Continuous-Time Distributed Filtering With Sensing and Communication Constraints

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

***Abstract*—Distributed filtering is crucial in many applications such as localization, radar, autonomy, and environmental monitoring. The aim of distributed filtering is to infer time-varying unknown states using data obtained via sensing and communication in a network. This paper analyzes continuous-time distributed filtering with sensing and communication constraints. In particular, the paper considers a building-block system of two nodes, where each node is tasked with inferring a time-varying unknown state. At each time, the two nodes obtain noisy observations of the unknown states via sensing and perform communication via a Gaussian feedback channel. The distributed filter of the unknown state is computed based on both the sensor observations and the received messages. We analyze the asymptotic performance of the distributed filter by deriving a necessary and sufficient condition of the sensing and communication capabilities under which the mean-square error of the distributed filter is bounded over time. Numerical results are presented to validate the derived necessary and sufficient condition.**

***Index Terms*—Distributed inference, Kalman–Bucy filter, channel capacity, stochastic differential equation.**

I. Introduction

INFERENCE of time-varying states, also referred to as filtering [1], [2], [3], is critical in various applications including localization and tracking [4], [5], [6], [7], [8], [9], autonomy [10], [11], [12], [13], Internet-of-Things [14], [15], [16], [17], and beyond 5G networks [18], [19], [20], [21]. In several network applications, it is preferable to perform filtering in a distributed manner. The accuracy of distributed filtering is affected by the sensing and communication capabilities of nodes in the network. A deep understanding of such effects is important for the efficient management of wireless resources in the network [22], [23], [24], [25].

Manuscript received 10 October 2022; revised 22 March 2023 and 26 June 2023; accepted 28 July 2023. Date of publication 10 August 2023; date of current version 11 December 2023. The fundamental research described in this paper was supported, in part, by the Office of Naval Research under Grant N62909-22-1-2009, by the Army Research Office through the MIT Institute for Soldier Nanotechnologies under Contract W911NF-13-D-0001, by the National Science Foundation under Grant CNS-2148251, and by federal agency and industry partners in the RINGS Program. *(Corresponding author: Moe Z. Win.)*

Zhenyu Liu is with the Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

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/JSAIT.2023.3304249

Theoretical studies [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38] and efficient algorithms [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51] for filtering have been studied in the literature. In particular, the boundedness of inference error over time is studied in [36], [37], [38]. The Kalman–Bucy filter [52], [53], [54] is investigated in [26] from an information-theoretical perspective. Specifically, a fundamental relationship between Shannon information and Fisher information is derived therein, and an analogy of the filter to a statistical mechanical system is established. Those results have been extended for nonlinear filtering in [27]. Distributed filtering is closely related to distributed control problems [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75], [76], [77], where control signals for stabilizing a dynamical system are generated in real time based on data received via a channel with communication constraints. In particular, the notion of anytime capacity is introduced in [60] for characterizing the channel quality in distributed control problems. This notion is then applied for investigating distributed filtering problems [78], [79], [80].

This paper analyzes distributed filtering in continuous-time scenarios. Specifically, a building-block system with two nodes is considered where each node is tasked with inferring a time-varying unknown state. At every time, each node obtains noisy sensor observations of both unknown states. Moreover, one node transmits encoded messages containing information of the unknown state that the other node aims to infer via a Gaussian feedback channel. The node on the receiving end of the channel performs filtering to infer its unknown state using both its sensor observations and its received messages. This paper aims to establish conditions under which the mean-square error (MSE) of the distributed filter is bounded over time. Key contributions of this paper can be summarized as the following; specifically, we

- derive a necessary and sufficient condition on sensing and communication capabilities under which the MSE of the distributed filter is bounded over time;
- establish an analogy between distributed filtering and statistical mechanical systems, as well as derive the evolution of energy and entropy of the system; and
- characterize the relationship between the accuracy of distributed filtering and the capabilities of sensing and communication.

The remaining sections are organized as follows. Section II describes the system model. Section III derives conditions for

TABLE I
NOTATION AND DEFINITIONS OF IMPORTANT QUANTITIES

Notation	Definition	Notation	Definition
$\mathsf{x}_t^{(i)}$	state of node i at time t	$A^{(i)}$	scalar that determines the evolution of $\mathsf{x}_t^{(i)}$
$\mathbf{v}_t^{(i)}$	random vector corresponding to disturbances to the state of node i	$\mathbf{z}_t^{(i)}$	observation obtained by the sensor of node i at time t
$\boldsymbol{\Gamma}^{(i)}$	sensor gain matrix for observations obtained by node i	$\mathbf{n}_t^{(i)}$	random vector corresponding to noise in the observations obtained by node i
m_t	message transmitted from node 2 to node 1 at time t	r_t	message received by node 1 from node 2 at time t
μ_t	encoding function employed for generating m_t	$\mu_{0:T}$	encoding strategy of horizon T consisting of encoding functions μ_t for $t \in [0, T]$
P	constraint on the transmit power	w_t	random variable corresponding to noise in the communication channel
κ	scalar determining the power of noise in the communication channel	C	capacity of 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
$\breve{e}_T$	infimum of the MSE for the distributed filter at time T over all encoding strategies	μ_t^{P}	encoding function at time t in the proposed encoding strategy
$\boldsymbol{A}$	diagonal matrix with $A^{(1)}$ and $A^{(2)}$ on the diagonal	$\boldsymbol{\Gamma}$	vertical concatenation of $\boldsymbol{\Gamma}^{(1)}$ and $\boldsymbol{\Gamma}^{(2)}$
$\mathbf{x}_t$	vertical concatenation of $\mathsf{x}_t^{(1)}$ and $\mathsf{x}_t^{(2)}$	$\mathbf{y}_t$	centralized minimum-mean-square-error estimator of $\mathbf{x}_t$ using $\mathbf{z}_{0:t}^{(1)}$ and $\mathbf{z}_{0:t}^{(2)}$
$\hat{\mathbf{x}}_t$	distributed minimum-mean-square-error estimator of $\mathbf{x}_t$ using $\mathbf{z}_{0:t}^{(1)}$ and $\mathsf{r}_{0:t}$	$\mathbf{s}_t$	vertical concatenation of $\mathbf{x}_t$, $\mathbf{y}_t$, and $\hat{\mathbf{x}}_t$
$\tilde{\mathbf{x}}_t$	conditional state defined as the difference between $\mathbf{x}_t$ and $\mathbf{y}_t$	$\tilde{\mathbf{y}}_t$	conditional estimator defined as the difference between $\mathbf{y}_t$ and $\hat{\mathbf{x}}_t$

