Continuous-Time Distributed Filtering via a Gaussian Feedback Channel

Zhenyu Liu, *Member, IEEE*, Andrea Conti, *Fellow, IEEE*, Sanjoy K. Mitter, *Life Fellow, IEEE*, and Moe Z. Win, *Fellow, IEEE*

***Abstract*—Filtering refers to the methods for inferring time-varying parameters and is a crucial task in cyber-physical systems. An important category of filtering is distributed filtering, where sensor nodes transmit observations via communication links to inference nodes that estimate the unknown states. Distributed filtering is challenging in the sense that the communication constraint of the sensor nodes limits the amount of information available to the inference node, calling for the co-design of communication and computing. This paper establishes a theoretical framework for the co-design of communication and computing in distributed filtering, building on an information-theoretic view of the Kalman-Bucy filtering. In particular, this paper considers a networked system consisting of two nodes, where each node aims to infer its own time-varying state in continuous-time scenarios. The two nodes are connected by a Gaussian feedback channel. Via the feedback link, one of the nodes can obtain the sensor observations and received signals of the other node. This paper develops an optimal linear strategy, namely the information difference encoding strategy, for generating signals transmitted via the Gaussian feedback channel. This paper also presents an inequality that relates Shannon information with Fisher information in distributed filtering. The inference accuracy and power efficiency of the information difference encoding strategy are quantified via simulations.**

***Index Terms*—Distributed inference, Kalman-Bucy filter, mutual information, Fisher information, Shannon information.**

I. Introduction

INFERENCE of environmental states is a critical task for cyber-physical systems (CPSs) [1], [2], [3], [4], [5], [6], [7], [8]. For example, estimating and tracking the positions of machines and workers is important for industrial CPSs [9], [10], [11]. Distributed filtering is an important type of inference where an inference node estimates a time-varying unknown state based on communication data that the node receives from a remote sensor node.

Accurate distributed filtering is challenging as the amount of data that the sensor node can transmit to the inference node is limited due to communication constraints. The maximization of inference accuracy requires the co-design of communication strategy for generating the transmitted data and the computing algorithm for estimating the unknown state.

Filtering and inference have been investigated in the literature from different perspectives. From the perspective of theoretical foundation, the calculation of mutual information in filtering problems is investigated in [12] and [13]. Relationships between mutual information, Fisher information, and minimum mean-square error (MSE) for filtering and estimation problems are derived in [14], [15], and [16]. Connections between filtering in dynamical systems with statistical mechanical systems are established in [15], [16], and [17]. The problem of distributed filtering over lossy channels is studied in [18] and [19]. For this type of problem, it is shown that the notion of anytime capacity [20] plays an important role [21], [22], [23]. In addition, the accuracy of spatiotemporal signal reconstruction is derived in [24]. From the perspective of algorithm design, consensus methods for distributed filtering are presented in [25], [26], and [27], whereas diffusion methods are presented in [28], [29], and [30]. Distributed estimation and filtering algorithms are also developed for industrial CPS applications such as target localization and tracking [31], [32], [33].

The design of encoding strategies for distributed filtering has been investigated for channels with feedback in [34], [35], and [36], and without feedback in [37], [38], and [39]. Such design is closely related to the problem of control under communication constraints [40]. One line of research considers conditions of the channel under which a networked control system is stabilizable [41], [42], [43]. An important result is that the system is stabilizable if the data rate of the channel exceeds the intrinsic entropy rate of the system [43], [44], [45]. Another line of research considers minimization of mutual information and of directed mutual information under constraints of linear-quadratic-Gaussian control costs [46], [47], [48]. While most existing works address discrete-time scenarios, only a few papers study control under communication constraints in continuous-time scenarios [49], [50]. Existing works typically assume that the receiver node does not observe the state of the system and relies on the received signal for inferring the state. Moreover, previous study on the connection between mutual information and Fisher information is rather

Received 30 September 2024; revised 17 January 2025; accepted 28 February 2025. Date of publication 6 October 2025; date of current version 14 October 2025. The fundamental research described in this paper was supported, in part, by the European Union through the NextGenerationEU Italian NRRP under Grant PE00000001 Program RESTART, the National Science Foundation under Grant CNS-2148251, by funds from federal agency and industry partners in the RINGS program, and by Robert R. Taylor Professorship. An earlier version of this paper was presented in part at the American Control Conference, Atlanta, GA, USA, in June 2022 [DOI: 10.23919/ACC53348.2022.9867546]. *(Corresponding author: Moe Z. Win.)*

Zhenyu Liu was with the Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China.

Andrea Conti is with the Department of Engineering and CNIT, University of Ferrara, 44122 Ferrara, Italy.

Sanjoy K. Mitter and Moe Z. Win are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: moewin@mit.edu).

Digital Object Identifier 10.1109/JSAC.2025.3574610

TABLE I
NOTATION AND DEFINITIONS OF QUANTITIES

Notation	Definition	Notation	Definition
$\mathsf{x}_t^{(i)}$	state of node i at time t	$a_t^{(i)}$	scalar that determines the evolution of $\mathsf{x}_t^{(i)}$
$\mathbf{v}_t^{(i)}$	random vector corresponding to noise in node i's state process	$\mathbf{z}_t^{(i)}$	observation obtained by the sensor of node i at time t
$\mathring{\boldsymbol{\gamma}}_t^{(1)}$	sensor gain vector for observations obtained by node 1	$\boldsymbol{\Gamma}_t^{(2)}$	sensor gain matrix for observations obtained by node 2
$\mathbf{n}_t^{(i)}$	random vector corresponding to noise in the observations obtained by node i	s_t	signal transmitted from node 2 to node 1 at time t
r_t	signal received by node 1 from node 2 at time t	μ_t	encoding function employed for generating s_t
$\mu_{0:T}$	encoding strategy of horizon T consisting of encoding functions μ_t for $t \in [0,T]$	$\mathring{P}_t$	constraint on the transmit power at time t
w_t	random variable corresponding to noise in the communication channel	κ_t	scalar determining the power of noise in the communication channel
$\hat{\mathsf{x}}_t^{(1)}$	distributed filter of $\mathsf{x}_t^{(1)}$ computed by node 1 at time t	$e_t(\mu_{0:t})$	mean-square error of $\hat{\mathsf{x}}_t^{(1)}$ if encoding strategy $\mu_{0:t}$ is employed

limited for distributed filtering. Fundamental questions on distributed filtering include (i) how to conduct co-design of the encoding strategy and the filtering algorithm to maximize the filtering accuracy; and (ii) what is the relationship between information-related quantities such as Shannon and Fisher information? Answers to these questions will deepen the understanding of distributed filtering and provide methods for achieving desirable inference performance under communication constraints.

This paper investigates continuous-time distributed filtering for a two-node system, where each node attempts to infer an evolving unknown state in real-time. In this system, both nodes obtain noisy sensor observations of the unknown states via sensing and communicate via a Gaussian channel with feedback. Using the feedback link, one node can obtain the sensor observations as well as the received signals of the other node. The goals of this paper include performing co-design of the encoding strategy and distributed filtering algorithm as well as establishing connections between information-related quantities in distributed filtering. The key contributions of this paper include the following, we

- propose the information difference encoding (IDE) strategy and prove that it is an optimal linear strategy;
- establish a relationship between Shannon information and Fisher information for distributed filtering; and
- compare the performance of IDE with the performance of existing encoding strategies.

The remaining sections are organized as follows. Section II presents the system model. Section III introduces the co-design of IDE and distributed filtering. Section IV presents extensions for average power constraints and for multivariate unknown state scenarios. Section V presents an information-theoretical interpretation of the distributed filtering problem. Section VI shows case studies. Section VII concludes the paper.

Notation: Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. The m-by-n matrix of zeros is denoted by $\mathbf{0}_{m\times n}$; when $n=1$, the m-dimensional vector of zeros is simply denoted by $\mathbf{0}_m$. The subscript is removed if the dimension of the matrix is clear from the context. The entry on the ith row and jth column of a matrix $\boldsymbol{A}$ is denoted by $[\boldsymbol{A}]_{i,j}$. The transpose, trace, determinant, and the column space of $\boldsymbol{A}$ are denoted by $\boldsymbol{A}^{\mathrm{T}}$, $\mathrm{tr}\{\boldsymbol{A}\}$, $\det(\boldsymbol{A})$, and $\mathcal{C}(\boldsymbol{A})$, respectively. Expression $\mathrm{diag}\{\boldsymbol{A}_1, \boldsymbol{A}_2\}$ represents the block diagonal matrix $\begin{bmatrix} \boldsymbol{A}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_2 \end{bmatrix}$. The gradient of function $g(\boldsymbol{x})$ is denoted by a column vector $\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}$. The σ-algebra generated by a random quantity $\mathbf{x}$ is denoted by $\sigma(\mathbf{x})$. The expectation of a random vector $\mathbf{x}$ is denoted by $\mathbb{E}\{\mathbf{x}\}$. The conditional expectation of $\mathbf{x}$ given the sub-σ-algebra generated by random quantity $\mathbf{y}$ is denoted by $\mathbf{y}$. The covariance matrix of $\mathbf{x}$ and the conditional covariance matrix of $\mathbf{x}$ given $\mathbf{y}$ are denoted by $\mathbb{V}\{\mathbf{x}\}$ and $\mathbb{V}\{\mathbf{x}\,|\,\mathbf{y}\}$, respectively. Independence between $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\mathbf{x} \perp\!\!\!\perp \mathbf{y}$. Conditional independence between $\mathbf{x}$ and $\mathbf{y}$ given a random quantity $\mathbf{z}$ is denoted by $\mathbf{x} \perp\!\!\!\perp \mathbf{y} | \mathbf{z}$. Given a stochastic process $\{\mathbf{x}_t\}_{t\geqslant 0}$, the set $\{\mathbf{x}_\tau\}_{\tau\in[s,t]}$ is denoted by $\mathbf{x}_{s:t}$ for any $0 \leqslant s \leqslant t$. Notation and definitions of quantities used in the paper are summarized in Table I.

II. SYSTEM MODEL

Consider a networked system involving two nodes, node 1 and node 2, where each node aims to infer a distinct state (e.g., position or temperature) that varies with time (see Fig. 1). At each time, node 1 obtains a noisy sensor observation of its own unknown state, whereas node 2 obtains a noisy sensor observation of the unknown states of both nodes. Moreover, node 2 transmits a signal to node 1 with the aim of facilitating the inference of node 1 via a noisy channel at each time. Node 1 infers its state of current time based on all the observations and received signals it has obtained up to that time. In addition, node 1 can send data back to node 2 via noiseless feedback. With such feedback, the sensor observations and signals obtained by node 1 are available to node 2. These data can be employed by node 2 for generating signals at future times as indicated by the dashed arrows entering the encoder in Fig. 1a. This paper considers noisy communication from node 2 to node 1 and noiseless feedback from node 1 to node 2. This corresponds to scenarios where node 1 has a

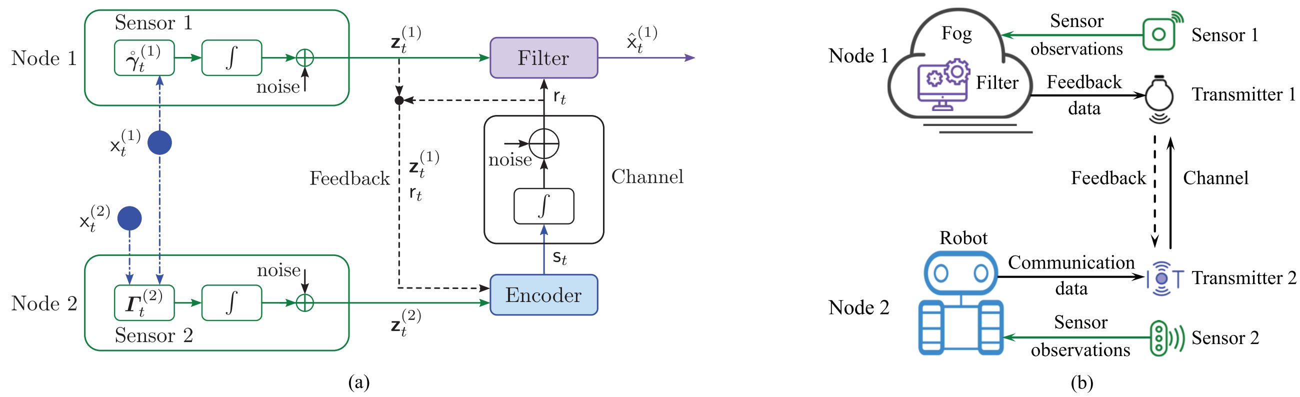


Fig. 1. System model and application in industrial CPS of distributed filtering: (a) System model. Node $i = 1, 2$ is associated with a state $\mathsf{x}_t^{(i)}$ and obtains a noisy sensor observation $\mathbf{z}_t^{(i)}$ at time t. Node 2 transmits encoded signal s_t to node 1, and node 1 determines a distributed filter $\hat{\mathsf{x}}_t^{(1)}$ based on its observations and the received signals. Using the feedback link, node 2 can obtain the sensor observations as well as the received signals of node 1. These data are combined with node 2's observations for generating the transmitted signal. (b) Application in industrial CPS. Sensor 1 is connected to the fog where a filter infers the unknown state of node 1 based on sensor 1's observations and received communication data. Node 2 is a mobile robot that transmits data to node 1 at a low data rate (e.g., using narrowband IoT communication) and receives feedback data from node 1 at a high data rate (e.g., via visible light communication).

more reliable communication link than node 2 or the noise at the receiver of node 2 is less strong than that at node 1.

An application in industrial CPS that suits this model is shown in Fig. 1b. In particular, node 1 contains a sensor and wireless transceivers connected to the fog. A filter at the fog infers the unknown state of node 1 at each time based on observations obtained by sensor 1 and signals received from node 2 via the channel. Node 2 is a mobile robot equipped with a sensor and wireless transceivers. The robot transmits data via the channel to node 1 at each time based on its sensor observations and the data received via the feedback link. The transmission capability of node 1 is significantly stronger than node 2. This is the case, for example, when node 1 employs visible light communication whereas node 2 employs narrowband IoT communication.

Details of the system model are presented below. Let $\mathsf{x}_t^{(i)}$ represent the state of node i at time $t \in [0, \infty)$ for $i = 1, 2$. The state process $\{\mathsf{x}_t^{(i)}\}_{t \geqslant 0}$ is a Gaussian process described by the following stochastic differential equation (SDE) [51], [52], [53]

$$d\mathsf{x}_t^{(i)} = a_t^{(i)} \mathsf{x}_t^{(i)} dt + \big(\boldsymbol{b}_t^{(i)}\big)^{\mathrm{T}} d\mathbf{v}_t^{(i)} \quad \forall t \in [0, \infty) \tag{1}$$

where $a_t^{(i)}$ is a scalar and $\boldsymbol{b}_t^{(i)}$ is a non-zero column vector. In particular, $\boldsymbol{b}_t^{(i)}$ determines the level of noise in node i's state process. The process $\{\mathbf{v}_t^{(i)}\}_{t \geqslant 0}$ is a Brownian motion. The state $\mathsf{x}_0^{(i)}$ at time 0 is unknown and modeled as a zero-mean Gaussian random variable. Processes $\{\mathbf{v}_t^{(1)}\}_{t \geqslant 0}$ and $\{\mathbf{v}_t^{(2)}\}_{t \geqslant 0}$ are independent, whereas $\mathsf{x}_0^{(1)}$ and $\mathsf{x}_0^{(2)}$ are also independent. Consequently, $\{\mathsf{x}_t^{(1)}\}_{t \geqslant 0}$ and $\{\mathsf{x}_t^{(2)}\}_{t \geqslant 0}$ are independent. Such independence is realistic, especially in scenarios where the two nodes' states do not interfere with each other. An example scenario is that a state corresponds to the position of a node, and the motions of the two nodes are independent.

Observations obtained by node i's sensor are represented by a stochastic process $\{\mathbf{z}_t^{(i)}\}_{t \geqslant 0}$ that satisfies

$$d\mathbf{z}_t^{(1)} = \mathring{\boldsymbol{\gamma}}_t^{(1)} \mathsf{x}_t^{(1)}\, dt + \boldsymbol{\Xi}_t^{(1)}\, d\mathbf{n}_t^{(1)} \tag{2a}$$

$$d\mathbf{z}_t^{(2)} = \boldsymbol{\Gamma}_t^{(2)} \Big[\mathsf{x}_t^{(1)} \quad \mathsf{x}_t^{(2)}\Big]^{\mathrm{T}} dt + \boldsymbol{\Xi}_t^{(2)}\, d\mathbf{n}_t^{(2)} \tag{2b}$$

for all $t \in [0, \infty)$, where $\mathring{\boldsymbol{\gamma}}_t^{(1)}$ and $\boldsymbol{\Gamma}_t^{(2)}$ are referred to as a sensor gain vector and a sensor gain matrix, respectively. Process $\{\mathbf{n}_t^{(i)}\}_{t \geqslant 0}$ in (2) is a Brownian motion and represents noise in node i's sensor observations. The level of such noise is affected by the deterministic matrix $\boldsymbol{\Xi}_t^{(i)}$. It is considered that $\boldsymbol{\Xi}_t^{(i)}\big(\boldsymbol{\Xi}_t^{(i)}\big)^{\mathrm{T}}$ is invertible for all $t \geqslant 0$. The observation $\mathbf{z}_0^{(i)}$ of node i at time 0 is given by

$$\mathbf{z}_0^{(1)} = \boldsymbol{g}_0^{(1)} \mathsf{x}_0^{(1)} + \boldsymbol{\zeta}^{(1)} \tag{3a}$$

$$\mathbf{z}_0^{(2)} = \boldsymbol{G}_0^{(2)} \Big[\mathsf{x}_0^{(1)} \quad \mathsf{x}_0^{(2)}\Big]^{\mathrm{T}} + \boldsymbol{\zeta}^{(2)} \tag{3b}$$

where $\boldsymbol{g}_0^{(1)}$ and $\boldsymbol{G}_0^{(2)}$ are a deterministic quantities. Vector $\boldsymbol{\zeta}^{(i)}$ is a zero-mean Gaussian random vector that represents noise in the observation at time 0 of node $i = 1, 2$.

