Joint Communication and Parameter Estimation in MIMO Channels

Gökhan Yılmaz, Franz Lampel, Hamdi Joudeh and Giuseppe Caire

Abstract

We study a joint communication and sensing setting comprising a transmitter, a receiver, and a sensor, all equipped with multiple antennas. The transmitter sends an encoded signal over the channel with the dual purpose of communicating an information message to the receiver, and enabling the sensor to estimate a target parameter vector by generating back-scattered signals. We assume that the transmitter and sensor are co-located, or fully connected, giving the latter access to the transmitted signal. The target parameter vector is randomly drawn from a continuous distribution, yet remains fixed throughout the transmission block. We establish the fundamental performance trade-off between the communication and sensing tasks, captured in terms of a capacity-MSE function. In doing so, we identify optimal coding schemes for this multi-antenna joint communication and sensing setting. Moreover, we particularize our result to two practically-inspired scenarios where we showcase optimal schemes and trade-offs.

I. Introduction

Joint communication and sensing (JCAS), also known as integrated sensing and communication (ISAC), is expected to be a key feature of next-generation wireless technology. This emerging paradigm aims at designing efficient wireless systems with transceivers that can utilize the same physical resources (e.g., hardware, spectrum, and power) to carry out both communication and sensing tasks [1]–[5]. Envisaged use cases include cellular networks, where base stations will have the ability to communicate with active devices and simultaneously detect and track passive targets from back-scattered signals [6]; and automotive, where vehicles will be able to sense their

Gökhan Yılmaz, Franz Lampel, and Hamdi Joudeh are with the Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: g.yilmaz@tue.nl; f.lampel@tue.nl; h.joudeh@tue.nl).

Giuseppe Caire is with the Faculty of Electrical Engineering and Computer Science, Technical University of Berlin, 10623 Berlin, Germany (e-mail: caire@tu-berlin.de).

This work was supported by the European Research Council (ERC) under Grant 101116550.

surroundings to identify and track road obstacles, while communicating with other vehicles and the infrastructure [7], [8]. In both of these use cases, and several others, the underlying JCAS physical channel will very likely be a multiple-input and multiple-output (MIMO) channel, due to the employment of multi-antenna terminals, multiple sub-carriers, or both.

In recent years, a number of studies have emerged investigating various aspects of MIMO in JCAS; see, e.g., [2]–[5], [9]–[14]. With the exception of [13], [14], the vast majority of studies on MIMO JCAS focus mainly on signal processing and optimization aspects. In particular, a transmission scheme with a predetermined signal structure, often based on linear precoding and Gaussian signaling, is adopted; reducing the problem to the design of signal parameters (e.g., covariance matrices) to achieve different performance trade-offs. While these works have provided useful insights into the potential benefits of co-designed MIMO JCAS systems, without an information-theoretic treatment involving coding theorems and converse bounds, the optimality of these predetermined signal structures and schemes remains unclear. Moreover, we still lack performance benchmarks that indicate how much more can be gained.

Goal and considered setting: Our primary goal in this work is to shed light on the fundamental performance limits of MIMO JCAS systems by taking an information-theoretic approach. To this end, we focus on an elemental setting comprising a transmitter, a receiver, and a sensor, all equipped with multiple antennas (or any vector dimension, e.g., sub-carriers). The transmitter and sensor are co-located or strongly connected, enabling mono-static sensing, and henceforth can be thought of as components of the same cellular base station (BS). The receiver, on the other hand, belongs to an active user equipment (UE) served by the BS. The transmitter sends an encoded signal with the dual purpose of communicating a message to the receiver and, at the same time, probing the surrounding environment and generating a back-scatter signal to enable target sensing. We refer to this setting as a *MIMO JCAS channel*.

Information-theoretic formulations and performance limits for MIMO *communication* channels are well established and extensively studied [15]–[17]. To treat MIMO JCAS channels with the same degree of rigor, it is crucial to concretely formulate the *sensing* aspect of the problem. In the current work, we focus on the case where the sensing task involves estimating a *continuous* parameter vector of the back-scatter channel. The parameter vector is drawn at random according to a known prior distribution, yet remains *fixed* throughout the transmission block. This *abstraction* is meant to model scenarios where parameters of interest change at a much slower timescale compared to channel uses (i.e., symbols). Practical examples include direction-of-arrival (DoA),

range (delay), and velocity (Doppler) estimation, all scenarios in which only slight changes in parameters occur over the span of a typical transmission block (see Remark 2).

Fixed vs. varying parameter: Our choice to adopt a fixed random parameter model for sensing stands in sharp contrast to a prevalent approach in the information-theoretic literature on JCAS, where the parameter of interest is assumed to vary in an i.i.d. fashion from one channel symbol to the other. The i.i.d. model was proposed by Kobayashi *et al.* in [18], where the focus is on discrete memoryless and basic Gaussian channels; and has since then been widely adopted in several follow-up works [19]–[23]. In [13], Xiong *et al.* study a MIMO JCAS channel that is very similar to ours, with the exception that their sensing model also follows the i.i.d. philosophy in [18]. In particular, they consider the estimation of a continuous parameter vector that varies across channel symbols (or sub-blocks of symbols) in an i.i.d. fashion. An asymptotically large number of symbols (or sub-blocks) is considered, resembling ergodic fading in wireless communication. Moreover, to measure the sensing performance, a form of Bayesian Cramér-Rao bound (BCRB) [24] is used in [13] as a proxy for the mean square error (MSE) averaged over parameter states.

The BCRB is known to be generally not tight, even asymptotically, except in certain special cases [25]; hence, the performances characterized in [13] are not exact in general. Nevertheless, high signal-to-noise-ratio (SNR) analysis reveals two distinct types of trade-offs in the MIMO JCAS channel model of [13]. The first is referred to as the *deterministic–random trade-off (DRT)*, which arises when the family of Gaussian input distributions—optimal for communication alone—fails to achieve all Pareto-optimal performance trade-off points. In this case, "less random" input distributions are required to achieve trade-off points that favor sensing. The second is referred to as the *subspace trade-off (ST)*, which arises from the fact that the communication user and the sensing target often occupy distinct subspaces in the MIMO vector space. Hence, signals more aligned with the user (resp. target) favor communication (resp. sensing).

The i.i.d. model does not fully capture practical scenarios where the parameters of interest remain relatively constant within transmission blocks, as the ones highlighted earlier. The idea to study fundamental performance limits in JCAS channels with a fixed parameter originated from Joudeh and Willems in [26], where the focus is on the basic sensing task of target detection (i.e., a binary parameter). Extensions later followed in [14], [27]–[29], where the focus remained on discrete (mostly binary) parameters. The extension to MIMO JCAS channels in [14] is particularly relevant, and can be seen as a detection counterpart to the estimation problem we consider in the present paper. In a recent contribution, Seo [30] extended the fixed discrete

parameter JCAS setting in [26]–[28] to the continuous parameter regime. Our current paper follows in the footsteps of [30], and further extends the approach therein in several directions, as discussed in detail below in light of our contributions.

Overview and contribution: We consider a MIMO JCAS channel where the sensing task is to estimate a random yet fixed parameter vector, as described earlier. We adopt a Bayesian framework for parameter estimation, with an MSE-based risk function [25], [31]. For this setting, we establish the fundamental trade-off between communication and sensing efficiencies in the asymptotic regime of large block-length. Communication efficiency is measured in terms of the rate of reliable communication, in bits per channel use; whose ultimate asymptotic limit is the celebrated Shannon capacity [32]. Sensing efficiency, on the other hand, is measured in terms of the MSE risk scaled by the block-length, which we refer to as the *asymptotic MSE*. This choice of scaling is commonplace in asymptotic statistics, as pointed out in [33, Sec. I], since the MSE risk in such settings decays at a rate proportional to one over the block-length. Therefore, a larger asymptotic MSE implies a slower decay rate and vice versa.

In the main result of this paper, presented in Section III, we derive the exact optimal trade-off between the rate of reliable communication and the sensing asymptotic MSE, described in terms of a capacity-MSE function. The rate is characterized by a mutual information term, while the asymptotic MSE is characterized by a conditional variant of the Expected Cramér-Rao bound (ECRB) [25], adapted to our sensing setting. It is interesting to note that while the ECRB does not generally yield a valid lower bound at finite block-lengths, it is asymptotically tight [25], [34]. We extend ECRB's asymptotic tightness to the case of independent non-identical measurements arising in JCAS, from which the conditional ECRB variant emerges. Unlike the BCRB used in [13], the ECRB enables an exact asymptotic characterization in our setting. Moreover, we showcase the established fundamental trade-off through two practically-inspired use cases, one involving DoA estimation and another involving OFDM signaling (see Section V).

An interesting feature revealed by our analysis is that all optimal rate-MSE trade-off points are characterized by Gaussian input distributions with different choices of covariance matrices. In the terminology of Xiong *et al.* [13], the MIMO JCAS setting considered here exhibits only an ST and no DRT, with Gaussian-like communication waveforms also being optimal for sensing. This is mainly because the sensing channel parameter is fixed. Arriving at this conclusion, however, is nontrivial, since i.i.d. Gaussian codebook ensembles widely used in standard achievability proofs do not directly apply in our case. Intuitively, i.i.d.-generated codebooks occasionally

contain codewords that are unfavorable for sensing. Alternatively, we rely on *almost-constant covariance codes* [14], which yield uniformly favorable sensing waveforms while asymptotically exhibiting Gaussian-like properties that preserve their communication optimality.

Related work: A capacity-MSE trade-off, similar to the one we present here, was recently derived by Seo [30], albeit in a much more restricted setting. A major difference is that the approach of Seo is only applicable to discrete memoryless channels (DMCs), where inputs and outputs come from finite sets (despite the parameter being continuous). In particular, Seo follows the approach developed by Wu and Joudeh in [27], [35], which relies on constant composition codes to prove achievability and type-by-type analysis to prove the converse. Such an approach is not applicable to Gaussian (or other continuous) channels, and cannot be used to analyze the MIMO JCAS setting at hand. In contrast, our achievability proof is based on Feinstein's lemma, which is applicable to channels with general alphabets [36]. Moreover, our converse proof avoids the type-based argument in [30] and instead relies on a convexity property of the conditional ECRB, which we also prove in this paper. Another smaller difference is that Seo considers a scalar sensing channel parameter only, while we consider a general vector parameter.

The MIMO JCAS channel in the fixed-parameter regime has also been studied in the prior literature, e.g., [9]–[12]. With the exception of [12], however, these works adopt the classical CRB within a non-Bayesian framework. As noted earlier, they also focus primarily on signal processing and optimization aspects without providing information-theoretic optimality guarantees. Moreover, the reliance on single-letter models, prevalent in the wireless communication literature, typically involves an informal notion of empirical input covariance convergence under the generic pretext of "Gaussian signaling" (see, e.g., [9, eq. (3)] and [10]). We address this issue by adopting a multi-letter perspective and employing almost-constant covariance codes.

Finally, although Xiong *et al.* [13] consider the i.i.d. varying-parameter regime, the obtained rate-BCRB trade-off can be applied to the Bayesian fixed-parameter regime we consider in the current paper. As we show, this yields an outer bound that is strictly loose, and hence not achievable, in general. This has also been observed by Seo [30] in the context of the DMC.

Notation: We adopt basic notation, which is mostly clear from the context. Upper-case bold letters as $\boldsymbol{A}$ denote matrices; lower-case bold letters as $\boldsymbol{a}$ denote (column) vectors; while A and a are scalars. Calligraphic letters as $\mathcal{A}$ denote sets. Random variables are denoted as $\mathbf{A}$, $\mathbf{a}$ and a, while $\boldsymbol{A}$, $\boldsymbol{a}$ and a denote deterministic quantities (e.g., realizations).

II. Problem Setting

We consider a vector (i.e., MIMO) JCAS channel comprising an M-antenna transmitter, an L-antenna receiver (or user), and a T-antenna sensor.[1] The transmitter sends an encoded signal over the channel with the dual purpose of *communicating* an information message to the user, and enabling *sensing* of a target or the environment by generating back-scattered signals for the sensor. We assume that the sensor has access to the transmitted encoded signal and can perform coherent sensing. This is the case in mono-static sensing, where the transmitter and sensor are co-located; or in bi-static sensing with a strong link between the transmitter and sensor.

A. Signal model

We consider a base-band discrete-time model, in which each channel use can be interpreted as one symbol period. In the n-th use, the transmitter sends a vector channel symbol $\mathbf{x}_n \in \mathbb{C}^M$ through its M antennas. The user receives a linearly-transformed noisy vector $\mathbf{y}_n \in \mathbb{C}^L$ given as

$$\mathbf{y}_n = \boldsymbol{H}\mathbf{x}_n + \boldsymbol{\omega}_n. \tag{1}$$

$\boldsymbol{H} \in \mathbb{C}^{L\times M}$ is the channel matrix between the transmitter and the user and $\boldsymbol{\omega}_n \sim \mathcal{CN}(\mathbf{0}, \sigma_{\boldsymbol{\omega}}^2 \boldsymbol{I})$ is a zero-mean circularly symmetric complex Gaussian noise vector, independent over channel uses n. We refer to $\boldsymbol{H}$ as the *user channel matrix*. This remains fixed over channel uses in the setting we consider in this paper, and it is assumed to be known to both the transmitter and the user. An underlying assumption here is that initial access, including channel training, between the transmitter and the user has already been performed.

The sensor receives a noisy echo signal $\mathbf{z}_n \in \mathbb{C}^T$ through a back-scatter channel modeled by

$$\mathbf{z}_n = \boldsymbol{G}(\theta)\mathbf{x}_n + \boldsymbol{\upsilon}_n \tag{2}$$

where $\boldsymbol{G}(\theta) \in \mathbb{C}^{T\times M}$ is the channel response matrix from the transmitter to the target (or environment) and back to the sensor, and $\boldsymbol{\upsilon}_n \sim \mathcal{CN}(\mathbf{0}, \sigma_{\boldsymbol{\upsilon}}^2 \boldsymbol{I})$ is the corresponding additive noise vector. We refer to $\boldsymbol{G}(\theta)$ as the *sensor channel matrix*. It is a function of a real *target parameter vector* $\theta \in \mathbb{R}^K$, drawn from a continuous distribution with probability density function $f_\theta(\boldsymbol{\theta})$ on $\mathbb{R}^K$. The sensor is interested in estimating the target parameter vector, as it contains the relevant information about the target or environment. We assume that the mapping $\boldsymbol{G} : \mathbb{R}^K \to \mathbb{C}^{T\times M}$ from

[1]The vector dimension is interpreted as antennas, but can also represent time extension or sub-carriers (e.g., OFDM).