the boundedness of MSE in distributed filtering. Section IV establishes an analogy of the distributed filtering problem to a statistical mechanics system. Section V provides numerical results. Section VI 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. For example, a random variable and its realization are denoted by x and x; a random vector and its realization are denoted by $\mathbf{x}$ and $\boldsymbol{x}$, 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, and the column space of $\boldsymbol{A}$ are denoted by $\boldsymbol{A}^{\mathrm{T}}$, $\mathrm{tr}\{\boldsymbol{A}\}$, and $\mathcal{C}(\boldsymbol{A})$, respectively. Notation $\mathrm{diag}\{\boldsymbol{A}_1, \boldsymbol{A}_2\}$ represents a block diagonal matrix with $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ being its diagonal blocks from top left to bottom right. All random quantities in this paper are defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$, unless otherwise mentioned, where Ω is a non-empty set, $\mathscr{F}$ is a σ-algebra over Ω, and $\mathbb{P}$ is a probability measure on the measurable space $(\Omega, \mathscr{F})$. The probability of $\mathcal{A} \in \mathscr{F}$ is denoted by $\mathbb{P}\{\mathcal{A}\}$. Notation $\sigma(\cdot)$ represents the σ-algebra generated by the random quantities (e.g., a random vector or a collection of random vectors) in the parentheses. The distribution of random vector $\mathbf{x}$ is denoted by $P_{\mathbf{x}}$. The Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The expectation and covariance matrix of a random vector $\mathbf{x}$ are denoted by $\mathbb{E}\{\mathbf{x}\}$ and $\mathbb{V}\{\mathbf{x}\}$, respectively. The cross-covariance matrix of random vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\mathbb{V}\{\mathbf{x},\mathbf{y}\} := \mathbb{E}\{(\mathbf{x} - \mathbb{E}\{\mathbf{x}\})(\mathbf{y} - \mathbb{E}\{\mathbf{y}\})^{\mathrm{T}}\}$. The probability density function of $\mathbf{x}$ and the conditional probability density function of $\mathbf{x}$ given $\mathbf{y}$ are denoted by $f_{\mathbf{x}}(\boldsymbol{x})$ and $f_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y})$, respectively. The conditional expectation of $\mathbf{x}$ given $\mathscr{F}_1 \subseteq \mathscr{F}$ is denoted by $\mathbb{E}\{\mathbf{x} \,|\, \mathscr{F}_1\}$. If $\mathscr{F}_1$ is the sub-σ-algebra generated by a collection of random vectors $\{\mathbf{x}_\alpha\}_{\alpha\in\mathcal{A}}$, such conditional expectation is also denoted by $\mathbb{E}\{\mathbf{x} \,\big|\, \{\mathbf{x}_\alpha\}_{\alpha\in\mathcal{A}}\}$. The conditional cross-covariance matrix of random vectors $\mathbf{x}$ and $\mathbf{y}$ given $\mathscr{F}_1 \subseteq \mathscr{F}$ is denoted by $\mathbb{V}\{\mathbf{x},\mathbf{y} \,\big|\, \mathscr{F}_1\} := \mathbb{E}\{(\mathbf{x} - \mathbb{E}\{\mathbf{x} \,|\, \mathscr{F}_1\})(\mathbf{y} - \mathbb{E}\{\mathbf{y} \,|\, \mathscr{F}_1\})^{\mathrm{T}} \,\big|\, \mathscr{F}_1\}$, and $\mathbb{V}\{\mathbf{x} \,|\, \mathscr{F}_1\}$ is a short notation for $\mathbb{V}\{\mathbf{x},\mathbf{x} \,\big|\, \mathscr{F}_1\}$. The relationship that sub-σ-algebras $\mathscr{F}_1$ and $\mathscr{F}_2$ are independent conditioned on $\mathscr{F}_3$ is denoted by $\mathscr{F}_1 \perp\!\!\!\perp \mathscr{F}_2 \,|\, \mathscr{F}_3$. If $\mathscr{F}_i$ is the sub-σ-algebra generated by a collection of random vectors $\{\mathbf{x}_\alpha\}_{\alpha\in\mathcal{A}_i}$ for $i = 1, 2, 3$, such conditional independence is also denoted by $\{\mathbf{x}_\alpha\}_{\alpha\in\mathcal{A}_1} \perp\!\!\!\perp \{\mathbf{x}_\alpha\}_{\alpha\in\mathcal{A}_2} \,|\, \{\mathbf{x}_\alpha\}_{\alpha\in\mathcal{A}_3}$. If $\mathscr{F}_3$ is the trivial σ-algebra $\{\varnothing, \Omega\}$, such conditional independence is denoted by $\mathscr{F}_1 \perp\!\!\!\perp \mathscr{F}_2$. 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 important quantities used in the paper are summarized in Table I.

II. SYSTEM MODEL

Consider a system consisting of two nodes, where each node is associated with a time-varying unknown state. Each node has a sensor that generates a noisy observation of both unknown states at every time. The two nodes also communicate with each other: node 2 transmits encoded messages to node 1 via a noisy channel and receives noiseless feedback from node 1 (see Fig. 1). The aim of node 1 is to infer its unknown state using observations obtained by its own sensor as well as the messages received from node 2.

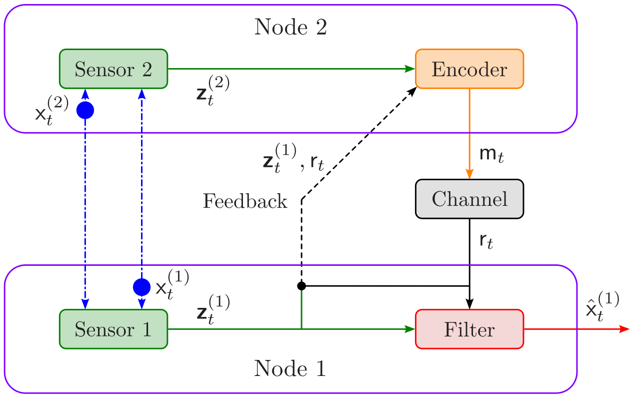


Fig. 1. Distributed filtering in a two-node system: each node $i \in \{1, 2\}$ obtains a sensor observation $\mathbf{z}_t^{(i)}$ of the unknown states associated with the two nodes at time t.

The state and observation of node i at time $t \geqslant 0$ are denoted by $\mathsf{x}_t^{(i)}$ and $\mathbf{z}_t^{(i)}$, respectively, for $i = 1, 2$. In particular, the state process $\{\mathsf{x}_t^{(i)}\}_{t\geqslant 0}$ is described by the following stochastic differential equation (SDE) [81], [82], [83]

$$d\mathsf{x}_t^{(i)} = A^{(i)}\mathsf{x}_t^{(i)}\,dt + \boldsymbol{B}^{(i)}\,d\mathbf{v}_t^{(i)} \qquad \forall t \in [0, \infty) \tag{1}$$

where $A^{(i)}$ is a scalar and $\boldsymbol{B}^{(i)}$ is a row vector. Both $A^{(i)}$ and $\boldsymbol{B}^{(i)}$ are deterministic quantities known to both nodes. Process $\{\mathbf{v}_t^{(i)}\}_{t\geqslant 0}$ is a Brownian motion and represents disturbance to the state of node i. The initial states $\mathsf{x}_0^{(1)}$ and $\mathsf{x}_0^{(2)}$ are zero-mean Gaussian random variables. The observation process $\{\mathbf{z}_t^{(i)}\}_{t\geqslant 0}$ satisfies

$$d\mathbf{z}_t^{(i)} = \boldsymbol{\Gamma}^{(i)}\left[\mathsf{x}_t^{(1)} \quad \mathsf{x}_t^{(2)}\right]^{\mathrm{T}} dt + \boldsymbol{\varXi}^{(i)}\,d\mathbf{n}_t^{(i)} \quad \forall t \in [0, \infty) \tag{2}$$

where $\boldsymbol{\Gamma}^{(i)}$ and $\boldsymbol{\varXi}^{(i)}$ are deterministic matrices known to both nodes. Process $\{\mathbf{n}_t^{(i)}\}_{t\geqslant 0}$ is a Brownian motion and represents noise in the sensor observations. At time 0, observation $\mathbf{z}_0^{(i)}$ is given by $\mathbf{z}_0^{(i)} = \boldsymbol{G}^{(i)}\left[\mathsf{x}_0^{(1)} \quad \mathsf{x}_0^{(2)}\right]^{\mathrm{T}} + \boldsymbol{\zeta}^{(i)}$ where $\boldsymbol{G}^{(i)}$ is a deterministic matrix known to both nodes, and $\boldsymbol{\zeta}^{(i)}$ is a zero-mean Gaussian random vector with invertible covariance matrix. We consider scenarios where $\boldsymbol{B}^{(i)}$ is non-zero and $\boldsymbol{\varXi}^{(i)}\left(\boldsymbol{\varXi}^{(i)}\right)^{\mathrm{T}}$ is invertible.