At time t, node 2 transmits a signal $\mathsf{s}_t \in \mathbb{R}$ to node 1 via a Gaussian channel, and the received signal at time t is denoted by r_t. The transmitted signal s_t is generated by node 2 based on observations $\mathbf{z}_{0:t}^{(2)}$ obtained by its own sensor as well as observations $\mathbf{z}_{0:t}^{(1)}$ and signals $\mathsf{r}_{0:t}$ obtained via feedback. Consequently, s_t is $\sigma\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\big)$-measurable [53]. Therefore, s_t can be written as

$$\mathsf{s}_t = \mu_t\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\big) \tag{4}$$

for a real function μ_t referred to as the encoding function at time t. Note that "encoding" refers to the generation of signals as channel input based on information sources. The range of an encoding function is determined by the type of the channel. Since this paper considers a Gaussian channel, the range of an encoding function consists of all the real numbers. This is different from communication over a discrete memoryless channel (DMC), where an encoding function maps the information source to an element in a finite set via quantization and channel coding procedures. Consequently, the range of an encoding function is a finite set in DMC.

A collection of encoding functions $\mu_{0:T} := \{\mu_t\}_{t\in[0,T]}$ from time 0 to $T \geqslant 0$ is referred to as an encoding strategy of horizon T if the following constraint on transmit power is satisfied

$$\mathbb{E}\{\mathsf{s}_t^2\} \leqslant \mathring{P}_t \qquad \forall t \in [0,T] \tag{5}$$

where $\mathring{P}_t \geqslant 0$ represents a predefined power constraint at time t. Note that $\mathring{P}_t$ can vary with t, indicating that node 2's transmitter is allowed to adjust its power budget based on its available resources and system requirement at different times. This improves the flexibility and efficiency of the system and is especially suited for Internet of Things and beyond 5G network applications.

The signals received by node 1 are represented by a stochastic process $\{\mathsf{r}_t\}_{t\geqslant 0}$ given by

$$d\mathsf{r}_t = \mathsf{s}_t\, dt + \kappa_t\, d\mathsf{w}_t \tag{6}$$

with $\mathsf{r}_0 = 0$, where $\{\mathsf{w}_t\}_{t\geqslant 0}$ is a one-dimensional Brownian motion corresponding to additive Gaussian noise in the channel. The deterministic scalar κ_t affects the power of noise in the communication channel.

The following scenario is considered in the paper. The deterministic quantities $a_t^{(i)}$, $\boldsymbol{b}_t^{(i)}$, $\mathring{\gamma}_t^{(1)}$, $\boldsymbol{\Gamma}_t^{(2)}$, $\boldsymbol{\Xi}_t^{(i)}$, $\boldsymbol{g}_0^{(1)}$, $\boldsymbol{G}_0^{(2)}$ in (1)–(3) are known to both nodes. That is to say, the parameters determining the evolution of the state processes, sensor gain vectors and matrices, as well as levels of noise are known to both nodes. Random quantities $\mathsf{x}_0^{(1)}$, $\mathsf{x}_0^{(2)}$, $\boldsymbol{\zeta}^{(1)}$, $\boldsymbol{\zeta}^{(2)}$, $\{\mathbf{v}_t^{(1)}\}_{t\geqslant 0}$, $\{\mathbf{v}_t^{(2)}\}_{t\geqslant 0}$, $\{\mathbf{n}_t^{(1)}\}_{t\geqslant 0}$, $\{\mathbf{n}_t^{(2)}\}_{t\geqslant 0}$, and $\{\mathsf{w}_t\}_{t\geqslant 0}$ are independent. For any finite horizon $T \geqslant 0$, the quantities $\mathring{P}_t$, κ_t, and $a_t^{(i)}$, as well as each element of $\boldsymbol{b}_t^{(i)}$, $\mathring{\gamma}_t^{(1)}$, $\boldsymbol{\Gamma}_t^{(2)}$, and $\boldsymbol{\Xi}_t^{(i)}$ with $i = 1, 2$, considered as functions of t, are bounded and have only finitely many points of discontinuity in the interval $[0,T]$.

At each time $t \geqslant 0$, node 1 infers its unknown state $\mathsf{x}_t^{(1)}$ at time t based on its local observations $\mathbf{z}_{0:t}^{(1)}$ and signals $\mathsf{r}_{0:t}$ received from node 2. Therefore, the estimator of $\mathsf{x}_t^{(1)}$ computed by node 1 is $\sigma(\mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t})$-measurable. For an arbitrary encoding strategy $\mu_{0:t}$, the minimum-mean-square-error (MMSE) estimator $\hat{\mathsf{x}}_t^{(1)}$ of $\mathsf{x}_t^{(1)}$ at node 1 is the conditional expectation given by

$$\hat{\mathsf{x}}_t^{(1)} := \mathbb{E}\Big\{\mathsf{x}_t^{(1)} \,\big|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\Big\}. \tag{7}$$

This estimator is referred to as the distributed filter and is the focus of this paper. The MSE of the distributed filter at time t is determined by the encoding strategy employed by node 2 and is denoted by $e_t(\mu_{0:t})$ if strategy $\mu_{0:t}$ is employed, i.e.,

$$e_t(\mu_{0:t}) := \mathbb{E}\Big\{\big(\mathsf{x}_t^{(1)} - \hat{\mathsf{x}}_t^{(1)}\big)^2\Big\}.$$

On the other hand, inference of $\mathsf{x}_t^{(2)}$ at node 2 is less interesting in this paper and is thus not discussed. This is because node 2 can infer $\mathsf{x}_t^{(2)}$ optimally using the well-known Kalman-Bucy filter [53] as observations of both nodes are available to node 2 in the presence of feedback.

Remark 1: Equations (4)–(7) indicate that the encoder at node 2 only uses observations and feedback data obtained up to time t for generating the signal to be used by node 1 for inferring $\mathsf{x}_t^{(1)}$. In other words, the encoder does not wait to collect observations and feedback data after time t for inferring the state $\mathsf{x}_t^{(1)}$ at time t. This minimizes encoding and inference delay and is thus especially suitable for real-time communication, a critical requirement for modern CPSs [54], [55]. □

Remark 2: The distributed filtering problem in this paper is a generalization of the Gaussian process transmission problem in [56]. There, the aim is to design an encoding strategy for transmitting a known Gaussian process accurately via a Gaussian channel. For the special case where $\mathring{\gamma}_t^{(1)} = \mathbf{0}$, $\boldsymbol{\Gamma}_t^{(2)} = [1 \quad 0]$ and $\boldsymbol{\Xi}_t^{(2)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$, the distributed filtering problem considered in this paper reduces to the Gaussian process transmission problem in the literature. In particular, node 2 and node 1 play the roles of the transmitter and the receiver, respectively, and $\{\mathsf{x}_t^{(1)}\}_{t\geqslant 0}$ is the process to be sent from the transmitter to the receiver. Note that $\{\mathsf{x}_t^{(1)}\}_{t\geqslant 0}$ is known to node 2 in this special case as (2) simplifies to $d\mathbf{z}_t^{(2)} = \mathsf{x}_t^{(1)}\, dt$ for $i = 2$. □

III. Co-Design of Encoding and Filtering

This section introduces definitions for encoding strategies and presents the co-design of IDE and distributed filtering.

A. Definitions for Encoding Strategies

The definition of a linear encoding strategy is presented in the following.

Definition 1 (Linear encoding strategy) An encoding strategy $\mu_{0:T}$ is called a linear encoding strategy if μ_t is a linear function for all $t \in [0,T]$. □

If a linear encoding strategy of horizon T is employed, then $\mathsf{x}_{0:T}^{(1)}$, $\mathsf{x}_{0:T}^{(2)}$, $\mathbf{z}_{0:T}^{(1)}$, $\mathbf{z}_{0:T}^{(2)}$, and $\mathsf{r}_{0:T}$ are jointly Gaussian. Consequently, the distributed filter $\hat{\mathsf{x}}_T^{(1)}$ and its MSE are more tractable compared to those in cases where a nonlinear encoding strategy is employed.

We next define the notions of optimal linear encoding strategy (OLES) and optimal encoding strategy (OES). To this end, define $\mathring{P}_{0:T}$ as a short notation for $\{\mathring{P}_t\}_{0\leqslant t\leqslant T}$. Moreover, let $\mathcal{L}(\mathring{P}_{0:T})$ and $\mathcal{M}(\mathring{P}_{0:T})$ represent the set of linear encoding strategies and the set of encoding strategies of horizon T, respectively, that satisfy the power constraints $\mathring{P}_{0:T}$. Definitions of OLES and OES are presented in the following.

Definition 2 (OLES and OES) An OLES for $\mathsf{x}_T^{(1)}$ is an encoding strategy that minimizes the MSE of the distributed filter $\hat{\mathsf{x}}_T^{(1)}$ among all the strategies in $\mathcal{L}(\mathring{P}_{0:T})$. In other words, an OLES for $\mathsf{x}_T^{(1)}$ is a solution to the optimization problem

$$\underset{\mu_{0:T}\in\mathcal{L}(\mathring{P}_{0:T})}{\text{minimize}} \;\; e_T(\mu_{0:T}).$$

Similarly, an OES for $\mathsf{x}_T^{(1)}$ is a solution to the following optimization problem

$$\underset{\mu_{0:T}\in\mathcal{M}(\mathring{P}_{0:T})}{\text{minimize}} \;\; e_T(\mu_{0:T}). \qquad \square$$

Finally, two encoding strategies $\mu_{0:T}$ and $\tilde{\mu}_{0:T'}$ of horizon T and T', respectively, are called consistent if $\mu_t = \tilde{\mu}_t$ for all

$0 \leqslant t \leqslant \min\{T, T'\}$. The co-design of encoding and filtering is inherent in the definitions of OLES and OES as the error of the MMSE distributed filter is used as the metric for evaluating the encoding strategy.

B. IDE for Distributed Filtering

We now design IDE, an encoding strategy for minimizing the MSE in distributed filtering. To this end, define $\mathsf{y}_t^{(1)}$ as the MMSE estimator of $\mathsf{x}_t^{(1)}$ based on observations obtained by node 1 and node 2 from time 0 to time t, i.e.,

$$\mathsf{y}_t^{(1)} := \mathbb{E}\Big\{\mathsf{x}_t^{(1)} \,\big|\, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}\Big\} \qquad \text{for } t \in [0, T]. \tag{8}$$

The IDE strategy designed in this paper is described below.

Definition 3 (IDE) Consider an encoding function μ_t^{i} at time t given by

$$\mu_t^{\mathrm{i}}\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\big) = \alpha_t\Big(\mathsf{y}_t^{(1)} - \mathbb{E}\Big\{\mathsf{x}_t^{(1)} \,\big|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\Big\}\Big) \tag{9}$$

where $\alpha_t > 0$ is a scalar chosen to ensure that $\mathbb{E}\{\mathsf{s}_t^2\} = \mathring{P}_t$. Then $\mu_{0:T}^{\mathrm{i}}$ is called an IDE strategy of horizon T. □

As shown by (9), the transmitted signal at time t is proportional to the difference between two MMSE estimators of $\mathsf{x}_t^{(1)}$. In particular, MMSE estimator $\mathsf{y}_t^{(1)}$ can be computed by node 2 since it has access to $\mathbf{z}_{0:t}^{(1)}$ and $\mathbf{z}_{0:t}^{(2)}$. This estimator is a sufficient statistic of $\mathbf{z}_{0:t}^{(1)}$ and $\mathbf{z}_{0:t}^{(2)}$ for $\mathsf{x}_t^{(1)}$. Therefore, this estimator contains all the information in the observations of $\mathsf{x}_t^{(1)}$. On the other hand, $\mathbb{E}\{\mathsf{x}_t^{(1)} \,|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\}$ is the MMSE estimator of $\mathsf{x}_t^{(1)}$ based on observations and received signals available to node 1 up to time t. This estimator contains the information of $\mathsf{x}_t^{(1)}$ available to node 1. The difference between the two estimators is scaled by α_t to employ all the available transmit power at node 2.

IDE is explained in the following. By subtracting $\mathbb{E}\{\mathsf{x}_t^{(1)} \,|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\}$ from $\mathsf{y}_t^{(1)}$, node 2 only transmits information of $\mathsf{x}_t^{(1)}$ that has not been obtained by node 1. This increases the power efficiency of node 2 and improves the robustness of the transmitted signals against channel noise. To see this, consider a reference encoding strategy that scales $\mathsf{y}_t^{(1)}$ directly without the subtraction. In particular, the encoding function at time t of the reference strategy is given by $\mu_t^{\mathrm{r}}\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\big) = \alpha_t^{\mathrm{r}} \mathsf{y}_t^{(1)}$, where α_t^{r} is a scalar such that $\mathbb{E}\{\mu_t^{\mathrm{r}}\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\big)^2\} = \mathring{P}_t$. It can be verified that $\alpha_t > \alpha_t^{\mathrm{r}}$. This shows that under a same power constraint, IDE employs a larger scaling factor compared to the reference one, and thus more transmit power is used for transmitting the sufficient statistic that contains the useful information.

The scaling factor α_t can be computed by each node based on $a_\tau^{(i)}$, $\boldsymbol{b}_\tau^{(i)}$, $\boldsymbol{\Gamma}_\tau^{(i)}$, $\boldsymbol{\Xi}_\tau^{(i)}$, $\mathring{P}_\tau$, and κ_τ in the system model for $i = 1, 2$ and $0 \leqslant \tau \leqslant t$. Therefore, node 2 does not need to transmit information about α_t to node 1. Expressions of the conditional expectations in (9) and α_t are presented in Appendix A.

The next proposition shows that IDE is an OLES. It is also an OES in the special case where $\mathring{\boldsymbol{\gamma}}_t^{(1)} = \mathbf{0}$, i.e., node 1 observes only noise.

Proposition 1: For any $T \geqslant 0$, the IDE $\mu_{0:T}^{\mathrm{i}}$ is an OLES for $\mathsf{x}_T^{(1)}$. Furthermore, $\mu_{0:T}^{\mathrm{i}}$ is an OES for $\mathsf{x}_T^{(1)}$ if $\mathring{\boldsymbol{\gamma}}_t^{(1)} = \mathbf{0}$ at all $0 \leqslant t \leqslant T$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$. □

Proof: For an arbitrary encoding strategy $\mu_{0:t}$, define $\varepsilon_t(\mu_{0:t})$ as the MMSE for inferring $\mathsf{y}_t^{(1)}$ at node 1 based on $\mathbf{z}_{0:t}^{(1)}$ and $\mathsf{r}_{0:t}$ if $\mu_{0:t}$ is employed, i.e.,

$$\varepsilon_t(\mu_{0:t}) := \mathbb{V}\Big\{\mathsf{y}_t^{(1)} - \mathbb{E}\Big\{\mathsf{y}_t^{(1)} \,\big|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\Big\}\Big\}. \tag{10}$$

Since $\mathsf{x}_t^{(1)} - \mathsf{y}_t^{(1)} \perp\!\!\!\perp \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}$, the MSE $e_t(\mu_{0:t})$ of the distributed filter for inferring $\mathsf{x}_t^{(1)}$ satisfies

$$e_t(\mu_{0:t}) = \varepsilon_t(\mu_{0:t}) + \mathbb{V}\Big\{\mathsf{x}_t^{(1)} - \mathsf{y}_t^{(1)}\Big\}. \tag{11}$$

In (11), the first term in the sum is the MMSE for inferring $\mathsf{y}_t^{(1)}$ at node 1, whereas the second term in the sum is the MSE of $\mathsf{y}_t^{(1)}$ as an estimator of $\mathsf{x}_t^{(1)}$. The second term is determined by the quality of sensor observations and not by the encoding strategy. As a result, the encoding strategy design for minimizing $e_T(\mu_{0:T})$ is equivalent to that for minimizing $\varepsilon_T(\mu_{0:T})$. Appendix B shows that $\mu_{0:T}^{\mathrm{i}}$ minimizes $\varepsilon_T(\mu_{0:T})$ among all linear encoding strategies. In addition, we can show that $\mu_{0:T}^{\mathrm{i}}$ minimizes $\varepsilon_T(\mu_{0:T})$ among all encoding strategies if $\mathring{\boldsymbol{\gamma}}_t^{(1)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$, thus completing the proof. ⊠

Remark 3: The IDE strategies $\mu_{0:T}^{\mathrm{i}}$ and $\mu_{0:T'}^{\mathrm{i}}$ are consistent for any non-negative T and T'. Consequently, Proposition 1 shows the following favorable property of IDE: if $\mu_{0:T}^{\mathrm{i}}$ is an OLES (resp. OES) of horizon T, then for any $0 \leqslant T' < T$, strategy $\mu_{0:T'}^{\mathrm{i}}$ is an OLES (resp. OES) for $\mathsf{x}_{T'}^{(1)}$. □

In Proposition 1, the condition $\mathring{\boldsymbol{\gamma}}_t^{(1)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$ indicates that observations obtained by node 1 contain only noise and are thus independent of states of both node 1 and node 2. Under this condition, all the useful information is contained in the observations of node 2, i.e., the node that performs encoding. Note that the condition in Proposition 1 does not depend on $\boldsymbol{\Gamma}_t^{(2)}$.

The problem investigated in this paper under the condition $\mathring{\boldsymbol{\gamma}}_t^{(1)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$ is closely related to that studied in [35] and [56] for continuous-time scenarios and in [34] and [44] for discrete-time scenarios. Specifically, recall from (11) that an encoding strategy for minimizing $e_T(\mu_{0:T})$ is equivalent to that for minimizing $\varepsilon_T(\mu_{0:T})$. Using results for Kalman-Bucy filtering [52], [57], [58], $\{\mathsf{y}_t^{(1)}\}_{t\geqslant 0}$ can be shown to be a Gaussian Markov process. Moreover, if $\mathring{\boldsymbol{\gamma}}_t^{(1)} = \mathbf{0}$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$, then node 1 can only obtain information from received signals and not sensor observations. Consequently, the problem becomes designing an encoding strategy at node 2 so that the MSE at node 1 for inferring $\{\mathsf{y}_t^{(1)}\}_{t\geqslant 0}$ based only on node 1's received signals is minimized. This problem is referred to as transmission of a Gaussian process through a channel in [56], and the results therein are used in [35] for the design of encoding strategy and control policy for linear-quadratic stochastic systems.[1] The difference between

[1]Transmitted signals are viewed as measurements in [35], and an encoding strategy is referred to as a measurement strategy therein.

the problem investigated in this paper under the condition $\mathring{\gamma}_t^{(1)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$ as well as $\boldsymbol{g}_0^{(1)} = \mathbf{0}$ and that studied in [35] and [56] lies in the set of encoding strategies being considered. Specifically, recall that the transmitted signal s_t in this paper is $\sigma\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\big)$-measurable, whereas the transmitted signal in [35] and [56] are $\sigma\big(\mathsf{y}_t^{(1)}, \mathsf{r}_{0:t}\big)$-measurable. By definition (8), $\sigma\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\big) \supseteq \sigma\big(\mathsf{y}_t^{(1)}, \mathsf{r}_{0:t}\big)$, which indicates that a larger class of encoding strategies is considered here. Therefore, the results in this paper under the condition $\mathring{\gamma}_t^{(1)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$ are extensions of the results in [35] and [56].