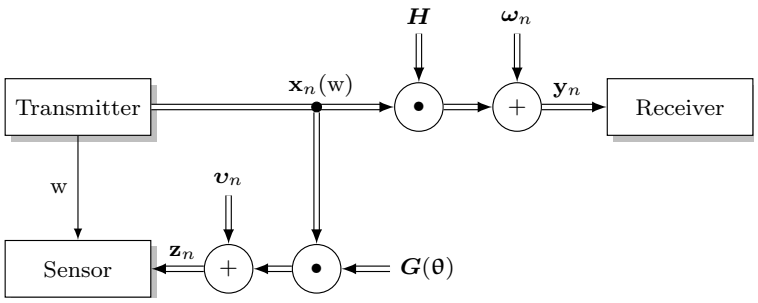


Fig. 1: The considered MIMO JCAS system model.

the target parameter vector to the sensor channel matrix is differentiable and invertible (almost surely on the support of θ). As we shall see in Section V, where we discuss particularized examples of this model, these assumptions are justifiable in many settings of practical interest. An illustration of our model is shown in Fig. 1.

Note that while θ (and therefore $\boldsymbol{G}(\theta)$) is drawn at random, it remains fixed throughout the transmission block in the setting we consider. This represents scenarios where the target parameter changes at a much slower time scale compared to channel uses within a single transmission block, which is the case in many practical scenarios, as explained earlier.

Transmission occurs over a block of N channel uses, where a signal (i.e., a sequence of vector channel symbols) given by $\boldsymbol{x}^N \triangleq \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ is sent. As we specify shortly, the transmitted signal $\boldsymbol{x}^N$ is a codeword selected from a codebook, and encodes a message intended for the user. It also serves the additional purpose of probing, to enable parameter estimation at the sensor. The transmitted signal is subject to an average (over symbols) power constraint expressed by

$$\frac{1}{N}\sum_{n=1}^{N}\|\boldsymbol{x}_n\|^2 \leq P \tag{3}$$

where P is the average transmit power budget.

B. Encoding, decoding, and estimation

We now formally describe a general JCAS scheme for the above setting, which involves encoding at the transmitter, decoding at the user, and estimation at the sensor.

1) Encoding: In a given transmission block of length N, the transmitter wishes to communicate a message w to the user, where w is drawn uniformly at random from a message set

of M_N indices given by $\mathcal{M}_N \triangleq \{1, 2, \ldots, M_N\}$. To this end, an encoded signal (or codeword) $\boldsymbol{x}^N(w) = \boldsymbol{x}_1(w), \ldots, \boldsymbol{x}_N(w)$ is sent. The set of all codewords, i.e., codebook, is given by

$$\boldsymbol{\mathcal{C}}_N \triangleq \left\{\boldsymbol{x}^N(1), \boldsymbol{x}^N(2), \ldots, \boldsymbol{x}^N(M_N)\right\}.$$

Each codeword in the codebook must satisfy the power constraint in (3). Moreover, since w is uniform, then the random transmitted signal $\mathbf{x}^N \triangleq \boldsymbol{x}^N(\mathrm{w})$ is uniform on $\boldsymbol{\mathcal{C}}_N$. Note that the signals received by the user and the sensor depend on w through $\boldsymbol{x}^N(\mathrm{w})$, and hence conditioning on $\mathrm{w} = w$ in what follows is often replaced with conditioning on $\mathbf{x}^N = \boldsymbol{x}^N(w)$.

2) Decoding: The user receives a random signal given by an N-sequence of vectors $\mathbf{y}^N \triangleq \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ from which it produces a decoded message $\hat{w}(\mathbf{y}^N) \in \mathcal{M}_N$. Recall the assumption that the user has access to its channel matrix $\boldsymbol{H}$, on which the decoder may depend. Given that $\mathrm{w} = w$ is sent, and hence $\mathbf{x}^N = \boldsymbol{x}^N(w)$ is transmitted, the probability of decoding error is

$$\varepsilon_N(w) \triangleq \mathbb{P}\left[\hat{w}(\mathbf{y}^N) \neq \mathrm{w} \mid \mathbf{x}^N = \boldsymbol{x}^N(w)\right]. \tag{4}$$

The *communication reliability* is measured by the maximal probability of decoding error

$$\varepsilon_N \triangleq \max_{w \in \mathcal{M}_N} \varepsilon_N(w). \tag{5}$$

We shall refer to this as the *error probability* for short. The maximum over all messages in (5) reflects an underlying assumption that all messages are equally important, i.e., there is no reason to distinguish between different message reliabilities. It is sometimes technically more convenient to work with the average (over messages) decoding error probability $\varepsilon_N^{\mathrm{av}} = \frac{1}{M_N} \sum_{w \in \mathcal{M}_N} \varepsilon_N(w)$. However, as is often the case in asymptotic analysis of point-to-point communication channels, capacity results remain unchanged regardless of whether we choose ε_N or $\varepsilon_N^{\mathrm{av}}$ as the measure of decoding error probability. This is also the case for the results we present in this paper.

3) Estimation: The sensor receives the random signal given by $\mathbf{z}^N \triangleq \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$, and with knowledge of the codeword $\boldsymbol{x}^N(w)$, it produces an estimate $\hat{\boldsymbol{\theta}}_N(\mathbf{z}^N, \boldsymbol{x}^N(w))$ of the random target parameter vector $\boldsymbol{\theta}$. Note that the estimation function $\hat{\boldsymbol{\theta}}_N : \mathbb{C}^{T \times N} \times \boldsymbol{\mathcal{C}}_N \to \mathbb{R}^K$ itself is a deterministic mapping, while the estimate $\hat{\boldsymbol{\theta}}_N(\mathbf{z}^N, \boldsymbol{x}^N(w))$ is a random variable due to the randomness of the observation $\mathbf{z}^N$. For brevity, we sometimes write $\hat{\boldsymbol{\theta}}_N(\mathbf{z}^N)$ as $\boldsymbol{x}^N(w)$ is given.

A quadratic loss criterion is used to measure the estimation error, leading to an MSE-based estimation risk described next. To formulate the estimation risk, we shall first define the MSE matrix. Given an estimator $\hat{\boldsymbol{\theta}}_N$ and that $\boldsymbol{x}^N(w)$ is sent, the MSE matrix is defined as

$$\boldsymbol{\Sigma}\left(\boldsymbol{x}^N(w), \hat{\boldsymbol{\theta}}_N\right) \triangleq \mathbb{E}\left[\left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N(\mathbf{z}^N)\right]\left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N(\mathbf{z}^N)\right]^{\mathsf{T}} \,\middle|\, \mathbf{x}^N = \boldsymbol{x}^N(w)\right] \tag{6}$$

where the expectation in (6) is with respect to $\boldsymbol{\theta}$ and $\mathbf{z}^N$, given $\mathbf{x}^N = \boldsymbol{x}^N(w)$.[2] The diagonal elements of $\boldsymbol{\Sigma}(\boldsymbol{x}^N(w), \hat{\boldsymbol{\theta}}_N)$ represent the MSEs of individual target parameters in $\boldsymbol{\theta}$. The error in estimating the target parameter vector, given codeword $\boldsymbol{x}^N(w)$, is measured in terms of the sum-MSE risk (or *MSE risk* for short), defined as

$$\xi_N(w) \triangleq \operatorname{tr}\left(\boldsymbol{\Sigma}\left(\boldsymbol{x}^N(w), \hat{\boldsymbol{\theta}}_N\right)\right). \tag{7}$$

The *sensing risk* is chosen to be the *maximal MSE risk* (maximal over messages) defined as

$$\xi_N \triangleq \max_{w \in \mathcal{M}_N} \xi_N(w). \tag{8}$$

The maximum over messages in (8) ensures a uniform sensing risk across different messages, or codewords in the codebook. This reflects a natural operational requirement that the sensing performance should be independent of which message is being communicated, or equivalently, which codeword is being transmitted. Such a formulation gives rise to the interesting design problem of finding *good* channel codebooks with the important property that all codewords also constitute *equally good* probing signals for the sensing task.

A less stringent measure of sensing risk is the average (over messages) MSE risk, defined as $\xi_N^{\text{av}} \triangleq \frac{1}{M_N}\sum_{w \in \mathcal{M}_N} \xi_N(w)$. This, however, does not provide the same sensing performance guarantees. Interestingly, as we shall see further on, regardless of whether we choose ξ_N or ξ_N^{av}, the fundamental asymptotic trade-off we establish in the next section remains unchanged.

Remark 1. In many practical settings, individual target parameters may carry different importance. For such scenarios, we may introduce the weights $a_1, a_2, \ldots, a_K$, where $a_i \geq 0$ signifies the "importance" of the i-th parameter in $\boldsymbol{\theta}$. Weights are assumed to be fixed and known beforehand. In this case, we modify the MSE in (7) into a weighted MSE (WMSE)

$$\xi_N(w) \triangleq \sum_{i=1}^{K} a_i \left[\boldsymbol{\Sigma}\left(\boldsymbol{x}^N(w), \hat{\boldsymbol{\theta}}_N\right)\right]_{ii} \tag{9}$$

$$= \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{\Sigma}\left(\boldsymbol{x}^N(w), \hat{\boldsymbol{\theta}}_N\right)\right) \tag{10}$$

where $\boldsymbol{A} = \operatorname{diag}(a_1, a_2, \ldots, a_K)$ is a diagonal matrix induced by the weights. The WMSE sensing risk is then obtained from (9) by taking the maximum or average over messages. For ease of exposition, we will present our results in terms of the unweighted MSE risk. Nevertheless, all results and proofs can be easily extended to account for the weighted case.

[2]Our choice of MSE is sometimes referred to as the Bayesian MSE [31], since it involves an expectation over the prior of $\boldsymbol{\theta}$.

C. Asymptotic rate-MSE trade-off

A JCAS scheme, as defined above, is referred to as an $(N, M_N, \varepsilon_N, \xi_N)$-scheme if it has a codebook $\boldsymbol{\mathcal{C}}_N$ of M_N length-N codewords, a decoder with maximal error probability ε_N, and an estimator with maximal MSE risk ξ_N. We are interested in characterizing the fundamental performance limits of such schemes in the asymptotic regime of $N \to \infty$.

As the block-length N grows, it is known from the channel coding theorem that the size of the message set M_N can be made to grow exponentially in N while guaranteeing reliable decoding, i.e., $\varepsilon_N \to 0$. Therefore, communication efficiency is often measured by the *rate* $\frac{1}{N}\log M_N$, which is understood as the number of communicated bits per channel use.

On the other hand, we expect the MSE sensing risk in (8) to decay to zero as N grows large at a rate of $O(1/N)$, since $\boldsymbol{\theta}$ is fixed and the measurements are conditionally independent [25]. Therefore, an appropriate asymptotic measure of sensing risk is the scaled MSE $N\xi_N$. We will refer to the scaled MSE as the *asymptotic MSE*, or *MSE* for short, where it should be clear from the context whether we are referring to ξ_N or the scaled version $N\xi_N$. We are now in position to introduce the notion of asymptotically achievable rate-MSE pairs.

Definition 1. *The rate-MSE tuple (R, Δ) is said to be achievable if for every $\epsilon > 0$ and $N \geq N_\epsilon$ (possibly large and may depend on ϵ), there exists an $(N, M_N, \varepsilon_N, \xi_N)$-scheme with*

$$\begin{aligned} \varepsilon_N &\leq \epsilon \\ \frac{1}{N}\log M_N &\geq R - \epsilon \\ N\xi_N &\leq \Delta + \epsilon. \end{aligned}$$

The capacity-MSE function $C(\Delta)$ is the maximum rate R for which (R, Δ) is achievable.

Stated differently, the above definition says that if a rate-MSE tuple (R, Δ) is achievable, then there exists an N-block scheme for the above setting that can reliably communicate any one of $M_N = 2^{N(R-\epsilon)}$ messages, with error probability $\varepsilon_N = \epsilon$; while simultaneously achieving a sensing risk of $\xi_N = (\Delta + \epsilon)/N$. Note that ϵ can be made as small as desired by choosing a large-enough block-length N. Our main result in this paper is to fully characterize $C(\Delta)$, and then showcase the trade-off through a couple of canonical examples.

Remark 2. Before we proceed, we justify some of our assumptions in this remark. First, for the fixed target parameter, this is practically motivated by the fact that many parameters of interest

in sensing remain unchanged or change very slightly over a typical transmission block. For instance, during a transmission period of 0.5 ms, a fast-traveling car at 100 km/h moves by just over 1.3 cm, and hence parameters of interest (e.g., range, velocity, angle) remain almost fixed.

Now moving on to the asymptotic regime ($N \to \infty$), we argue that this does not pose a contradiction when considered together with a fixed parameter. For instance, consider a wireless JCAS system operating with $B = 250$ MHz of bandwidth. A transmission block spanning $\tau = 0.5$ ms corresponds to roughly $N = 125000$ channel uses (i.e. $N \approx \tau B$ complex base-band samples). Even if we consider smaller bandwidths, shorter intervals, or lower bandwidth utilization due to pulse shaping, we will still obtain a large enough N that justifies asymptotic analysis.

III. Main Result

A. Preliminaries

To facilitate the presentation of our main result, we start by introducing common information measures, namely the mutual information and the Fisher information. We also introduce a function of the Fisher information matrix, related to the ECRB, which will be used to characterize the asymptotic MSE in the main theorem. In what follows, we use $f_{\mathbf{y}|\mathbf{x}}(\boldsymbol{y}|\boldsymbol{x})$ to denote the density of $\mathcal{CN}(\boldsymbol{H}\boldsymbol{x}, \sigma_{\boldsymbol{\omega}}^2\boldsymbol{I})$, i.e., the transition law of the user channel (1). Similarly, $f_{\mathbf{z}|\mathbf{x},\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{\theta})$ denotes the density of $\mathcal{CN}(\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{x}, \sigma_{\boldsymbol{v}}^2\boldsymbol{I})$, associated with the sensor channel (2).