At each time t, node 2 transmits an encoded message $\mathsf{m}_t \in \mathbb{R}$ to node 1 via a Gaussian feedback channel. The message received by node 1 at time t is denoted by r_t. In particular, m_t is generated by node 2 based on its sensor observations $\mathbf{z}_{0:t}^{(2)}$ as well as node 1's sensor observations $\mathbf{z}_{0:t}^{(1)}$ and received messages $\mathsf{r}_{0:t}$ up to time t. Consequently, m_t can be written as $\mathsf{m}_t = \mu_t\left(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\right)$, where the real function μ_t is referred to as the encoding function at time t. A collection of encoding functions $\mu_{0:T} := \{\mu_t\}_{t\in[0,T]}$ from time 0 to time T is referred to as an encoding strategy of horizon T if the following constraint on transmit power is satisfied

$$\mathbb{E}\left\{\mu_t\left(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\right)^2\right\} \leqslant P \qquad \forall t \in [0, T] \tag{3}$$

where P is a constant representing the power constraint. The set of encoding strategies of horizon T is denoted by $\mathcal{M}_T$.

The process of the received messages satisfies

$$d\mathsf{r}_t = \mathsf{m}_t\,dt + \kappa\,d\mathsf{w}_t \qquad \mathsf{r}_0 = 0 \tag{4}$$

where $\{\mathsf{w}_t\}_{t\geqslant 0}$ is one-dimensional Brownian motion, which represents additive Gaussian noise in the channel, and κ determines the power of noise. The capacity of this continuous-time Gaussian channel is [84, Chapter 16]

$$C := P/\left(2\kappa^2\right). \tag{5}$$

Node 1 aims to infer its unknown state in real time based on its own sensor observations and the messages received from node 2. Specifically, node 1 computes an estimator of $\mathsf{x}_t^{(1)}$ at time t based on $\mathbf{z}_{0:t}^{(1)}$ and $\mathsf{r}_{0:t}$. This estimator is thus $\sigma\left(\mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\right)$-measurable. For an arbitrary encoding strategy $\mu_{0:t}$ employed by node 2, the minimum-mean-square-error (MMSE) estimator $\hat{\mathsf{x}}_t^{(1)}$ of $\mathsf{x}_t^{(1)}$ at node 1 is

$$\hat{\mathsf{x}}_t^{(1)} := \mathbb{E}\left\{\mathsf{x}_t^{(1)} \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\right\}.$$

This estimator is referred to as the distributed filter and its MSE $e_t(\mu_{0:t})$ is given by

$$e_t\left(\mu_{0:t}\right) := \mathbb{E}\left\{\left(\mathsf{x}_t^{(1)} - \mathbb{E}\left\{\mathsf{x}_t^{(1)} \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\right\}\right)^2\right\}.$$

The MSE $e_t(\mu_{0:t})$ is affected by the encoding strategy employed by node 2. Define $\breve{e}_T$ as the infimum of the MSE for the distributed filter at time T over all encoding strategies that satisfy the power constraints (3), i.e.,

$$\breve{e}_T := \inf_{\mu_{0:T}\in\mathcal{M}_T} e_T\left(\mu_{0:T}\right).$$

The next section studies conditions under which $\{\breve{e}_T\}_{T\geqslant 0}$ is bounded.

III. Boundedness of MSE for Distributed Filtering

The section presents a necessary and sufficient condition for $\{\breve{e}_T\}_{T\geqslant 0}$ to be bounded. Before presenting this condition, some definitions are introduced. First, given an n-by-n real matrix $\boldsymbol{F}$ and a real matrix $\boldsymbol{C}$ with n columns, define the unobservable subspace for $(\boldsymbol{C}, \boldsymbol{F})$ as the kernel of the observability matrix $\boldsymbol{O}(\boldsymbol{C}, \boldsymbol{F}) := \left[\boldsymbol{C}^{\mathrm{T}} \quad \boldsymbol{F}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}} \quad \ldots \quad (\boldsymbol{F}^{n-1})^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\right]^{\mathrm{T}}$. In other words, this unobservable subspace is given by $\{\boldsymbol{x} : \boldsymbol{O}(\boldsymbol{C}, \boldsymbol{F})\boldsymbol{x} = \boldsymbol{0}\}$. Second, define

$$\boldsymbol{A} := \mathrm{diag}\left\{A^{(1)}, A^{(2)}\right\} \tag{6a}$$

$$\boldsymbol{\Gamma} := \left[\left(\boldsymbol{\Gamma}^{(1)}\right)^{\mathrm{T}} \quad \left(\boldsymbol{\Gamma}^{(2)}\right)^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{6b}$$

Third, define an encoding function μ_t^{p} at time t as

$$\mu_t^{\mathrm{p}}\left(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\right) = \alpha_t\left(\mathbb{E}\left\{\mathsf{x}_t^{(1)} \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}\right\} - \mathbb{E}\left\{\mathsf{x}_t^{(1)} \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\right\}\right) \tag{7}$$

where $\alpha_t > 0$ is a scalar such that

$$\mathbb{E}\left\{\mu_t^{\mathrm{p}}\left(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\right)^2\right\} = P.$$

The intuition for this encoding function is that node 2 transmits to node 1 the knowledge of $\mathsf{x}_t^{(1)}$ that is available to node 2 but not available to node 1. Detailed interpretation of this encoding function can be found in [85].

The necessary and sufficient condition is presented in the next proposition.

Proposition 1: Set $\{\breve{e}_T\}_{T\geqslant 0}$ is bounded if and only if at least one of the following two conditions holds:

1) Vector $[1 \quad 0]^{\mathrm{T}}$ is orthogonal to the unobservable subspace of $(\boldsymbol{\Gamma}^{(1)}, \boldsymbol{A})$, or $A^{(1)} < 0$.
2) Vector $[1 \quad 0]^{\mathrm{T}}$ is orthogonal to the unobservable subspace of $(\boldsymbol{\Gamma}, \boldsymbol{A})$ and $C > A^{(1)}$, where C is the capacity of the channel given by (5).

If Condition 1 holds, then $\{e_T(\mu_{0:T})\}_{T\geqslant 0}$ is bounded for arbitrary encoding strategies $\mu_{0:T}$. If Condition 2 holds, then $\{e_T(\mu^{\mathrm{p}}_{0:T})\}_{T\geqslant 0}$ is bounded for the encoding strategy $\mu^{\mathrm{p}}_{0:T}$.

Proof: See Appendix A. ⊠

Proposition 1 shows that the MSE of a distributed filter is affected by the sensing and communication capabilities of the system as well as the variation rate of node 1's unknown state indicated by $A^{(1)}$. In particular, if Condition 1 of Proposition 1 holds, then node 1 can construct an estimator of $\mathsf{x}_t^{(1)}$ with bounded MSE using only its own observations $\mathbf{z}_{0:t}^{(1)}$ and not the received messages. If Condition 1 does not hold, then node 1 also needs to use messages received from node 2 to ensure the boundedness of the MSE, and the channel capacity is required to be larger than the variation rate of node 1's unknown state. One method to meet this capacity requirement is to allocate more communication resources to node 2 and increase its transmit power.