In general cases where $\mathring{\gamma}_t^{(1)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$ or $\boldsymbol{g}_0^{(1)} = \mathbf{0}$ does not hold, IDE is not necessarily an OES. In fact, an OES is typically nonlinear. Specifically, the observations obtained by node 1 and node 2 are correlated in general cases, indicating that observations obtained by the two nodes contain common information about the unknown state $\mathsf{x}_t^{(1)}$. To maximize the encoding efficiency, node 2 needs to remove such common information in the transmitted signal and only transmit knowledge of node 1's unknown state not contained in node 1's observations. This in general requires employing nonlinear encoding. An example where a nonlinear encoding strategy outperforms linear ones is given in [59] for discrete-time scenarios. Indeed, it has been noted that if observations of different sensors are correlated, then linear encoding strategies may not be optimal [44], [60].

The MSE of the centralized estimator can be used as a benchmark on the performance of the distributed filter in general cases where an OES is difficult to find. In particular, (11) shows that $e_T(\mu_{0:T}) \geqslant \mathbb{V}\{\mathsf{x}_T^{(1)} - \mathsf{y}_T^{(1)}\}$ for any encoding strategy $\mu_{0:T}$. As a result, the gap between the MSE corresponding to IDE and an OES can be upper bounded by $e_T(\mu_{0:T}^{\mathrm{i}}) - \mathbb{V}\{\mathsf{x}_T^{(1)} - \mathsf{y}_T^{(1)}\}$, where $\mathbb{V}\{\mathsf{x}_t^{(1)} - \mathsf{y}_t^{(1)}\}$ can be obtained using classical Kalman-Bucy filtering results [53, Chapter 6].

The channel feedback affects the information pattern, i.e., what data are available, for the encoder at node 2. Since noiseless feedback is considered in this paper, node 2 has access to sensor observations and received signals obtained by node 1. In IDE, node 2 uses these data to remove from the transmitted signals the knowledge that has already been obtained by node 1. If the feedback is noisy, then the information pattern is changed, and the optimal encoding strategy would be generally nonlinear even if $\mathring{\gamma}_t^{(1)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$. The effects of information pattern on the design of encoding strategies for control and communication have been investigated in the literature [61], [62], [63]. For example, nonlinear control strategies have been shown to outperform linear ones in Witsenhausen's counterexample and in control over lossy channels with user-datagram-like communication protocols [61], [62].

IV. Extensions for Results on Encoding Design

This section discusses extensions of results in Section III for scenarios with average transmit power constraints and for multivariate unknown states.

A. Case of Average Power Constraints

Section II defines an encoding strategy as a collection of encoding functions under constraints (5) of the instantaneous transmit power at each time. In some applications such as event-triggered control [64], a node can choose to transmit signals intermittently in order to reserve communication resources for critical time instants. For such applications, constraints on the transmit power averaged over the time domain instead of on instantaneous power should be considered. Specifically, define an encoding strategy of horizon T under average power constraint $\bar{P}$ as a collection of encoding functions $\mu_{0:T} := \{\mu_t\}_{t\in[0,T]}$ that satisfy the following inequalities

$$\frac{1}{t}\int_{\tau=0}^{t} \mathbb{E}\Big\{\mu_\tau\big(\mathbf{z}_{0:\tau}^{(1)}, \mathbf{z}_{0:\tau}^{(2)}, \mathsf{r}_{0:\tau}\big)^2\Big\} d\tau \leqslant \bar{P} \quad \forall t \in (0, T]. \quad (12)$$

In other words, the average transmit power in the time window $[0, t]$ does not exceed $\bar{P}$ for all $t \in (0, T]$. We define OLES under average power constraints as follows. Let $\mathcal{L}_T(\bar{P})$ represent the set of all the linear encoding strategies of horizon T that satisfy constraints (12). An OLES for $\mathsf{x}_T^{(1)}$ under average power constraints is an encoding strategy that minimizes the MSE for inferring $\mathsf{x}_T^{(1)}$ among all strategies belonging to $\mathcal{L}_T(\bar{P})$. In other words, an OLES for $\mathsf{x}_T^{(1)}$ under average power constraints is a solution to the optimization problem

$$\underset{\mu_{0:T}\in\mathcal{L}_T(\bar{P})}{\text{minimize}}\; e_T\left(\mu_{0:T}\right). \quad (13)$$

An OES under average power constraints is defined similarly.

An OLES under average power constraints can be found via a two-step procedure. To this end, write (13) equivalently as

$$\underset{\mathring{P}_{0:T}}{\text{minimize}}\; \underset{\mu_{0:T}\in\mathcal{L}(\mathring{P}_{0:T})}{\text{minimize}}\; e_T\left(\mu_{0:T}\right) \quad (14a)$$

$$\text{subject to} \quad \frac{1}{t}\int_{\tau=0}^{t} \mathring{P}_\tau d\tau \leqslant \bar{P} \quad \forall t \in (0, T] \quad (14b)$$

where $\mathcal{L}(\mathring{P}_{0:T})$ represents the set of linear encoding strategies of horizon T that satisfy power constraints $\mathring{P}_{0:T}$. The first step for finding an OLES under average power constraints is to solve the inner minimization problem of (14), i.e., finding an OLES under instantaneous power constraints (5) specified by $\mathring{P}_{0:T}$. Note that $\mathring{P}_{0:T}$ can be viewed as an allocation scheme of the transmit power from time 0 to T. The second step is to solve the outer minimization problem of (14), i.e., finding the optimal power allocation scheme $P_{0:T}^*$ that leads to the minimum MSE. The OLES under instantaneous power constraints specified by $P_{0:T}^*$ is the OLES under average power constraints. An OES under average power constraints can also be found via a two-step procedure by replacing $\mathcal{L}(\mathring{P}_{0:T})$ in (14a) with $\mathcal{M}(\mathring{P}_{0:T})$, i.e., the set of encoding strategies of horizon T that satisfy power constraints $\mathring{P}_{0:T}$. Note that a similar two-step procedure has been used in the literature for the joint design of transmission and control strategies [34], [35]. Specifically, these works first design transmission and control strategies under hard power constraints, and then employ dynamic programming approaches for finding the optimal power allocation scheme.

OLES and OES under average power constraints can be constructed using IDE strategies. To this end, let $\mu^{\mathrm{i}}_{0:T}(\mathring{P}_{0:T})$ represent IDE strategies given in Definition 3 under constraint (5). Moreover, let $P^{*}_{0:T}$ represent the solution to the following optimization problem

$$\underset{\mathring{P}_{0:T}}{\text{minimize}} \qquad e_T\big(\mu^{\mathrm{i}}_{0:T}(\mathring{P}_{0:T})\big) \tag{15a}$$

$$\text{subject to} \quad \frac{1}{t}\int_{\tau=0}^{t} \mathring{P}_\tau \, d\tau \leqslant \bar{P} \qquad \forall t \in (0,T]\ . \tag{15b}$$

The following corollary shows that IDE is an OLES or an OES under average power constraints when certain conditions are satisfied.

Corollary 1: For any $T \geqslant 0$, the strategy $\mu^{\mathrm{i}}_{0:T}(P^{*}_{0:T})$ is an OLES for $\mathsf{x}^{(1)}_T$ under average power constraints. Furthermore, $\mu^{\mathrm{i}}_{0:T}(P^{*}_{0:T})$ is an OES for $\mathsf{x}^{(1)}_T$ under average power constraints if $\mathring{\boldsymbol{\gamma}}^{(1)}_t = \mathbf{0}$ at all $0 \leqslant t \leqslant T$ and $\boldsymbol{g}^{(1)}_0 = \mathbf{0}$. □

Proof: The IDE strategy $\mu^{\mathrm{i}}_{0:T}(\mathring{P}_{0:T})$ is a solution for the first step according to Proposition 1, and thus the first claim of Corollary 1 holds. The second claim of the corollary can be proved in the same manner. ⊠

B. Case of Multivariate Unknown States

Consider scenarios where the unknown states associated with the two nodes are vectors and the signal transmitted from node 2 to node 1 is also a vector. For $i = 1, 2$, let $\mathbf{x}^{(i)}_t \in \mathbb{R}^{n_i}$ represent the state of node i at time t, where n_i is the number of entries in $\mathbf{x}^{(i)}_t$. The state process $\{\mathbf{x}^{(i)}_t\}_{t\geqslant 0}$ follows the following SDE

$$d\mathbf{x}^{(i)}_t = \boldsymbol{A}^{(i)}_t \mathbf{x}^{(i)}_t dt + \boldsymbol{B}^{(i)}_t d\mathbf{v}^{(i)}_t \quad \forall t \in [0,\infty) \tag{16}$$

where $\boldsymbol{A}^{(i)}_t$ and $\boldsymbol{B}^{(i)}_t$ are deterministic matrices. The sensor observation $\mathbf{z}^{(1)}_t$ at node 1 satisfies

$$d\mathbf{z}^{(1)}_t = \mathring{\boldsymbol{\Gamma}}^{(1)}_t \mathbf{x}^{(1)}_t \, dt + \boldsymbol{\Xi}^{(1)}_t \, d\mathbf{n}^{(1)}_t \tag{17}$$

where $\mathring{\boldsymbol{\Gamma}}_t$ is a deterministic matrix representing the sensor gain of node 1 at time t.

Communication from node 2 to node 1 is described as follows. At time t, node 2 transmits a signal $\mathbf{s}_t \in \mathbb{R}^{n_{\mathrm{c}}}$ to node 1, where n_{c} is the number of entries in $\mathbf{s}_t$. The signal $\mathbf{s}_t$ can be written as (4) with $\mu_t(\cdot)$ replaced by a vector encoding function $\boldsymbol{\mu}_t(\cdot)$. The process $\{\mathbf{r}_t\}_{t\geqslant 0}$ of the signal received by node 1 satisfies

$$d\mathbf{r}_t = \mathbf{s}_t \, dt + \boldsymbol{K}_t \, d\mathbf{w}_t \tag{18}$$

where $\boldsymbol{K}_t$ is a diagonal matrix determining the level of channel noise, and $\{\mathbf{w}_t\}_{t\geqslant 0}$ is a Brownian motion. The transmitted signal satisfies the power constraint $\mathbb{E}\{\mathbf{s}^{\mathrm{T}}_t \mathbf{s}_t\} \leqslant \mathring{P}_t$ for all $t \in [0,T]$.

The aim of node 1 is to determine a distributed filter of $\mathbf{x}^{(1)}_t$ based on $\mathbf{z}^{(1)}_{0:t}$ and $\mathbf{r}_{0:t}$. The performance metric for such an estimator is its MSE. The MMSE estimator $\hat{\mathbf{x}}^{(1)}_t$ of $\mathbf{x}^{(1)}_t$ is given by the right hand side of (7), with $\mathsf{x}^{(1)}_t$ replaced by $\mathbf{x}^{(1)}_t$. The MSE of this estimator is affected by the encoding strategy employed by node 2. For an encoding strategy $\boldsymbol{\mu}_{0:t}$, the MSE of $\hat{\mathbf{x}}^{(1)}_t$ is defined as

$$e_t(\boldsymbol{\mu}_{0:t}) := \mathrm{tr}\Big\{\mathbb{V}\Big\{\mathbf{x}^{(1)}_t - \hat{\mathbf{x}}^{(1)}_t\Big\}\Big\}\,. \tag{19}$$

We present a lower bound on $e_T(\boldsymbol{\mu}_{0:T})$ over all linear encoding strategies. To this end, let $[\boldsymbol{M}]_{1:n}$ represent the n-by-n submatrix of a general matrix $\boldsymbol{M}$ consisting of the first n rows and columns of $\boldsymbol{M}$. A lower bound on $e_T(\boldsymbol{\mu}_{0:T})$ is presented in the following proposition.

Proposition 2: For any linear encoding strategy $\boldsymbol{\mu}_{0:T}$, a lower bound on $e_T(\boldsymbol{\mu}_{0:T})$ is given as follows

$$e_T(\boldsymbol{\mu}_{0:T}) \geqslant \mathrm{tr}\Big\{\big[\boldsymbol{E}^{\mathrm{c}}_T\big]_{1:n_1}\Big\} + n_1 \breve{\varepsilon}_T^{1/n_1} \tag{20}$$

where $\boldsymbol{E}^{\mathrm{c}}_t$ is defined in (36), and function $\breve{\varepsilon}_t$ solves the initial value problem (69) presented in Appendix C. □

Proof: See Appendix C. ⊠

Remark 4: Characterizing the MSE of optimal linear encoding strategies for multivariate unknown states is more challenging than for scalar states as the one-to-one mapping between Shannon information and MSE for scalar states does not exist. Specifically, if a linear encoding strategy is employed, then

$$\begin{aligned}\det&\Big(\mathbb{V}\Big\{\mathbf{x}^{(1)}_t - \hat{\mathbf{x}}^{(1)}_t\Big\}\Big) \\ &= \det\Big(\mathbb{V}\Big\{\mathbf{x}^{(1)}_t\Big\}\Big)\exp\Big\{-2I\big(\mathbf{x}^{(1)}_t; \mathbf{z}^{(1)}_{0:t}, \mathbf{r}_{0:t}\big)\Big\}\,.\end{aligned} \tag{21}$$

For scenarios where $\mathsf{x}^{(1)}_t$ is a scalar, the determinant of $\mathbb{V}\{\mathsf{x}^{(1)}_t - \hat{\mathsf{x}}^{(1)}_t\}$ equals the trace of this matrix, which is the MSE $e_t(\mu_{0:t})$. This shows a one-to-one mapping between the MSE and the mutual information term in (21), and thus the MSE of the optimal linear encoding strategies can be determined by maximizing the mutual information. By contrast, for scenarios where $\mathbf{x}^{(1)}_t$ is a vector, the determinant and the trace of $\mathbb{V}\{\mathbf{x}^{(1)}_t - \hat{\mathbf{x}}^{(1)}_t\}$ are different in general, and thus the one-to-one mapping between the MSE and the mutual information term does not exist. Therefore, the MSE of the optimal linear encoding strategies for multivariate states is more difficult to characterize, and a lower bound on such MSE is derived in this subsection. □

Next, we present extension of IDE for multivariate unknown states. In particular, the encoding function $\boldsymbol{\mu}^{\mathrm{i}}_t$ at time t is

$$\boldsymbol{\mu}^{\mathrm{i}}_t\big(\mathbf{z}^{(1)}_{0:t}, \mathbf{z}^{(2)}_{0:t}, \mathbf{r}_{0:t}\big) := \boldsymbol{\Phi}^{\mathrm{T}}_t\Big(\mathbf{y}_t - \mathbb{E}\Big\{\mathbf{x}_t \,\Big|\, \mathbf{z}^{(1)}_{0:t}, \mathbf{r}_{0:t}\Big\}\Big) \tag{22}$$

where $\boldsymbol{\Phi}_t \in \mathbb{R}^{(n_1+n_2)\times n_{\mathrm{c}}}$ is an encoding matrix; $\mathbf{x}_t$ and $\mathbf{y}_t$ are defined as

$$\mathbf{x}_t := \Big[\,\big(\mathbf{x}^{(1)}_t\big)^{\mathrm{T}} \quad \big(\mathbf{x}^{(2)}_t\big)^{\mathrm{T}}\Big]^{\mathrm{T}} \tag{23}$$

$$\mathbf{y}_t := \mathbb{E}\Big\{\mathbf{x}_t \,\Big|\, \mathbf{z}^{(1)}_{0:t}, \mathbf{z}^{(2)}_{0:t}\Big\}\,. \tag{24}$$

Equation (24) indicates that $\mathbf{y}_t$ can be viewed as the centralized MMSE estimator of $\mathbf{x}_t$ based on sensor observations of both nodes. Matrix $\boldsymbol{\Phi}_t$ is chosen to satisfy that the transmit power at time t equals $\mathring{P}_t$. This condition is written as $\mathrm{tr}\big\{\boldsymbol{\Phi}^{\mathrm{T}}_t \boldsymbol{Q}_t \boldsymbol{\Phi}_t\big\} = \mathring{P}_t$, where $\boldsymbol{Q}_t$ is defined in (36). Note that (9) is a special case of (22) with $\boldsymbol{\Phi}_t = [\alpha_t \quad 0]^{\mathrm{T}}$.

The MSE of the distributed filter is given by $e_t(\boldsymbol{\mu}^{\mathrm{i}}_{0:t}) = \mathrm{tr}\{[\boldsymbol{E}^{\mathrm{c}}_t + \boldsymbol{Q}_t]_{1:n_1}\}$, where $\boldsymbol{E}^{\mathrm{c}}_t$ is defined in (36). In particular, the expression of $\boldsymbol{Q}_t$ is given by (41) with ϕ_t and κ_t replaced by $\boldsymbol{\Phi}_t$ and $\boldsymbol{K}_t$, respectively, whereas the expression of $\boldsymbol{E}^{\mathrm{c}}_t$ is given by (37). Note that for scenarios of multivariate unknown states, definition of $\boldsymbol{B}_t$ in (38) is changed to $\boldsymbol{B}_t = \mathrm{diag}\{\boldsymbol{B}^{(1)}_t, \boldsymbol{B}^{(2)}_t\}$.

