I(U; S) = 0$ and the minimum in the right-hand side of (69) evaluates to 0. Moreover, for $\Xi = S$ the right-hand side of (68) is larger than the right-hand side of (69) because $I(V; SZ|U) \geq I(V; S|U)$. Thus, for $\Xi = S$, Constraint (68) is less stringent than Constraint (69) where U only plays the role of a convexification random variable.

Comparing the results where both message and state have to be kept secret with the results with no secrecy constraint is applied, the price for the double state-and-message secrecy in the proposed scheme seems to be independence of S with Z and the rate-reduction of $I(V; Z|S) = I(SV; Z)$, see [73, Corollaries 1–3].

Example 3. *From [73]. Consider the interesting examples with Gaussian channels*

$$Y_i = X_i + S_i + N_i, \tag{71}$$

$$Z_i = aX_i + bS_i + N_{e,i}, \tag{72}$$

for some given parameters a, b and $\{N_i\}$ and $\{N_{e,i}\}$ memoryless standard Gaussian noise sequences. Let further

$$\Xi = S + A, \tag{73}$$

for $A \sim \mathcal{N}(0, \sigma_A^2 \geq 0)$ independent of all other r.vs. This setup covers the scenario where the entire state-sequence has to be kept secret, with the choice $\sigma_A^2 = 0$, and (with a slight abuse of notation) the scenario where the state does not have to be kept secret at all, with the choice $\sigma_A^2 \to \infty$. The following Figure 18 shows an achievable set of rate-distortion pairs according to above Theorem 13 for $a = 0.7$

and $b = 0.3$ and $S \sim \mathcal{N}(0,3)$ and with a input block-power constraint of $P = 30$. The set of achievable rate-distortion pairs is provided without any secrecy constraints neither on messages nor state, with full secrecy constraints on both ($\Xi = S$), and with security constraints only on the message but not on the state (Ξ const). As shown at hand of this example, the additional secrecy constraint on the state is clearly harmful and significantly reduces the set of achievable rates and distortions.
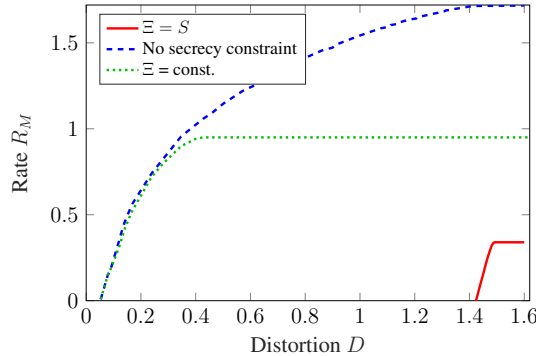

Fig. 18. Comparison of the achievable rate-distortion tradeoffs under different secrecy constraints.

VIII. ISAC WITH DETECTION-ERROR EXPONENTS

Radar is not only used to estimate parameters such as vehicle velocities, arrival angles, etc., but is also extensively employed to detect obstacles, or more generally, other terminals. In this context, the sensing problem must be framed as a detection or hypothesis testing problem. This framework allows us to incorporate memory into the channel transition law, analogous to slow-fading channels, thereby making the model more reflective of realistic scenarios. This involves considering multiple hypotheses (e.g., the presence or absence of one or more obstacles), and to ensure the model is fully general, we allow the communication channel to depend on the chosen hypothesis. Such a scenario arises, for example, when an obstacle obstructs the line of sight between the Tx and Rx, thereby significantly altering the communication channel characteristics. Of course, the general model also accommodates simpler cases where the communication channel remains independent of the hypothesis.

In this section, the sensing performance is quantified by detection error probabilities under the different hypotheses. Since these error probabilities can be made to vanish asymptotically with infinite observation lengths, the focus here will be on the exponential decay rate of these error probabilities. Much of the existing literature on the fundamental limits of ISAC with detection error exponents has focused on monostatic radar, where sensing (detection) is performed at the communication Tx [59], [74]–[77], [118]. A notable exception is [118], where detection is carried out at the communication receiver.

From an information-theoretic perspective, the detection-error ISAC problem is considerably more challenging than the previously presented distortion-based ISAC setup. The primary difficulty stems from the sensing aspect, as the optimal performance of hypothesis testing systems is not well understood, even without the additional communication component present in ISAC systems. In particular, the simple estimation and communication strategies employed in the distortion-ISAC setup are suboptimal. Instead, the Tx can improve its detection performance by generating inputs according to a smart, sequential strategy that adapts based on previously observed outputs. Specifically, in a multi-hypothesis testing problem, the Tx might initially produce inputs based on a particular distribution, and once the observed signals provide sufficient evidence to discriminate one hypothesis with high confidence, it can switch to a different input distribution that better discriminates among the remaining hypotheses.

Adaptive systems are complex to implement, which motivates the practical interest in excluding them from certain considerations. When restricting attention to non-adaptive strategies, similar closed-form

expressions for the fundamental ISAC performance limits and trade-offs can be derived, as in the distortion ISAC setup.

The next subsection explains the model both for the adaptive and non-adaptive coding scenarios, and is followed by a section presenting the existing information-theoretic results on ISAC with detection exponents. The last subsection has a slightly different flavour for the sensing task and the goal is to detect a change point.
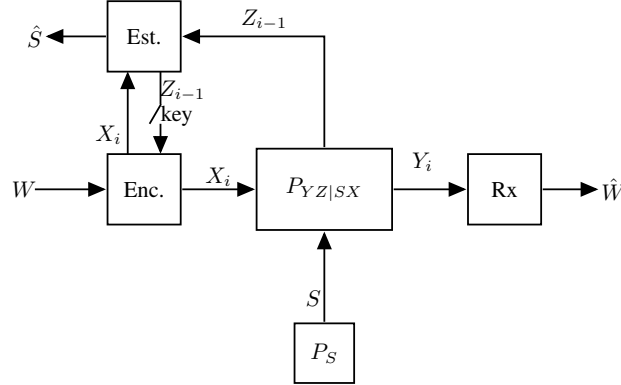
A. The Memoryless Block-Model

Fig. 19. ISAC with a detection sensing problem.