Remark 1: Condition 2 is analogous to the data rate theorem for control under constraint problems, which states that a linear system can be stabilized based on messages received via a channel if the data rate or channel capacity is above a threshold determined by the system dynamics [59], [60], [61], [62], [63], [64], [70], [72], [73]. Different from existing works where the receiver does not perform sensing, node 1 in this paper combines both the received messages with its own sensing observations for computing the distributed filter. Under this scenario, a data rate theorem in terms of the channel's Shannon capacity is established in Proposition 1. Note that another information-theoretical notion for studying control under communication constraints problems and sequential rate distortion problems is the directed mutual information [86], [87], [88], [89], [90] introduced in [91].

If Condition 2 holds, then node 2 can employ the encoding strategy given by (7) and the MSE of the distributed filter is guaranteed to be bounded over time. The following corollary shows a favorable property of this encoding strategy.

Corollary 1: If there exists an encoding strategy $\mu_{0:T}$ for every $T \geqslant 0$ such that $\{e_T(\mu_{0:T})\}_{T\geqslant 0}$ is bounded, then $\{e_T(\mu^{\mathrm{p}}_{0:T})\}_{T\geqslant 0}$ is also bounded.

Proof: If $\{e_T(\mu_{0:T})\}_{T\geqslant 0}$ is bounded, then $\{\breve{e}_T\}_{T\geqslant 0}$ is also bounded since $\breve{e}_T \leqslant e_T(\mu_{0:T})$ by definition. Therefore, at least one of the two conditions in Proposition 1 holds, and thus $\{e_T(\mu^{\mathrm{p}}_{0:T})\}_{T\geqslant 0}$ is bounded. ⊠

Corollary 1 shows that if the aim of the encoding strategy is to ensure the MSE of the distributed filter is bounded, then the proposed encoding strategy can be employed without considering other strategies, including nonlinear ones. In particular, if the MSE of the distributed filter is unbounded when the proposed encoding strategy is employed, then such MSE would also be unbounded for any other encoding strategy.

IV. Analogy to a Statistical Mechanical System

This section introduces a stochastic process $\{\mathbf{s}_t\}_{t\geqslant 0}$ associated with the distributed filtering problem and establishes an analogy of this process to a statistical mechanical system. To this end, define $\mathbf{x}_t$, $\mathbf{y}_t$, and $\hat{\mathbf{x}}_t$ as

$$\mathbf{x}_t := \left[\mathsf{x}_t^{(1)} \quad \mathsf{x}_t^{(2)}\right]^{\mathrm{T}} \tag{8a}$$

$$\mathbf{y}_t := \mathbb{E}\left\{\mathbf{x}_t \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}\right\} \tag{8b}$$

$$\hat{\mathbf{x}}_t := \mathbb{E}\left\{\mathbf{x}_t \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\right\}. \tag{8c}$$

In other words, $\mathbf{x}_t$ represents the joint unknown state of both nodes. Random vectors $\mathbf{y}_t$ and $\hat{\mathbf{x}}_t$ are both estimators of $\mathbf{x}_t$. In particular, $\mathbf{y}_t$ is the MMSE estimator of $\mathbf{x}_t$ based on observations $\mathbf{z}_{0:t}^{(1)}$ and $\mathbf{z}_{0:t}^{(2)}$. In other words, $\mathbf{y}_t$ is the centralized MMSE estimator of $\mathbf{x}_t$ based on sensor observations of both nodes. On the other hand, $\hat{\mathbf{x}}_t$ represents the distributed MMSE estimator of $\mathbf{x}_t$ based on sensor observations and received messages obtained by node 1. Define $\mathbf{s}_t$ as

$$\mathbf{s}_t := \left[\mathbf{x}_t^{\mathrm{T}} \quad \mathbf{y}_t^{\mathrm{T}} \quad \hat{\mathbf{x}}_t^{\mathrm{T}}\right]^{\mathrm{T}}.$$

An analogy of $\{\mathbf{s}_t\}_{t\geqslant 0}$ to a statistical mechanical system is established in scenarios where the following conditions hold. First, the encoding strategy employed by node 2 belongs to a class of linear encoding strategies with the encoding function μ_t at time t given by

$$\mu_t\big(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\big) = \boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{y}_t + g_t\big(\mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\big) \tag{9}$$

where $\boldsymbol{\beta}_t$ is a deterministic vector and g_t is an affine function. In particular, $\boldsymbol{\beta}_t$ and g_t are design parameters for the encoding strategy. In fact, the encoding strategy in (7) belongs to this class of strategies and is obtained by setting $\boldsymbol{\beta}_t = [\alpha_t \quad 0]^{\mathrm{T}}$ and $g_t(\mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}) = -\alpha_t\mathbb{E}\{\mathsf{x}_t^{(1)} \,|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\}$. Second, $A^{(i)}$ is negative for $i = 1, 2$. Third, $\boldsymbol{\beta}_t$ and $\mathbb{V}\{\mathbf{x}_t \,|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\}$ converge as t approaches infinity. In particular, the second and third conditions ensure that $\mathbb{V}\{\mathbf{s}_t\}$ converges as t approaches infinity. Note that $\mathbf{x}_t$, $\mathbf{z}_{0:t}^{(1)}$, and $\mathsf{r}_{0:t}$ are jointly Gaussian if linear encoding strategies are employed, and thus instantiations of these random quantities do not affect the value of $\mathbb{V}\{\mathbf{x}_t \,|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\}$.

Consider a statistical mechanical system $\Pi^{\mathbf{s}}$ associated with the process $\{\mathbf{s}_t\}_{t\geqslant 0}$. A microstate of the statistical mechanical system is represented by a vector $\boldsymbol{s} \in \mathbb{R}^{3n}$, where n is the dimension of $\mathbf{x}_t$. The Hamiltonian $H^{\mathbf{s}}(\boldsymbol{s})$ of $\boldsymbol{s}$ is defined as

$$H^{\mathbf{s}}(\boldsymbol{s}) := \frac{1}{2}\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\boldsymbol{s}$$

where $\boldsymbol{\Sigma}_{\mathbf{s}} := \lim_{t\to\infty}\mathbb{V}\{\mathbf{s}_t\}$. Macroscopic properties of $\Pi^{\mathbf{s}}$ at time t, including average energy $E_t^{\mathbf{s}}$, entropy $S_t^{\mathbf{s}}$, and free energy $F_t^{\mathbf{s}}$, are defined as

$$E_t^{\mathbf{s}} := \mathbb{E}\{H^{\mathbf{s}}(\mathbf{s}_t)\} \tag{10a}$$

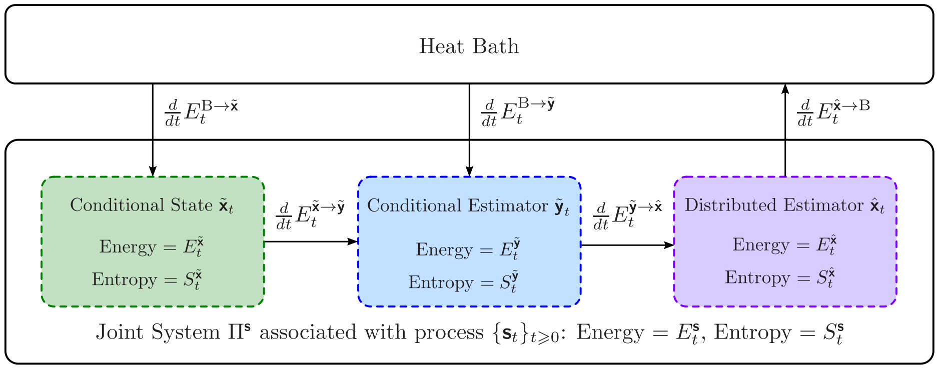


Fig. 2. Decomposition of the statistical mechanical system $\Pi^{\mathbf{s}}$ into three subsystems associated with the conditional state $\{\tilde{\mathbf{x}}_t\}_{t\geqslant 0}$, the conditional estimator $\{\tilde{\mathbf{y}}_t\}_{t\geqslant 0}$, and the distributed estimator $\{\hat{\mathbf{x}}_t\}_{t\geqslant 0}$, respectively. Arrows in the figure indicate the energy flows among the three subsystems and a heat bath.