Now suppose that the random vector $\mathbf{x} \in \mathbb{C}^M$ is used as input in a single use of the JCAS channel under consideration. The mutual information between $\mathbf{x}$ and the user output $\mathbf{y}$ is

$$I(\mathbf{x};\mathbf{y}) \triangleq \mathbb{E}\left[\log\left(\frac{f_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})}{f_{\mathbf{y}}(\mathbf{y})}\right)\right] \tag{11}$$

where $f_{\mathbf{y}}(\boldsymbol{y})$ is the probability density of the output distribution of $\mathbf{y}$ induced by the input $\mathbf{x}$. Note that the quantity inside the expectation in (11) is known as the information density, defined for any given input-output pair of realizations $\boldsymbol{x}$ and $\boldsymbol{y}$ as

$$\imath(\boldsymbol{x};\boldsymbol{y}) \triangleq \log\left(\frac{f_{\mathbf{y}|\mathbf{x}}(\boldsymbol{y}|\boldsymbol{x})}{f_{\mathbf{y}}(\boldsymbol{y})}\right). \tag{12}$$

For $\tilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{Q})$, the mutual information $I(\tilde{\mathbf{x}};\mathbf{y})$ evaluates to

$$I(\tilde{\mathbf{x}};\mathbf{y}) = \log\det\left(\boldsymbol{I} + \frac{1}{\sigma_{\boldsymbol{\omega}}^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathsf{H}}\right). \tag{13}$$

It is well known that (13) is non-decreasing, and concave in $\boldsymbol{Q} \in \mathbb{S}_+^M$, where $\mathbb{S}_+^M$ denotes the cone of $M \times M$ complex, symmetric, positive semi-definite matrices.

Now let us condition on $\mathsf{\theta} = \boldsymbol{\theta}$ and $\mathbf{x} = \boldsymbol{x}$. The Fisher information matrix (FIM) associated with estimating $\boldsymbol{\theta}$ from the random sensor observation $\mathbf{z} \sim f_{\mathbf{z}|\mathbf{x},\mathsf{\theta}}(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{\theta})$ is given by

$$\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{x}) \triangleq \mathbb{E}\left[\left[\nabla_{\boldsymbol{\theta}} \ln f_{\mathbf{z}|\mathbf{x},\mathsf{\theta}}(\mathbf{z}|\boldsymbol{x},\boldsymbol{\theta})\right]\left[\nabla_{\boldsymbol{\theta}} \ln f_{\mathbf{z}|\mathbf{x},\mathsf{\theta}}(\mathbf{z}|\boldsymbol{x},\boldsymbol{\theta})\right]^{\mathsf{T}}\right]. \tag{14}$$

For our particular case, the (i,j)-th element of $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{x})$ evaluates to

$$[\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{x})]_{ij} = \frac{2}{\sigma_{\boldsymbol{v}}^2}\,\mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}}\frac{\partial \boldsymbol{G}(\boldsymbol{\theta})^{\mathsf{H}}}{\partial \theta_i}\frac{\partial \boldsymbol{G}(\boldsymbol{\theta})}{\partial \theta_j}\boldsymbol{x}\right\} \tag{15}$$

$$= \frac{2}{\sigma_{\boldsymbol{v}}^2}\,\mathrm{Re}\left\{\mathrm{tr}\left(\frac{\partial \boldsymbol{G}(\boldsymbol{\theta})}{\partial \theta_j}\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\frac{\partial \boldsymbol{G}(\boldsymbol{\theta})^{\mathsf{H}}}{\partial \theta_i}\right)\right\}. \tag{16}$$

A quantity derived from the FIM which is very relevant to us is the following:

$$\boldsymbol{K}(\mathsf{\theta};\mathbf{z}|\mathbf{x}) \triangleq \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{J}(\mathsf{\theta}|\mathbf{x})|\mathsf{\theta}\right]^{-1}\right]. \tag{17}$$

To parse (17), given $\mathsf{\theta} = \boldsymbol{\theta}$, we take the expectation $\mathbb{E}\left[\boldsymbol{J}(\boldsymbol{\theta}|\mathbf{x})\right]$ with respect to a random input $\mathbf{x}$, which we will refer to as the *conditional FIM*.[3] We then invert the conditional FIM and take its expectation with respect to the random parameter $\mathsf{\theta}$. As shown in Section IV, $\boldsymbol{K}(\mathsf{\theta};\mathbf{z}|\mathbf{x})$ is closely related to the ECRB for estimating $\mathsf{\theta}$ from the sensor observations with knowledge of the transmitted codeword, and exactly characterizes the asymptotic MSE. In fact, we can think of (17) as a form of *conditional ECRB* matrix.

Now suppose that we have an input $\tilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{Q})$. We can see from (16) that the conditional FIM $\mathbb{E}\left[\boldsymbol{J}(\boldsymbol{\theta}|\tilde{\mathbf{x}})\right]$ in this case only depends on the input distribution through $\boldsymbol{Q}$. Therefore, we denote $\mathbb{E}\left[\boldsymbol{J}(\boldsymbol{\theta}|\tilde{\mathbf{x}})\right]$ by $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q})$, in which the (i,j)-th element is given by

$$[\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q})]_{ij} = \frac{2}{\sigma_{\boldsymbol{v}}^2}\,\mathrm{Re}\left\{\mathrm{tr}\left(\frac{\partial \boldsymbol{G}(\boldsymbol{\theta})}{\partial \theta_j}\boldsymbol{Q}\frac{\partial \boldsymbol{G}(\boldsymbol{\theta})^{\mathsf{H}}}{\partial \theta_i}\right)\right\}. \tag{18}$$

In this case, the conditional ECRB matrix in (17) becomes

$$\boldsymbol{K}(\mathsf{\theta};\mathbf{z}|\tilde{\mathbf{x}}) = \mathbb{E}\left[\boldsymbol{J}(\mathsf{\theta}|\boldsymbol{Q})^{-1}\right]. \tag{19}$$

As we will see shortly, the asymptotic MSE is characterized in terms of the function

$$\mathrm{tr}\left(\boldsymbol{K}(\mathsf{\theta};\mathbf{z}|\tilde{\mathbf{x}})\right) = \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\mathsf{\theta}|\boldsymbol{Q})^{-1}\right]\right) \tag{20}$$

which is perhaps not surprising in view of $\boldsymbol{K}(\mathsf{\theta};\mathbf{z}|\tilde{\mathbf{x}})$ being effectively a conditional ECRB matrix. The exact relationship is discussed in more detail in Section IV.

[3]This is similar, in a sense, to the conditional entropy [32], which involves conditioning and then taking an expectation. Note that for fixed $\boldsymbol{\theta}$, the conditional FIM $\mathbb{E}\left[\boldsymbol{J}(\boldsymbol{\theta}|\mathbf{x})\right]$ appears in the *modified CRB* [37].

Lemma 1. $\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\mathsf{\theta}|\boldsymbol{Q})^{-1}\right]\right)$ *is monotonically non-increasing and convex in* $\boldsymbol{Q} \in \mathbb{S}_{+}^{M}$.

The above Lemma will be useful for our analysis. Its proof is presented in Appendix A-A.

B. Result and insights

We are now in position to present the main result.

Theorem 1. *The capacity-MSE function* $C(\Delta)$ *is characterized by*

$$\begin{aligned} \max_{\boldsymbol{Q}\succeq\boldsymbol{0}} \ & \log\det\left(\boldsymbol{I}+\frac{1}{\sigma_{\omega}^{2}}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathsf{H}}\right) \\ \text{s.t.} \ & \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\mathsf{\theta}|\boldsymbol{Q})^{-1}\right]\right) \leq \Delta \\ & \mathrm{tr}\left(\boldsymbol{Q}\right) \leq P. \end{aligned} \tag{21}$$

A detailed proof of Theorem 1 is presented in Section IV. The rest of this section is dedicated to dissecting the above result and deriving some analytical insights.

1) Trade-off: First, we observe that any optimal rate-MSE pair $(R,\Delta)=(C(\Delta),\Delta)$, which must necessarily lie on the capacity-MSE curve $C(\Delta)$, is characterized by

$$R = I(\tilde{\mathbf{x}};\mathbf{y}) \quad \text{and} \quad \Delta = \mathrm{tr}\left(\boldsymbol{K}(\mathsf{\theta};\mathbf{z}|\tilde{\mathbf{x}})\right) \tag{22}$$

for some Gaussian input vector $\tilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{0},\boldsymbol{Q})$ with an optimized choice of input covariance matrix $\boldsymbol{Q}$ that satisfies the power constraint $\mathrm{tr}\left(\boldsymbol{Q}\right) \leq P$. That is, Gaussian input distributions are sufficient for characterizing all optimal rate-MSE trade-off points. In the terminology of [13], the optimal characterization in Theorem 1 only exhibits a so-called ST, but no DRT. This is due to the fact that $\mathsf{\theta}$, albeit randomly chosen, remains fixed throughout the transmission.

2) Almost-constant covariance codes: Despite the fact that the capacity-MSE function $C(\Delta)$ is characterized using Gaussian input distributions, the common random coding approach based on i.i.d. Gaussian codebook ensembles [15], [32] seems to be insufficient for proving the achievability of $C(\Delta)$. This is mainly due to the worst-case nature of the sensing risk in (8). The intuition here is that an i.i.d. Gaussian process will occasionally generate codewords that are very bad for sensing, which will dominate the maximal MSE risk in (8). Alternatively, in our achievability proof, we rely on analyzing codebooks in which all codewords have approximately the same sample covariance matrix. In particular, we impose that $\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n(w)\boldsymbol{x}_n(w)^{\mathsf{H}} \approx \boldsymbol{Q}$ must hold for all $\boldsymbol{x}^N(w) \in \boldsymbol{\mathcal{C}}_N$, in a sense that will be made precise in the proof. This restriction guarantees a uniform MSE risk performance across all codewords (or messages); and at the same

time yields an achievable rate equal to the mutual information $I(\tilde{\mathbf{x}};\mathbf{y})$, which is identical to the one achieved with i.i.d. Gaussian codebook ensembles with covariance $\boldsymbol{Q}$.

Note that in previous works, it is commonly assumed that $\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n(w)\boldsymbol{x}_n(w)^{\mathsf{H}} \approx \boldsymbol{Q}$ holds under "Gaussian signaling" as N grows large [9], [10]. The almost-constant covariance codes framework adopted here can be viewed as a rigorous formalization of this assumption.

3) Analytical properties: (21) constitutes a convex optimization problem, since the objective to be maximized is concave in $\boldsymbol{Q}$, while the constraints are convex and linear. The convex constraint function $\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{Q})^{-1}\right]\right)$, however, involves an expectation with respect to the prior of θ. Hence, solving (21) may require approximating this expectation by its sample average. In the special case where $\boldsymbol{G}(\theta)$ is a linear map, the dependency of $\boldsymbol{J}(\theta|\boldsymbol{Q})$ on the prior washes out, and the problem becomes simpler. This is seen in Section V, where we investigate case studies and present numerical examples.

It is also useful to note that the function $C(\Delta)$, as described in (21), is continuous, non-decreasing, and concave in Δ (see Appendix A-B). Consequently, the rate increase is expected to slow down as the MSE grows, and exhibit a "diminishing returns" behavior due to concavity. This trend is indeed confirmed by the examples in Section V.

4) BCRB-based bound: When proving a converse for $C(\Delta)$, instead of the ECRB-based approach we follow in Section IV-C, the BCRB (i.e., van Trees inequality) [25] can be used to lower bound the MSE. This leads to an upper bound on the capacity-MSE function given by

$$\begin{aligned}\max_{\boldsymbol{Q}\succeq\boldsymbol{0}}\ & \log\det\left(\boldsymbol{I}+\frac{1}{\sigma_{\boldsymbol{\omega}}^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathsf{H}}\right)\\ \text{s.t.}\ & \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{Q})\right]^{-1}\right)\le\Delta\\ & \mathrm{tr}\left(\boldsymbol{Q}\right)\le P\end{aligned} \tag{23}$$

which we denote by $C^{\mathrm{ub}}(\Delta)$ (see Appendix C for details). Note that the prior information term in the BCRB vanishes since we scale by N and take $N\to\infty$. It is readily seen that

$$\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{Q})\right]^{-1}\right)\le\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{Q})^{-1}\right]\right) \tag{24}$$

which holds due to Jensen's inequality and convexity of the matrix inverse. In fact, due to strict convexity, equality in (24) holds only if $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q})$ is invariant under $\boldsymbol{\theta}$. From (24), it is clear that (23) gives an upper bound on $C(\Delta)$, which is not achievable in general. To illustrate this relation, we include the BCRB-based upper bound in our case studies in Section V.

Remark 3. In [13], Xiong *et al.* used the Miller-Chang type BCRB [24] as a proxy for the MSE, where the random input signal $\mathbf{x}^N$ is treated as a so-called "nuisance parameter". Adapting this approach to our setting leads to the same upper bound in (23), see Appendix C.

IV. Proof of Theorem 1

This section is dedicated to proving Theorem 1. We start by presenting essential preliminaries on Bayesian parameter estimation, which will be used further on in our proof.

A. Bayesian MMSE estimation and ECRB

Suppose that a codeword $\boldsymbol{x}^N \in \boldsymbol{\mathcal{C}}^N$ is sent (we suppress the dependency on w in this part) and $\boldsymbol{z}^N$ is observed by the sensor. The optimal estimator that minimizes the MSE risk in (7) is the minimum MSE (MMSE) estimator, which is given by the conditional mean

$$\hat{\boldsymbol{\theta}}^{\star}_N(\boldsymbol{z}^N, \boldsymbol{x}^N) = \mathbb{E}\left[\theta | \mathbf{z}^N = \boldsymbol{z}^N, \mathbf{x}^N = \boldsymbol{x}^N\right]. \tag{25}$$

The corresponding MSE matrix satisfies

$$\boldsymbol{\Sigma}\left(\boldsymbol{x}^N, \hat{\boldsymbol{\theta}}^{\star}_N\right) \preceq \boldsymbol{\Sigma}\left(\boldsymbol{x}^N, \hat{\boldsymbol{\theta}}_N\right) \tag{26}$$

compared to any estimator $\hat{\boldsymbol{\theta}}_N$, where $\boldsymbol{\Sigma}\left(\boldsymbol{x}^N, \hat{\boldsymbol{\theta}}^{\star}_N\right)$ is known to be equal to the conditional covariance matrix $\boldsymbol{K}_{\theta|\mathbf{z}^N, \mathbf{x}^N = \boldsymbol{x}^N}$ [25], [31]. Note that (26) directly implies

$$\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Sigma}\left(\boldsymbol{x}^N, \hat{\boldsymbol{\theta}}^{\star}_N\right)\right) \leq \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Sigma}\left(\boldsymbol{x}^N, \hat{\boldsymbol{\theta}}_N\right)\right) \tag{27}$$

for any non-negative diagonal weighting matrix $\boldsymbol{A}$, and hence optimality is maintained for the WMSE risk in Remark 1. We are interested in the asymptotic MSE performance.