Finally, the connection between distributed filtering in continuous-time scenarios and that in discrete-time scenarios is discussed. A discrete-time counterpart of the system described by (1)–(6) can be obtained via time discretization. Specifically, let $\tau > 0$ represent a sampling interval, then the mth sample of the state, observation, and received signal in continuous-time scenarios are $\mathsf{x}^{(i)}_{m\tau}$ $(i = 1, 2)$, $\mathbf{z}^{(i)}_{m\tau}$, and $\mathsf{r}_{m\tau}$, respectively. These quantities are approximated by $\breve{\mathsf{x}}^{(i)}_m$, $\breve{\mathbf{z}}^{(i)}_m$, and $\breve{\mathsf{r}}_m$, respectively, using the Euler-Maruyama method [65]. These quantities satisfy the following difference equations

$$\breve{\mathsf{x}}^{(i)}_{m+1} = \left(1 + a^{(i)}_{m\tau}\,\tau\right)\breve{\mathsf{x}}^{(i)}_m + \sqrt{\tau}\left(\boldsymbol{b}^{(i)}_{m\tau}\right)^{\mathrm{T}}\breve{\mathbf{v}}^{(i)}_m \quad (25)$$

$$\frac{1}{\tau}\left(\breve{\mathbf{z}}^{(1)}_{m+1} - \breve{\mathbf{z}}^{(1)}_m\right) = \mathring{\gamma}^{(1)}_{m\tau}\,\breve{\mathsf{x}}^{(1)}_m + \frac{1}{\sqrt{\tau}}\boldsymbol{\Xi}^{(1)}_{m\tau}\,\breve{\mathbf{n}}^{(1)}_m \quad (26)$$

$$\frac{1}{\tau}\left(\breve{\mathbf{z}}^{(2)}_{m+1} - \breve{\mathbf{z}}^{(2)}_m\right) = \boldsymbol{\Gamma}^{(2)}_{m\tau}\left[\breve{\mathsf{x}}^{(1)}_m \quad \breve{\mathsf{x}}^{(2)}_m\right]^{\mathrm{T}} + \frac{1}{\sqrt{\tau}}\boldsymbol{\Xi}^{(2)}_{m\tau}\,\breve{\mathbf{n}}^{(2)}_m \quad (27)$$

$$\frac{1}{\tau}\left(\breve{\mathsf{r}}_{m+1} - \breve{\mathsf{r}}_m\right) = \mu_{m\tau}\left(\breve{\mathbf{z}}^{(1)}_{0:m}, \breve{\mathbf{z}}^{(2)}_{0:m}, \breve{\mathsf{r}}_{0:m}\right) + \frac{1}{\sqrt{\tau}}\kappa_{m\tau}\,\breve{\mathsf{w}}_m \quad (28)$$

where $\breve{\mathbf{v}}^{(i)}_m$, $\breve{\mathbf{n}}^{(1)}_m$, $\breve{\mathbf{n}}^{(2)}_m$, and $\breve{\mathsf{w}}_m$ are independent zero-mean Gaussian random vectors and variables. Specifically, $\breve{\mathbf{n}}^{(1)}_m$ and $\breve{\mathbf{n}}^{(2)}_m$ have identity covariance matrices, whereas $\breve{\mathsf{w}}_m$ has unit variance. Vector $\breve{\mathbf{z}}^{(i)}_{0:m}$ represents the concatenation of $\breve{\mathbf{z}}^{(i)}_0, \breve{\mathbf{z}}^{(1)}_i, \ldots, \breve{\mathbf{z}}^{(i)}_m$, and $\breve{\mathsf{r}}_{0:m}$ is defined similarly. Equations (25)-(28) specify a discrete-time version of the distributed filtering problem, where the left-hand sides of these equations represent the state of node i at time step $m+1$, the observation of sensor 1 at time step m, the observation of sensor 2 at time step m, and the received message at time step m, respectively. Such discrete-time distributed filtering problem has been investigated in [59].

V. Connection Between Shannon Information and Fisher Information

Shannon information and Fisher information are both related to MSE. The Shannon information quantities related to filtering problems include information supply, information storage, and information dissipation, introduced in [15] and [16]. The information-theoretical approach presented in those works is adopted in this paper to investigate the relationship between Shannon information and Fisher information for the distributed filtering problem described in Section IV-B. While results in this section are not used directly for deriving encoding strategies of Sections III and IV, these results demonstrate the difference in the information structure between distributed filtering and centralized filtering. In terms of notation, we denote Shannon's mutual information between general random quantities $\mathbf{x}$ and $\mathbf{y}$ by $I(\mathbf{x}; \mathbf{y})$, and denote the Fisher information matrix (FIM) for $\mathbf{x}$ from $\mathbf{y}$ by $\boldsymbol{J}(\mathbf{x}; \mathbf{y})$. Definitions of these quantities can be found in [15] and [16].

Consider the scenario where the following class of linear encoding strategies is employed

$$\mu_t\left(\mathbf{z}^{(1)}_{0:t}, \mathbf{z}^{(2)}_{0:t}, \mathbf{r}_{0:t}\right) = \boldsymbol{\Phi}^{\mathrm{T}}_t\,\mathbf{y}_t + \boldsymbol{g}_t\left(\mathbf{z}^{(1)}_{0:t}, \mathbf{r}_{0:t}\right) \quad (29)$$

where $\boldsymbol{\Phi}_t$ is a deterministic matrix, and $\boldsymbol{g}_t$ is an affine function. Consequently, the signal transmitted at time t is a linear function of the centralized MMSE estimator $\mathbf{y}_t$ plus a function of the local observations and received signals obtained by node 1. Note that (29) describes a class of encoding strategies since $\boldsymbol{\Phi}_t$ and $\boldsymbol{g}_t$ are design parameters and can be set to different values. Moreover, the class of encoding strategies specified by (22) is contained in the class of strategies (29) with $\boldsymbol{g}_t\left(\mathbf{z}^{(1)}_{0:t}, \mathbf{r}_{0:t}\right) = -\boldsymbol{\Phi}^{\mathrm{T}}_t\,\mathbb{E}\{\mathbf{x}_t \,|\, \mathbf{z}^{(1)}_{0:t}, \mathbf{r}_{0:t}\}$.

Information supply, information storage, and information dissipation for distributed filtering are defined as follows. The information supply $S(t)$ to the distributed filter up to time t is defined as the mutual information between the path of node 1's state $\mathbf{x}^{(1)}_{0:t}$ from time 0 to time t and the observations $\mathbf{z}^{(1)}_{0:t}$ as well as received signals $\mathbf{r}_{0:t}$ obtained by node 1, i.e.,

$$S(t) := I\left(\mathbf{x}^{(1)}_{0:t}; \mathbf{z}^{(1)}_{0:t}, \mathbf{r}_{0:t}\right).$$

The information supply $S(t)$ consists of two parts: one part is useful only for inferring the unknown state in the past, i.e., $\mathsf{x}^{(1)}_\tau$ with $\tau < t$, and can be discarded by the filter; the other part is useful for inferring the current unknown state $\mathsf{x}^{(1)}_t$ and should be stored by the filter. In particular, the information storage $C(t)$ of the distributed filter at time t is defined as the mutual information between node 1's state $\mathbf{x}^{(1)}_t$ at time t and $\mathbf{z}^{(1)}_{0:t}$ as well as $\mathsf{r}_{0:t}$, i.e.,

$$C(t) := I\left(\mathbf{x}^{(1)}_t; \mathbf{z}^{(1)}_{0:t}, \mathbf{r}_{0:t}\right).$$

The difference $S(t) - C(t)$ between information supply and information storage is referred to as information dissipation up to time t and its derivative $\frac{d}{dt}(S(t) - C(t))$ is called the dissipation rate of mutual information.

Next, a proposition on mutual information is presented, and a connection between information dissipation rate and Fisher information for distributed filtering is derived. To this end, recall that the FIM $\boldsymbol{J}(\mathbf{x}; \mathbf{y})$ for $\mathbf{x}$ from $\mathbf{y}$ is defined as $\boldsymbol{J}(\mathbf{x}; \mathbf{y}) := \mathbb{E}\{\jmath(\mathbf{x}; \mathbf{y})\}$. Here $\jmath : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}^{n \times n}$ is defined as

$$\jmath(\boldsymbol{x}; \boldsymbol{y}) := \frac{\partial}{\partial \boldsymbol{x}} \ln\left(\frac{f_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y})}{f_{\mathbf{x}}(\boldsymbol{x})}\right) \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}} \ln\left(\frac{f_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y})}{f_{\mathbf{x}}(\boldsymbol{x})}\right)$$

where $f_{\mathbf{x}}$ and $f_{\mathbf{x}|\mathbf{y}}$ represent the probability distribution function (PDF) of $\mathbf{x}$ and the conditional PDF of $\mathbf{x}$ given $\mathbf{y}$, respectively. Consider processes $\{\boldsymbol{\theta}_t\}_{t \geqslant 0}$, $\{\boldsymbol{\varphi}_t\}_{t \geqslant 0}$, and $\{\boldsymbol{\xi}_t\}_{t \geqslant 0}$ as given by

$$d\boldsymbol{\theta}_t = \boldsymbol{A}_t \boldsymbol{\theta}_t\, dt + \boldsymbol{B}_t\, d\mathbf{v}_t \quad (30a)$$

$$d\boldsymbol{\varphi}_t = \left(\boldsymbol{D}_t \boldsymbol{\theta}_t + \boldsymbol{E}_t \boldsymbol{\varphi}_t\right) dt + \boldsymbol{F}_t\, d\mathbf{v}_t + \boldsymbol{K}_t\, d\boldsymbol{\omega}_t \quad (30b)$$

$$d\boldsymbol{\xi}_t = \left(\boldsymbol{G}_t \boldsymbol{\theta}_t + \boldsymbol{H}_t \boldsymbol{\varphi}_t + \boldsymbol{g}_t(\boldsymbol{\xi}_{0:t})\right) dt + \boldsymbol{L}_t\, d\boldsymbol{\omega}_t \quad (30c)$$

where $\boldsymbol{A}_t$, $\boldsymbol{B}_t$, $\boldsymbol{D}_t$, $\boldsymbol{E}_t$, $\boldsymbol{F}_t$, $\boldsymbol{K}_t$, $\boldsymbol{G}_t$, $\boldsymbol{H}_t$, and $\boldsymbol{L}_t$ are deterministic matrices such that $\boldsymbol{L}_t \boldsymbol{L}^{\mathrm{T}}_t$ is invertible for any $t \geqslant 0$; $\boldsymbol{g}_t$ is a linear function; $\{\mathbf{v}_t\}_{t \geqslant 0}$ and $\{\boldsymbol{\omega}_t\}_{t \geqslant 0}$ are independent Brownian motions. In addition, $\boldsymbol{\theta}_0$, $\boldsymbol{\varphi}_0$, and $\boldsymbol{\xi}_0$ are Gaussian and are independent of $\{\mathbf{v}_t\}_{t \geqslant 0}$ as well as

$\{\boldsymbol{\omega}_t\}_{t\geqslant 0}$. The following proposition presents an inequality between dissipation rate of mutual information and a function of FIM.

Proposition 3: Consider processes $\{\boldsymbol{\theta}_t\}_{t\geqslant 0}$, $\{\boldsymbol{\varphi}_t\}_{t\geqslant 0}$, and $\{\boldsymbol{\xi}_t\}_{t\geqslant 0}$ given by (30). It holds that

$$\frac{d}{dt}\Big(I\big(\boldsymbol{\theta}_{0:t};\boldsymbol{\xi}_{0:t}\big)-I\big(\boldsymbol{\theta}_t;\boldsymbol{\xi}_{0:t}\big)\Big)\geqslant\frac{1}{2}\,\mathrm{tr}\Big\{\boldsymbol{B}_t\boldsymbol{B}_t^{\mathrm{T}}\boldsymbol{J}\big(\boldsymbol{\theta}_t;\boldsymbol{\xi}_{0:t}\big)\Big\} \tag{31}$$

for all $t\in(0,T]$, where equality is achieved if $\boldsymbol{H}_t=\mathbf{0}$. □

Proof: See Appendix D. ⊠

Next, an inequality between the information dissipation rate and Fisher information in distributed filtering is presented.

Proposition 4: If an encoding strategy that belongs to the class of strategies given by (29) is employed by node 2, then the information dissipation rate for distributed filtering satisfies

$$\frac{d}{dt}\big(S(t)-C(t)\big)\geqslant\frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{B}_t^{(1)}\big(\boldsymbol{B}_t^{(1)}\big)^{\mathrm{T}}\boldsymbol{J}\big(\mathbf{x}_t^{(1)};\mathbf{z}_{0:t}^{(1)},\mathbf{r}_{0:t}\big)\Big\}. \tag{32}$$

□

Proof: Applying Proposition 3 with $\boldsymbol{\theta}_t=\mathbf{x}_t^{(1)}$, $\boldsymbol{\varphi}_t=\big[(\mathbf{x}_t^{(2)})^{\mathrm{T}}\quad\mathbf{y}_t^{\mathrm{T}}\big]^{\mathrm{T}}$, and $\boldsymbol{\xi}_t=\big[\big(\mathbf{z}_t^{(1)}\big)^{\mathrm{T}}\quad\mathbf{r}_t^{\mathrm{T}}\big]^{\mathrm{T}}$, we obtain (32). ⊠

Remark 5: Proposition 4 can be extended to cases where observations of sensor 1 depend on states of both nodes. □

Fundamental equalities connecting Shannon information and Fisher information have been established for centralized filtering problems [15], [16]. In particular, it was shown that

$$\frac{d}{dt}\big(I(\mathbf{x}_{0:t};\mathbf{z}_{0:t})-I(\mathbf{x}_t;\mathbf{z}_{0:t})\big)=\frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{B}_t\boldsymbol{B}_t^{\mathrm{T}}\boldsymbol{J}\big(\mathbf{x}_t;\mathbf{z}_{0:t}\big)\Big\}$$

where $\mathbf{x}_t$ is defined in (23), $\mathbf{z}_t:=\big[\big(\mathbf{z}_t^{(1)}\big)^{\mathrm{T}}\quad\big(\mathbf{z}_t^{(2)}\big)^{\mathrm{T}}\big]^{\mathrm{T}}$, and $\boldsymbol{B}_t:=\mathrm{diag}\big\{\boldsymbol{B}_t^{(1)},\boldsymbol{B}_t^{(2)}\big\}$. A similar equality is derived in [56] for the transmission of $\big\{\mathsf{x}_t^{(1)}\big\}_{t\geqslant 0}$ given by (1) via a Gaussian channel. There, the received signal process $\{\mathsf{r}_t\}_{t\geqslant 0}$ satisfies

$$d\mathsf{r}_t=\mu_t\big(\mathsf{x}_t^{(1)},\mathsf{r}_{0:t}\big)\,dt+\kappa_t\,d\mathsf{w}_t \tag{33}$$

where $\mu_t(\cdot)$ is the encoding function at time t, quantity κ_t is a deterministic scalar, and w_t represents the additive Gaussian noise in the channel. The equality between Shannon information and Fisher information in [56] is derived using the fact that the drift coefficients $a_t^{(1)}\mathsf{x}_t^{(1)}$ and $\mu_t\big(\mathsf{x}_t^{(1)},\mathsf{r}_{0:t}\big)$ at time t for $d\mathsf{x}_t^{(1)}$ and $d\mathsf{r}_t$ (see (1) and (33)) are $\sigma\big(\mathsf{x}_{0:t}^{(1)},\mathsf{r}_{0:t}\big)$-measurable. However, this is not the case for the distributed filtering problem studied in this paper. In particular, the drift coefficient $\boldsymbol{\mu}_t\big(\mathbf{z}_{0:t}^{(1)},\mathbf{z}_{0:t}^{(2)},\mathbf{r}_{0:t}\big)$ at time t for $d\mathbf{r}_t$ (see (4) and (18)) is not $\sigma\big(\mathbf{x}_{0:t}^{(1)},\mathbf{z}_{0:t}^{(1)},\mathbf{r}_{0:t}\big)$-measurable. As a result, an inequality, instead of an equality, between Shannon information and Fisher information is derived in this paper.

This section considers linear encoding strategies, which make $\big\{\big[\big(\mathbf{x}_t^{(1)}\big)^{\mathrm{T}}\quad\big(\mathbf{z}_t^{(1)}\big)^{\mathrm{T}}\quad\mathbf{r}_t^{\mathrm{T}}\big]^{\mathrm{T}}\big\}_{t\geqslant 0}$ a Gaussian process. Note that calculation of Shannon information and Fisher information is not limited to Gaussian processes. For example, mutual information between non-Gaussian processes is calculated in [13] and [16]. However, the calculation of mutual information can be simplified significantly for Gaussian processes.

VI. Case Studies

This section presents numerical results for scalar unknown states in the setup described in Section II and for multivariate unknown states in the setup described in Section IV-A. Time-invariant systems are considered such that $a_t^{(i)}$ and $\boldsymbol{b}_t^{(i)}$ in (1), $\mathring{\gamma}_t^{(1)}$, $\boldsymbol{\Gamma}_t^{(2)}$, and $\boldsymbol{\Xi}_t^{(i)}$ in (2), $\mathring{P}_t$ in (5), κ_t in (6), $\boldsymbol{A}_t^{(i)}$ and $\boldsymbol{B}_t^{(i)}$ in (16), $\mathring{\boldsymbol{\Gamma}}_t^{(1)}$ in (17), and $\boldsymbol{K}_t$ in (18) do not vary with time t. Subscripts t for these quantities are thus omitted.