$$S_t^{\mathbf{s}} := -D\big(P_{\mathbf{s}_t}\,\big\|\,\lambda_{\mathrm{L}}\big) \tag{10b}$$
$$F_t^{\mathbf{s}} := E_t^{\mathbf{s}} - S_t^{\mathbf{s}} \tag{10c}$$

where $D\big(P_{\mathrm{d}}\,\big\|\,\lambda_{\mathrm{L}}\big)$ represents the relative entropy, namely the Kullback-Leibler divergence, of the probability distribution P_{d} with respect to the Lebesgue measure λ_{L}. The macroscopic properties defined in (10) depend on the distribution of $\mathbf{s}_t$, which evolves according to the Fokker-Planck equation (i.e., Kolmogorov forward equation) since $\{\mathbf{s}_t\}_{t\geqslant 0}$ is a diffusion Markov process. The evolution of the distribution of $\mathbf{s}_t$ corresponds to the interaction between $\Pi^{\mathbf{s}}$ and a unit-temperature heat bath, which drives the process $\{\mathbf{s}_t\}_{t\geqslant 0}$ to its invariant distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{s}})$. The free energy of $\Pi^{\mathbf{s}}$ can be shown to be non-increasing with time and achieve minimal value when $\{\mathbf{s}_t\}_{t\geqslant 0}$ converges to its invariant distribution [26], [27].

The statistical mechanical system $\Pi^{\mathbf{s}}$ is decomposed into three physically distinct subsystems. To this end, define "conditional state" $\tilde{\mathbf{x}}_t$ and "conditional estimator" $\tilde{\mathbf{y}}_t$ as

$$\tilde{\mathbf{x}}_t := \mathbf{x}_t - \mathbf{y}_t \tag{11a}$$
$$\tilde{\mathbf{y}}_t := \mathbf{y}_t - \hat{\mathbf{x}}_t\,. \tag{11b}$$

In particular, $-\tilde{\mathbf{x}}_t$ is the estimation error of the centralized MMSE estimator of $\mathbf{x}_t$; random vector $-\tilde{\mathbf{y}}_t$ is the additional estimation error of the distributed filter of $\mathbf{x}_t$ compared to the centralized MMSE estimator. The next proposition shows properties related to $\tilde{\mathbf{x}}_t$, $\tilde{\mathbf{y}}_t$, and $\hat{\mathbf{x}}_t$ that enable the decomposition of $\Pi^{\mathbf{s}}$.

Proposition 2: The following statements hold:

1) Processes $\{\tilde{\mathbf{x}}_t\}_{t\geqslant 0}$, $\{\tilde{\mathbf{y}}_t\}_{t\geqslant 0}$, and $\{\hat{\mathbf{x}}_t\}_{t\geqslant 0}$ are all Markovian.
2) (Energy is additive): $E_t^{\mathbf{s}} = E_t^{\tilde{\mathbf{x}}} + E_t^{\tilde{\mathbf{y}}} + E_t^{\hat{\mathbf{x}}}$, with $E_t^{\tilde{\mathbf{x}}}$, $E_t^{\tilde{\mathbf{y}}}$, and $E_t^{\hat{\mathbf{x}}}$ defined as

$$E_t^{\tilde{\mathbf{x}}} := \frac{1}{2}\mathbb{E}\Big\{\tilde{\mathbf{x}}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\tilde{\mathbf{x}}}^{-1}\tilde{\mathbf{x}}_t\Big\} \tag{12a}$$
$$E_t^{\tilde{\mathbf{y}}} := \frac{1}{2}\mathbb{E}\Big\{\tilde{\mathbf{y}}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}}^{-1}\tilde{\mathbf{y}}_t\Big\} \tag{12b}$$
$$E_t^{\hat{\mathbf{x}}} := \frac{1}{2}\mathbb{E}\Big\{\hat{\mathbf{x}}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}^{-1}\hat{\mathbf{x}}_t\Big\}\,. \tag{12c}$$

Here $\boldsymbol{\Sigma}_{\tilde{\mathbf{x}}} := \lim_{t\to\infty}\mathbb{V}\{\tilde{\mathbf{x}}_t\}$, $\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}} := \lim_{t\to\infty}\mathbb{V}\{\tilde{\mathbf{y}}_t\}$, and $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}} := \lim_{t\to\infty}\mathbb{V}\{\hat{\mathbf{x}}_t\}$.

3) (Entropy is additive): $S_t^{\mathbf{s}} = S_t^{\tilde{\mathbf{x}}} + S_t^{\tilde{\mathbf{y}}} + S_t^{\hat{\mathbf{x}}}$, with $S_t^{\tilde{\mathbf{x}}}$, $S_t^{\tilde{\mathbf{y}}}$, and $S_t^{\hat{\mathbf{x}}}$ defined as

$$S_t^{\tilde{\mathbf{x}}} := -D\big(P_{\tilde{\mathbf{x}}_t}\,\big\|\,\lambda_{\mathrm{L}}\big) \tag{13a}$$
$$S_t^{\tilde{\mathbf{y}}} := -D\big(P_{\tilde{\mathbf{y}}_t}\,\big\|\,\lambda_{\mathrm{L}}\big) \tag{13b}$$
$$S_t^{\hat{\mathbf{x}}} := -D\big(P_{\hat{\mathbf{x}}_t}\,\big\|\,\lambda_{\mathrm{L}}\big)\,. \tag{13c}$$

Proof: See Appendix B. ⊠

Proposition 2 shows that the statistical mechanical system $\Pi^{\mathbf{s}}$ can be decomposed into three subsystems $\Pi^{\tilde{\mathbf{x}}}$, $\Pi^{\tilde{\mathbf{y}}}$, and $\Pi^{\hat{\mathbf{x}}}$, which are associated with the conditional state $\{\tilde{\mathbf{x}}_t\}_{t\geqslant 0}$, the conditional estimator $\{\tilde{\mathbf{y}}_t\}_{t\geqslant 0}$, and the distributed estimator $\{\hat{\mathbf{x}}_t\}_{t\geqslant 0}$, respectively (see Fig. 2). Quantities $E_t^{\tilde{\mathbf{x}}}$, $E_t^{\tilde{\mathbf{y}}}$, and $E_t^{\hat{\mathbf{x}}}$ given in (12) represent the average energy of $\Pi^{\tilde{\mathbf{x}}}$, $\Pi^{\tilde{\mathbf{y}}}$, and $\Pi^{\hat{\mathbf{x}}}$, respectively, whereas $S_t^{\tilde{\mathbf{x}}}$, $S_t^{\tilde{\mathbf{y}}}$, and $S_t^{\hat{\mathbf{x}}}$ given in (13c) represent the entropy of $\Pi^{\tilde{\mathbf{x}}}$, $\Pi^{\tilde{\mathbf{y}}}$, and $\Pi^{\hat{\mathbf{x}}}$, respectively. The energy and entropy of the three subsystems vary with time as they interact with each other and with the heat bath. In particular, the variation rates of energy associated with the three subsystems are given by (82) in Appendix C.