By extending (14), let $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{x}^N)$ be the FIM associated with estimating $\theta = \boldsymbol{\theta}$ from the random sequence of sensor observations $\mathbf{z}^N$ given that $\mathbf{x}^N = \boldsymbol{x}^N$ is transmitted. From the additive property of the FIM under conditionally independent measurements, we have

$$\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{x}^N) = \sum_{i=1}^{N} \boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{x}_i). \tag{28}$$

The classical CRB is obtained by inverting the FIM $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{x}^N)$, while the ECRB is obtained by taking the expectation of the CRB with respect to θ, and is given by

$$\boldsymbol{K}(\theta; \mathbf{z}^N | \mathbf{x}^N = \boldsymbol{x}^N) = \mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{x}^N)^{-1}\right]. \tag{29}$$

Unlike the BCRB [25], the ECRB does not always yield a lower bound (in the semi-definite sense) on the MSE matrix for every value of N. Nevertheless, in the i.i.d. measurement regime recovered by setting $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N = \boldsymbol{x}$ in (28), the ECRB provides an exact characterization of the asymptotic MSE as $N \to \infty$. This was stated by Van Trees *et al.* in [25, Sec. 4.2–4.3] and studied more recently by Aharon and Tabrikian in [34]. In particular, [34] investigates a weighted variant of the BCRB, showing its validity as a lower bound for all N and its asymptotic tightness, with convergence to the ECRB as N grows large (under regularity conditions). Next, we adapt this conclusion to the sensing setting at hand, where measurements are independent but non-identical, arising from the fact that the codeword symbols are not identical in general.

To this end, we define the empirical covariance matrix of an input sequence $\boldsymbol{x}^N$ as

$$\hat{\boldsymbol{Q}}(\boldsymbol{x}^N) \triangleq \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathsf{H}}. \tag{30}$$

From (18), it is clear that the N-letter FIM in (28) is equal to a scaled conditional FIM:

$$\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{x}^N) = N \boldsymbol{J}(\boldsymbol{\theta}|\hat{\boldsymbol{Q}}(\boldsymbol{x}^N)). \tag{31}$$

This relation is key to formulating the conditional ECRB tightness result presented next.

Lemma 2. *Fix $\epsilon > 0$. There exists a (possibly large) block-length N_ϵ such that*

$$\operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\hat{\boldsymbol{Q}}(\boldsymbol{x}^N))^{-1}\right]\right) - \epsilon \leq N \operatorname{tr}\left(\boldsymbol{\Sigma}\left(\boldsymbol{x}^N, \hat{\boldsymbol{\theta}}_N^\star\right)\right) \leq \operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\hat{\boldsymbol{Q}}(\boldsymbol{x}^N))^{-1}\right]\right) + \epsilon \tag{32}$$

holds for every $N \geq N_\epsilon$, provided that some regularity conditions hold.

The above lemma follows by slightly adapting existing results for i.i.d. measurements, as shown by Seo [30] in the scalar parameter case. In particular, the lower bound in (32) can be obtained from the weighted BCRB in [34, Theorem 8], which is shown to approach the ECRB as N grows large (see [34, Prop. 3]). The upper bound in (32) can be obtained using the sub-optimal maximum likelihood (ML) estimator, whose asymptotic MSE is characterized exactly by the ECRB as discussed in [25, Sec. 4.2–4.3] (see also [34, eq. (12)]).

B. Achievability

In order to prove achievability, it suffices to show that for an arbitrary fixed covariance matrix $\boldsymbol{Q} \succeq 0$ that satisfies $\operatorname{tr}(\boldsymbol{Q}) \leq P$, and for every $\epsilon > 0$ and N large enough, there exists an $(N, M_N, \varepsilon_N, \xi_N)$-scheme with $\varepsilon_N \leq \epsilon$ and

$$\frac{1}{N} \log M_N \geq \log\det\left(\boldsymbol{I} + \frac{1}{\sigma_{\boldsymbol{\omega}}^2} \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathsf{H}}\right) - \epsilon \tag{33}$$

$$N\xi_N \leq \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{Q})^{-1}\right]\right) + \epsilon. \tag{34}$$

By optimizing $\boldsymbol{Q}$ and making $\epsilon \to 0$ through $N \to \infty$, we achieve all capacity-MSE trade-off points. We now break down the achievability proof into three parts as follows.

1) Codebook constraint: For fixed N and codebook $\boldsymbol{\mathcal{C}}_N$, the sensing risk is dictated by the *worst* codeword, as clearly seen for (8). To guarantee a somewhat uniform MSE across codewords, we impose the constraint that codewords of $\boldsymbol{\mathcal{C}}_N$ must be drawn from

$$\boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q}) \triangleq \left\{\boldsymbol{x}^N \in \mathbb{C}^N : \boldsymbol{Q} - \delta\boldsymbol{I} \preceq \hat{\boldsymbol{Q}}(\boldsymbol{x}^N) \preceq \boldsymbol{Q} + \delta\boldsymbol{I}\right\} \tag{35}$$

where $\delta > 0$ is sufficiently small. All sequences $\boldsymbol{x}^N \in \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})$ have an empirical covariance matrix $\hat{\boldsymbol{Q}}(\boldsymbol{x}^N)$ that is sufficiently close to $\boldsymbol{Q}$, in the semi-definite sense. Next, we show that (i) by choosing all codewords from $\boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})$, the MSE in (34) is achieved; and (ii) there exists a codebook with codewords from $\boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})$ that achieves the rate in (33).

2) MSE: Suppose that $\boldsymbol{\mathcal{C}}_N \subset \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})$ and consider the codeword $\boldsymbol{x}^N(w) \in \boldsymbol{\mathcal{C}}_N$. Assuming MMSE estimation, and starting from the upper bound in (32), we obtain

$$N\xi_N(w) \leq \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\hat{\boldsymbol{Q}}(\boldsymbol{x}^N(w)))^{-1}\right]\right) + \epsilon' \tag{36}$$

$$\leq \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{Q} - \delta\boldsymbol{I})^{-1}\right]\right) + \epsilon' \tag{37}$$

$$\leq \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{Q})^{-1}\right]\right) + \delta' + \epsilon' \tag{38}$$

where (37) is due to monotonicity while (38) is due to continuity (see Lemma 1). Note that $\delta' + \epsilon'$ can be made small by controlling δ and N. Since (38) holds for all $w \in \mathcal{M}_N$, it holds by taking the maximum, from which we obtain (34).

3) Rate: We now wish to show that there exists a codebook $\boldsymbol{\mathcal{C}}_N \subset \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})$ of size M_N that satisfies (33) and has a sufficiently small maximal decoding error probability.

To this end, we invoke a refined version of the channel coding theorem known in the literature as Feinstein's lemma (see [36] for a comprehensive treatment). The advantage of using Feinstein's lemma compared to the abundant typicality-based approach [32], [38] is that it provides guarantees on the maximal (as opposed to average) decoding error probability, and more importantly, it is better suited for handling multi-letter input distributions (non i.i.d. codebooks) and arbitrary codeword constraint sets. The version we adopt incorporates a constraint set, and is referred to in [36] as the extended Feinstein's lemma. To this end, recall the information density defined in (12), and let $\imath(\boldsymbol{x}^N; \boldsymbol{y}^N)$ be the corresponding N-letter extension. Moreover, in what follows,

let $\tilde{\mathbf{x}}^N = \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_n$ be an i.i.d. sequence of Gaussian vectors, in which every $\tilde{\mathbf{x}}_n$ is drawn independently from $\mathcal{CN}(\mathbf{0}, \boldsymbol{Q})$; and $\mathbf{y}^N$ be a random output sequence induced by the input $\tilde{\mathbf{x}}^N$.

Lemma 3 (Extended Feinstein's lemma [36, Th. 20.7]). *Fix N and $\gamma_N > 0$. There exists a codebook $\boldsymbol{\mathcal{C}}_N \subset \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})$ of size M_N and maximal decoding error probability that satisfies*

$$\varepsilon_N \times \mathbb{P}[\tilde{\mathbf{x}}^N \in \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})] \leq \mathbb{P}\left[\imath(\tilde{\mathbf{x}}^N; \mathbf{y}^N) \leq \log \gamma_N\right] + \frac{M_N}{\gamma_N}. \tag{39}$$

Note that since $\tilde{\mathbf{x}}^N$ is i.i.d. and the channel is memoryless, the information density in our case is additive, and for every realization $(\tilde{\boldsymbol{x}}^N, \boldsymbol{y}^N)$ of $(\tilde{\mathbf{x}}^N, \mathbf{y}^N)$ we have

$$\imath(\tilde{\boldsymbol{x}}^N; \boldsymbol{y}^N) = \sum_{n=1}^{N} \imath(\tilde{\boldsymbol{x}}_n; \boldsymbol{y}_n) = \sum_{n=1}^{N} \log \frac{f_{\mathbf{y}|\mathbf{x}}(\boldsymbol{y}_n|\tilde{\boldsymbol{x}}_n)}{f_{\mathbf{y}}(\boldsymbol{y}_n)}. \tag{40}$$

Moreover, since the mutual information is the expected information density (see (11)), then

$$\mathbb{E}\left[\imath(\tilde{\mathbf{x}}^N; \mathbf{y}^N)\right] = I(\tilde{\mathbf{x}}^N; \mathbf{y}^N) = NI(\tilde{\mathbf{x}}; \mathbf{y}) \tag{41}$$

where $\tilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{Q})$. Going back to Lemma 3, we set

$$\frac{1}{N} \log \gamma_N = I(\tilde{\mathbf{x}}; \mathbf{y}) - \frac{1}{2}\epsilon \quad \text{and} \quad \frac{1}{N} \log M_N = I(\tilde{\mathbf{x}}; \mathbf{y}) - \epsilon.$$

The error probability bound becomes

$$\varepsilon_N \times \mathbb{P}[\tilde{\mathbf{x}}^N \in \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})] \leq \mathbb{P}\left[\frac{1}{N}\sum_{n=1}^{N} \imath(\tilde{\mathbf{x}}_n; \mathbf{y}_n) \leq I(\tilde{\mathbf{x}}; \mathbf{y}) - \frac{\epsilon}{2}\right] + 2^{-N\frac{\epsilon}{2}}. \tag{42}$$

Since the sequences involved are i.i.d., by the weak law of large numbers (WLLN), we get

$$\mathbb{P}\left[\frac{1}{N}\sum_{n=1}^{N} \imath(\tilde{\mathbf{x}}_n; \mathbf{y}_n) \leq I(\tilde{\mathbf{x}}; \mathbf{y}) - \frac{\epsilon}{2}\right] \to 0 \quad \text{as} \quad N \to \infty. \tag{43}$$

Moreover, we have $2^{-N\frac{\epsilon}{2}} \to 0$ as $N \to \infty$. Therefore, the right-hand side of (42) approaches zero as N grows large. The following lemma is key for completing the proof.

Lemma 4. *Let $\tilde{\mathbf{x}}^N$ be an i.i.d. sequence of Gaussian vectors with distribution $\mathcal{CN}(\mathbf{0}, \boldsymbol{Q})$. Then*

$$\mathbb{P}[\tilde{\mathbf{x}}^N \in \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})] \to 1 \quad \textit{as} \quad N \to \infty. \tag{44}$$

In words, the above lemma says that an i.i.d. $\mathcal{CN}(\mathbf{0}, \boldsymbol{Q})$ sequence will, with high probability, have an empirical covariance matrix that is close to $\boldsymbol{Q}$ in the positive semi-definite sense. A proof is presented in Appendix B. Putting everything together, we conclude that $\varepsilon_N \leq \epsilon$ is attainable by making $N \to \infty$ large enough, which concludes the proof of achievability.

C. Converse

Here we show that any achievable pair (R, Δ) must satisfy $R \leq C(\Delta)$, where $C(\Delta)$ is as characterized in Theorem 1. For this purpose, recall from Definition 1 that if (R, Δ) is achievable, then for large enough N, there exists an $(N, M_N, \varepsilon_N, \xi_N)$-scheme with $\varepsilon_N \leq \epsilon$, $\frac{1}{N}\log M_N \geq R - \epsilon$ and $N\xi_N \leq \Delta + \epsilon$. We now proceed to bound the sensing MSE risk as follows

$$\Delta + \epsilon \geq N \max_{w \in \mathcal{M}_N} \xi_N(w) \tag{45}$$

$$\geq \frac{N}{M_N} \sum_{w \in \mathcal{M}_N} \xi_N(w) \tag{46}$$

$$\geq \frac{1}{M_N} \sum_{w \in \mathcal{M}_N} \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta | \hat{\boldsymbol{Q}}(\boldsymbol{x}^N(w)))^{-1} \right] \right) - \epsilon \tag{47}$$

$$\geq \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}\left(\theta \,\middle|\, \frac{1}{M_N} \sum_{w \in \mathcal{M}_N} \hat{\boldsymbol{Q}}(\boldsymbol{x}^N(w)) \right)^{-1} \right] \right) - \epsilon \tag{48}$$

where (46) is since the maximum (over messages) is lower bounded by the average, (47) is from the lower bound in (32), and (48) is due to the convexity of (20) and Jensen's inequality (see Lemma 1). Next, we denote the empirical covariance matrix average over messages by

$$\bar{\boldsymbol{Q}} \triangleq \frac{1}{M_N} \sum_{w \in \mathcal{M}_N} \hat{\boldsymbol{Q}}(\boldsymbol{x}^N(w)) \tag{49}$$

which clearly must also satisfy $\bar{\boldsymbol{Q}} \succeq \mathbf{0}$ and $\mathrm{tr}\left(\bar{\boldsymbol{Q}}\right) \leq P$. The bound in (48) is rewritten as

$$\Delta \geq \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta | \bar{\boldsymbol{Q}})^{-1} \right] \right) - 2\epsilon. \tag{50}$$

We proceed to bound the rate. To this end, we invoke Fano's inequality [32], [38] and obtain

$$R - \epsilon \leq \frac{1}{N} \log M_N \tag{51}$$

$$\leq \frac{1}{N} I(\mathbf{x}^N; \mathbf{y}^N) + \frac{1}{N} + \frac{\varepsilon_N^{\mathrm{av}}}{N} \log M_N \tag{52}$$

where the term $\frac{1}{N} + \frac{\varepsilon_N^{\mathrm{av}}}{N} \log M_N$ vanishes as N grows large, and hence we bound it by ϵ from now on. Moreover, it is worthwhile noting that Fano's inequality yields a bound in terms of $\varepsilon_N^{\mathrm{av}}$. This is, however, valid for our purpose since $\varepsilon_N^{\mathrm{av}} \leq \varepsilon_N$.