Consider the model in Figure 19, where the Tx wishes to communicate a message to a Rx over a channel that depends on a single parameter $\theta \in \Theta$, for Θ a finite set, and at the same time aims to determine this parameter based on the backscattered (generalized-feedback) signals. The parameter is assumed to take value in a discrete and finite set, transforming the sensing problem into a hypothesis testing/detection problem. For a given parameter θ in a finite set Θ, the communication channel to the Rx as well as the generalized feedback to the Tx are assumed to be stationary and memoryless, and are described by a joint transition law $P^\theta_{YZ|X}$. In the information-theory literature, such a communication channel with a fixed but à prior unknown parameter θ is known as a *compound channel* [20], [119]

The switch indicates whether the Tx can employ adaptive/closed-loop coding or non-adaptive/open-loop coding. More precisely, if the switch is closed, the i-th channel input X_i can depend on the previous generalized feedback signals $Z_1, \ldots, Z_{i-1}$, and if the switch is open then all inputs only depend on the message W.

Formally, the Tx produces the channel inputs either as (in case of non-adaptive coding)

$$X^n = \phi^{(n)}(W) \tag{74}$$

or as (in case of adaptive coding)

$$X_i = \phi_i(W, Z_1, \ldots, Z_{i-1}), \quad i = 1, \ldots, n, \tag{75}$$

where $\phi^{(n)}$ and $\phi_1, \ldots, \phi_n$ are encoding functions on appropriate domains. The Tx further guesses the parameter θ as

$$\hat{\theta} = h(X_1, \ldots, X_n, Z_1, \ldots, Z_n), \tag{76}$$

using some appropriate detection function $h(\cdot)$. As before, the Rx decodes the transmitted message using an appropriate decoding function: $\hat{W} = g(Y_1, \ldots, Y_n)$.

Communication performance is measured as before in terms of rate R of message W, where the Tx and Rx have to be designed in a way that the decoding error probability $\Pr[\hat{W} \neq W | \theta]$ vanishes asymptotically

when the blocklength n increases under *any of the hypothesis θ*. In this sense, the communication rate is defined in the same way as for the compound channel [20], [119].

Sensing performance is measured in terms of the asymptotic detection-error exponents

$$E_\theta \triangleq -\frac{1}{n} \log \Pr[h(Z^n, X^n) \neq \theta \mid \theta], \quad \theta \in \boldsymbol{\Theta}, \tag{77}$$

where the conditioning on θ simply indicates that the Z^n sequence is generated from X^n according to the memoryless law $P^\theta_{Z|X}$.

Different requirements on the detection-error exponents have been considered in the literature [59], [74]–[77]. We summarize the requirements in the following definition.

Definition 7. *Let* $\Theta = \{0, 1\}$. *Then we say that a rate-detection-error exponent* (R, D) *is achievable in the Stein setup if there exists a sequence (in the blocklength n) of encoding, decoding and detection functions so that the following three conditions are satisfied simultaneously:*

$$\lim_{n \to \infty} \Pr[\hat{W} \neq W] = 0 \tag{78}$$

$$\lim_{n \to \infty} \Pr[h(Z^n, X^n) \neq 0 \mid \theta = 0] = 0 \tag{79}$$

$$E_1 \geq \mathsf{D}. \tag{80}$$

Similarly, the triple $(\mathsf{R}, \mathsf{D}_0, \mathsf{D}_1)$ *is achievable* in a exponents-region sense*, if above sequences exist so that* (78) *holds, as well as*

$$E_\theta \geq \mathsf{D}_\theta, \quad \theta \in \Theta. \tag{81}$$

Let now Θ *be arbitrary. Then, the rate-detection-exponent pair* (R, D) *is called achievable* in the symmetric setup *if encoding, decoding and detection functions exist so that* (78) *holds and*

$$\min_{\theta \in \Theta} E_\theta \geq \mathsf{D}. \tag{82}$$

To distinguish between the adaptive and non-adaptive case we will add the superscripts ad *and* nad *to the exponents and write* D^{ad} *and* D^{nad}.

In contrast to the model described here, the works in [74]–[76] imposed maximum error probability conditions over the messages both for the decoding error probabilities as well for the detection error probabilities. It turns out that the setup of achievable rate-detection-exponent(s) is the same under both average and maximum error probabilities as long as one requires that all error probabilities vanish asymptotically.

B. Results on the Block-Model

We first focus on the model where coding at the Tx is restricted to be non-adaptive. Combining the results in [59], [74]–[77], we obtain the following theorem:

Theorem 14. *Under non-adaptive coding we have the following information-theoretic results for the rate-detection-exponents regions in the Stein setup, the symmetric setup, and in the exponents region sense.*

1) *In the* Stein setup, *a nonnegative rate-detection-error pair* $(\mathsf{R}, \mathsf{D}^{\mathrm{nad}})$ *is achievable if, and only if,*

$$\mathsf{R} \quad \leq \quad \min_{\theta \in \Theta} I_{P^\theta_{Y|X}}(X; Y), \tag{83}$$

$$\mathsf{D}^{\mathrm{nad}} \quad \leq \quad \sum_x P_X(x) D\left(P^0_{Z|X}(\cdot|x) \big\| P^1_{Z|X}(\cdot|x)\right) \tag{84}$$

2) *In the* exponents region sense, *a nonnegative rate-detection-error pair* $(\mathsf{R}, \mathsf{D}^{\mathrm{nad}})$ *is achievable if, and only if, for some input distribution* P_X:

$$\mathsf{R} \quad \leq \quad \min_{\theta} I_{P_{Y|X}^\theta}(X;Y) \tag{85}$$

$$\mathsf{D}_1^{\mathrm{nad}} \quad \leq \quad \min_{\substack{\bar{P}_{Z|X}: \\ \mathbb{E}_{P_X}\left[D(\bar{P}_{Z|X}\|P_{Z|X}^0)\right] \leq \mathsf{D}_0^{\mathrm{nad}}}} \mathbb{E}_{P_X}\left[D\left(\bar{P}_{Z|X}\|P_{Z|X}^1\right)\right] \tag{86}$$