Figure 2 shows the energy flows among the three subsystems and the heat bath. The rates of these energy flows depend on the evolution of the distributions of $\tilde{\mathbf{x}}_t$, $\tilde{\mathbf{y}}_t$, and $\hat{\mathbf{x}}_t$. These rates are evaluated as follows. First, consider the rate $\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{x}}}$ of the energy flow from the heat bath to the conditional state. The conditional state absorbs energy from the heat bath and supplies energy to the conditional estimator. If the conditional state is disconnected with the conditional estimator at time t and only interacts with the heat bath, then $\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{x}}}$ would equal the variation rate $\frac{d}{dt}E_t^{\tilde{\mathbf{x}}}$ of the average energy of the conditional state. Meanwhile, such disconnection corresponds to the scenario where observations become unavailable at time t for the centralized MMSE estimator $\mathbf{y}_t$ of state $\mathbf{x}_t$. Since $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \mathbf{y}_t$, if the observations become unavailable, then the distribution of the conditional state $\tilde{\mathbf{x}}_t$ would evolve in the same manner as $\mathbf{x}_t$. Consequently, $\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{x}}}$ can be obtained by the following two methods. The first method is setting $\boldsymbol{\Gamma} = \mathbf{0}$ in (82a) to remove the effect of observations. The second method evaluates $\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{x}}}$ as

$$\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{x}}} = \frac{1}{2}\mathrm{tr}\Big\{\frac{d}{dt}\mathbb{V}\{\mathbf{x}_t\}\Big|_{\mathbb{V}\{\mathbf{x}_t\}=\mathbb{V}\{\tilde{\mathbf{x}}_t\}}\boldsymbol{\Sigma}_{\tilde{\mathbf{x}}}^{-1}\Big\}\,. \tag{14}$$

Both methods lead to the same result. The rate $\frac{d}{dt}E_t^{\tilde{\mathbf{x}}\to\tilde{\mathbf{y}}}$ of the energy flow from the conditional state to the conditional estimator is thus

$$\frac{d}{dt}E_t^{\tilde{\mathbf{x}}\to\tilde{\mathbf{y}}} = \frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{x}}} - \frac{d}{dt}E_t^{\tilde{\mathbf{x}}}\,. \tag{15}$$

The rate $\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{y}}}$ of the energy flow from the heat bath to the conditional estimator is computed as follows. If the conditional estimator is disconnected with the distributed estimator at time t, then the sum of $\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{y}}}$ and $\frac{d}{dt}E_t^{\tilde{\mathbf{x}}\to\tilde{\mathbf{y}}}$ would equal the variation rate $\frac{d}{dt}E_t^{\tilde{\mathbf{y}}}$ of the average energy of the conditional estimator. Meanwhile, such disconnection corresponds to the scenario where observations and received messages become unavailable to the distributed estimator at time t, and thus the distribution of the conditional estimator $\tilde{\mathbf{y}}_t$ would evolve in the same manner as $\mathbf{y}_t$. As a result,

$$\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{y}}} = \frac{1}{2}\mathrm{tr}\left\{\frac{d}{dt}\mathbb{V}\{\mathbf{y}_t\}\Big|_{\mathbb{V}\{\mathbf{y}_t\}=\mathbb{V}\{\tilde{\mathbf{y}}_t\}}\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}}^{-1}\right\} - \frac{d}{dt}E_t^{\tilde{\mathbf{x}}\to\tilde{\mathbf{y}}}\,. \tag{16}$$

The rates of energy flows from the conditional estimator to the distributed estimator $\frac{d}{dt}E_t^{\tilde{\mathbf{y}}\to\hat{\mathbf{x}}}$ and from the distributed estimator to the heat bath $\frac{d}{dt}E_t^{\hat{\mathbf{x}}\to\mathrm{B}}$ are given by

$$\frac{d}{dt}E_t^{\tilde{\mathbf{y}}\to\hat{\mathbf{x}}} = \frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{y}}} + \frac{d}{dt}E_t^{\tilde{\mathbf{x}}\to\tilde{\mathbf{y}}} - \frac{d}{dt}E_t^{\tilde{\mathbf{y}}} \tag{17}$$

$$\frac{d}{dt}E_t^{\hat{\mathbf{x}}\to\mathrm{B}} = \frac{d}{dt}E_t^{\tilde{\mathbf{y}}\to\hat{\mathbf{x}}} - \frac{d}{dt}E_t^{\hat{\mathbf{x}}}\,. \tag{18}$$

Derivation of the energy flow rates in (14)–(18) is presented in Appendix C. As time t approaches infinity, the distributions of $\tilde{\mathbf{x}}_t$, $\tilde{\mathbf{y}}_t$, and $\hat{\mathbf{x}}_t$ converge to $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{\tilde{\mathbf{x}}})$, $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}})$, and $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{\hat{\mathbf{x}}})$, respectively. Moreover, the average energy of $\Pi^{\tilde{\mathbf{x}}}$, $\Pi^{\tilde{\mathbf{y}}}$, and $\Pi^{\hat{\mathbf{x}}}$ also converge to their stationary values, respectively. This can be seen from the following equations

$$\lim_{t\to\infty}\left(\frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{x}}} - \frac{d}{dt}E_t^{\tilde{\mathbf{x}}\to\tilde{\mathbf{y}}}\right) = 0$$
$$\lim_{t\to\infty}\left(\frac{d}{dt}E_t^{\tilde{\mathbf{x}}\to\tilde{\mathbf{y}}} + \frac{d}{dt}E_t^{\mathrm{B}\to\tilde{\mathbf{y}}} - \frac{d}{dt}E_t^{\tilde{\mathbf{y}}\to\hat{\mathbf{x}}}\right) = 0$$
$$\lim_{t\to\infty}\left(\frac{d}{dt}E_t^{\tilde{\mathbf{y}}\to\hat{\mathbf{x}}} - \frac{d}{dt}E_t^{\hat{\mathbf{x}}\to\mathrm{B}}\right) = 0\,.$$

The linear assumption (9) on encoding functions simplifies the analogy of the distributed filtering problem to the statistical mechanical system. In particular, each subsystem of the statistical mechanical system is associated at time t with a random vector related to the conditional expectation of $\mathbf{x}_t$. This is possible because processes $\{\mathbf{x}_t\}_{t\geqslant 0}$, $\{\mathbf{z}_t^{(1)}\}_{t\geqslant 0}$, $\{\mathbf{z}_t^{(2)}\}_{t\geqslant 0}$, and $\{\mathsf{r}_t\}_{t\geqslant 0}$ are jointly Gaussian. Consequently, the conditional expectations become sufficient statistics for $\mathbf{x}_t$. For example, $\hat{\mathbf{x}}_t$ defined in (8c) is a sufficient statistic of $\mathbf{z}_{0:t}^{(1)}$ and $\mathsf{r}_{0:t}$ for $\mathbf{x}_t$. As a result, $\hat{\mathbf{x}}_t$ retains all the information about $\mathbf{x}_t$ contained in $\mathbf{z}_{0:t}^{(1)}$ and $\mathsf{r}_{0:t}$. However, this would not hold if nonlinear encoding functions are employed and thus $\{\mathbf{x}_t\}_{t\geqslant 0}$, $\{\mathbf{z}_t^{(1)}\}_{t\geqslant 0}$, and $\{\mathsf{r}_t\}_{t\geqslant 0}$ are longer jointly Gaussian. As a result, the conditional expectation $\hat{\mathbf{x}}_t$ alone cannot be used for establishing the analogy as it does not capture all the information of $\mathbf{x}_t$. Instead, the conditional probability distribution of $\mathbf{x}_t$ given $\{\mathbf{z}_t^{(1)}\}_{t\geqslant 0}$ and $\{\mathsf{r}_t\}_{t\geqslant 0}$ is needed in this case. Note that $\hat{\mathbf{x}}_t$ is only the first moment corresponding to such a conditional probability distribution and thus contains less information than this distribution. In fact, establishing analogies of nonlinear filtering problems to statistical mechanical systems is more complicated than that of linear filtering problems. For example, [27] derives an analogy of a centralized nonlinear filtering problem to a statistical mechanical system consisting of multiple subsystems. There, each statistical mechanical subsystem is associated with a random probability measure, which is an element in the space of probability measures. By contrast, in the distributed filtering problem in this paper with linear encoding functions, each subsystem is associated with a random vector, which is an element in a Euclidean space.