The next step is to bound the normalized multi-letter mutual information term $\frac{1}{N} I(\mathbf{x}^N; \mathbf{y}^N)$. To this end, it helps to remember that the random input sequence $\mathbf{x}^N = \boldsymbol{x}^N(\mathrm{w})$ is uniformly distributed on the codebook $\boldsymbol{\mathcal{C}}_N$. Using the fact that the channel is memoryless, we proceed as

$$\frac{1}{N} I(\mathbf{x}^N; \mathbf{y}^N) \leq \frac{1}{N} \sum_{n=1}^{N} I(\mathbf{x}_n; \mathbf{y}_n) \tag{53}$$

$$\leq \frac{1}{N}\sum_{n=1}^{N}\log\det\left(\boldsymbol{I}+\frac{1}{\sigma_{\boldsymbol{\omega}}^2}\boldsymbol{H}\mathbb{E}\left[\mathbf{x}_n\mathbf{x}_n^{\mathsf{H}}\right]\boldsymbol{H}^{\mathsf{H}}\right) \tag{54}$$

$$\leq \log\det\left(\boldsymbol{I}+\frac{1}{\sigma_{\boldsymbol{\omega}}^2}\boldsymbol{H}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[\mathbf{x}_n\mathbf{x}_n^{\mathsf{H}}\right]\boldsymbol{H}^{\mathsf{H}}\right). \tag{55}$$

The inequality in (54) follows from the extremal property of Gaussian inputs under a covariance (or correlation) matrix constraint [38, Sec. 2.2], which leads to $I(\mathbf{x}_n;\mathbf{y}_n)$ being maximized by an input distribution $\mathcal{CN}(\mathbf{0},\mathbb{E}\left[\mathbf{x}_n\mathbf{x}_n^{\mathsf{H}}\right])$. (55) is due to concavity and Jensen's inequality.

Using the fact that $\mathbf{x}^N=\boldsymbol{x}^N(\mathrm{w})$ is uniformly distributed on the codebook, we observe that

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[\mathbf{x}_n\mathbf{x}_n^{\mathsf{H}}\right]=\frac{1}{M_N}\sum_{w\in\mathcal{M}_N}\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n(w)\boldsymbol{x}_n(w)^{\mathsf{H}}=\bar{\boldsymbol{Q}}. \tag{56}$$

Plugging everything back into (52), we obtain

$$R\leq \log\det\left(\boldsymbol{I}+\frac{1}{\sigma_{\boldsymbol{\omega}}^2}\boldsymbol{H}\bar{\boldsymbol{Q}}\boldsymbol{H}^{\mathsf{H}}\right)+2\epsilon \tag{57}$$

$$\leq C(\Delta+2\epsilon)+2\epsilon \tag{58}$$

where the last inequality is obtained by maximizing over $\bar{\boldsymbol{Q}}$ subject to the MSE constraint in (50), and the power constraint. Taking $\epsilon\to 0$ as $N\to\infty$, we obtain the desired result.

Remark 4. It can be seen from our converse proof that we bounded the maximal (over messages) MSE risk ξ_N in terms of its average counterpart ξ_N^{av} in (46); and the maximal decoding error probability ε_N in terms of its average counterpart $\varepsilon_N^{\mathrm{av}}$ in (52). These are, of course, valid steps as they are applied in the converse part only. More importantly, they imply that the characterization obtained in Theorem 1 remains valid if we adopt less stringent average risk and error measures.

V. Case Studies

A. JCAS with DoA Estimation

In our first case study, we consider scenarios where the transmitter is communicating with a single antenna user, i.e., $L=1$; while estimating the DoA of a passive point target in its range. The transmitter and the sensor are equipped with uniform linear arrays (ULAs) of M and T antennas, respectively. Assuming half-wavelength element spacing for both arrays, the steering vectors at the transmitter and at the sensor are given by

$$\boldsymbol{v}_M(\theta)\triangleq\left[1\;e^{\mathrm{i}\pi\sin(\theta)}\;e^{\mathrm{i}2\pi\sin(\theta)}\;\ldots\;e^{\mathrm{i}(M-1)\pi\sin(\theta)}\right]^{\mathsf{T}} \tag{59}$$

$$\boldsymbol{v}_T(\theta) \triangleq \left[1 \; e^{\mathrm{i}\pi \sin(\theta)} \; e^{\mathrm{i}2\pi \sin(\theta)} \; \dots \; e^{\mathrm{i}(T-1)\pi \sin(\theta)}\right]^{\mathsf{T}} \tag{60}$$

respectively. We adopt a line-of-sight, single-ray model for the communication channel as

$$\boldsymbol{H} = \beta \, \boldsymbol{v}_M(\phi)^{\mathsf{H}} \tag{61}$$

where $\phi \in (-\frac{\pi}{2}, \frac{\pi}{2})$ denotes the angle of departure (AoD) from the transmitter array towards the user's direction, and $\beta \in \mathbb{C}$ denotes the channel gain. As stated in Section II, we assume that the user channel matrix is fixed and known to both the transmitter and the user through, e.g., an initial access phase. For simplicity, we set $\beta = 1$. As for the sensing channel, this is given as

$$\boldsymbol{G}(\theta) = \lambda \, \boldsymbol{v}_T(\theta) \, \boldsymbol{v}_M(\theta)^{\mathsf{H}} \tag{62}$$

where θ is the DoA parameter to be estimated, taking values from $(-\frac{\pi}{2}, \frac{\pi}{2})$, and $\lambda \in \mathbb{C}$ is the complex back-scatter channel gain (see, e.g., [9], among others). For simplicity, we assume that the channel gain is known *a priori* and set $\lambda = 1$. Furthermore, the angle of arrival (AoA) and the AoD of the sensing channel coincide in our monostatic setting, as the transmitter and the sensor arrays are co-located and parallel. θ is drawn at random according to a prior $f_\theta(\theta)$, yet remains fixed throughout the transmission. We examine two different prior distributions specified below. The procedure and analysis can be extended to the joint estimation of DoA and gain [39].

Capacity-MSE trade-off: To calculate the conditional FIM in accordance with (18), which reduces to a scalar for single-parameter estimation, the first-order derivative of the sensing channel matrix with respect to the target parameter θ is obtained as

$$\frac{\partial \boldsymbol{G}(\theta)}{\partial \theta} = \mathrm{i}\pi \cos(\theta) \left(\boldsymbol{T}\boldsymbol{G}(\theta) - \boldsymbol{G}(\theta)\boldsymbol{M}\right) \tag{63}$$

where $\boldsymbol{T} \triangleq \mathrm{diag}(0, \dots, T-1)$ and $\boldsymbol{M} \triangleq \mathrm{diag}(0, \dots, M-1)$ represent element positions in the sensor and transmitter arrays, respectively. Defining $\tilde{\boldsymbol{G}}(\theta) \triangleq \boldsymbol{T}\boldsymbol{G}(\theta) - \boldsymbol{G}(\theta)\boldsymbol{M}$, we obtain

$$J(\theta|\boldsymbol{Q}) = \frac{2\pi^2}{\sigma_{\boldsymbol{v}}^2} \cos^2(\theta) \, \mathrm{tr}\left(\tilde{\boldsymbol{G}}(\theta)\boldsymbol{Q}\tilde{\boldsymbol{G}}(\theta)^{\mathsf{H}}\right) \tag{64}$$

where the operator $\mathrm{Re}\{\cdot\}$ is dropped since the trace of a Hermitian matrix is real.

From Theorem 1, the capacity-MSE function $C(\Delta)$ specialized for the present use case is

$$\begin{aligned} \max_{\boldsymbol{Q}\succeq\boldsymbol{0}} \; & \log\left(1 + \frac{1}{\sigma_{\boldsymbol{\omega}}^2}\boldsymbol{v}_M(\phi)^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{v}_M(\phi)\right) \\ \text{s.t.} \; & \frac{\sigma_{\boldsymbol{v}}^2}{2\pi^2}\mathbb{E}\left[\frac{1}{\cos^2(\theta)\mathrm{tr}(\tilde{\boldsymbol{G}}(\theta)\boldsymbol{Q}\tilde{\boldsymbol{G}}(\theta))}\right] \le \Delta \\ & \mathrm{tr}\left(\boldsymbol{Q}\right) \le P. \end{aligned} \tag{65}$$

For Δ sufficiently large, such that the sensing constraint is inactive, the capacity is given by $\log\left(1+MP/\sigma_{\boldsymbol{\omega}}^2\right)$, where M is the array gain. In this case, beamforming along the user's AoD is optimal, i.e. $\boldsymbol{Q}=\frac{P}{M}\boldsymbol{v}_M(\phi)\boldsymbol{v}_M(\phi)^{\mathsf{H}}$. More generally, by optimizing $\boldsymbol{Q}$, which reflects the beamforming characteristics of the transmit array, a clear trade-off between the rate and the MSE is obtained, as demonstrated in our numerical results below.

In addition to $C(\Delta)$, we will also examine the BCRB-based upper bound $C^{\text{ub}}(\Delta)$ given in (23). This is obtained by replacing the sensing constraint in (65) with

$$\frac{\sigma_{\boldsymbol{v}}^2}{2\pi^2}\frac{1}{\mathbb{E}[\cos^2(\theta)\text{tr}(\tilde{\boldsymbol{G}}(\theta)\boldsymbol{Q}\tilde{\boldsymbol{G}}(\theta))]}\leq\Delta \tag{66}$$

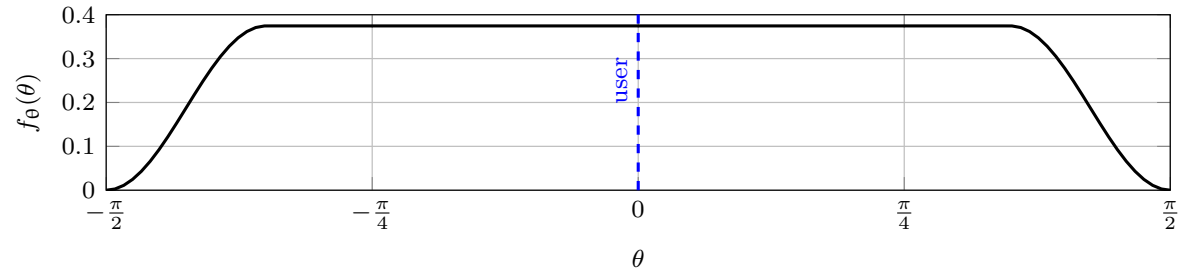
which is a looser constraint, as discussed in Section III-B4. In the numerical results presented below, we set $T=M=16$ for the number of transmitter and sensor antennas. The communication SNR is defined as $P/\sigma_{\boldsymbol{\omega}}^2$ and set to 15 dB. Similarly, the sensing SNR is defined as $P/\sigma_{\boldsymbol{v}}^2$, and set to -25 dB. We examine two types of prior distribution for θ: uniform and non-uniform. The uniform case is meant to represent scenarios where no prior information is available; while the non-uniform case represents scenarios where prior information suggests that the target is more likely to be in a certain angular sector as in, e.g., tracking from a previous estimate.

Tapered uniform prior: Under a uniform prior, one would set $f_\theta(\theta)=1/\pi$ for $\theta\in(-\frac{\pi}{2},\frac{\pi}{2})$ and zero otherwise. However, as highlighted in [34], [40], Bayesian variants of the CRB (e.g., BCRB and ECRB) are not defined for such a prior as it violates the regularity conditions. Therefore, as suggested in [34], we adopt a tapered uniform distribution model with a raised-cosine

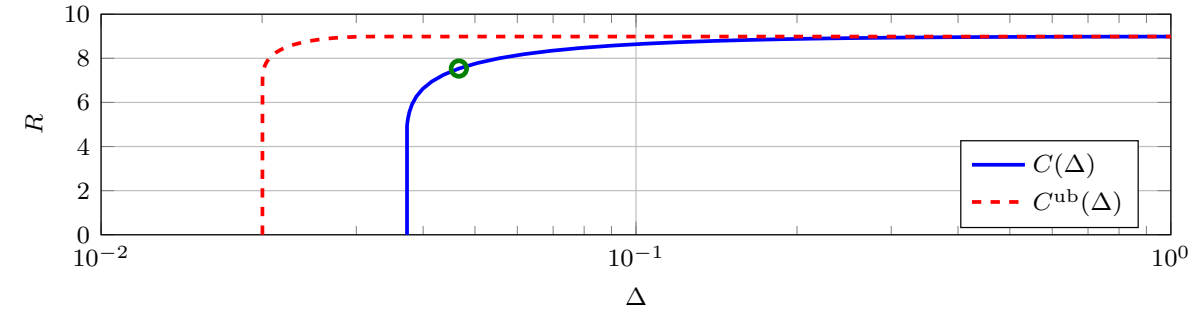
$$f_\theta(\theta)=\frac{1}{s(1+\kappa)}\begin{cases}1, & |\theta|<s\kappa\\ \frac{1}{2}+\left(1+\cos\left(\frac{\pi(|\theta|-s\kappa)}{s(1-\kappa)}\right)\right), & s\kappa\leq|\theta|\leq s\\ 0, & |\theta|>s\end{cases} \tag{67}$$

where $s\in\mathbb{R}^+$ determines the boundaries of the distribution and $\kappa\in[0,1]$ is the roll-off factor that controls the sharpness at the edges. $\kappa=0$ corresponds to a pure raised-cosine shaped distribution, whereas $\kappa=1$ corresponds to a perfect uniform distribution. We set $s=\pi/2$ to cover the entire range and $\kappa=0.7$ to have a tapered uniform distribution. Note that since ULAs have lower resolution in the endfire directions, i.e., $\theta\approx\pm\frac{\pi}{2}$, it is reasonable to bias the estimate away from these directions by setting a low prior value. We set $\phi=0$ for the user channel. The tapered uniform prior and the user location are shown in Fig. 2a.