3) *In the* symmetric setup, *a nonnegative rate-detection-exponent pair* $(\mathsf{R}, \mathsf{D}^{\mathrm{nad}})$ *is achievable if, and only if, for some input distribution* P_X:

$$\mathsf{R} \leq \min_{\theta \in \Theta} I_{P_{Y|X}^\theta}(X;Y), \tag{87}$$

$$\mathsf{D}^{\mathrm{nad}} \leq \min_{\theta \in \Theta} \min_{a' \in \Theta \setminus \theta} \max_{l \in [0,1]} -\sum_x P_X(x) \log\left(\sum_z \left(P_{Z|XS}^\theta(z|x)\right)^l \left(P_{Z|XS}^a(z|x)\right)^{1-l}\right). \tag{88}$$

As already mentioned, exactly characterizing the fundamental limits under adaptive coding seems a very challenging problem and for the moment only achievability results are known [76], which however prove the superiority of adaptive coding over non-adaptive coding. Notice that it has been known for a long time that for the compound channel adaptive coding increases communication rate, because it allows the Tx to learn the hypothesis with high probability and then adapt the input distribution (and thus the code construction) to the actual transition law of the communication channel. This idea allows to obtain the following result [76, Theorem 5].

Theorem 15. *Under adaptive coding, a rate-detection-exponent pair* $(\mathsf{R}, \mathsf{D}^{ad})$ *is achievable in the symmetric setup if for any* $\theta \in \Theta$ *there exists an input distribution* P_X *so that*

$$\mathsf{R} \leq I_{P_{Y|X}^\theta}(X;Y), \tag{89}$$

$$\mathsf{D}^{ad} \leq \min_{a \neq \theta} \max_{l \in [0,1]} -\sum_x P_X(x) \log\left(\sum_z \left(P_{Z|X}^\theta(z|x)\right)^l \left(P_{Z|X}^a(z|x)\right)^{1-l}\right). \tag{90}$$

As mentioned in [76], a further improved region can be achieved by using adaptive strategies also to improve the sensing parts, not only the communication parts. For a more detailed discussion, see [76].

C. Sequential (Variable-Length) ISAC with Detection-Exponents

In [120], a variable-length version of the ISAC setup with detection-error exponents in the symmetric setup is considered. In this variable-length version, the transmission duration is not fixed from the beginning, but varies in function of the generalized feedback signals, which in this case has to coincide with the Rx's channel outputs to ensure synchronization of the communication. After each time t, the Tx will decide based on the past channel outputs $Y_1, \ldots, Y_t$ whether to stop or whether to continue communication. Let T be the random time where transmission stops. The model in [120] imposes that T be smaller than a given threshold n with high probability.

Given that the communication duration is random, also the number of transmitted message bits, and thus the rate of communication, are allowed to be random. In fact, the message bits are supposed to consist of a stream of i.i.d. Bernoulli-1/2 bits $\{U_i\}_{i=1}^\infty$, and an increasing sequence of numbers $\{M_t\}_{t=1}^T$ which indicates at each time t how many information bits have been transmitted until then and have to be decoded at the Rx if transmission stops at time t. Since transmission stops at time T, the Rx has to decode W_T information bits. The rate is defined as

$$R_T = \frac{W_T}{n}, \tag{91}$$

where recall that T is the stopping time of the communication and n is the given constraint (upper bound) on this stopping time.

Formally, encoder, decoder, and state detector are described as follows:

- At each time $t = 1, 2, \ldots$, the Tx forms the channel input as $X_t = f_t(U_1, \ldots, U_{W_t}, Z^{t-1})$, for an appropriate encoding function f_t;
- At the end of the transmission, the Tx guesses the state as $\hat{\theta} = h(X^T, Z^T)$, for an appropriate guessing function h;
- At the end of transmission, the Rx decodes the transmitted message bits as $\hat{U}^{W_T} = g(Y_1, \ldots, Y_T)$ for an appropriate decoding function g.

Definition 8. *A rate-detection-exponent (R, D_{VL}) is achievable in this variable-length setup if there exists a sequence (in the blocklength constraint n) of stopping rules, encoding functions, decoding functions, and state guessing functions, as defined above, such that:*

$$\lim_{n \to \infty} \max_{\theta \in \Theta} \max_{\boldsymbol{u}} \Pr[T > n] = 0 \tag{92}$$

$$\lim_{n \to \infty} \min_{\theta \in \Theta} \min_{\boldsymbol{u}} \Pr\left[R^{(n)} \geq R\right] = 1 \tag{93}$$

$$\lim_{n \to \infty} \max_{\theta \in \Theta} \max_{\boldsymbol{u}} \Pr\left[\hat{U}^{W_T} \neq U^{W_T} \big| \theta, U^{W_T} = \boldsymbol{u}\right] = 0, \tag{94}$$

$$\lim_{n \to \infty} -\frac{1}{n} \log \max_{\theta \in \Theta} \max_{\boldsymbol{u}} P\left[\hat{\theta} \neq \theta \,\big|\, \theta, U^{W_T} = \boldsymbol{u}\right] \geq D_{\mathrm{VL}}. \tag{95}$$

A set of achievable rate-detection-exponent pairs (R, E_{VL}) for the described setup was presented in [120]:

Theorem 16. *All rate-detection-exponents (R, D_{VL}) that for each $\theta \in \Theta$ satisfy the following two conditions for some choice of P_X (which can depend on θ)*

$$R \leq I(P_X, P_{Y|X}^\theta) \tag{96}$$

$$D_{\mathrm{VL}} \leq \min_{\theta' \neq \theta} \mathbb{E}_{P_X}\left[D\left(P_{Z|X}^{\theta'} \| P_{Z|X}^\theta\right)\right] \tag{97}$$

are achievable.