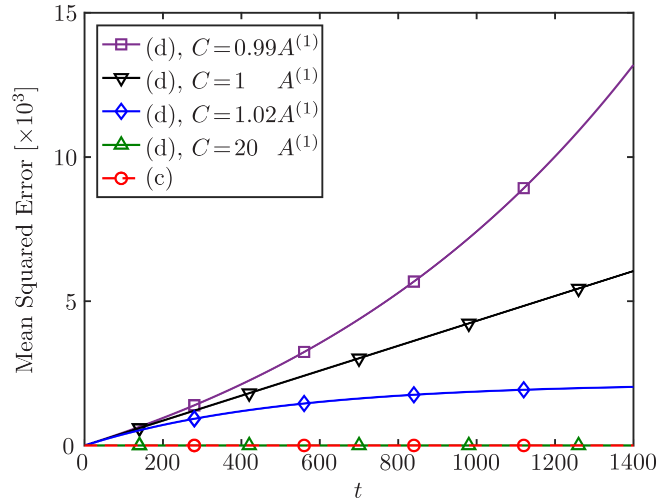


Fig. 3. MSE of the (d) distributed filter under different channel capacities when node 2 employs the proposed encoding strategy, and (c) the centralized MMSE estimator.

V. Numerical Results

This section presents a numerical example where the dimension of $\mathbf{v}_t^{(i)}$ is one, and thus $\boldsymbol{B}^{(i)}$ becomes a scalar for $i\in\{1,2\}$. Deterministic quantities in (1) and (2) are set to

$$\begin{aligned} A^{(1)} &= 0.05 & \boldsymbol{B}^{(1)} &= 2\\ A^{(2)} &= -0.05 & \boldsymbol{B}^{(2)} &= 1\\ \boldsymbol{\Gamma}^{(1)} &= [0 \quad 1] & \boldsymbol{\Xi}^{(1)} &= 2\\ \boldsymbol{\Gamma}^{(2)} &= [1 \quad -1] & \boldsymbol{\Xi}^{(2)} &= 1\,. \end{aligned}$$

At time 0, $\mathbb{V}\{\mathsf{x}_0^{(1)}\} = \mathbb{V}\{\mathsf{x}_0^{(2)}\} = 1.2$, $\boldsymbol{G}^{(i)}$ is set to $\boldsymbol{G}^{(i)} = 0.1\boldsymbol{\Gamma}^{(i)}$, and $\mathbb{V}\{\boldsymbol{\zeta}^{(i)}\} = \boldsymbol{\Xi}^{(i)}$ for $i=1,2$. It can be seen that $(\boldsymbol{\Gamma},\boldsymbol{A})$ is observable, where we recall that $\boldsymbol{\Gamma}$ and $\boldsymbol{A}$ are defined in (6). However, $[1 \quad 0]^{\mathrm{T}}$ is not orthogonal to the unobservable subspace of $(\boldsymbol{\Gamma}^{(1)},\boldsymbol{A})$, which is $\mathcal{C}([1 \quad 0]^{\mathrm{T}})$. As a result, Condition 1 of Proposition 1 does not hold. According to this proposition, when the proposed encoding strategy is employed, the MSE of the distributed filter is bounded if and only if Condition 2 holds, which translates to $C > A^{(1)}$. This section shows the accuracy of the distributed filter when C is chosen from the set $\{0.99\,A^{(1)}, A^{(1)}, 1.02\,A^{(1)}, 20\,A^{(1)}\}$ nats/s. Note that $A^{(1)}$ is the threshold of the channel capacity that determines whether the MSE of the distributed filter is bounded or not, whereas $0.99\,A^{(1)}$ and $1.02\,A^{(1)}$ are values close to this threshold.

Figure 3 shows the MSE of the distributed filter as a function of time t for different channel capacities when the proposed encoding strategy is employed. The MSE of the centralized MMSE estimator represented by curve (c) is also shown for comparison. Such an MSE converges after some time. The convergence of the MSE for the distributed filter depends on the channel capacity. Specifically, such MSE increases with time when $C = 0.99\,A^{(1)}$ or $C = A^{(1)}$. By contrast, the MSE of the distributed filter converges when $C = 1.02\,A^{(1)}$ or $C = 20\,A^{(1)}$, and the MSE is significantly smaller in the latter case when t is large. This example shows that higher channel capacity improves the performance of distributed filtering, which supports Proposition 1.

VI. Conclusion

This paper analyzed continuous-time distributed filtering with sensing and communication constraints. In particular, a building-block system with two nodes has been considered, where each node is tasked to infer a time-varying unknown state. In particular, node 2 transmits encoded messages containing information of the unknown state that node 1 attempts to infer via a Gaussian feedback channel. The paper derived a necessary and sufficient condition on the sensing and communication capabilities of the nodes under which the MSE of the distributed filter is bounded over time. Specifically, the condition indicates that if node 1 needs to rely on the received messages to achieve bounded MSE, then the capacity of the channel from node 2 to node 1 must be larger than a threshold determined by the dynamic model of node 1's unknown state. Moreover, the paper established an analogy between the distributed filtering problem and a statistical mechanical system. The paper shows the effects of sensing and communication capabilities on the accuracy of distributed filtering and provides insights for efficient allocation of sensing and communication resources in networked systems.

Appendix A
Proof of Proposition 1

This section first introduces a lemma and a few definitions used for proving Proposition 1. Then, Proposition 1 is proved.

A. Lemma and Definitions Used for Proving Proposition 1

A lemma used in the proof is presented as follows. Consider Gaussian processes $\{\boldsymbol{\theta}_t\}_{t\geqslant 0}$ and $\{\boldsymbol{\xi}_t\}_{t\geqslant 0}$ described by the following SDE

$$d\boldsymbol{\theta}_t = \boldsymbol{A}\boldsymbol{\theta}_t\,dt + \boldsymbol{B}\,d\mathbf{v}_t$$
$$d\boldsymbol{\xi}_t = \boldsymbol{G}\boldsymbol{\theta}_t\,dt + \boldsymbol{L}\,d\boldsymbol{\omega}_t$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{G}$, and $\boldsymbol{L}$ are deterministic matrices such that both $\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$ and $\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$ are invertible; $\boldsymbol{\theta}_0$ and $\boldsymbol{\xi}_0$ are jointly Gaussian; $\{\mathbf{v}_t\}_{t\geqslant 0}$ and $\{\boldsymbol{\omega}_t\}_{t\geqslant 0}$ are Brownian motions such that $\boldsymbol{\theta}_0$, $\boldsymbol{\xi}_0$, $\{\mathbf{v}_t\}_{t\geqslant 0}$, and $\{\boldsymbol{\omega}_t\}_{t\geqslant 0}$ are independent. The following lemma shows the relationship between the boundedness of the inference error for $\{\boldsymbol{\theta}_t\}_{t\geqslant 0}$ and unobservable subspaces.