A. Results for Scalar Unknown States

We evaluate the accuracy of distributed filtering for scenarios where $\mathbf{v}_t^{(i)}$ contains only one entry and thus $\boldsymbol{b}^{(i)}$ becomes a scalar for $i=1,2$. Quantities in (1) and (2) are set to

$$a^{(1)}=0.05\,,\qquad \boldsymbol{b}^{(1)}=2\,,\qquad a^{(2)}=-0.05$$
$$\boldsymbol{b}^{(2)}=1\,,\qquad \mathring{\gamma}^{(1)}=0,\qquad \boldsymbol{\Xi}^{(1)}=2$$
$$\boldsymbol{\Gamma}^{(2)}=\begin{bmatrix}0&1\\-1&1\end{bmatrix},\qquad \boldsymbol{\Xi}^{(2)}=\begin{bmatrix}4&0\\0&1\end{bmatrix}$$

with $\mathbb{V}\{\mathsf{x}_0^{(1)}\}=\mathbb{V}\{\mathsf{x}_0^{(2)}\}=1.2$. The value of $\mathring{P}/\kappa^2$, which characterizes the signal-noise-ratio (SNR), is set to 8.4dB and -8.6dB, and the capacities of the Gaussian channel corresponding to these SNRs are 5 bits/s (bps) and 0.1 bps, respectively [56]. Parameters in (3) are given by $\boldsymbol{g}_0^{(1)}=\mathbf{0}$, $\boldsymbol{G}_0^{(2)}=0.1\boldsymbol{\Gamma}^{(2)}$, and $\mathbb{V}\{\boldsymbol{\zeta}^{(i)}\}=\boldsymbol{\Xi}^{(i)}$ for $i=1,2$.

The following class of encoding strategies is considered

$$\mathsf{s}_t=\alpha_t\big[\rho\quad\sqrt{1-\rho^2}\,\big]\Big(\mathbf{y}_t-\mathbb{E}\Big\{\mathbf{x}_t\,\big|\,\mathbf{z}_{0:t}^{(1)},\mathsf{r}_{0:t}\Big\}\Big) \tag{34}$$

where $\alpha_t>0$ and $\rho\in[0,1]$ are parameters chosen to ensure that $\mathbb{E}\{\mathsf{s}_t^2\}=\mathring{P}$. This class of encoding strategy can be interpreted as follows. According to (24), vector $\mathbf{y}_t-\mathbb{E}\{\mathbf{x}_t\,|\,\mathbf{z}_{0:t}^{(1)},\mathsf{r}_{0:t}\}$ can be viewed as the knowledge of $\mathbf{x}_t$ that is available to node 2 but not to node 1. Node 2 employs the available transmit power to send a linear combination of such a vector to node 1. For the case where $\rho=1$, (34) is the IDE shown in (9), and for cases where $\rho<1$, (34) represents reference methods. According to Proposition 1, the strategy given in (34) with $\rho=1$ is an OES for $\mathsf{x}_t^{(1)}$.

In addition to the class of encoding strategies in (34), we also consider the scenario where there is no communication link from node 2 to node 1, which is equivalent to setting $\mathsf{s}_t=0$. Another encoding strategy we consider as a reference method is that node 2 does not use feedback and uses only its own observations for encoding. For this strategy, we set $\mathsf{s}_t=\alpha_t^{\mathrm{c}}\,\mathbb{E}\{\mathsf{x}_t^{(1)}\,|\,\mathbf{z}_{0:t}^{(2)}\}$, where $\alpha_t^{\mathrm{c}}>0$ is chosen to satisfy $\mathbb{E}\{\mathsf{s}_t^2\}=\mathring{P}$. Note that feedback information $\mathbf{z}_{0:t}^{(1)}$ and $\mathsf{r}_{0:t}$ is not used for encoding at time t.

The normalized MSE of the distributed filter $\hat{\mathsf{x}}_t^{(1)}$ is used as the performance metric in this subsection. This metric is defined as $\mathbb{E}\big\{\big(\hat{\mathsf{x}}_t^{(1)}-\mathsf{x}_t^{(1)}\big)^2\big\}/\mathbb{E}\big\{\big(\mathsf{y}_t^{(1)}-\mathsf{x}_t^{(1)}\big)^2\big\}$, namely the ratio between the MSE of $\hat{\mathsf{x}}_t^{(1)}$ and that of the centralized MMSE estimator $\mathsf{y}_t^{(1)}$ given in (8). Note that the normalized MSE of a distributed filter is always greater than or equal to one. This is because a centralized estimator uses sensor

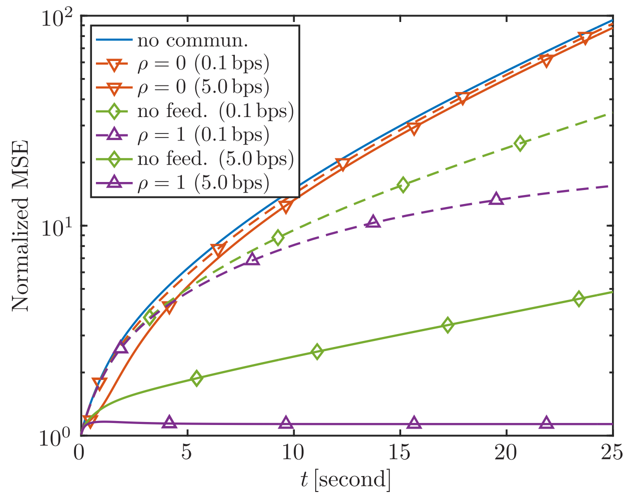


Fig. 2. Normalized MSE of the distributed filter in the log-scale as a function of time if node 2 employs the class of encoding strategy in (34) with $\rho = 1$ (proposed strategy) and $\rho = 0$. The normalized MSE for the scenario where there is no communication from node 2 to node 1 and the normalized MSE for the strategy that does not use feedback are also shown.

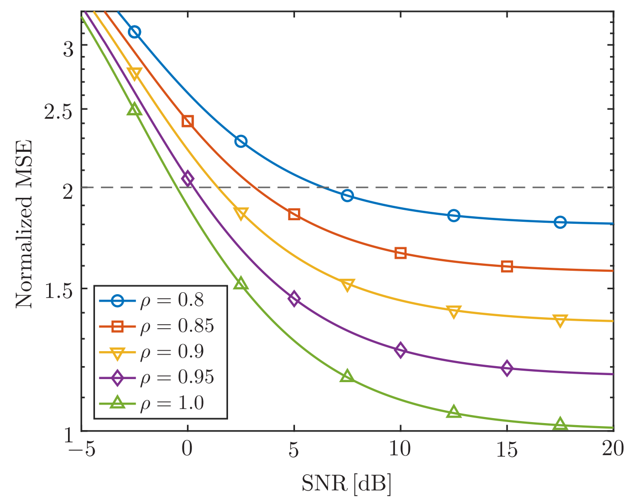


Fig. 3. Normalized MSEs of the distributed filter in the log-scale as a function of the SNR in the communication channel for different values of ρ. The IDE strategy corresponds to $\rho = 1$.

observations of both nodes and thus achieves smaller error than a distributed filter irrespective of the employed encoding strategy.

Figure 2 shows the normalized MSE of the distributed filter $\hat{\mathsf{x}}_t^{(1)}$ as a function of time t when different encoding strategies are employed. Note that if there is no communication, then the normalized MSE is not affected by the channel capacity, and thus only one curve is plotted for this strategy. One observation to make regarding Fig. 2 is that the normalized MSE of the distributed filter increases with time and is significantly larger than one for all the reference methods, indicating that the accuracy of the distributed filter degrades remarkably compared to the centralized MMSE estimator. By contrast, the normalized MSE of the distributed filter approaches one when IDE (i.e., $\rho = 1$) is employed, indicating that the performance gap between the distributed filter and the centralized MMSE estimator is marginal. This supports the claim of Proposition 1 that IDE is an OES in this setting.

Another observation to make regarding Fig. 2 is that the normalized MSEs of both the proposed strategy and reference methods are larger when the channel capacity is 0.1 bps compared to the scenario where the channel capacity is 5 bps, as the received signals suffer more from channel noise. In particular, the normalized MSE of the IDE keeps increasing with time when the channel capacity is 0.1 bps. Since IDE is an OES, the MSE of the distributed filter is unbounded as time approaches infinity irrespective of the encoding strategy employed by node 2. This is affirmed in [66], where a necessary and sufficient condition under which the MSE of the distributed filter is bounded over time is derived.

Next, we evaluate the power efficiency of IDE compared to reference methods specified by (34) with different values of ρ. In particular, we consider the stationary performance of the distributed filter, namely the error of the filter at a time sufficiently large such that the error covariance has converged, as a function of the SNRs $\mathring{P}/\kappa^2$ in the communication channel from node 2 to node 1. Figure 3 shows the normalized MSE of the distributed filter for different values of ρ. First, the normalized MSE decreases with ρ and achieves the minimum value at $\rho = 1$, i.e., when IDE is used, at all SNRs. In particular, to achieve a normalized MSE of 2 as indicated by the horizontal dashed line in Fig. 3, the transmit power required for IDE is 6.9 dB and 2.9 dB, respectively, smaller than that required for the reference strategy with $\rho = 0.8$ and $\rho = 0.9$. Second, the normalized MSEs decrease with SNR for all the encoding strategies as more information can be transmitted over the channel.

B. Results for Multivariate Unknown States

We consider the use case of 2-dimensional (2D) tracking, which is an important application in industrial CPSs. The state $\mathsf{x}_t^{(i)}$ of each node $i = 1, 2$ has four entries, where the first two entries represent the 2D position of node i, and the last two entries represent the velocity of this node. The evolution of each node's state follows the continuous white noise acceleration model [67], and parameters in (16)–(17) are given by

$$\boldsymbol{A}^{(i)} = \begin{bmatrix} \mathbf{0}_{2\times 2} & \boldsymbol{I}_2 \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \end{bmatrix}, \qquad \boldsymbol{B}^{(i)} = \boldsymbol{I}_4\,, \qquad \mathring{\boldsymbol{\Gamma}}^{(1)} = \begin{bmatrix} \boldsymbol{I}_2 & \mathbf{0}_{2\times 2} \end{bmatrix}$$
$$\boldsymbol{\Gamma}^{(2)} = \begin{bmatrix} \mathbf{0}_{4\times 4} & \boldsymbol{I}_4 \\ -\boldsymbol{I}_4 & \boldsymbol{I}_4 \end{bmatrix}, \qquad \boldsymbol{\Xi}^{(1)} = \xi \boldsymbol{I}_2\,, \qquad \boldsymbol{\Xi}^{(2)} = 0.2\boldsymbol{I}_8 \tag{35}$$

where $\boldsymbol{I}_n$ represents the n-by-n identity matrix. In (35), ξ is a parameter determining the noise level of sensor 1. The observation model specified by (35) is interpreted as follows. Node 1 can observe its 2D position. Node 2 can observe its own state, as well as the difference between the states of the two nodes. In terms of communication, strategy (22) is employed with $\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{I}_4 & \mathbf{0}_{4\times 4} \end{bmatrix}^{\mathrm{T}}$, and matrix $\boldsymbol{K}$ in (18) is set to $\boldsymbol{K} = \boldsymbol{I}_4$.

Figure 4 shows the MSE of the distributed filter and the lower bound (20) as functions of sensor noise level ξ for

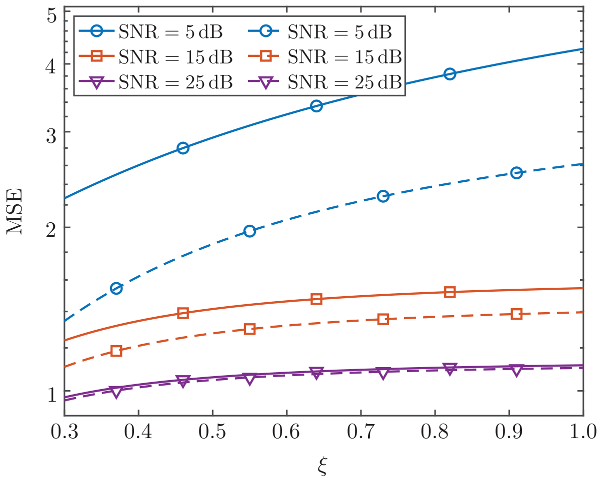


Fig. 4. MSEs of the distributed filter in the log-scale as a function of noise level of sensor 1 for different SNRs in the communication channel. Solid lines represent MSEs for the encoding strategy (22), whereas dashed lines represent lower bounds (20) on MSEs for all linear encoding strategies.

different SNRs in the communication channel. First, the MSEs of the distributed filter and their lower bounds increase with ξ as sensor 1 suffers more from noise. Such increase becomes less significant at higher SNR in the channel, since the distributed filter can employ the received communication data for inferring unknown state $\mathrm{x}^{(1)}_t$ and thus relies less on the observations of sensor 1. Second, there is a gap between the MSEs of the distributed filter and their lower bounds, and such gap is smaller at higher SNR. The gap is due to the effect of the bounding techniques for deriving the lower bound and also due to the non-optimality of the evaluated encoding strategy.

VII. Conclusion

This paper established theories and designed strategies for continuous-time distributed filtering. In particular, the paper investigated a two-node system in a network where one node transmits encoded signals to the other node via a Gaussian feedback channel. We designed an encoding strategy named IDE and proved that it is an optimal linear encoding strategy. IDE is also optimal among all encoding strategies in the special case where the node that receives encoded signals can only observe noise. Moreover, an inequality between the information dissipation rate and Fisher information in distributed filtering was established. This extends the fundamental relationship between Shannon information and Fisher information for Kalman-Bucy filtering to distributed settings. The findings in this paper serve as a building block for extension to general networks and provides insights for the co-design of communication and computing algorithms in applications such as industrial cyber-physical systems.

Appendix A
Transmitted Signals for IDE

For the simplicity of presentation, define

$$\hat{\mathbf{x}}_t := \mathbb{E}\left\{\mathbf{x}_t \,\middle|\, \mathbf{z}^{(1)}_{0:t}, \mathrm{r}_{0:t}\right\}, \qquad \boldsymbol{E}^{\mathrm{c}}_t := \mathbb{V}\left\{\mathbf{x}_t - \mathbf{y}_t\right\} \tag{36a}$$

$$\boldsymbol{E}_t := \mathbb{V}\left\{\mathbf{x}_t - \hat{\mathbf{x}}_t\right\}, \qquad \boldsymbol{Q}_t := \mathbb{V}\left\{\mathbf{y}_t - \hat{\mathbf{x}}_t\right\} \tag{36b}$$

where $\mathbf{x}_t$ and $\mathbf{y}_t$ are defined in (23) and (24), respectively. Note that $\boldsymbol{E}^{\mathrm{c}}_t$ is the error covariance matrix of the centralized estimator $\mathbf{y}_t$. Using results for optimal linear filtering [56, Chapter 10], we obtain

$$\frac{d\boldsymbol{E}^{\mathrm{c}}_t}{dt} = \boldsymbol{A}_t\boldsymbol{E}^{\mathrm{c}}_t + \boldsymbol{E}^{\mathrm{c}}_t\boldsymbol{A}^{\mathrm{T}}_t + \boldsymbol{B}_t\boldsymbol{B}^{\mathrm{T}}_t - \boldsymbol{E}^{\mathrm{c}}_t\boldsymbol{\Gamma}^{\mathrm{T}}_t\left(\boldsymbol{\Xi}_t\boldsymbol{\Xi}^{\mathrm{T}}_t\right)^{-1}\boldsymbol{\Gamma}_t\boldsymbol{E}^{\mathrm{c}}_t \tag{37}$$

where

$$\begin{aligned}\boldsymbol{A}_t &:= \mathrm{diag}\left\{a^{(1)}_t, a^{(2)}_t\right\}, \qquad \boldsymbol{B}_t := \mathrm{diag}\left\{\left(\boldsymbol{b}^{(1)}_t\right)^{\mathrm{T}}, \left(\boldsymbol{b}^{(2)}_t\right)^{\mathrm{T}}\right\}\\ \boldsymbol{\Gamma}_t &:= \left[\left(\boldsymbol{\Gamma}^{(1)}_t\right)^{\mathrm{T}} \ \left(\boldsymbol{\Gamma}^{(2)}_t\right)^{\mathrm{T}}\right]^{\mathrm{T}}, \quad \boldsymbol{\Xi}_t := \mathrm{diag}\left\{\boldsymbol{\Xi}^{(1)}_t, \boldsymbol{\Xi}^{(2)}_t\right\}.\end{aligned} \tag{38}$$

Here,

$$\boldsymbol{\Gamma}^{(1)}_t := \left[\mathring{\gamma}^{(1)}_t \quad \mathbf{0}\right]. \tag{39}$$

The term $\mathbb{E}\{\mathrm{x}^{(1)}_t \,|\, \mathbf{z}^{(1)}_{0:t}, \mathrm{r}_{0:t}\}$ in (9) is the first entry of $\hat{\mathbf{x}}_t$. If IDE is employed, then $\hat{\mathbf{x}}_t$ can be computed as

$$\begin{aligned}d\hat{\mathbf{x}}_t = {}& \left[\boldsymbol{A}_t - \boldsymbol{E}_t\left(\boldsymbol{\Gamma}^{(1)}_t\right)^{\mathrm{T}}\left(\boldsymbol{\Xi}^{(1)}_t\left(\boldsymbol{\Xi}^{(1)}_t\right)^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}^{(1)}_t\right]\hat{\mathbf{x}}_t\,dt\\ &+ \boldsymbol{E}_t\left(\boldsymbol{\Gamma}^{(1)}_t\right)^{\mathrm{T}}\left(\boldsymbol{\Xi}^{(1)}_t\left(\boldsymbol{\Xi}^{(1)}_t\right)^{\mathrm{T}}\right)^{-1}d\mathbf{z}^{(1)}_t + \boldsymbol{Q}_t\boldsymbol{\phi}_t\kappa^{-2}_t\,d\mathrm{r}_t\end{aligned} \tag{40}$$

with $\hat{\mathbf{x}}_0 = \mathbb{E}\{\mathbf{x}_0 \,|\, \mathbf{z}^{(1)}_0\}$. Here, we define $\boldsymbol{\phi}_t := [\alpha_t \quad 0]^{\mathrm{T}}$. The expression of α_t is $\alpha_t = \left(\mathring{P}_t/[\boldsymbol{Q}_t]_{1,1}\right)^{1/2}$. Moreover, $\boldsymbol{Q}_t$ and $\boldsymbol{E}_t$ satisfy

$$\begin{aligned}\frac{d\boldsymbol{Q}_t}{dt} = {}& \boldsymbol{A}_t\boldsymbol{Q}_t + \boldsymbol{Q}_t\boldsymbol{A}^{\mathrm{T}}_t + \boldsymbol{E}^{\mathrm{c}}_t\boldsymbol{\Gamma}^{\mathrm{T}}_t\left(\boldsymbol{\Xi}_t\boldsymbol{\Xi}^{\mathrm{T}}_t\right)^{-1}\boldsymbol{\Gamma}_t\boldsymbol{E}^{\mathrm{c}}_t\\ &- \left(\boldsymbol{E}^{\mathrm{c}}_t + \boldsymbol{Q}_t\right)\left(\boldsymbol{\Gamma}^{(1)}_t\right)^{\mathrm{T}}\left(\boldsymbol{\Xi}^{(1)}_t\left(\boldsymbol{\Xi}^{(1)}_t\right)^{\mathrm{T}}\right)^{-1}\\ &\times \boldsymbol{\Gamma}^{(1)}_t\left(\boldsymbol{E}^{\mathrm{c}}_t + \boldsymbol{Q}_t\right) - \boldsymbol{Q}_t\boldsymbol{\phi}_t\,\kappa^{-2}_t\,\boldsymbol{\phi}^{\mathrm{T}}_t\boldsymbol{Q}_t\end{aligned} \tag{41}$$

$$\boldsymbol{E}_t = \boldsymbol{E}^{\mathrm{c}}_t + \boldsymbol{Q}_t\,. \tag{42}$$

Note that $\boldsymbol{E}_t$, $\boldsymbol{Q}_t$, and α_t are determined by parameters in the system model $a^{(i)}_\tau$, $\boldsymbol{b}^{(i)}_\tau$, $\boldsymbol{\Gamma}^{(i)}_\tau$, $\boldsymbol{\Xi}^{(i)}_\tau$, $\mathring{P}_\tau$, and κ_τ for $i = 1, 2$ and $0 \leqslant \tau \leqslant t$. Consequently, $\boldsymbol{E}_t$, $\boldsymbol{Q}_t$, and α_t are not affected by the instantiations of $\mathrm{x}^{(1)}_{0:t}$, $\mathrm{x}^{(2)}_{0:t}$, $\mathbf{z}^{(1)}_{0:t}$, $\mathbf{z}^{(2)}_{0:t}$, and $\mathrm{r}_{0:t}$.