This result looks similar to the achiebability result in Theorem 15, where however the variable-length coding allows to improve the detection-error exponent from Chernoff information in (90) to the Kullback-Leibler divergence in (97). An example in [120] illustrates well this benefit of variable-length coding by means of a numerical plot.

D. Sequential (Variable-Length) ISAC with Change-Point Detection

A related model has also been considered in [78]. Communication again takes place over a fixed block of n channel uses. However, the channel starts in the state $\theta = 0$ and at a random time ν it will change to a state $\theta = 1$. The goal of the state-estimator is to detect this change-point ν with smallest delay. So, detection is variable-length as in the previous subsection, however data communication is fixed-length over n channel uses. Inputs have to be generated in an non-adaptive fashion.

Formally, the Tx generates its inputs as $X^n = f^{(n)}(W)$, where W is a uniform message of rate R and ϕ an encoding function of appropriate domain. The Rx guesses the message W as $\hat{W} = g(Y^n)$ using a guessing function g. We again assume perfect feedback $Z = Y$ and the Tx thus estimates the change-point using a stopping rule based on all inputs X^n and the past outputs $Y_1, \ldots, Y_t$. We denote the estimate of the change point by the random variable $N \in \{1, \ldots, n+1\}$, where $n+1$ indicates that

the channel did not change state. Notice that N being a stopping rule based on the inputs and previous outputs, formally means that each event $N = i$ is measurable with respect to Y^i and X^n.

As usual, decoding error probability $\Pr[\hat{W} \neq W]$ is required to tend to 0 as the blocklength $n \to \infty$. Typical performance measures for the change-point detection problem are the *false alarm rate (FAR)*, which in the present setup should be defined as

$$\text{FAR} = \overline{\lim_{n \to \infty}} \max_w \frac{1}{\mathbb{E}_\infty [N|\mathbf{x}^n(w)]}. \tag{98}$$

and the *worst-case average detection delay (WADD)*, which here is defined as:

$$\text{WADD} = \sup_{\nu \geq 1} \overline{\lim_{n \to \infty}} \max_w \text{ ess sup}_{\mathbf{Y}^{\nu-1}} \mathbb{E}_\nu \left[(N - \nu + 1)^+ |\mathbf{x}^n(w), \mathbf{Y}^{\nu-1} \right]. \tag{99}$$

Here, defined $\mathbb{E}_\nu[\cdot]$ as the expectation operator assuming that the change point is at time ν. Similarly, $\mathbb{E}_\infty[\cdot]$ denotes the expectation when the state of the channel never changes and remains $\theta = 0$ throughout.

Definition 9. *The pair* (R, Δ) *is called achievable in this quickest change point detection problem, if for arbitrary small* $\alpha > 0$ *there exists a sequence (in the blocklength* n*) of stopping rules* N*, encoding functions* $f^{(n)}$ *and decoding functions* g*, such that the following conditions are satisfied:*

$$\textit{FAR} \leq \alpha \tag{100}$$

$$\textit{WADD} \leq \frac{\alpha}{\Delta} \tag{101}$$

$$\lim_{n \to \infty} \Pr[\hat{W} \neq W] = 0. \tag{102}$$

In above definition, Δ describes the ratio between the FAR and the WADD.

The work in [78] establishes a set of achievable (R, Δ) pairs by using subblock-composition codes [121], which not only ensure a given empirical statistics (type) across any given codeword $x^n(w)$ but also within each subblock.

Theorem 17. *For any choice of* P_X*, all pairs of* (R, Δ) *are achievable that satisfy*

$$R \leq I(X;Y) \tag{103}$$

$$\Delta \leq \mathbb{E}_{P_X} \left[D \left(P_{Y|X}^1 \| P_{Y|X}^0 \right) \right]. \tag{104}$$

IX. CONCLUSION AND FUTURE RESEARCH DIRECTION

In this work, we revisited several models of Integrated Sensing and Communication (ISAC) and information-theoretic results on their fundamental performance limits and the tradeoffs between sensing and communication. These results emphasize the dual role of signals in estimating channel characteristics and enabling communication. We began by analyzing a point-to-point communication setup, where a simple modification of the traditional telecommunication framework allows for integrated sensing at the transmitter or receiver. We then reviewed the extended models for network scenarios such as the broadcast channel, multiple-access channels, interference channels and device-to-device communication. As we have seen, in these network scenarios, the backscattered feedback signals not only enable sensing at the Txs but can also be leveraged for collaborative transmissions of sensing and communication data in future blocks. This improves both the communication and sensing performance metrics and allows for new improved tradeoffs and improved overall efficiency. To fully exploit the concept of collaborative communication and sensing in these network ISAC scenarios, advanced coding schemes involving joint source-channel coding schemes are needed. While a large amount of results has already been obtained on ISAC systems with distortion constraints, important problems remain open in particular single-letter or numerical solutions for channels with memory as well as improved coding schemes and matching converse results for network ISAC systems.

Emerging ISAC systems often face security constraints and the transmitted data or properties of the sensing targets have to be kept secure from external or internal eavesdroppers. Information-theorists have studied such ISAC systems with secrecy constraints and determined bounds on the fundamental performance limits, with a focus on secure ISAC coding schemes. We presented these secure ISAC coding schemes and the corresponding securely achievable rate-distortion tuples. Results for different setups were reported: the first setup only requires that part of the data be kept secret, while the second setup imposes the more stringent constraint that besides the data also part of the sensing information should remain unknown to an external eavesdropper. Various future research directions are still open on the information-theoretic framework of secure ISAC, starting with improved achievability and converse results, channels with memory, multi-user systems, and other security constraints regarding the information the eavesdropper obtains about the state-sequence.