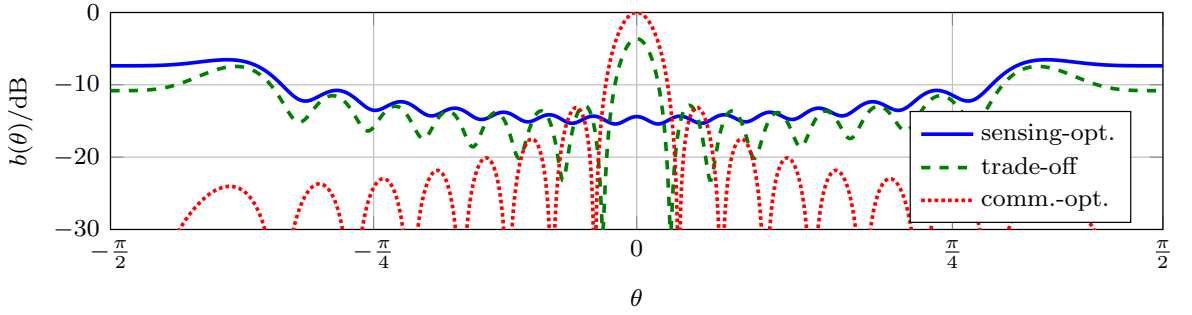
Under the above specified assumptions, the capacity-MSE function and its BCRB-based upper bound are shown in Fig. 2b. By varying Δ in (65), and solving the corresponding optimization



(a) prior distribution



(b) capacity–MSE tradeoff

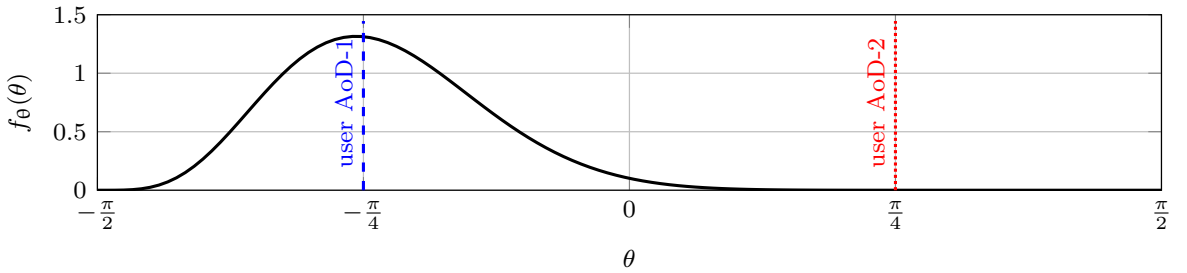


(c) beamformer evolvement

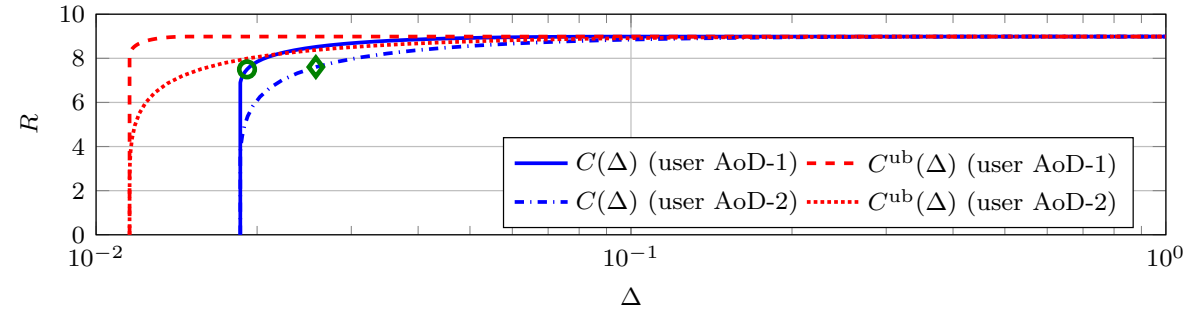
Fig. 2: Capacity–MSE trade-off with a uniform prior in JCAS with DoA estimation.

problem, different rate-MSE trade-off points are obtained. Moreover, the looseness of the BCRB-based upper bound compared to the ECRB-based exact characterization is clearly seen.

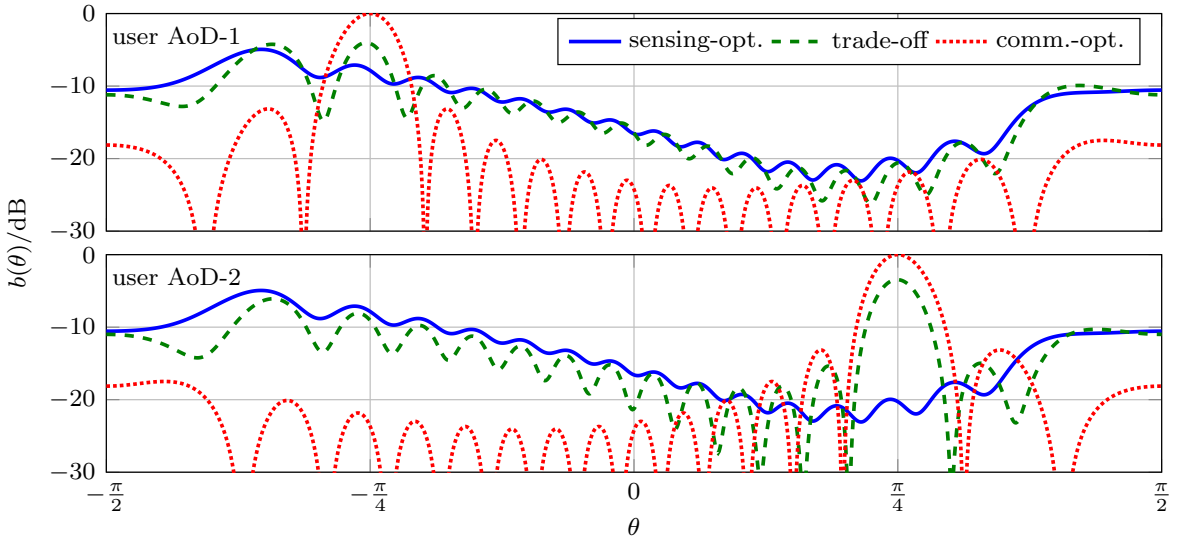
Next, we also utilize the projection $b(\theta) \triangleq \frac{1}{MP}\boldsymbol{v}_M(\theta)^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{v}_M(\theta)$ to calculate the normalized beamforming gain in any given angular direction θ, and show the result in Fig. 2c. To illustrate a specific trade-off choice, we select a point on the $C(\Delta)$ curve indicated with a marker in Fig. 2b, whose optimal beamforming characteristics are also illustrated in Fig. 2c. While the sensing-optimal beamforming favors a wide scan of the entire range, due to the almost uniform prior; beam steering towards the user's direction gradually turns into the optimal beamformer.



(a) prior distribution



(b) capacity-MSE tradeoff



(c) beamformer evolvement

Fig. 3: Capacity-MSE trade-off with a non-uniform prior in JCAS with DoA estimation.

Non-uniform prior: For this case, we utilize the Beta distribution to model a concentrated prior for which $f_\theta(\theta) = 0$ when $\theta = \pm\pi/2$. In particular, we have

$$f_\theta(\theta) = \frac{(\theta - \theta_{\min})^{s_1-1}(\theta_{\max} - \theta)^{s_2-1}}{(\theta_{\max} - \theta_{\min})^{s_1+s_2-1}B(s_1, s_2)} \tag{68}$$

for $\theta \in [\theta_{\min}, \theta_{\max}]$ and zero otherwise, where $B(s_1, s_2) \triangleq \int_0^1 t^{s_1-1}(1-t)^{s_2-1}\,dt$ is the Beta function, and s_1 and s_2 are the shape parameters of the distribution. We set $\theta_{\min} = -\pi/2$ and $\theta_{\max} = \pi/2$ so that the support of the distribution spans the full angular range. The shape parameters are chosen as $s_1 = 5.5$ and $s_2 = 15$, which results in a distribution that is more concentrated around a specific region that peaks around $-\pi/4$. This selection ensures that the prior still satisfies the regularity conditions while favoring a particular region in the angular domain. With the non-uniform prior, we consider two scenarios for the communication channel: first, $\phi = -\pi/4$, i.e., the user is located in the region where the target is most likely present, and second, $\phi = \pi/4$, to represent the scenario where the target is well separated from the user in the angular domain. The prior and the user AoD in the two scenarios are shown in Fig. 3a.

The capacity-MSE function and its BCRB-based upper bound are depicted in Fig. 3b, for two different user AoD scenarios. The optimal beamforming gains for the two scenarios are also shown in Fig. 3c. For both user AoD scenarios, a point is selected on the capacity-MSE curve to illustrate the beamforming characteristics of an intermediate trade-off point.

For user AoD-1, since the prior already favors the user direction, the capacity-MSE function exhibits a limited trade-off. In this case, the user and the target are highly aligned, and therefore, a $\boldsymbol{Q}$ which is good for communication is also good for sensing. This is in contrast to the user AoD-2 scenario, where the target's prior concentrates further away from the user. The evolution of beam patterns in Fig. 3c is also a clear indication of the trade-off in the misaligned case.

B. JCAS with OFDM

We now consider a cyclic-prefix OFDM scenario, where the vector dimension represents sub-carriers instead of antennas. The system communicates with a user and simultaneously senses the back-scatter channel characterized by the frequency response vector $\boldsymbol{\theta} \in \mathbb{C}^K$, where K is the number of sub-carriers being utilized. Assuming that the effect of the cyclic prefix is removed at the receiver, the user channel matrix in this case is given as

$$\boldsymbol{H} = \mathrm{diag}(\boldsymbol{F}\boldsymbol{h}) \tag{69}$$

where $\boldsymbol{F} \in \mathbb{C}^{K\times K}$ is the unitary discrete Fourier transform matrix (DFT) and $\boldsymbol{h} \in \mathbb{C}^K$ is the impulse response of the communication channel in the time-domain. We assume that $\boldsymbol{H}$ is known to the transmitter and receiver. Similarly, the sensor channel matrix in (2) can be written as

$$\boldsymbol{G}(\boldsymbol{\theta}) = \mathrm{diag}(\boldsymbol{F}\boldsymbol{\theta}). \tag{70}$$

Note that in this use case, we have $M = L = T = K$. It is also clear that the sensing channel matrix is linear in the target parameter vector $\mathsf{\theta}$. To allow for concise representation, we allow the parameter vector, i.e. $\mathsf{\theta}$, to be complex-valued. The model can be equivalently expressed using a $2K$-dimensional real parameter vector, consistent with (1) and (2). For more on the real-valued representation of complex parameters in estimation, see [31, Ch. 15].

As we shall see shortly, for this use case, the FIM is invariant under different realizations of the target parameter. Hence, the prior density $f_{\mathsf{\theta}}(\boldsymbol{\theta})$ does not influence the JCAS trade-off.

Capacity-MSE trade-off: Since the user channel matrix is diagonal, the mutual information in (13) is maximized by a diagonal input covariance matrix. Moreover, the conditional FIM and ECRB only depend on the input covariance matrix through its diagonal elements. To see this, we evaluate the conditional FIM, i.e., (18), for this setting. We first calculate the partial derivative of the sensor channel matrix with respect to the i-th entry of the parameter as

$$\frac{\partial \boldsymbol{G}(\theta)}{\partial \theta_i} = \operatorname{diag}(\boldsymbol{f}_i) \tag{71}$$

where $\boldsymbol{f}_i$ denotes the i-th column of the unitary DFT matrix $\boldsymbol{F}$. From this, we obtain

$$[\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q})]_{ij} = \frac{1}{\sigma_{\boldsymbol{v}}^2} \operatorname{tr}\left(\frac{\partial \boldsymbol{G}(\theta)}{\partial \theta_j} \boldsymbol{Q} \frac{\partial \boldsymbol{G}(\theta)}{\partial \theta_i}\right) \tag{72}$$

$$= \frac{1}{\sigma_{\boldsymbol{v}}^2} \operatorname{tr}\left(\operatorname{diag}(\boldsymbol{f}_j) \boldsymbol{Q} \operatorname{diag}(\boldsymbol{f}_i^{\mathsf{H}})\right) \tag{73}$$

$$= \frac{1}{\sigma_{\boldsymbol{v}}^2} \boldsymbol{f}_i^{\mathsf{H}} \left(\boldsymbol{I} \circ \boldsymbol{Q}\right) \boldsymbol{f}_j \tag{74}$$

where $\circ$ denotes the Hadamard (element-wise) product between two matrices. The resulting conditional FIM can be expressed more compactly as

$$\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}) = \frac{1}{\sigma_{\boldsymbol{v}}^2} \boldsymbol{F}^{\mathsf{H}} (\boldsymbol{I} \circ \boldsymbol{Q}) \boldsymbol{F}. \tag{75}$$

Note that this also directly follows from the FIM for linear models [31, Ex. 15.9]. From (75), we see that the conditional FIM does not depend on the off-diagonal terms of $\boldsymbol{Q}$. Hence, restriction to diagonal matrices does not influence the sensing performance either. Therefore, in what follows, we assume without loss of generality that $\boldsymbol{Q} = \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{x}}^2)$, where $\boldsymbol{\sigma}_{\mathbf{x}}^2 \triangleq (\sigma_{\mathbf{x},0}^2, \ldots, \sigma_{\mathbf{x},K-1}^2)$.

Continuing with a diagonal input covariance matrix, (75) becomes

$$\boldsymbol{J}(\mathsf{\theta}|\boldsymbol{\sigma}_{\mathbf{x}}^2) = \frac{1}{\sigma_{\boldsymbol{v}}^2} \boldsymbol{F}^{\mathsf{H}} \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{x}}^2) \boldsymbol{F}. \tag{76}$$

and the asymptotic MSE is given by

$$\operatorname{tr}\left(\boldsymbol{J}(\mathsf{\theta}|\boldsymbol{\sigma}_{\mathbf{x}}^2)^{-1}\right) = \sigma_{\boldsymbol{v}}^2 \sum_{k=0}^{K-1} \frac{1}{\sigma_{\mathbf{x},k}^2}. \tag{77}$$

On the other hand, the rate expression is easily obtained by specializing (13), leading to the well-known sum-rate expression of parallel Gaussian channels [32, Sec. 9.4]. Combining this with Theorem 1, we obtain the following capacity-MSE trade-off

$$\begin{aligned} \max_{\mathrm{diag}(\boldsymbol{\sigma}_{\mathbf{x}}^2)\succeq\mathbf{0}} \quad & \sum_{k=0}^{K-1} \log\left(1+\frac{\sigma_{\mathbf{x},k}^2}{\sigma_{\boldsymbol{\omega}}^2}|[\boldsymbol{H}]_{kk}|^2\right) \\ \text{s.t.} \quad & \sigma_{\boldsymbol{\upsilon}}^2 \sum_{k=0}^{K-1} \frac{1}{\sigma_{\mathbf{x},k}^2} \leq \Delta \\ & \sum_{k=0}^{K-1} \sigma_{\mathbf{x},k}^2 \leq P \end{aligned} \tag{78}$$

which boils down to power allocation across parallel channels under an ECRB constraint.