Appendix B
Proof of Proposition 1

We first present a few lemmas to be used in the proof. The first lemma provides a lower bound on MSEs of estimators for Gaussian random variables.

Lemma 1: Let θ be a Gaussian random variable and $\boldsymbol{\xi}$ be a random vector. For an arbitrary estimator $\hat{\theta}$ of θ based on $\boldsymbol{\xi}$, i.e., θ is $\sigma(\boldsymbol{\xi})$-measurable, its MSE can be lower bounded as

$$\mathbb{E}\left\{\left(\theta - \hat{\theta}\right)^2\right\} \geqslant \mathbb{V}\{\theta\}\exp\left\{-2I\left(\theta; \boldsymbol{\xi}\right)\right\}. \tag{43}$$

Equality in (43) is achieved if θ and $\boldsymbol{\xi}$ are jointly Gaussian and $\hat{\theta}$ is the MMSE estimator of θ. □

Proof: Lemma 1 can be proved using similar arguments as those used for proving Lemma 11.3.1 of [44]. ⊠

The next lemma shows conditional independence relations for the distributed learning problem investigated by this paper.

Lemma 2: The following conditional independence relations hold for all $t \geqslant 0$

$$\mathsf{x}_t^{(1)} \perp\!\!\!\perp \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t} \,\big|\, \mathsf{y}_t^{(1)} \tag{44a}$$

$$\mathsf{x}_t^{(1)}, \mathsf{x}_t^{(2)} \perp\!\!\!\perp \mathsf{r}_{0:t} \,\big|\, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)} \tag{44b}$$

where $\mathsf{y}_t^{(1)}$ is defined in (8). □

Proof: See Appendix A.1 of [68]. ⊠

Next, Proposition 1 is proved.

Proof: First, it is proved that $\mu_{0:T}^{\mathrm{i}}$ is an OLES for $\mathsf{x}_T^{(1)}$. As discussed in Section III-B, we only need to show that $\mu_{0:T}^{\mathrm{i}}$ minimizes $\varepsilon_T(\mu_{0:T})$ among all linear encoding strategies of horizon T. To this end, consider an arbitrary linear encoding strategy $\mu_{0:T}$ of horizon T. We prove that $\mu_{0:T}^{\mathrm{i}}$ is an OLES by deriving a lower bound $\breve{\varepsilon}(T)$ on $\varepsilon_T(\mu_{0:T})$ and showing that $\mu_{0:T}^{\mathrm{i}}$ achieves this bound, i.e., $\varepsilon_T(\mu_{0:T}^{\mathrm{i}}) = \breve{\varepsilon}(T)$.

We derive $\breve{\varepsilon}(T)$ using the relationship between MSE and mutual information shown in Lemma 1. To this end, choose arbitrary numbers t and τ such that $0 \leqslant t \leqslant \tau \leqslant T$, and define $\tilde{\varepsilon}_t(\tau)$ as the MMSE for inferring $\mathsf{y}_\tau^{(1)}$ based on $\mathbf{z}_{0:\tau}^{(1)}$ and $\mathsf{r}_{0:t}$ if linear encoding strategy $\mu_{0:t}$ is employed by node 2, i.e.,

$$\tilde{\varepsilon}_t(\tau) := \mathbb{V}\Big\{\mathsf{y}_\tau^{(1)} - \mathbb{E}\Big\{\mathsf{y}_\tau^{(1)} \,\big|\, \mathbf{z}_{0:\tau}^{(1)}, \mathsf{r}_{0:t}\Big\}\Big\}. \tag{45}$$

Choose an arbitrary s such that $t < s \leqslant T$. Applying Lemma 1, we obtain

$$\tilde{\varepsilon}_t(s) = \mathbb{V}\big\{\mathsf{y}_s^{(1)}\big\} \exp\Big\{-2I\big(\mathsf{y}_s^{(1)}; \mathbf{z}_{0:s}^{(1)}, \mathsf{r}_{0:t}\big)\Big\} \tag{46}$$

$$\varepsilon_s(\mu_{0:s}) = \mathbb{V}\big\{\mathsf{y}_s^{(1)}\big\} \exp\Big\{-2I\big(\mathsf{y}_s^{(1)}; \mathbf{z}_{0:s}^{(1)}, \mathsf{r}_{0:s}\big)\Big\}. \tag{47}$$

Next, an inequality between the two mutual information terms in (46) and (47) is derived. Applying the chain rule of mutual information gives

$$I(\mathsf{y}_s^{(1)}; \mathbf{z}_{0:s}^{(1)}, \mathsf{r}_{0:s}) = I(\mathsf{y}_s^{(1)}; \mathbf{z}_{0:s}^{(1)}, \mathsf{r}_{0:t}) + I\big(\mathsf{y}_s^{(1)}; \mathsf{r}_{t:s} \,\big|\, \mathbf{z}_{0:s}^{(1)}, \mathsf{r}_{0:t}\big). \tag{48}$$

Define $\mathring{\mathsf{r}}_\tau := \mathsf{r}_\tau - \mathsf{r}_t$ for $t \leqslant \tau \leqslant T$, then

$$\begin{aligned} I\big(\mathsf{y}_s^{(1)}; \mathsf{r}_{t:s} \,\big|\, \mathbf{z}_{0:s}^{(1)}, \mathsf{r}_{0:t}\big) &= I\big(\mathsf{y}_s^{(1)}; \mathring{\mathsf{r}}_{t:s} \,\big|\, \mathbf{z}_{0:s}^{(1)}, \mathsf{r}_{0:t}\big) \\ &\leqslant I\big(\mathsf{y}_s^{(1)}, \mathbf{z}_{0:s}^{(1)}, \mathsf{r}_{0:t}; \mathring{\mathsf{r}}_{t:s}\big) \\ &\leqslant I\big(\mathbf{z}_{0:s}^{(1)}, \mathbf{z}_{0:s}^{(2)}, \mathsf{r}_{0:t}; \mathring{\mathsf{r}}_{t:s}\big) \end{aligned} \tag{49}$$

where the last inequality is because $\mathsf{y}_s^{(1)}$ is $\sigma\big(\mathbf{z}_{0:s}^{(1)}, \mathbf{z}_{0:s}^{(2)}\big)$-measurable. Note that $\mathring{\mathsf{r}}_\tau$ satisfies $d\mathring{\mathsf{r}}_\tau = \mathsf{s}_\tau \, d\tau + \kappa_\tau \, d\mathring{\mathsf{w}}_\tau$, where $\mathring{\mathsf{w}}_\tau := \mathsf{w}_\tau - \mathsf{w}_t$. Recall that s_τ is $\sigma\big(\mathbf{z}_{0:\tau}^{(1)}, \mathbf{z}_{0:\tau}^{(2)}, \mathsf{r}_{0:\tau}\big)$-measurable, and is thus also $\sigma\big(\mathbf{z}_{0:\tau}^{(1)}, \mathbf{z}_{0:\tau}^{(2)}, \mathsf{r}_{0:t}, \mathring{\mathsf{r}}_{t:\tau}\big)$-measurable. Moreover, the relationship $\mathbf{z}_{0:s}^{(1)}, \mathbf{z}_{0:s}^{(2)}, \mathsf{r}_{0:t} \perp\!\!\!\perp \mathring{\mathsf{w}}_{t:s}$ holds. Using similar arguments as in [56, Chapter 16.3], we obtain

$$I\big(\mathbf{z}_{0:s}^{(1)}, \mathbf{z}_{0:s}^{(2)}, \mathsf{r}_{0:t}; \mathring{\mathsf{r}}_{t:s}\big) \leqslant \int_t^s C_\tau \, d\tau \tag{50}$$

where $C_\tau := \mathring{P}_\tau / (2\kappa_\tau^2)$. Combining (46)–(50) gives

$$\varepsilon_s(\mu_{0:s}) \geqslant \tilde{\varepsilon}_t(s) \exp\left\{-2\int_t^s C_\tau \, d\tau\right\}. \tag{51}$$

Define a function $\psi(\Delta, t)$ as

$$\psi(\Delta, t) := \tilde{\varepsilon}_t(t+\Delta) \exp\Big\{-2\int_t^{t+\Delta} C_\tau \, d\tau\Big\} \quad \text{for } t \in [0,T],\ \Delta \in [0, T-t]. \tag{52}$$

Using this definition, we obtain

$$\psi(0,t) = \tilde{\varepsilon}_t(t) = \varepsilon_t(\mu_{0:t}) \tag{53}$$

where the second equality is obtained using (10) and (45). Subtracting both sides of (51) by $\psi(0,t)$ gives

$$\varepsilon_s(\mu_{0:s}) - \varepsilon_t(\mu_{0:t}) \geqslant \psi(s-t, t) - \psi(0,t).$$

Therefore,

$$\begin{aligned} \frac{d}{dt}\varepsilon_t(\mu_{0:t}) &\geqslant \lim_{s\to t^+} \frac{1}{s-t}\big(\psi(s-t,t) - \psi(0,t)\big) \\ &= \frac{\partial}{\partial\Delta}\psi(\Delta,t)\Big|_{\Delta=0} = \frac{d}{d\tau}\tilde{\varepsilon}_t(\tau)\Big|_{\tau=t} - 2C_t\,\varepsilon_t(\mu_{0:t}) \end{aligned} \tag{54}$$

where $\lim_{s\to t^+}$ is the one-sided limit as s approaches t from above. The second equality in (54) is obtained using the formula for derivatives of composite functions and (53).

Next, $\frac{d}{d\tau}\tilde{\varepsilon}_t(\tau)$ in (54) is derived using the notion of innovation processes. Define innovation process $\{\tilde{\boldsymbol{\eta}}_\tau^{(i)}\}_{\tau\geqslant 0}$ with respect to $\{\mathbf{z}_\tau^{(i)}\}_{\tau\geqslant 0}$ for $i = 1, 2$ as

$$d\tilde{\boldsymbol{\eta}}_\tau^{(i)} = d\mathbf{z}_\tau^{(i)} - \boldsymbol{\Gamma}_\tau^{(i)}\mathbf{y}_\tau \, d\tau, \qquad \tilde{\boldsymbol{\eta}}_0^{(i)} = \mathbf{z}_0^{(i)}$$

where $\boldsymbol{\Gamma}_t^{(1)}$ is defined in (39), and $\mathbf{y}_t$ is defined in (24). In addition, define a process $\{\boldsymbol{\eta}_\tau^{(i)}\}_{\tau\geqslant 0}$ as

$$d\boldsymbol{\eta}_\tau^{(i)} = \Big(\boldsymbol{\Xi}_\tau^{(i)}\big(\boldsymbol{\Xi}_\tau^{(i)}\big)^{\mathrm{T}}\Big)^{-1/2} d\tilde{\boldsymbol{\eta}}_\tau^{(i)}, \qquad \boldsymbol{\eta}_0^{(i)} = \mathbf{z}_0^{(i)}. \tag{55}$$

The process $\{\boldsymbol{\eta}_\tau^{(i)}\}_{\tau\geqslant 0}$ is named a scaled innovation process. Processes $\{\boldsymbol{\eta}_\tau^{(1)} - \boldsymbol{\eta}_0^{(1)}\}_{\tau\geqslant 0}$ and $\{\boldsymbol{\eta}_\tau^{(2)} - \boldsymbol{\eta}_0^{(2)}\}_{\tau\geqslant 0}$ are independent Brownian motions [69], and $\sigma\big(\boldsymbol{\eta}_{0:t}^{(1)}, \boldsymbol{\eta}_{0:t}^{(2)}\big) = \sigma\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}\big)$ [52]. Using Kalman-Bucy filtering results and (55), $d\mathbf{y}_\tau$ can be written as

$$d\mathbf{y}_\tau = \boldsymbol{A}_\tau \mathbf{y}_\tau \, d\tau + \boldsymbol{E}_\tau^{\mathrm{c}} \sum_{i=1}^{2} \big(\boldsymbol{\Gamma}_\tau^{(i)}\big)^{\mathrm{T}} \Big(\boldsymbol{\Xi}_\tau^{(i)}\big(\boldsymbol{\Xi}_\tau^{(i)}\big)^{\mathrm{T}}\Big)^{-1/2} d\boldsymbol{\eta}_\tau^{(i)} \tag{56}$$

where $\boldsymbol{A}_\tau := \mathrm{diag}\{a_\tau^{(1)}, a_\tau^{(2)}\}$. Combining (55) and (56), we can write $\mathsf{y}_\tau^{(1)}$ and $\mathbf{z}_\tau^{(1)}$ as

$$d\mathsf{y}_\tau^{(1)} = a_\tau^{(1)}\mathsf{y}_\tau^{(1)} \, d\tau + \big(\boldsymbol{h}_\tau^{(1)}\big)^{\mathrm{T}} d\boldsymbol{\eta}_\tau^{(1)} + \big(\boldsymbol{h}_\tau^{(2)}\big)^{\mathrm{T}} d\boldsymbol{\eta}_\tau^{(2)} \tag{57a}$$

$$d\mathbf{z}_\tau^{(1)} = \mathring{\gamma}_\tau^{(1)}\mathsf{y}_\tau^{(1)} \, d\tau + \Big(\boldsymbol{\Xi}_\tau^{(1)}\big(\boldsymbol{\Xi}_\tau^{(1)}\big)^{\mathrm{T}}\Big)^{1/2} d\boldsymbol{\eta}_\tau^{(1)} \tag{57b}$$

where vectors $\boldsymbol{h}_\tau^{(1)}$ and $\boldsymbol{h}_\tau^{(2)}$ are defined as

$$\boldsymbol{h}_\tau^{(1)} := \Big(\boldsymbol{\Xi}_\tau^{(1)}\big(\boldsymbol{\Xi}_\tau^{(1)}\big)^{\mathrm{T}}\Big)^{-1/2} \mathring{\gamma}_\tau^{(1)} \big[\boldsymbol{E}_\tau^{\mathrm{c}}\big]_{1,1} \tag{58a}$$

$$\boldsymbol{h}_\tau^{(2)} := \Big(\boldsymbol{\Xi}_\tau^{(2)}\big(\boldsymbol{\Xi}_\tau^{(2)}\big)^{\mathrm{T}}\Big)^{-1/2} \boldsymbol{\Gamma}_\tau^{(2)} \big[\boldsymbol{E}_\tau^{\mathrm{c}}\big]_{1,:}^{\mathrm{T}} \tag{58b}$$

with $\big[\boldsymbol{E}_\tau^{\mathrm{c}}\big]_{1,:}$ representing the first row of $\boldsymbol{E}_\tau^{\mathrm{c}}$. Using (57), we can derive

$$\frac{d}{d\tau}\tilde{\varepsilon}_t(\tau) = v(\tau, \tilde{\varepsilon}_t(\tau)) \tag{59}$$

where the function $v(\tau, x)$ is defined as

$$\begin{aligned} v(\tau, x) :=& -\Big[x\,\big(\mathring{\gamma}_\tau^{(1)}\big)^{\mathrm{T}} + \big(\boldsymbol{h}_\tau^{(1)}\big)^{\mathrm{T}}\Big(\boldsymbol{\Xi}_\tau^{(1)}\big(\boldsymbol{\Xi}_\tau^{(1)}\big)^{\mathrm{T}}\Big)^{1/2}\Big] \\ &\times \Big(\boldsymbol{\Xi}_\tau^{(1)}\big(\boldsymbol{\Xi}_\tau^{(1)}\big)^{\mathrm{T}}\Big)^{-1} \\ &\times \Big[x\,\big(\mathring{\gamma}_\tau^{(1)}\big)^{\mathrm{T}} + \big(\boldsymbol{h}_\tau^{(1)}\big)^{\mathrm{T}}\Big(\boldsymbol{\Xi}_\tau^{(1)}\big(\boldsymbol{\Xi}_\tau^{(1)}\big)^{\mathrm{T}}\Big)^{1/2}\Big]^{\mathrm{T}} \\ &+ 2a_\tau^{(1)}\, x + \sum_{i=1}^{2}\big(\boldsymbol{h}_\tau^{(i)}\big)^{\mathrm{T}}\boldsymbol{h}_\tau^{(i)}. \end{aligned} \tag{60}$$

Substituting (59) into (54) and using (53) gives

$$\frac{d}{dt}\varepsilon_t(\mu_{0:t}) \geqslant v\big(t, \varepsilon_t(\mu_{0:t})\big) - 2C_t\,\varepsilon_t(\mu_{0:t}) \tag{61}$$

and $\varepsilon_0(\mu_0) = \mathbb{V}\{\mathsf{y}_0^{(1)} \,|\, \mathbf{z}_0^{(1)}\}$ at time 0. Consider function $\breve{\varepsilon}(t)$ that solves the following initial value problem

$$\frac{d}{dt}\breve{\varepsilon}(t) = v(t, \breve{\varepsilon}(t)) - 2C_t\,\breve{\varepsilon}(t) \tag{62a}$$

$$\breve{\varepsilon}(0) = \mathbb{V}\Big\{\mathsf{y}_0^{(1)} \,\Big|\, \mathbf{z}_0^{(1)}\Big\}. \tag{62b}$$

Substitution of (60) into (62a) shows that (62a) is a Riccati equation, and (62) has a unique solution [69]. Comparing (61) with (62) and applying Theorem 4.1 of [70] gives $\varepsilon_T(\mu_{0:T}) \geqslant \breve{\varepsilon}(T)$. This shows that $\breve{\varepsilon}(T)$ is a lower bound on $\varepsilon_T(\mu_{0:T})$ for arbitrary linear encoding strategy $\mu_{0:T}$, as claimed in the beginning of the proof. On the other hand, if the IDE strategy $\mu_{0:T}^{\mathrm{i}}$ is employed, then $\varepsilon_t\big(\mu_{0:t}^{\mathrm{i}}\big)$ can be shown to solve (62). Therefore,

$$\varepsilon_T\big(\mu_{0:T}^{\mathrm{i}}\big) = \breve{\varepsilon}(T) \leqslant \varepsilon_T(\mu_{0:T}).$$

This shows that $\mu_{0:T}^{\mathrm{i}}$ minimizes $\varepsilon_T(\mu_{0:T})$ among all linear encoding strategies of horizon T, thus proving that $\mu_{0:T}^{\mathrm{i}}$ is an OLES for $\mathsf{x}_T^{(1)}$.