Besides the mentioned ISAC scenarios where sensing performance is measured in distortion, this overview article has also considered a fundamentally different ISAC model where the sensing task consists in a detection/hypothesis testing problem. In this setup the relevant property of the sensing target is characterized by a single finite-valued parameter and the goal of the sensing task is to correctly guess this parameter. The main focus here was on the tradeoff between the achievable data rates and the exponential decay rate of the detection-error exponents. Full characterizations of the set of achievable rate-exponents tuples were obtained under the assumption that the Tx produces its channel inputs in a non-adaptive way. Only preliminary results are available under adaptive coding, thus leaving an interesting field of future research directions. Additional possible directions for future research include also network scenarios or setups with memory.

Furthermore, key research topics include the role of Reconfigurable Intelligent Surfaces (RIS), high-frequency systems (Terahertz and beyond), computational paradigms, and holographic technologies in ISAC systems. Additionally, practical wireless propagation aspects, such as near-field effects, require focused attention. Relevant recent works addressing these challenges include [122]–[128].

In conclusion, the convergence of sensing and communication in ISAC systems promises significant advancements in both fields, but it also presents new technical challenges. As highlighted in [129], these challenges span multiple domains and require a comprehensive and interdisciplinary approach. Addressing these challenges will require innovative adaptations of existing models and the development of new schemes that harmonize communication, sensing, and security requirements in increasingly complex environments. One notable direction is the combination of Artificial Intelligence (AI) and ISAC, which has been extensively discussed in [130] and [131]. This integration demands not only algorithmic innovation but also a rigorous theoretical foundation. Information-theoretic literature has proved extremely successful in tackling first standard ISAC models, and one can hope that it will also provide a fertile background obtaining these advanced theoretical foundations.

X. Acknowledgement

References

[1] R. Du, H. Hua, H. Xie, X. Song, Z. Lyu, M. Hu, Narengerile, Y. Xin, S. McCann, M. Montemurro, T. X. Han, and J. Xu, "An overview on ieee 802.11bf: Wlan sensing," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 1, pp. 184–217, 2025.

[2] Y. Wu and M. Wigger, "Coding schemes with rate-limited feedback that improve over the nofeedback capacity for a large class of broadcast channels," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2009–2033, Apr. 2016.

[3] Y. Steinberg, "Instances of the relay-broadcast channel and cooperation strategies," Jun. 2015, pp. 2653–2657.

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

[5] 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.

[6] 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.

[7] H. Kim, Y.-K. Chia, and A. E. Gamal, "A note on the broadcast channel with stale state information at the transmitter," *IEEE Transactions on Information Theory*, vol. 61, no. 7, pp. 3622–3631, 2015.

[8] S. Watanabe, "Neyman–pearson test for zero-rate multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4923–4939, 2018.

[9] C. Tian and J. Chen, "Successive refinement for hypothesis testing and lossless one-helper problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4666–4681, Oct. 2008.

[10] H. Shimokawa, T. Han, and S. I. Amari, "Error bound for hypothesis testing with data compression," in *Proc. ISIT*, Jul. 1994, p. 114.

[11] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing: cooperation and concurrent detection," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7550–7564, 2020.

[12] N. Weinberger, Y. Kochman, and M. Wigger, "Exponent trade-off for hypothesis testing over noisy channels," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1852–1856.

[13] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via korner-marton coding," in *Proc. IEEE Info. Theory Work. (ITW)*, 2016.

[14] W. Zhao and L. Lai, "Distributed testing with cascaded encoders," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7339–7348, 2018.

[15] S. Salehkalaibar and V. Y. Tan, "Distributed sequential hypothesis testing with zero-rate compression," in *2021 IEEE Information Theory Workshop (ITW)*. IEEE, 2021, pp. 1–5.

[16] G. Katz, P. Piantanida, and M. Debbah, "Distributed binary detection with lossy data compression," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5207–5227, 2017.

[17] T. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 2–14, 1989.

[18] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, Nov. 1987.

[19] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6282–6303, Oct. 2012.

[20] R. Ahlswede, "Certain results in coding theory for compound channels," in *Proc. Coll. Information Theory*, vol. 1, 1967, pp. 35–60.

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

[22] B. Isik, W.-N. Chen, A. Ozgur, T. Weissman, and A. No, "Exact optimality of communication-privacy-utility tradeoffs in distributed mean estimation," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. [Online]. Available: https://openreview.net/forum?id=7ETbK9lQd7

[23] U. Hadar and O. Shayevitz, "Distributed estimation of gaussian correlations," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 511–515.

[24] T. Berg, O. Ordentlich, and O. Shayevitz, "Statistical inference with limited memory: A survey," *IEEE Journal on Selected Areas in Information Theory*, vol. 5, pp. 623–644, 2024.

[25] A. Wyner, "On source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 21, no. 3, pp. 294–300, 1975.

[26] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[27] E. Tuncel, "Slepian-wolf coding over broadcast channels," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1469–1–482, 2006.

[28] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1509–1523, 2015.

[29] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *The Bell System Technical Journal*, vol. 53, no. 9, pp. 1681–1721, 1974.

[30] N. Merhav, "Universal decoding for source-channel coding with side information," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 1093–1097.

[31] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 727–734, 1985.

[32] 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.

[33] A. Liu, Z. Huang, M. Li, Y. Wan, W. Li, T. X. Han, C. Liu, R. Du, D. K. P. Tan, J. Lu, Y. Shen, F. Colone, and K. Chetty, "A survey on fundamental limits of integrated sensing and communication," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 994–1034, 2022.

[34] R. W. Heath, "Communications and sensing: An opportunity for automotive systems [from the editor]," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 3–13, 2020.

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

[36] D. Ma, N. Shlezinger, T. Huang, Y. Liu, and Y. C. Eldar, "Joint radar-communication strategies for autonomous vehicles: Combining two key automotive technologies," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 85–97, 2020.