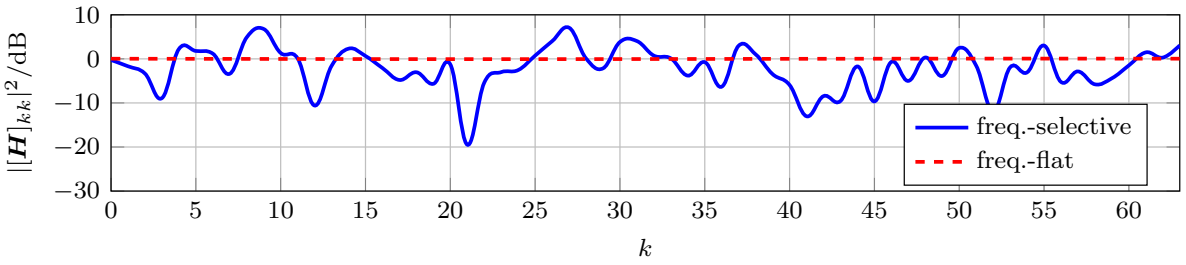
Remark 5. In this special case, where the sensor observations are a linear function of the parameter vector corrupted by Gaussian noise, the conditional FIM in (76) is invariant with respect to the parameter realization $\boldsymbol{\theta}$. Therefore, equality holds in (24) and the capacity-MSE function $C(\Delta)$ coincides with the BCRB-based upper bound $C^{\mathrm{ub}}(\Delta)$. Moreover, $C(\Delta)$ remains the same if one adopts a non-Bayesian framework with the standard CRB instead of the ECRB, as in [10]. In that sense, the characterization in (78) can be seen as a special case of [10, eq. (18)], notwithstanding that the Bayesian framework in the present paper is more general.

The communication-optimal trade-off point—arising when Δ in (78) is sufficiently large so that the sensing constraint becomes inactive—is achieved by a water-filling sub-carrier power allocation [32, Sec. 9.6]. On the other hand, the sensing optimal trade-off point is achieved with an equal power allocation [10, Prop. 2]. This follows from the fact that the MSE in (77) is convex in the power allocation vector. Specifically, by Jensen's inequality, we have

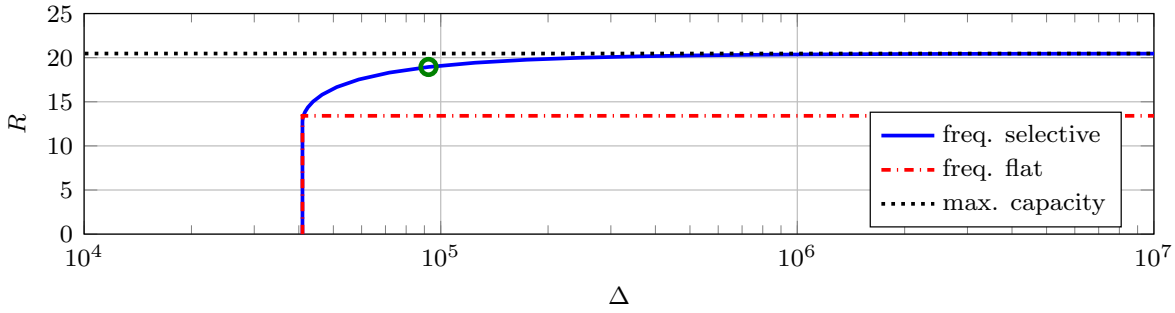
$$\sigma_{\boldsymbol{\upsilon}}^2 \sum_{k=0}^{K-1} \frac{1}{\sigma_{\mathbf{x},k}^2} \geq \sigma_{\boldsymbol{\upsilon}}^2 \frac{K}{\frac{1}{K}\sum_{k=0}^{K-1}\sigma_{\mathbf{x},k}^2} \tag{79}$$

$$= \sigma_{\boldsymbol{\upsilon}}^2 K^2/P \tag{80}$$

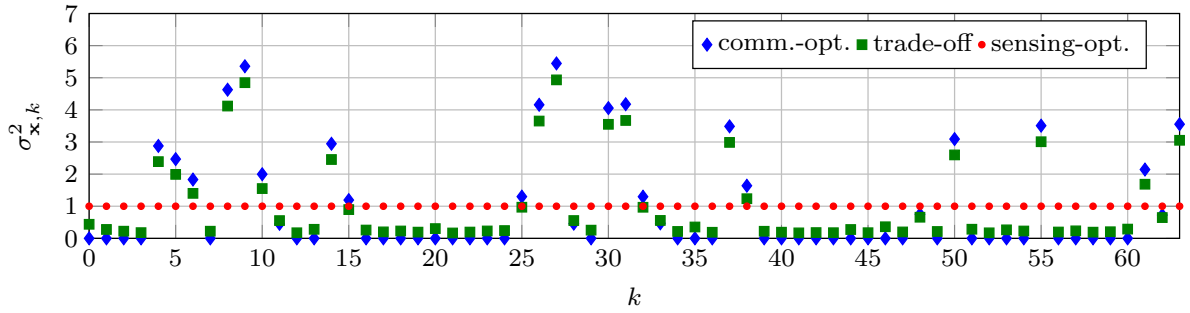
with equality when $\sigma_{\mathbf{x},k}^2 = P/K$ for all k. More generally, the structure of the optimal solution beyond these two extremes has been studied extensively in [10]. It is worth noting that the sensing constraint in (78) requires non-zero power allocation across all sub-carriers in order to yield a finite MSE Δ. This contrasts with the communication-optimal water-filling solution, which may deactivate some sub-carriers. This phenomenon is further explored next.



(a) communication channels



(b) capacity–MSE trade-off



(c) power allocation (we set $P = 64$)

Fig. 4: Capacity–MSE trade-off in JCAS with OFDM.

Frequency-selective vs. frequency-flat communication channels: Here, we numerically evaluate the capacity-MSE function. We set $K = 64$ and study two cases: frequency-selective communication channel and frequency-flat communication channel. The former is meant to illustrate the trade-off between communication and sensing, while the latter will lead to no trade-off. To this end, we set the entries of the time-domain channel impulse response $\boldsymbol{h}$ as

$$h_k = \exp\left(-\frac{1}{2\alpha}k + \mathrm{i}\varphi_k\right), \quad k = 0, 1, \ldots, K-1 \tag{81}$$

where φ_k is a uniformly distributed phase term. The channel model has an exponential power delay profile with delay spread α. For the frequency-selective channel, we set $\alpha = 10$. To get an

almost frequency-flat channel, on the other hand, we chose $\alpha = 0.1$. In both cases, we normalize the channel impulse response to have unit energy per sub-carrier, i.e., $\|\boldsymbol{h}\|^2/K = 1$. We define the communication and sensing SNR as $P/\sigma_{\boldsymbol{\omega}}^2 = 10$ dB and $P/\sigma_{\boldsymbol{v}}^2 = -10$ dB, respectively.

The magnitude responses of the two channels are depicted in Fig. 4a. The resulting capacity-MSE functions are illustrated in Fig. 4b. We can see a clear trade-off between the capacity and the MSE in the frequency-selective channel case. In the frequency-flat channel, on the other hand, there is no trade-off between optimal communication and sensing, as one would expect.

In Fig. 4c, we show the allocation of power among sub-carriers for different points on the capacity-MSE curve. Interestingly, in the frequency-selective scenarios we consider, the communication-optimal water-filling solution allocates zero power to sub-carriers with insufficient SNR. As a result, the communication-optimal point on the Pareto boundary is characterized by $\Delta \to \infty$, as shown in Fig. 4b. For the intermediate trade-off point, indicated by a marker in Fig. 4b, non-zero power is allocated to all sub-carriers to guarantee a finite MSE, while still following the water-filling trend by allocating more power to sub-carriers with higher SNR.

VI. Conclusion

We considered the problem of joint communication and sensing in a basic Gaussian MIMO setting, where the sensing task is to estimate a random parameter vector. The random parameter vector, drawn according to a known prior distribution, is assumed to remain fixed throughout the transmission block. We studied the fundamental trade-off between communication and sensing, formulated in terms of the rate of reliable communication (i.e., channel coding rate) against the MSE of parameter estimation. The optimal rate-MSE trade-off in the asymptotic regime of large block-lengths, characterized in terms of a capacity-MSE function, was established and a number of analytical properties were shown. Finally, case studies of exemplary scenarios illustrated this trade-off: one involving beamforming and DoA estimation, the other involving power allocation over OFDM sub-carriers and tapped-delay channel response estimation.

The problem we considered here can be extended in several directions. In the current paper, we focused our attention on non-adaptive (i.e., open-loop) transmission strategies. It is of interest to investigate the potential gains of adaptive (i.e. closed-loop) transmission strategies, where the next channel input is chosen depending on previous sensor observations. Progress along these lines in the case of discrete parameters (i.e., hypothesis testing) has been reported in [28], and it is of interest to extend some of these insights to problems involving continuous parameter

estimation. Another possible direction would be to extend the setting to multiple users, as well as multiple distinct targets, and study trade-off regions involving multiple rates and MSEs. It is also of interest to refine the results and derive characterizations valid in the finite block-length regime, as has been done for channel coding problems [41]–[43]. As a first step in this direction, one can investigate refined asymptotics, such as second-order coding rates and tighter asymptotic MSE characterizations. Our achievability approach, based on Feinstein's lemma, provides a starting point for enabling such analysis. Finally, relaxing the fixed-parameter assumption—without going as far as the i.i.d. varying regime—is also of interest. Ideally, one would like to investigate models where parameters evolve slowly, as in practical sensing scenarios. Initial attempts to model and study such scenarios in basic settings have been reported in [44]–[46].

APPENDIX A

A. *Proof of Lemma 1*

Here we show that the function $d(\boldsymbol{Q}) \triangleq \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\boldsymbol{Q})^{-1}\right]\right)$ is monotonically non-increasing and convex in $\boldsymbol{Q} \in \mathbb{S}_+^M$. To this end, we first note that for any $\boldsymbol{\theta}$, the conditional FIM $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q})$ is positive semi-definite and linear in $\boldsymbol{Q}$, which can be verified directly from the definition, i.e., (18). Let $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ be two covariance matrices such that $\boldsymbol{Q}_1 \succeq \boldsymbol{Q}_2$. We have

$$\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}_1) = \boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}_1 + \boldsymbol{Q}_2 - \boldsymbol{Q}_2) \tag{82}$$

$$= \boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}_2) + \boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}_1 - \boldsymbol{Q}_2) \tag{83}$$

$$\succeq \boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}_2) \tag{84}$$

where (83) is from linearity, while (84) holds since $\boldsymbol{Q}_1 - \boldsymbol{Q}_2 \succeq \boldsymbol{0}$ and hence $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}_1 - \boldsymbol{Q}_2) \succeq \boldsymbol{0}$. Therefore, we have shown that $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q})$ is monotonically non-decreasing in $\boldsymbol{Q} \in \mathbb{S}_+^M$. Note that (84) directly implies that $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}_2)^{-1} \succeq \boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q}_1)^{-1}$, which in turn implies

$$d(\boldsymbol{Q}_2) \geq d(\boldsymbol{Q}_1). \tag{85}$$

Therefore, we conclude that $d(\boldsymbol{Q})$ is non-increasing. For convexity, this holds since $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q})$ is linear in $\boldsymbol{Q}$, the inverse function $\boldsymbol{J}(\boldsymbol{\theta}|\boldsymbol{Q})^{-1}$ is convex, and then the expectation and trace are linear. Therefore, it can be easily verified that the overall function $d(\boldsymbol{Q})$ is convex.

B. Properties of $C(\Delta)$

We now show that the characterization of $C(\Delta)$ in (21) is continuous, non-decreasing, and concave in Δ. First, it is clear from (21) that the objective is non-decreasing in Δ, since a solution under a smaller MSE constraint Δ is also feasible under a larger constraint. To show concavity, let us consider two MSE constraint values Δ_1 and Δ_2, and let $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ be optimal solutions achieving $C(\Delta_1)$ and $C(\Delta_1)$, respectively, in (21). We have

$$\alpha C(\Delta_1) + (1-\alpha)C(\Delta_1) \leq \log\det\left(\boldsymbol{I} + \frac{1}{\sigma_{\boldsymbol{\omega}}^2}\boldsymbol{H}\big(\alpha\boldsymbol{Q}_1 + (1-\alpha)\boldsymbol{Q}_2\big)\boldsymbol{H}^{\mathsf{H}}\right) \tag{86}$$

which holds due to the concavity of the Gaussian mutual information function. We observe that

$$d\big(\alpha\boldsymbol{Q}_1 + (1-\alpha)\boldsymbol{Q}_2\big) \leq \alpha d\big(\boldsymbol{Q}_1\big) + (1-\alpha)d(\boldsymbol{Q}_2) \tag{87}$$

$$\leq \alpha\Delta_1 + (1-\alpha)\Delta_2 \tag{88}$$

where the first inequality is by convexity of $d\big(\boldsymbol{Q}\big)$, while the second inequality follows from the optimality (and hence feasibility) of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ for their corresponding problems. The inequality in (88) implies that $\alpha\boldsymbol{Q}_1 + (1-\alpha)\boldsymbol{Q}_2$ is a feasible solution for (21) with an MSE constraint $\Delta = \alpha\Delta_1 + (1-\alpha)\Delta_2$. Therefore, (86) is further upper bounded by $C\big(\alpha\Delta_1 + (1-\alpha)\Delta_2\big)$, which proves concavity. Continuity, on interior points of the domain, follows from concavity.