Next, it is shown that $\mu_{0:T}^{\mathrm{i}}$ is an OES for $\mathsf{x}_T^{(1)}$ if $\mathring{\gamma}_t^{(1)} = \mathbf{0}$ for $0 \leqslant t \leqslant T$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$. Recall that an OES minimizes $\varepsilon_T(\mu_{0:T})$ among all encoding strategies of horizon T, i.e., designing an OES is equivalent to finding an encoding strategy that minimizes the MSE for inferring $\mathsf{y}_T^{(1)}$ at node 1. The expression of $\mathsf{y}_t^{(1)}$ can be obtained by substituting $\boldsymbol{h}_\tau^{(1)}$ in (57a) with $\mathbf{0}$, since $\mathring{\gamma}_\tau^{(1)} = \mathbf{0}$ in (58). As a result,

$$d\mathsf{y}_\tau^{(1)} = a_\tau^{(1)}\mathsf{y}_\tau^{(1)}\,d\tau + \big(\boldsymbol{h}_\tau^{(2)}\big)^{\mathrm{T}}\,d\boldsymbol{\eta}_\tau^{(2)}. \tag{63}$$

If $\mathring{\gamma}_t^{(1)} = \mathbf{0}$ for all $0 \leqslant t \leqslant T$ and $\boldsymbol{g}_0^{(1)} = \mathbf{0}$, then $\mathbf{z}_{0:T}^{(1)}$ are $\sigma\big(\mathbf{n}_{0:T}^{(1)}, \boldsymbol{\zeta}^{(1)}\big)$-measurable, i.e., observations of sensor 1 consist of only noise. Consequently, these observations are not useful for inferring $\mathsf{x}_t^{(1)}$ and can be omitted in both encoding and inference. Specifically, the MSE $\varepsilon_T(\mu_{0:T})$ becomes

$$\varepsilon_T(\mu_{0:T}) = \mathbb{V}\Big\{\mathsf{y}_T^{(1)} - \mathbb{E}\Big\{\mathsf{y}_T^{(1)} \,|\, \mathsf{r}_{0:T}\Big\}\Big\}. \tag{64}$$

Combining (63) and (64), we observe that an OES should minimize the MSE for inferring the Gaussian process $\mathsf{y}_{0:T}^{(1)}$ using signals $\mathsf{r}_{0:T}$ received via the Gaussian channel. This problem has been studied in [56] and it was shown that the OES is the encoding strategy presented in Definition 3, and thus the proof is complete. Note that the proof uses an equality between mutual information and Fisher information involving processes $\{\mathsf{y}_t^{(1)}\}_{t\geqslant 0}$ and $\{\mathsf{r}_t\}_{t\geqslant 0}$. ⊠

Appendix C
Proof of Proposition 2

Proof: Let $\boldsymbol{\mu}_{0:T}$ represent an arbitrary linear encoding strategy. Define $\boldsymbol{E}_t^{\mathrm{c}}$, $\boldsymbol{E}_t$, and $\boldsymbol{Q}_t$ as in (36). It can be verified that (42) holds. Combining this with (19) and (36), we obtain

$$e_T(\boldsymbol{\mu}_{0:T}) = \mathrm{tr}\Big\{[\boldsymbol{E}_T]_{1:n_1}\Big\} + \mathrm{tr}\Big\{[\boldsymbol{Q}_T]_{1:n_1}\Big\}. \tag{65}$$

We next derive a lower bound on the last term of (65).

For any $0 \leqslant t \leqslant s \leqslant T$, define

$$\tilde{\varepsilon}_t(s) := \det\Big(\mathbb{V}\Big\{\mathbf{y}_s^{(1)} \big| \mathbf{z}_{0:s}^{(1)}, \mathbf{r}_{0:t}\Big\}\Big),\ \varepsilon_t(\boldsymbol{\mu}_{0:t}) := \det\Big([\boldsymbol{Q}_t]_{1:n_1}\Big).$$

In particular, $\tilde{\varepsilon}_t(s)$ and $\varepsilon_t(\mu_{0:t})$ reduce to (46) and (47), respectively, for the special case where $\mathbf{x}_t^{(1)}$ contains only one entry. Using similar arguments as shown in (48)–(53), we can prove that (54) holds, where the capacity C_t is given by

$$C_t := \frac{1}{2}\mathring{P}_t \big/ \min\big\{[\boldsymbol{K}]_{1,1}, [\boldsymbol{K}]_{2,2}, \ldots, [\boldsymbol{K}]_{n_\mathrm{c},n_\mathrm{c}}\big\}.$$

Next, we derive $\frac{d}{d\tau}\tilde{\varepsilon}_t(\tau)\Big|_{\tau=t}$ in (54). To this end, define

$$\tilde{\boldsymbol{E}}_t(s) := \mathbb{V}\left\{\mathbf{y}_s^{(1)} \,|\, \mathbf{z}_{0:s}^{(1)}, \mathbf{r}_{0:t}\right\}.$$

Note that $\tilde{\varepsilon}_t(s) = \det(\tilde{\boldsymbol{E}}_t(s))$. Using similar arguments as shown in (55)–(60), we can derive the expression of $\frac{d}{ds}\tilde{\boldsymbol{E}}_t(s)$. Combining such an expression with the equality $\frac{d}{ds}\tilde{\varepsilon}_t(s) = \det\big(\tilde{\boldsymbol{E}}_t(s)\big)\mathrm{tr}\big\{\tilde{\boldsymbol{E}}_t(s)^{-1}\frac{d}{ds}\,\tilde{\boldsymbol{E}}_t(s)\big\}$, we obtain

$$\begin{aligned} &\frac{d}{d\tau}\tilde{\varepsilon}_t(\tau)\Big|_{\tau=t} \\ &= \mathrm{tr}\bigg\{2\boldsymbol{A}_t^{(1)} + \tilde{\boldsymbol{E}}_t(t)^{-1}\,\big(\boldsymbol{H}_t^{(2)}\big)^{\mathrm{T}}\boldsymbol{H}_t^{(2)} \\ &\quad - 2(\mathring{\boldsymbol{\Gamma}}_t^{(1)})^{\mathrm{T}}\Big(\boldsymbol{\Xi}_t^{(1)}\big(\boldsymbol{\Xi}_t^{(1)}\big)^{\mathrm{T}}\Big)^{-1}\mathring{\boldsymbol{\Gamma}}_t^{(1)}\big[\boldsymbol{E}_t^{\mathrm{c}}\big]_{1:n_1} \\ &\quad - \tilde{\boldsymbol{E}}_t(t)(\mathring{\boldsymbol{\Gamma}}_t^{(1)})^{\mathrm{T}}\Big(\boldsymbol{\Xi}_t^{(1)}\big(\boldsymbol{\Xi}_t^{(1)}\big)^{\mathrm{T}}\Big)^{-1}\mathring{\boldsymbol{\Gamma}}_t^{(1)}\bigg\}\,\varepsilon_t(\boldsymbol{\mu}_{0:t}) \end{aligned} \tag{66}$$

where $\boldsymbol{H}_t^{(2)}$ is defined by replacing τ with t in (58b), and by replacing $\big[\boldsymbol{E}_t^{\mathrm{c}}\big]_{1,:}$ there with the matrix consisting of the top n_1 rows of $\boldsymbol{E}_t^{\mathrm{c}}$.

To proceed, we derive an inequality for the right-hand side of (66). First, note that

$$\tilde{\boldsymbol{E}}_t(t) - \mathbb{V}\Big\{\mathbf{y}_t^{(1)} \,\Big|\, \mathbf{z}_{0:t}^{(1)}\Big\} \preccurlyeq \mathbf{0} \tag{67}$$

where $\boldsymbol{M} \preccurlyeq 0$ indicates that $\boldsymbol{M}$ is negative semidefinite. Second, using inequality between arithmetic and geometric means of eigenvalues for a square matrix, we obtain

$$\mathrm{tr}\Big\{\tilde{\boldsymbol{E}}_t(t)^{-1}\Big\} \geqslant n_1\det\Big(\tilde{\boldsymbol{E}}_t(t)^{-1}\Big)^{1/n_1} = n_1\varepsilon_t(\boldsymbol{\mu}_{0:t})^{-1/n_1}.$$

Therefore,

$$\mathrm{tr}\Big\{\tilde{\boldsymbol{E}}_t(t)^{-1}\big(\boldsymbol{H}_t^{(2)}\big)^{\mathrm{T}}\boldsymbol{H}_t^{(2)}\Big\} \geqslant \underset{\smile}{\lambda}\; n_1\varepsilon_t(\boldsymbol{\mu}_{0:t})^{-1/n_1} \tag{68}$$

$$\frac{d}{dt}I\big(\boldsymbol{\theta}_t;\boldsymbol{\xi}_{0:t}\big)+\frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{B}_t\boldsymbol{B}_t^{\mathrm{T}}\boldsymbol{J}\big(\boldsymbol{\theta}_t;\boldsymbol{\xi}_{0:t}\big)\Big\}=\frac{1}{2}\mathrm{tr}\Big\{\big(\boldsymbol{L}_t\boldsymbol{L}_t^{\mathrm{T}}\big)^{-1}\big[\boldsymbol{G}_t\;\;\boldsymbol{H}_t\big]\Big(\boldsymbol{M}_t-\mathrm{diag}\Big\{\boldsymbol{0},\mathbb{V}\Big\{\boldsymbol{\varphi}_t\,\big|\,\boldsymbol{\theta}_t,\boldsymbol{\xi}_{0:t}\Big\}\Big\}\Big)\big[\boldsymbol{G}_t\;\;\boldsymbol{H}_t\big]^{\mathrm{T}}\Big\}. \tag{74}$$

$$\frac{d}{dt}I\big(\boldsymbol{\theta}_{0:t};\boldsymbol{\xi}_{0:t}\big)=\frac{1}{2}\mathrm{tr}\Big\{\big(\boldsymbol{L}_t\boldsymbol{L}_t^{\mathrm{T}}\big)^{-1}\big[\boldsymbol{G}_t\;\;\boldsymbol{H}_t\big]\Big(\boldsymbol{M}_t-\mathrm{diag}\Big\{\boldsymbol{0},\mathbb{V}\Big\{\boldsymbol{\varphi}_t\,\big|\,\boldsymbol{\theta}_{0:t},\boldsymbol{\xi}_{0:t}\Big\}\Big\}\Big)\big[\boldsymbol{G}_t\;\;\boldsymbol{H}_t\big]^{\mathrm{T}}\Big\}. \tag{75}$$

where $\underset{\smile}{\lambda}$ is the smallest eigenvalue of $\big(\boldsymbol{H}_t^{(2)}\big)^{\mathrm{T}}\boldsymbol{H}_t^{(2)}$, i.e.,

$$\underset{\smile}{\lambda}:=\min\Big\{\lambda\in\mathbb{R}:\exists\boldsymbol{x}\neq\boldsymbol{0}\quad\text{s.t. }\big(\boldsymbol{H}_t^{(2)}\big)^{\mathrm{T}}\boldsymbol{H}_t^{(2)}\boldsymbol{x}=\lambda\boldsymbol{x}\Big\}.$$

Substituting (67) and (68) into (66), using (54) and applying Theorem 4.1 of [70, Chapter 3], we obtain that $\varepsilon_t(\boldsymbol{\mu}_{0:t})\geqslant\breve{\varepsilon}_t$, where $\breve{\varepsilon}_t$ solves the following initial value problem

$$\begin{aligned}\frac{d}{dt}\breve{\varepsilon}_t=\breve{\varepsilon}_t\,\mathrm{tr}\Big\{&2\boldsymbol{A}_t^{(1)}-(\mathring{\boldsymbol{\Gamma}}_t^{(1)})^{\mathrm{T}}\Big(\boldsymbol{\Xi}_t^{(1)}\big(\boldsymbol{\Xi}_t^{(1)}\big)^{\mathrm{T}}\Big)^{-1}\mathring{\boldsymbol{\Gamma}}_t^{(1)}\\&\times\Big(2\big[\boldsymbol{E}_t^{\mathrm{c}}\big]_{1:n_1}+\mathbb{V}\Big\{\mathbf{y}_t^{(1)}\,\big|\,\mathbf{z}_{0:t}^{(1)}\Big\}\Big)\Big\}\\&+\underset{\smile}{\lambda}\;n_1\breve{\varepsilon}_t^{1-1/n_1}-2C_t\breve{\varepsilon}_t\end{aligned}\tag{69a}$$

$$\breve{\varepsilon}_0=\det\Big(\mathbb{V}\Big\{\mathbf{y}_0\,\big|\,\mathbf{z}_0^{(1)}\Big\}\Big). \tag{69b}$$

Therefore, $\mathrm{tr}\big\{[\boldsymbol{Q}_T]_{1:n_1}\big\}\geqslant n_1\varepsilon_T(\boldsymbol{\mu}_{0:T})^{1/n_1}\geqslant n_1\breve{\varepsilon}_T^{1/n_1}$. By substituting this into (65), the desired result (20) is obtained. ⊠

Appendix D
Proof of Proposition 3

Proof: Define matrix $\boldsymbol{M}_t:=\mathbb{V}\Big\{\big[\boldsymbol{\theta}_t^{\mathrm{T}}\quad\boldsymbol{\varphi}_t^{\mathrm{T}}\big]^{\mathrm{T}}\,\big|\,\boldsymbol{\xi}_{0:t}\Big\}$. Note that $\boldsymbol{M}_t$ can be partitioned as

$$\boldsymbol{M}_t=\begin{bmatrix}\boldsymbol{M}_t^{\boldsymbol{\theta}} & \boldsymbol{M}_t^{\boldsymbol{\theta\varphi}}\\ (\boldsymbol{M}_t^{\boldsymbol{\theta\varphi}})^{\mathrm{T}} & \boldsymbol{M}_t^{\boldsymbol{\varphi}}\end{bmatrix}$$

where $\boldsymbol{M}_t^{\boldsymbol{\theta}}$ is a square matrix such that the number of its rows equals the number of entries in $\boldsymbol{\theta}_t$. Since $\boldsymbol{\theta}_t$, $\boldsymbol{\varphi}_t$, and $\boldsymbol{\xi}_{0:t}$ are jointly Gaussian, it holds that

$$\mathbb{V}\Big\{\boldsymbol{\varphi}_t\,\big|\,\boldsymbol{\theta}_t,\boldsymbol{\xi}_{0:t}\Big\}=\boldsymbol{M}_t^{\boldsymbol{\varphi}}-\big(\boldsymbol{M}_t^{\boldsymbol{\theta\varphi}}\big)^{\mathrm{T}}\big(\boldsymbol{M}_t^{\boldsymbol{\theta}}\big)^{-1}\boldsymbol{M}_t^{\boldsymbol{\theta\varphi}} \tag{70}$$

$$I\big(\boldsymbol{\theta}_t;\boldsymbol{\xi}_{0:t}\big)=\frac{1}{2}\ln\frac{\det\big(\mathbb{V}\{\boldsymbol{\theta}_t\}\big)}{\det\big(\boldsymbol{M}_t^{\boldsymbol{\theta}}\big)} \tag{71}$$

$$\boldsymbol{J}\big(\boldsymbol{\theta}_t;\boldsymbol{\xi}_{0:t}\big)=\big(\boldsymbol{M}_t^{\boldsymbol{\theta}}\big)^{-1}-\mathbb{V}\big\{\boldsymbol{\theta}_t\big\}^{-1}. \tag{72}$$

According to results for optimal filtering [56, Chapter 12],

$$\begin{aligned}\frac{d}{dt}\boldsymbol{M}_t^{\boldsymbol{\theta}}=\;&\boldsymbol{A}_t\boldsymbol{M}_t^{\boldsymbol{\theta}}+\big(\boldsymbol{M}_t^{\boldsymbol{\theta}}\big)^{\mathrm{T}}\boldsymbol{A}_t^{\mathrm{T}}+\boldsymbol{B}_t\boldsymbol{B}_t^{\mathrm{T}}\\&-\big(\boldsymbol{M}_t^{\boldsymbol{\theta}}\boldsymbol{G}_t^{\mathrm{T}}+\boldsymbol{M}_t^{\boldsymbol{\theta\varphi}}\boldsymbol{H}_t^{\mathrm{T}}\big)\big(\boldsymbol{L}_t\boldsymbol{L}_t^{\mathrm{T}}\big)^{-1}\\&\times\big(\boldsymbol{M}_t^{\boldsymbol{\theta}}\boldsymbol{G}_t^{\mathrm{T}}+\boldsymbol{M}_t^{\boldsymbol{\theta\varphi}}\boldsymbol{H}_t^{\mathrm{T}}\big)^{\mathrm{T}}.\end{aligned}\tag{73}$$

On the one hand, combining (70)–(73) gives as in (74), shown at the top of the page. On the other hand, applying Girsanov's theorem [71], we obtain as in (75), shown at the top of the page. Note that $\mathbb{V}\big\{\boldsymbol{\varphi}_t\,\big|\,\boldsymbol{\theta}_t,\boldsymbol{\xi}_{0:t}\big\}-\mathbb{V}\big\{\boldsymbol{\varphi}_t\,\big|\,\boldsymbol{\theta}_{0:t},\boldsymbol{\xi}_{0:t}\big\}$ is positive semidefinite. Combining this with (74) and (75) gives the desired inequality (31). Finally, if $\boldsymbol{H}_t=\boldsymbol{0}$, then

$$\begin{aligned}&\big[\boldsymbol{G}_t\;\;\boldsymbol{H}_t\big]\,\mathrm{diag}\Big\{\boldsymbol{0},\mathbb{V}\Big\{\boldsymbol{\varphi}_t\,\big|\,\boldsymbol{\theta}_t,\boldsymbol{\xi}_{0:t}\Big\}\Big\}\big[\boldsymbol{G}_t\;\;\boldsymbol{H}_t\big]^{\mathrm{T}}\\&\quad=\big[\boldsymbol{G}_t\;\;\boldsymbol{H}_t\big]\,\mathrm{diag}\Big\{\boldsymbol{0},\mathbb{V}\Big\{\boldsymbol{\varphi}_t\,\big|\,\boldsymbol{\theta}_{0:t},\boldsymbol{\xi}_{0:t}\Big\}\Big\}\big[\boldsymbol{G}_t\;\;\boldsymbol{H}_t\big]^{\mathrm{T}}=\boldsymbol{0}\,.\end{aligned}$$

Combining this relationship with (74) and (75), we conclude that equality in (31) holds if $\boldsymbol{H}_t=\boldsymbol{0}$. ⊠

Acknowledgment

The authors wish to thank M. Clouâtré and A. Giani for their helpful suggestions and careful reading of the manuscript.