[37] K. V. Mishra, M. B. Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Towards millimeter wave joint radar-communications: A signal processing perspective," vol. 36, no. 5, pp. 100–114, Sep. 2019.

[38] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang, "Radar and communication co-existence: An overview: A review of recent methods," vol. 36, no. 5, pp. 85–99, Sep. 2019.

[39] S. Lu, F. Liu, Y. Li, K. Zhang, H. Huang, J. Zou, X. Li, Y. Dong, F. Dong, J. Zhu, Y. Xiong, W. Yuan, Y. Cui, and L. Hanzo, "Integrated sensing and communications: Recent advances and ten open challenges," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19 094–19 120, 2024.

[40] Z. Wei, J. Jia, Y. Niu, L. Wang, H. Wu, H. Yang, and Z. Feng, "Integrated sensing and communication channel modeling: A survey," *IEEE Internet of Things Journal*, pp. 1–1, 2024.

[41] M. Temiz, Y. Zhang, Y. Fu, C. Zhang, C. Meng, O. Kaplan, and C. Masouros, "Deep learning-based techniques for integrated sensing and communication systems: State-of-the-art, challenges, and opportunities," *TechRxiv*, December 2024, e-Prints posted on TechRxiv are preliminary reports that are not peer-reviewed. [Online]. Available: https://doi.org/10.36227/techrxiv.173532272.22441421/v1

[42] M. Kobayashi and G. Caire, "Information theoretic aspects of joint sensing and communications," in *Book Title*.

[43] A. Liu, M. Li, M. Kobayashi, and G. Caire, *Fundamental Limits for ISAC: Information and Communication Theoretic Perspective*.   Singapore: Springer Nature Singapore, 2023, pp. 23–52. [Online]. Available: https://doi.org/10.1007/978-981-99-2501-8_2

[44] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," vol. 17, no. 4, pp. 2755–2770, 2018.

[45] J. Li and P. Stoica, "MIMO radar with colocated antennas," vol. 24, no. 5, pp. 106–114, Sep. 2007.

[46] C. Xu and S. Zhang, "MIMO integrated sensing and communication exploiting prior information," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 9, pp. 2306–2321, September 2024.

[47] R. Zhang, L. Cheng, S. Wang, Y. Lou, Y. Gao, and W. Wu, "Integrated sensing and communication with massive MIMO: A unified tensor approach for channel and target parameter estimation," *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 8571–8587, August 2024.

[48] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "Joint radar target detection and parameter estimation with MIMO otfs," *arXiv preprint arXiv:2004.11035*, 2020.

[49] Y. Liu, X. Liu, and Y. Chen, "Cell-Free ISAC MIMO Systems: Joint Sensing and Communication Design," *arXiv preprint*, vol. arXiv:2301.11328, 2023, available: https://arxiv.org/abs/2301.11328.

[50] O. Özdogan, E. Björnson, and J. Hoydis, "MIMO-ISAC: Performance Analysis and Rate Region Characterization," *IEEE Wireless Communications Letters*, vol. 12, no. 2, pp. 147–150, Feb. 2023.

[51] J. Zhang, L. Dai, and Z. Wang, "Interference Management in MIMO-ISAC Systems," *arXiv preprint*, vol. arXiv:2407.05391, 2024, available: https://arxiv.org/abs/2407.05391.

[52] H. Zhang, B. Di, and L. Song, "Sensing-Efficient Transmit Beamforming for ISAC with MIMO Radar," *Remote Sensing*, vol. 16, no. 16, p. 3028, Aug. 2023, available: https://www.mdpi.com/2072-4292/16/16/3028.

[53] X. Wang, Y. Li, and M. Tao, "Information and Sensing Beamforming Optimization for Multi-User MIMO-ISAC Systems," *EURASIP Journal on Advances in Signal Processing*, vol. 2023, no. 72, pp. 1–13, 2023, available: https://asp-eurasipjournals.springeropen.com/articles/10.1186/s13634-023-00972-w.

[54] B. Wang, C. Huang, Y. Song *et al.*, "Joint Waveform and Beamforming Design for ISAC: A Survey," *IEEE Transactions on Signal Processing*, vol. 71, pp. 1962–1977, March 2023.

[55] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, "On the fundamental tradeoff of integrated sensing and communications under gaussian channels," *IEEE Transactions on Information Theory*, vol. 69, no. 9, pp. 5723–5751, 2023.

[56] X. Li, H. Min, Y. Zeng, S. Jin, L. Dai, Y. Yuan, and R. Zhang, "Sparse mimo for isac: New opportunities and challenges," 2024. [Online]. Available: https://arxiv.org/abs/2406.12270

[57] C. Ouyang, Y. Liu, and H. Yang, "Mimo-isac: Performance analysis and rate region characterization," *IEEE Wireless Communications Letters*, vol. 12, pp. 669–673, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:249152014

[58] 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.

[59] M. Ahmadipour, M. Wigger, and S. Shamai, "Strong converses for memoryless bi-static isac," in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 1818–1823.

[60] H. Joudeh and G. Caire, "Joint communication and state sensing under logarithmic loss," in *2024 4th IEEE International Symposium on Joint Communications & Sensing (JC&S)*.   Institute of Electrical and Electronics Engineers, 2024, in press.

[61] H. Nikbakht, M. Wigger, S. Shamai, and H. Poor, "Integrated sensing and communication in the finite blocklength regime," in *ISIT 2024*, Athens, Greece, July 7–12 2024.

[62] Y. Chen, T. Oechtering, M. Skoglund, and Y. Luo, "On general capacity-distortion formulas of integrated sensing and communication," 2023. [Online]. Available: https://arxiv.org/abs/2310.11080

[63] 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.

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

[65] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. Info. Theory*, vol. 51, no. 4, pp. 1486–1495, 2005.