APPENDIX B
PROOF OF LEMMA 4

Starting with the semi-positive definite constraint on the sequence $\tilde{\mathbf{x}}^N \in \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})$, i.e.,

$$\boldsymbol{Q} - \delta\boldsymbol{I} \preceq \hat{\boldsymbol{Q}}(\tilde{\mathbf{x}}^N) \preceq \boldsymbol{Q} + \delta\boldsymbol{I} \tag{89}$$

the following scalar inequality holds by the definition of positive semi-definiteness

$$\boldsymbol{a}^{\mathsf{H}}\left(\boldsymbol{Q} - \delta\boldsymbol{I}\right)\boldsymbol{a} \leq \boldsymbol{a}^{\mathsf{H}}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^{\mathsf{H}}\right)\boldsymbol{a} \leq \boldsymbol{a}^{\mathsf{H}}\left(\boldsymbol{Q} + \delta\boldsymbol{I}\right)\boldsymbol{a}. \tag{90}$$

This is equivalently written as

$$\boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a} - \delta\|\boldsymbol{a}\|^2 \leq \frac{1}{N}\sum_{i=1}^{N}|\boldsymbol{a}^{\mathsf{H}}\tilde{\mathbf{x}}_i|^2 \leq \boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a} + \delta\|\boldsymbol{a}\|^2 \tag{91}$$

for any $\boldsymbol{a} \in \mathbb{C}^M$ and some $\delta > 0$. Since $\boldsymbol{Q} \in \mathbb{S}_+^M$ is Hermitian positive semi-definite, there exists some matrix $\boldsymbol{D} \in \mathbb{C}^{r\times M}$ with $r \triangleq \mathrm{rank}(\boldsymbol{Q}) \leq M$ such that $\boldsymbol{Q} = \boldsymbol{D}^{\mathsf{H}}\boldsymbol{D}$.

Let $\mathbf{u}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I})$ and let $\tilde{\mathbf{x}}_i \triangleq \boldsymbol{D}^{\mathsf{H}}\mathbf{u}_i$. For $r < M$, i.e., $\boldsymbol{Q}$ is rank-deficient, the nullspace of $\boldsymbol{D}$ contains non-zero vectors. When $\boldsymbol{a}$ is in the nullspace of $\boldsymbol{D}$, i.e., $\boldsymbol{D}\boldsymbol{a} = \mathbf{0}$, (91) becomes

$$-\delta\|\boldsymbol{a}\|^2 \leq 0 \leq \delta\|\boldsymbol{a}\|^2 \tag{92}$$

which trivially holds true. When $\boldsymbol{Q}$ is full-rank, $\boldsymbol{D}$ is a full-rank square matrix, and the null space consists of only the zero vector. When a non-zero vector $\boldsymbol{a}$ is not in the nullspace of $\boldsymbol{D}$, i.e., $\boldsymbol{D}\boldsymbol{a} \neq \mathbf{0}$, we can divide (91) by $\boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a}$, which yields

$$1 - \frac{\delta\|\boldsymbol{a}\|^2}{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a}} \leq \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\boldsymbol{a}^{\mathsf{H}}\tilde{\mathbf{x}}_i}{\sqrt{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a}}}\right|^2 \leq 1 + \frac{\delta\|\boldsymbol{a}\|^2}{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a}}. \tag{93}$$

Since (91) must be satisfied for any $\boldsymbol{a} \in \mathbb{C}^M$, a sufficient condition is obtained using the bound

$$\boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a} \leq \lambda_{\max}(\boldsymbol{Q})\|\boldsymbol{a}\|^2 \tag{94}$$

where $\lambda_{\max}(\boldsymbol{Q})$ is the largest eigenvalue of $\boldsymbol{Q}$. Then, we get

$$1 - \frac{\delta}{\lambda_{\max}(\boldsymbol{Q})} \leq \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\boldsymbol{a}^{\mathsf{H}}\tilde{\mathbf{x}}_i}{\sqrt{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a}}}\right|^2 \leq 1 + \frac{\delta}{\lambda_{\max}(\boldsymbol{Q})}. \tag{95}$$

Defining $\alpha_i \triangleq (\boldsymbol{a}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{a})^{-\frac{1}{2}}\boldsymbol{a}^{\mathsf{H}}\tilde{\mathbf{x}}_i$, we observe that $\alpha_i \sim \mathcal{CN}(0,1)$. Setting $\delta' \triangleq \delta/\lambda_{\max}(\boldsymbol{Q}) > 0$, we can rewrite (95) as

$$\left|\frac{1}{N}\sum_{i=1}^{N}|\alpha_i|^2 - 1\right| \leq \delta' \tag{96}$$

where $|\alpha_i|^2 \sim \mathrm{Exp}(1)$ has an exponential distribution. Let $\mathrm{S}_N \triangleq \frac{1}{N}\sum_{i=1}^{N}|\alpha_i|^2$ denote the empirical mean of N i.i.d. exponential random variables. The mean and the variance of S_N are given by $\mathbb{E}[\mathrm{S}_N] = 1$ and $\mathbb{E}[(\mathrm{S}_N - 1)^2] = 1/N$, respectively. Finally, for any $\delta > 0$

$$\mathbb{P}[\tilde{\mathbf{x}}^N \notin \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})] \leq \mathbb{P}[|\mathrm{S}_N - 1| \geq \delta'] \tag{97}$$

$$\leq \frac{1}{\delta'^2 N} \tag{98}$$

using the sufficient condition in (95) as an upper bound in (97) and Chebyshev's inequality in (98). As $N \to \infty$, we get $\mathbb{P}[\tilde{\mathbf{x}}^N \in \boldsymbol{\mathcal{B}}_{N,\delta}(\boldsymbol{Q})] \to 1$, which completes the proof.

APPENDIX C
BCRB-BASED CAPACITY-MSE BOUND

In the appendix, we show how to obtain the capacity-MSE upper bound in (23) using the BCRB lower bound [25]. This derivation in this appendix builds on the converse proof in Section IV-C.

To define the BCRB, we first introduce the Bayesian information matrix associated with estimating θ from $\mathbf{z}^N$, given that $\mathbf{x}^N = \boldsymbol{x}^N$ is transmitted. This is written as

$$\boldsymbol{J}_{\mathrm{B}}(\boldsymbol{x}^N) \triangleq \mathbb{E}[\boldsymbol{J}(\theta|\boldsymbol{x}^N)] + \boldsymbol{J}_{\mathrm{P}} \tag{99}$$

where $\boldsymbol{J}_{\mathrm{P}} \triangleq \mathbb{E}\left[[\nabla_{\boldsymbol{\theta}} \ln f_\theta(\theta)]\,[\nabla_{\boldsymbol{\theta}} \ln f_\theta(\theta)]^{\mathsf{T}}\right]$ is the prior information matrix. The BCRB inequality of van Trees [25, 4.3] states the the MSE is lower bounded as

$$N \operatorname{tr}\left(\boldsymbol{\Sigma}\left(\boldsymbol{x}^N, \hat{\boldsymbol{\theta}}^\star_N\right)\right) \geq \operatorname{tr}\left(\left[\frac{1}{N}\boldsymbol{J}_{\mathrm{B}}(\boldsymbol{x}^N)\right]^{-1}\right) \tag{100}$$

$$= \operatorname{tr}\left(\left[\mathbb{E}\left[\boldsymbol{J}\big(\theta|\hat{\boldsymbol{Q}}(\boldsymbol{x}^N)\big)\right] + \frac{1}{N}\boldsymbol{J}_{\mathrm{P}}\right]^{-1}\right). \tag{101}$$

Unlike the lower bound in (32), this bound holds not only asymptotically but for every N. Adopting this in the converse proof in Section IV-C, and continuing from (46), we have

$$\Delta + \epsilon \geq \frac{1}{M_N}\sum_{w\in\mathcal{M}_N} \operatorname{tr}\left(\left[\mathbb{E}\left[\boldsymbol{J}\big(\theta|\hat{\boldsymbol{Q}}(\boldsymbol{x}^N(w))\big)\right] + \frac{1}{N}\boldsymbol{J}_{\mathrm{P}}\right]^{-1}\right) \tag{102}$$

$$\geq \operatorname{tr}\left(\left[\mathbb{E}\left[\boldsymbol{J}(\theta|\bar{\boldsymbol{Q}})\right] + \frac{1}{N}\boldsymbol{J}_{\mathrm{P}}\right]^{-1}\right) \tag{103}$$

$$\geq \operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{J}(\theta|\bar{\boldsymbol{Q}})\right]^{-1}\right) - \epsilon \tag{104}$$

where (103) holds by convexity and Jensen's inequality, and (104) holds for sufficiently large N by continuity and the fact that $\boldsymbol{J}_{\mathrm{P}}$ is constant, and hence $\frac{1}{N}\boldsymbol{J}_{\mathrm{P}}$ vanishes.

It is also possible to adopt the Miller–Chang version of the BCRB [24], as done in [13]. In this formulation, the random codeword $\mathbf{x}^N$ is treated as a nuisance parameter, and the average BCRB with respect to $\mathbf{x}^N$ is considered. Adapting this to our setting amounts to replacing the maximal MSE risk (over codewords) ξ_N with its average counterpart ξ_N^{av} (see Section II-B3). Nevertheless, since $\mathbf{x}^N$ is uniformly distributed on the codebook as the message is uniform, the previous result still applies. In particular, (102) onward remain valid in this case.

References

[1] 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.

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

[3] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4264–4279, 2018.

[4] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, "Joint transmit beamforming for multiuser MIMO communications and MIMO radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 3929–3944, 2020.

[5] L. Chen, F. Liu, W. Wang, and C. Masouros, "Joint radar-communication transmission: A generalized Pareto optimization framework," *IEEE Trans. Signal Process.*, vol. 69, pp. 2752–2765, 2021.

[6] T. Wild, V. Braun, and H. Viswanathan, "Joint design of communication and sensing for beyond 5G and 6G systems," *IEEE Access*, vol. 9, pp. 30 845–30 857, 2021.

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

[8] 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.

[9] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér-Rao bound optimization for joint radar-communication beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 240–253, 2022.

[10] H. Hua, T. X. Han, and J. Xu, "MIMO integrated sensing and communication: CRB-rate tradeoff," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 2839–2854, 2024.

[11] Z. Ren, Y. Peng, X. Song, Y. Fang, L. Qiu, L. Liu, D. W. K. Ng, and J. Xu, "Fundamental CRB-rate tradeoff in multi-antenna ISAC systems with information multicasting and multi-target sensing," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3870–3885, 2024.

[12] C. Xu and S. Zhang, "MIMO integrated sensing and communication exploiting prior information," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 9, pp. 2306–2321, 2024.

[13] 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 Trans. Inf. Theory*, vol. 69, no. 9, pp. 5723–5751, 2023.

[14] H. Joudeh, "Joint communication and target detection with multiple antennas," in *26th Int. ITG Workshop Smart Antennas and 13th Conf. on Syst., Commun. and Coding (WSA & SCC)*, 2023, pp. 1–6.

[15] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, 1999.

[16] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless personal commun.*, vol. 6, no. 3, pp. 311–335, 1998.

[17] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, 2003.

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

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

[20] M. Ahmadipour and M. Wigger, "An information-theoretic approach to collaborative integrated sensing and communication for two-transmitter systems," *IEEE J. Sel. Areas Inf. Theory*, vol. 4, pp. 112–127, 2023.

[21] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, "An information-theoretic approach to joint sensing and communication," *IEEE Trans. Inf. Theory*, vol. 70, no. 2, pp. 1124–1146, 2024.

[22] O. Günlü, M. R. Bloch, R. F. Schaefer, and A. Yener, "Secure integrated sensing and communication," *IEEE J. Sel. Areas Inf. Theory*, vol. 4, pp. 40–53, 2023.

[23] H. Joudeh and G. Caire, "Joint communication and state sensing under logarithmic loss," in *Proc. 4th IEEE Int. Symp. Joint Commun. Sensing (JC&S)*, 2024, pp. 1–6.

[24] R. Miller and C. Chang, "A modified Cramér-Rao bound and its applications (corresp.)," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 398–400, 1978.

[25] H. L. Van Trees, K. L. Bell, and Z. Tian, *Detection, Estimation, and Modulation Theory Part I*, 2nd ed. Hoboken, NJ, USA: John Wiley & Sons, 2013.

[26] H. Joudeh and F. M. J. Willems, "Joint communication and binary state detection," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 1, pp. 113–124, 2022.

[27] H. Wu and H. Joudeh, "Joint communication and channel discrimination," *Entropy*, vol. 26, no. 12, 2024.

[28] 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 J. Sel. Areas Inf. Theory*, vol. 4, pp. 245–259, 2023.

[29] D. Seo and S. H. Lim, "Integrated communication and binary state detection under unequal error constraints," *IEEE Trans. Commun.*, vol. 73, no. 8, pp. 5729–5744, 2025.

[30] D. Seo, "Integrated communication and Bayesian estimation of fixed channel states," *IEEE Commun. Lett.*, vol. 29, no. 3, 2025.

[31] S. M. Kay, "Fundamentals of statistical signal processing: Estimation theory," 1993.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[33] Y. Wu and S. Verdu, "MMSE dimension," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4857–4879, 2011.

[34] O. Aharon and J. Tabrikian, "Asymptotically tight Bayesian Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 72, pp. 3333–3346, 2024.

[35] H. Wu and H. Joudeh, "On joint communication and channel discrimination," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2022, pp. 3321–3326.

[36] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning*. Cambridge University Press, 2025.

[37] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchronization problems," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1391–1399, 2002.

[38] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[39] J. Li, L. Xu, P. Stoica, K. W. Forsythe, and D. W. Bliss, "Range compression and waveform optimization for MIMO radar: A Cramér–Rao bound based study," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 218–232, 2008.

[40] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing*. New York, USA: John Wiley & Sons, 2002.

[41] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[42] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov 2009.

[43] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, 2014.

[44] S. Li and G. Caire, "On the capacity and state estimation error of "beam-pointing" channels: The binary case," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5752–5770, 2023.

[45] H. Nikbakht, M. Wigger, S. S. Shitz, and H. V. Poor, "A memory-based reinforcement learning approach to integrated sensing and communication," in *Proc. 58th Asilomar Conf. Sig. Syst. Comp.*, 2024, pp. 433–437.

[46] C. P. Lindstrom and M. R. Bloch, "Rate distortion approach to joint communication and sensing with Markov states: Open loop case," *arXiv:2501.15652*, 2025.