References

[1] M. Afrin, J. Jin, A. Rahman, A. Gasparri, Y.-C. Tian, and A. Kulkarni, "Robotic edge resource allocation for agricultural cyber-physical system," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 6, pp. 3979–3990, Nov. 2022.

[2] J. Jin, K. Yu, J. Kua, N. Zhang, Z. Pang, and Q.-L. Han, "Cloud–fog automation: Vision, enabling technologies, and future research directions," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 1039–1054, Feb. 2024.

[3] H. Lyu, J. Yan, J. Zhang, Z. Pang, G. Yang, and A. J. Isaksson, "Cloud–fog automation: Heterogeneous applications over new-generation infrastructure of virtualized computing and converged networks," *IEEE Ind. Electron. Mag.*, vol. 18, no. 4, pp. 30–42, Dec. 2024.

[4] S. Mohanti, D. Roy, M. Eisen, D. Cavalcanti, and K. Chowdhury, "L-NORM: Learning and network orchestration at the edge for robot connectivity and mobility in factory floor environments," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, pp. 2898–2914, Apr. 2024.

[5] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "A hybrid stochastic game for secure control of cyber-physical systems," *Automatica*, vol. 93, pp. 55–63, Jul. 2018.

[6] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 106–117, Mar. 2017.

[7] Q. Zhu and T. Başar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 46–65, Feb. 2015.

[8] N. Marchenko, T. Andre, G. Brandner, W. Masood, and C. Bettstetter, "An experimental study of selective cooperative relaying in industrial wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 10, no. 3, pp. 1806–1816, Aug. 2014.

[9] M. Z. Win et al., "Network localization and navigation via cooperation," *IEEE Commun. Mag.*, vol. 49, no. 5, pp. 56–62, May 2011.

[10] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.

[11] A. Conti et al., "Location awareness in beyond 5G networks," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 22–27, Nov. 2021.

[12] T. E. Duncan, "On the calculation of mutual information," *SIAM J. Appl. Math.*, vol. 19, no. 1, pp. 215–220, Jul. 1970.

[13] T. E. Duncan, "Mutual information for stochastic differential equations," *Inf. Control*, vol. 19, no. 3, pp. 265–271, Oct. 1971.

[14] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[15] S. K. Mitter and N. J. Newton, "Information and entropy flow in the Kalman–Bucy filter," *J. Stat. Phys.*, vol. 118, nos. 1–2, pp. 145–176, Jan. 2005.

[16] N. J. Newton, "Interactive statistical mechanics and nonlinear filtering," *J. Stat. Phys.*, vol. 133, no. 4, pp. 711–737, Sep. 2008.

[17] C. D. Charalambous and F. Rezaei, "Stochastic uncertain systems subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games," *IEEE Trans. Autom. Control*, vol. 52, no. 4, pp. 647–663, Apr. 2007.

[18] J. Wu, G. Shi, B. D. O. Anderson, and K. H. Johansson, "Kalman filtering over Gilbert–Elliott channels: Stability conditions and critical curve," *IEEE Trans. Autom. Control*, vol. 63, no. 4, pp. 1003–1017, Apr. 2018.

[19] S. Dey, A. Chiuso, and L. Schenato, "Remote estimation with noisy measurements subject to packet loss and quantization noise," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 3, pp. 204–217, Sep. 2014.

[20] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link—Part I: Scalar systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3369–3395, Aug. 2006.

[21] Z. Liu, A. Conti, S. K. Mitter, and M. Z. Win, "Filtering over non-Gaussian channels: The role of anytime capacity," *IEEE Control Syst. Lett.*, vol. 7, pp. 472–477, 2023.

[22] Z. Liu, A. Conti, S. K. Mitter, and M. Z. Win, "Communication-efficient distributed learning over networks—Part I: Sufficient conditions for accuracy," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 4, pp. 1081–1101, Apr. 2023.

[23] Z. Liu, A. Conti, S. K. Mitter, and M. Z. Win, "Communication-efficient distributed learning over networks—Part II: Necessary conditions for accuracy," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 4, pp. 1102–1119, Apr. 2023.

[24] F. Zabini and A. Conti, "Inhomogeneous Poisson sampling of finite-energy signals with uncertainties in $\mathbb{R}^d$," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4679–4694, Sep. 2016.

[25] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[26] S. Kar, J. M. Moura, and H. V. Poor, "Distributed linear parameter estimation: Asymptotically efficient adaptive strategies," *SIAM J. Control Optim.*, vol. 51, no. 3, pp. 2200–2229, Jan. 2013.

[27] X. He, X. Ren, H. Sandberg, and K. H. Johansson, "How to secure distributed filters under sensor attacks," *IEEE Trans. Autom. Control*, vol. 67, no. 6, pp. 2843–2856, Jun. 2022.

[28] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.

[29] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[30] R. Abdolee, S. Saur, B. Champagne, and A. H. Sayed, "Diffusion LMS localization and tracking algorithm for wireless cellular networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 4598–4602.

[31] P. Addesso, S. Marano, and V. Matta, "Estimation of target location via likelihood approximation in sensor networks," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1358–1368, Mar. 2010.

[32] M. Z. Win, Y. Shen, and W. Dai, "A theoretical foundation of network localization and navigation," *Proc. IEEE*, vol. 106, no. 7, pp. 1136–1165, Jul. 2018.

[33] A. Conti, S. Mazuelas, S. Bartoletti, W. C. Lindsey, and M. Z. Win, "Soft information for localization-of-things," *Proc. IEEE*, vol. 107, no. 11, pp. 2240–2264, Sep. 2019.

[34] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 5, pp. 679–694, Sep. 1989.

[35] T. Başar and R. Bansal, "Optimum design of measurement channels and control policies for linear-quadratic stochastic systems," *Eur. J. Oper. Res.*, vol. 73, no. 2, pp. 226–236, Mar. 1994.

[36] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.

[37] T. Tanaka, K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1896–1910, Apr. 2017.

[38] A. Mahajan and D. Teneketzis, "Optimal design of sequential real-time communication systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5317–5338, Nov. 2009.

[39] D. Teneketzis, "On the structure of optimal real-time encoders and decoders in noisy communication," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4017–4035, Sep. 2006.

[40] V. S. Borkar and S. K. Mitter, "Markov control problems under communication constraints," *Commun. Inf. Syst.*, vol. 1, no. 1, pp. 15–32, 2001.

[41] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1056–1068, Jul. 2004.

[42] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proc. IEEE*, vol. 95, no. 1, pp. 108–137, Jan. 2007.

[43] P. Minero, M. Franceschetti, S. Dey, and G. N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 243–255, Feb. 2009.

[44] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. New York, NY, USA: Springer, 2013.

[45] M. Franceschetti, M. J. Khojasteh, and M. Z. Win, "The many facets of information in networked estimation and control," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 6, no. 1, pp. 233–259, 2023.

[46] R. Fox and N. Tishby, "Minimum-information LQG control—Part I: Memoryless controllers," in *Proc. IEEE Conf. Decis. Control*, Las Vegas, NV, USA, Dec. 2016, pp. 5610–5616.

[47] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4525–4540, Nov. 2019.

[48] T. Tanaka, P. M. Esfahani, and S. K. Mitter, "LQG control with minimum directed information: Semidefinite programming approach," *IEEE Trans. Autom. Control*, vol. 63, no. 1, pp. 37–52, Jan. 2018.

[49] J. Pearson, J. P. Hespanha, and D. Liberzon, "Control with minimal cost-per-symbol encoding and quasi-optimality of event-based encoders," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2286–2301, May 2017.

[50] C. D. Charalambous, A. Farhadi, and S. Z. Denic, "Control of continuous-time linear Gaussian systems over additive Gaussian wireless fading channels: A separation principle," *IEEE Trans. Autom. Control*, vol. 53, no. 4, pp. 1013–1019, May 2008.

[51] I. Karatzas and S. E. Shreve, *Brownian Motion Stochastic Calculus*, 2nd ed., New York, NY, USA: Springer, 1998.

[52] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes I—General Theory*, 2nd ed., Heidelberg, Germany: Springer, 2001.

[53] B. Øksendal, *Stochastic Differential Equations*, 6th ed., Heidelberg, Germany: Springer, 2013.

[54] C. Lu et al., "Real-time wireless sensor-actuator networks for industrial cyber-physical systems," *Proc. IEEE*, vol. 104, no. 5, pp. 1013–1024, May 2016.

[55] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proc. IEEE*, vol. 100, no. SpecialCentennial Issue, pp. 1287–1308, May 2012.

[56] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes II—Applications*, 2nd ed., Heidelberg, Germany: Springer, 2001.

[57] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Fluids Eng.*, vol. 83, no. 1, pp. 95–108, Mar. 1961.

[58] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in Mathematical System Theory*. New York, NY, USA: McGraw-Hill, 1969.

[59] Z. Liu, A. Conti, S. K. Mitter, and M. Z. Win, "Networked filtering with feedback for discrete-time observations," in *Proc. IEEE Conf. Decis. Control* , Austin, TX, USA, Dec. 2021, pp. 5882–5889.

[60] S. Yüksel and S. Tatikonda, "A counterexample in distributed optimal sensing and control," *IEEE Trans. Autom. Control*, vol. 54, no. 4, pp. 841–844, Apr. 2009.

[61] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM J. Control*, vol. 6, no. 1, pp. 131–147, 1968.

[62] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proc. IEEE*, vol. 95, no. 1, pp. 163–187, Jan. 2007.

[63] Y.-H. Kim, A. Lapidoth, and T. Weissman, "Error exponents for the Gaussian channel with active noisy feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1223–1236, Mar. 2011.

[64] M. J. Khojasteh, P. Tallapragada, J. Cortés, and M. Franceschetti, "The value of timing information in event-triggered control," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 925–940, Mar. 2020.

[65] D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations," *SIAM Rev.*, vol. 43, no. 3, pp. 525–546, Jan. 2001.

[66] Z. Liu, A. Conti, S. K. Mitter, and M. Z. Win, "Continuous-time distributed filtering with sensing and communication constraints," *IEEE J. Sel. Areas Inf. Theory*, vol. 4, pp. 667–681, 2023.

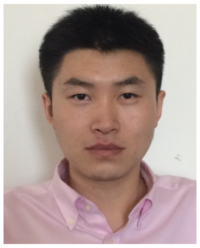
[67] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. Hoboken, NJ, USA: Wiley, Jul. 2001.

[68] Z. Liu, "Decentralized inference and its application to network localization and navigation," Ph.D. dissertation, Dept. Aeronaut. Astronaut., Massachusetts Inst. Technol., Cambridge, MA, USA, May 2022.

[69] M. H. Davis, *Linear Estimation Stochastic Control*, 1st ed., London, U.K.: Chapman & Hall, 1977.

[70] P. Hartman, *Ordinary Differential Equations*, 2nd ed., Philadelphia, PA, USA: SIAM, 2002.

[71] J.-F. Le Gall, *Brownian Motion, Martingales, Stochastic Calculus*. Cham, Switzerland: Springer, 2016.



Zhenyu Liu (Member, IEEE) is an Assistant Professor at Tsinghua Shenzhen International Graduate School, Tsinghua University, China. He received the Ph.D. degree in networks and statistics and the S.M. degree in aeronautics and astronautics both from the Massachusetts Institute of Technology (MIT) in 2022. He received the M.S. degree and the B.S. degree (with honor) in electronic engineering from Tsinghua University, China.

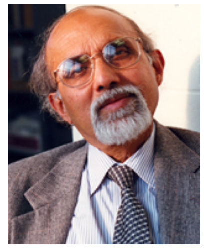
Dr. Liu was a Post-Doctoral Associate in the Wireless Information and Network Sciences Laboratory at MIT from 2022 to 2023. His research interests include wireless communications, network localization, distributed inference, networked control, and quantum information science.

Dr. Liu received the first prize of the IEEE Communications Society's Student Competition in 2019 and 2016, the Research and Development 100 Award for Peregrine System in 2018, and the Best Paper Award at the IEEE Latin-American Conference on Communications in 2017.

Andrea Conti (Fellow, IEEE) is a Professor and founding director of the Wireless Communication and Localization Networks Laboratory at the University of Ferrara, Italy. Prior to joining the University of Ferrara, he was with CNIT and with IEIIT-CNR.

In Summer 2001, he was with the Wireless Systems Research Department at AT&T Research Laboratories. Since 2003, he has been a frequent visitor to the Wireless Information and Network Sciences Laboratory at the Massachusetts Institute of Technology, where he presently holds the Research Affiliate appointment. His research interests involve theory and experimentation of wireless communication and localization systems. His current topics include network localization and navigation, distributed sensing, adaptive diversity communications, and quantum information science.

Dr. Conti has served as editor for IEEE journals and chaired international conferences. He was elected Chair of the IEEE Communications Society's Radio Communications Technical Committee and is Co-founder of the IEEE Quantum Communications & Information Technology Emerging Technical Subcommittee. He received the HTE Puskás Tivadar Medal, the IEEE Communications Society's Fred W. Ellersick Prize, and the IEEE Communications Society's Stephen O. Rice Prize in the field of Communications Theory. He is an elected Fellow of the IEEE and of the IET, and a member of Sigma Xi. He has been selected as an IEEE Distinguished Lecturer.



Sanjoy K. Mitter (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering (automatic control) from the Imperial College London, London, U.K., in 1965.

He taught at Case Western Reserve University from 1965 to 1969. He joined the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1969, where he has been a Professor of electrical engineering since 1973. He was the Director of the MIT Laboratory for Information and Decision Systems from 1981 to 1999. He has also been a Professor of mathematics at the Scuola Normale, Pisa, Italy, from 1986 to 1996. He has held visiting positions at Imperial College London; University of Groningen, The Netherlands; INRIA, France; Tata Institute of Fundamental Research, India; ETH, Zurich, Switzerland; and several American universities. He was the McKay Professor at the University of California, Berkeley, CA, USA, in March 2000, and held the Russell-Severance-Springer Chair in Fall 2003. His current research interests include communication and control in a networked environment, the relationship of statistical and quantum physics to information theory and control, and autonomy and adaptiveness for integrative organization.

Dr. Mitter received the AACC Richard E. Bellman Control Heritage Award in 2007 and the IEEE Eric E. Sumner Award in 2015. He is a member of the National Academy of Engineering. He has received the 2000 IEEE Control Systems Award.

Moe Z. Win (Fellow, IEEE) is the Robert R. Taylor Professor at the Massachusetts Institute of Technology (MIT) and the founding director of the Wireless Information and Network Sciences Laboratory. Prior to joining MIT, he was with AT&T Research Laboratories and with the NASA Jet Propulsion Laboratory.

His research encompasses fundamental theories, algorithm design, and network experimentation for a broad range of real-world problems. His current research topics include ultra-wideband systems, network localization and navigation, network interference exploitation, and quantum information science. He has served the IEEE Communications Society as an elected Member-at-Large on the Board of Governors, as elected Chair of the Radio Communications Committee, and as an IEEE Distinguished Lecturer. Over the last two decades, he held various editorial positions for IEEE journals and organized numerous international conferences. He has served on the SIAM Diversity Advisory Committee.

Dr. Win is an elected Fellow of the AAAS, the EURASIP, the IEEE, and the IET. He was honored with two IEEE Technical Field Awards: the IEEE Kiyo Tomiyasu Award (2011) and the IEEE Eric E. Sumner Award (2006, jointly with R. A. Scholtz). His publications, co-authored with students and colleagues, have received several awards. Other recognitions include the MIT Frank E. Perkins Award (2024), the MIT Everett Moore Baker Award (2022), the IEEE Vehicular Technology Society James Evans Avant Garde Award (2022), the IEEE Communications Society Edwin H. Armstrong Achievement Award (2016), the Cristoforo Colombo International Prize for Communications (2013), the Copernicus Fellowship (2011) and the *Laurea Honoris Causa* (2008) from the Università degli Studi di Ferrara, and the U.S. Presidential Early Career Award for Scientists and Engineers (2004).