[66] M. Ahmadipour, "An Information-theoretic approach to integrated sensing and communication," Theses, Institut Polytechnique de Paris, Nov. 2022. [Online]. Available: https://theses.hal.science/tel-03910385

[67] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, "An information-theoretic approach to joint sensing and communication," *IEEE Transactions on Information Theory*, pp. 1–1, 2022.

[68] 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.

[69] Y. Liu, M. Li, A. Liu, L. Ong, and A. Yener, "Fundamental limits of multiple-access integrated sensing and communication systems," *arXiv preprint arXiv:2205.05328v3*, 2023.

[70] M. Ahmadipour, M. Wigger, and M. Kobayashi, "Coding for sensing: An improved scheme for integrated sensing and communication over macs," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 3025–3030.

[71] M. Ahmadipour and M. Wigger, "An information-theoretic approach to collaborative integrated sensing and communication for two-transmitter systems," *IEEE Journal on Selected Areas in Information Theory*, vol. 4, pp. 112–127, 2023.

[72] O. Günlü, M. R. Bloch, R. F. Schaefer, and A. Yener, "Secure integrated sensing and communication," *IEEE Journal on Selected Areas in Information Theory*, vol. 4, pp. 40–53, 2023.

[73] M. Ahmadipour, M. Wigger, and S. Shamai, "Strong converses for memoryless bi-static isac," in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 1818–1823.

[74] H. Joudeh and F. M. J. Willems, "Joint communication and binary state detection," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 1, pp. 113–124, 2022.

[75] H. Wu and H. Joudeh, "On joint communication and channel discrimination," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 3321–3326.

[76] M.-C. Chang, , Erdogan, S.-Y. Wang, and M. R. Bloch, "Rate and detection error-exponent tradeoffs of joint communication and sensing," in *2022 2nd IEEE International Symposium on Joint Communications and Sensing (JCS)*, 2022, pp. 1–6.

[77] M. Ahmadipour, M. Wigger, and S. Shamai, "Strong converse for bi-static isac with two detection-error exponents," in *Proceedings of the International Zurich Seminar on Information and Communication (IZS 2024)*, March 2024, p. 45.

[78] D. Seo and S. H. Lim, "On the fundamental tradeoff of joint communication and quickest change detection," *arXiv preprint arXiv:2401.12499*, 2024, available at http://arxiv.org/abs/2401.12499v1.

[79] R. M. Mealey, "A method for calculating error probabilities in a radar communication system," *IEEE Transactions on Space Electronics and Telemetry*, vol. 9, no. 2, pp. 37–42, 1963.

[80] M. R.Winkler, "Chirp signals for communications," *IEEE WESCON Cony. Rec*, 1962.

[81] A. I. Bemi and W. Grcgg, "On the utility of chirp modulation for digital signaling," *IEEE Trans. Cnmmun*, vol. COM-21, pp. 748–751, I973.

[82] Q. Zhang, H. Sun, X. Gao, X. Wang, and Z. Feng, "Time-division isac enabled connected automated vehicles cooperation algorithm design and performance evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 7, pp. 2206–2218, 2022.

[83] C. Shi, F. Wang, M. Sellathurai, J. Zhou, and S. Salous, "Power minimization-based robust ofdm radar waveform design for radar and communication systems in coexistence," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1316–1330, 2018.

[84] S. K. Mohammed, R. Hadani, A. Chockalingam, and R. Calderbank, "Otfs—a mathematical foundation for communication and radar sensing in the delay-doppler domain," *IEEE BITS the Information Theory Magazine*, vol. 2, no. 2, pp. 36–55, 2022.

[85] J. Wu, W. Yuan, Z. Wei, K. Zhang, F. Liu, and D. Wing Kwan Ng, "Low-complexity minimum ber precoder design for isac systems: A delay-doppler perspective," *IEEE Transactions on Wireless Communications*, vol. 24, no. 2, pp. 1526–1540, 2025.

[86] X. Lin, "3gpp evolution from 5g to 6g: A 10-year retrospective," *arXiv preprint arXiv:2412.21077*, 2024.

[87] S. Sodagari, A. Khawar, T. C. Clancy, and R. McGwier, "A projection based approach for radar and telecommunication systems coexistence," in *2012 IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 5010–5014.

[88] O. Günlü, M. Bloch, R. F. Schaefer, and A. Yener, "Nonasymptotic performance limits of low-latency secure integrated sensing and communication systems," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 12 971–12 975.

[89] H. Nikbakht, M. Wigger, S. Shamai, and H. V. Poor, "A memory-based reinforcement learning approach to integrated sensing and communication," 2024. [Online]. Available: https://arxiv.org/abs/2412.01077

[90] Z. Aharoni, O. Sabag, and H. H. Permuter, "Computing the feedback capacity of finite state channels using reinforcement learning," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 837–841.

[91] W. Zhang, S. Vedantam, and U. Mitra, "A constrained channel coding approach to joint communication and channel estimation," in *2008 IEEE International Symposium on Information Theory*, 2008, pp. 930–934.

[92] C. Choudhuri and U. M. Ming, "On non-causal side information at the encoder," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 648–655.

[93] A. Salimi, W. Zhang, S. Vedantam, and U. Mitra, "The capacity-distortion function for multihop channels with state," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2228–2232.

[94] S. I. Gelfand and M. S. Pinsker, "Coding for channels with random parameters," *Problem of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[95] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. USA: Cambridge University Press, 2012.

[96] V. Koivunen, M. F. Keskin, H. Wymeersch, M. Valkama, and N. González-Prelcic, "Multicarrier isac: Advances in waveform design, signal processing, and learning under nonidealities [special issue on signal processing for the integrated sensing and communications revolution]," *IEEE Signal Processing Magazine*, vol. 41, no. 5, pp. 17–30, 2024.

[97] A. Gamal, "The feedback capacity of degraded broadcast channels (corresp.)," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 379–381, 1978.

[98] F. Willems, E. van der Meulen, and J. Schalkwijk, "Achievable rate region for the multiple access channel with generalized feedback," in *Proc. Annual Allerton Conf. on Communication, Control and Computing*, 1983, pp. 284–292.

[99] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 4–21, 2003.

[100] L. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 623–629, 1984.

[101] F. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels (corresp.)," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 93–95, 1982.

[102] G. N. T. and W. J. K. , "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Transactions on Information Theory*, vol. 21, no. 1, p. 100–102, 1975.

[103] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Transactions on Information Theory*, vol. 27, no. 3, pp. 292–298, 1981.

[104] A. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 841–850, 1982.

[105] A. Hekstra and F. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 44–53, 1989.

[106] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Double state," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1379–1393, 2013.

[107] ——, "The multiple-access channel with causal side information: Common state," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 32–50, 2013.

[108] A. Somekh-Baruch, S. Shamai, and S. Verdu, "Cooperative multiple-access encoding with states available at one transmitter," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4448–4469, 2008.

[109] S. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to one encoder: A case of degraded message sets," in *2007 IEEE International Symposium on Information Theory*, June 2007, pp. 1566–1570.

[110] M. Li, O. Simeone, and A. Yener, "Multiple access channels with states causally known at transmitters," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1394–1404, 2013.

[111] Y. Liu, M. Li, Y. Han, and L. Ong, "Information- theoretic limits of integrated sensing and communication over interference channels," in *ICC 2024 - IEEE International Conference on Communications*, 2024, pp. 3561–3566.

[112] T. Han, "A general coding scheme for the two-way channel," *IEEE Transactions on Information Theory*, vol. 30, no. 1, pp. 35–44, 1984.

[113] G. Kramer, "Directed information for channels with feedback," *Ph.D. dissertation, Swiss Federal Institute of Technology Zurich*, 1998.

[114] O. Gunlu, M. Bloch, R. F. Schaefer, and A. Yener, "Secure integrated sensing and communication for binary input additive white gaussian noise channels," in *2023 IEEE 3rd International Symposium on Joint Communications & Sensing (JC & S)*, 2023, pp. 1–6.

[115] M. Mittelbach, R. F. Schaefer, M. Bloch, A. Yener, and O. Gunlu, "Secure integrated sensing and communication under correlated rayleigh fading," Cryptology ePrint Archive, Paper 2024/290, 2024, https://eprint.iacr.org/2024/290. [Online]. Available: https://eprint.iacr.org/2024/290

[116] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6760–6786, 2014.

[117] C. Choudhuri and U. M. Ming, "On non-causal side information at the encoder," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 648–655.

[118] M.-C. Chang, S.-Y. Wang, T. Erdoğan, and M. R. Bloch, "Rate and detection-error exponent tradeoff for joint communication and sensing of fixed channel states," *IEEE Journal on Selected Areas in Information Theory*, vol. 4, pp. 245–259, 2023.

[119] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.

[120] M.-C. Chang, S.-Y. Wang, and M. R. Bloch, "Sequential joint communication and sensing of fixed channel states," in *2023 IEEE Information Theory Workshop (ITW)*, 2023, pp. 462–467.

[121] A. Tandon, M. Motani, and L. R. Varshney, "Subblock-constrained codes for real-time simultaneous energy and information transfer," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 4212–4227, 2016.

[122] D. Wen, Y. Zhou, X. Li, Y. Shi, K. Huang, and K. B. Letaief, "A survey on integrated sensing, communication, and computation," *IEEE Communications Surveys & Tutorials*, early Access.

[123] J. Cong, C. You, J. Li, L. Chen, B. Zheng, and Y. Liu, "Near-field integrated sensing and communication: Opportunities and challenges," *IEEE Wireless Communications*, vol. 31, no. 6, pp. 162–169, December 2024.

[124] W. Jiang, Q. Zhou, J. He, M. A. Habibi, S. Melnyk, and M. El-Absi, "Terahertz communications and sensing for 6g and beyond: A comprehensive review," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 4, pp. 2326–2381, Fourthquarter 2024.

[125] A. M. Elbir, K. V. Mishra, S. Chatzinotas, and M. Bennis, "Terahertz-band integrated sensing and communications: Challenges and opportunities," *IEEE Aerospace and Electronic Systems Magazine*, vol. 39, no. 12, pp. 38–49, December 2024.

[126] H. Zhang, H. Zhang, B. Di, M. D. Renzo, Z. Han, and H. V. Poor, "Holographic integrated sensing and communication," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 7, pp. 2114–2130, July 2022.

[127] R. Liu, M. Li, H. Luo, Q. Liu, and A. L. Swindlehurst, "Integrated sensing and communication with reconfigurable intelligent surfaces: Opportunities, applications, and future directions," *IEEE Wireless Communications*, vol. 30, no. 1, pp. 50–57, February 2023.

[128] Q. Ye, Y. Huang, Q. Luo, Z. Hu, Z. Zhang, Q. Zhao, Y. Su, S. Hu, and G. Yang, "A general integrated sensing and communication channel model combined with scattering clusters," *IEEE Transactions on Vehicular Technology*, pp. 1–14, 2024.

[129] S. Lu, F. Liu, Y. Li, K. Zhang, H. Huang, J. Zou, X. Li, Y. Dong, F. Dong, J. Zhu, Y. Xiong, W. Yuan, Y. Cui, and L. Hanzo, "Integrated sensing and communications: Recent advances and ten open challenges," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19 094–19 120, 2024.

[130] N. Wu, R. Jiang, X. Wang, L. Yang, K. Zhang, W. Yi, and A. Nallanathan, "Ai-enhanced integrated sensing and communications: Advancements, challenges, and prospects," *IEEE Communications Magazine*, vol. 62, no. 9, pp. 144–150, 2024.

[131] X. Liu, H. Zhang, K. Sun, K. Long, and G. K. Karagiannidis, "Ai-driven integration of sensing and communication in the 6g era," *IEEE Network*, vol. 38, no. 3, pp. 210–217, 2024.