Lemma 1: For any vector $\boldsymbol{h}$ orthogonal to the unobservable space of $(\boldsymbol{G}, \boldsymbol{A})$, the set $\left\{\boldsymbol{h}^{\mathrm{T}}\mathbb{V}\{\boldsymbol{\theta}_t \,|\, \boldsymbol{\xi}_{0:t}\}\boldsymbol{h}\right\}_{t\geqslant 0}$ is bounded.

Proof: See [92, Appendix A.3.1]. ⊠

Next, some definitions are introduced. Recall the concatenated state $\mathbf{x}_t$ and its estimator $\mathbf{y}_t$ defined in (8a) and (8b), respectively. Define the error covariance matrix of $\mathbf{y}_t$ as

$$\boldsymbol{E}_t^{\mathrm{c}} := \mathbb{V}\{\mathbf{x}_t - \mathbf{y}_t\}\,. \tag{19}$$

Using Kalman–Bucy filtering results [84], $\{\mathbf{y}_t\}_{t\geqslant 0}$ can be shown to satisfy

$$\begin{aligned} d\mathbf{y}_t = &\left(\boldsymbol{A} - \boldsymbol{E}_t^{\mathrm{c}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}\right)\mathbf{y}_t\,dt \\ &+ \boldsymbol{E}_t^{\mathrm{c}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1}d\mathbf{z}_t \end{aligned} \tag{20}$$

where

$$\boldsymbol{\Xi} := \mathrm{diag}\{\boldsymbol{\Xi}^{(1)}, \boldsymbol{\Xi}^{(2)}\} \tag{21a}$$
$$\mathbf{z}_t := \left[\left(\mathbf{z}_t^{(1)}\right)^{\mathrm{T}} \quad \left(\mathbf{z}_t^{(2)}\right)^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{21b}$$

Process $\{\mathbf{y}_t\}_{t\geqslant 0}$ can be represented in terms of an innovation process. To this end, define processes $\{\tilde{\boldsymbol{\eta}}_t^{(i)}\}_{t\geqslant 0}$ and $\{\boldsymbol{\eta}_t^{(i)}\}_{t\geqslant 0}$ for $i = 1, 2$ as

$$d\tilde{\boldsymbol{\eta}}_t^{(i)} = d\mathbf{z}_t^{(i)} - \boldsymbol{\Gamma}^{(i)}\mathbf{y}_t\,dt \qquad \tilde{\boldsymbol{\eta}}_0^{(i)} = \mathbf{z}_0^{(i)} \tag{22a}$$
$$d\boldsymbol{\eta}_t^{(i)} = \left(\boldsymbol{\Xi}^{(i)}\left(\boldsymbol{\Xi}^{(i)}\right)^{\mathrm{T}}\right)^{-1/2} d\tilde{\boldsymbol{\eta}}_t^{(i)} \qquad \boldsymbol{\eta}_0^{(i)} = \mathbf{z}_0^{(i)}\,. \tag{22b}$$

Moreover, define

$$\boldsymbol{\eta}_t := \left[\left(\boldsymbol{\eta}_t^{(1)}\right)^{\mathrm{T}} \quad \left(\boldsymbol{\eta}_t^{(2)}\right)^{\mathrm{T}}\right]^{\mathrm{T}} \qquad \text{for } t \geqslant 0\,. \tag{23}$$

The process $\{\boldsymbol{\eta}_t\}_{t\geqslant 0}$ is referred to as a scaled innovation process in this paper. Processes $\{\boldsymbol{\eta}_t - \boldsymbol{\eta}_0\}_{t\geqslant 0}$ is a Brownian motion [93, Ch. 4.4], and $\sigma(\boldsymbol{\eta}_{0:t}) = \sigma\left(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}\right)$ [83, Ch. 7.5]. Combining (20)–(23) gives

$$d\mathbf{y}_t = \boldsymbol{A}\mathbf{y}_t\,dt + \boldsymbol{E}_t^{\mathrm{c}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1/2} d\boldsymbol{\eta}_t\,. \tag{24}$$

Note that $\hat{\mathbf{x}}_t$ defined in (8c) is the MMSE estimator of $\mathbf{y}_t$ based on $\mathbf{z}_{0:t}^{(1)}$ and $\mathrm{r}_{0:t}$. In particular, it can be verified that $\mathbf{x}_t - \mathbf{y}_t \perp\!\!\!\perp \mathbf{z}_{0:t}^{(1)}, \mathrm{r}_{0:t}$. Consequently,

$$\hat{\mathbf{x}}_t = \mathbb{E}\left\{\mathbf{y}_t \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathrm{r}_{0:t}\right\}$$

which shows that $\hat{\mathbf{x}}_t$ is the MMSE estimator of $\mathbf{y}_t$ based on $\mathbf{z}_{0:t}^{(1)}$ and $\mathrm{r}_{0:t}$. Define $\boldsymbol{Q}_t$ as the error covariance matrix of $\hat{\mathbf{x}}_t$ as an estimator of $\mathbf{y}_t$, i.e.,

$$\boldsymbol{Q}_t := \mathbb{V}\{\mathbf{y}_t - \hat{\mathbf{x}}_t\}\,. \tag{25}$$

Matrix $\boldsymbol{Q}_t$ is affected by the encoding strategy employed by node 2. In particular, if the encoding strategy $\mu_{0:T}^{\mathrm{P}}$ is employed with μ_t^{P} given by (7), then $\boldsymbol{Q}_t$ can be shown to satisfy the following ordinary differential equation

$$\begin{aligned} \frac{d\boldsymbol{Q}_t}{dt} = &\ \boldsymbol{A}\boldsymbol{Q}_t + \boldsymbol{Q}_t\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{E}_t^{\mathrm{c}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}\boldsymbol{E}_t^{\mathrm{c}} \\ &- \frac{P}{\kappa^2}\frac{1}{[\boldsymbol{Q}_t]_{1,1}}\boldsymbol{Q}_t\,\mathrm{diag}\{1,0\}\boldsymbol{Q}_t \\ &- \left(\boldsymbol{E}_t^{\mathrm{c}} + \boldsymbol{Q}_t\right)\left(\boldsymbol{\Gamma}^{(1)}\right)^{\mathrm{T}}\left(\boldsymbol{\Xi}^{(1)}\left(\boldsymbol{\Xi}^{(1)}\right)^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}^{(1)} \\ &\times \left(\boldsymbol{E}_t^{\mathrm{c}} + \boldsymbol{Q}_t\right) \end{aligned} \tag{26}$$

for all $t \in [0, T]$.

Define $\mathsf{y}_t^{(1)}$ as

$$\mathsf{y}_t^{(1)} := \mathbb{E}\left\{\mathsf{x}_t^{(1)} \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}\right\} \qquad \text{for } t \in [0, T]\,. \tag{27}$$

Note that $\hat{\mathsf{x}}_t^{(1)}$ and $\mathsf{y}_t^{(1)}$ are first entries of $\hat{\mathbf{x}}_t$ and $\mathbf{y}_t$, respectively. Moreover, define $\varepsilon_t(\mu_{0:t})$ as

$$\varepsilon_t(\mu_{0:t}) := \mathbb{V}\left\{\mathsf{y}_t^{(1)} - \hat{\mathsf{x}}_t^{(1)}\right\}. \tag{28}$$

The quantity $\varepsilon_t(\mu_{0:t})$ is related to $e_t(\mu_{0:t})$ as follows

$$e_t(\mu_{0:t}) = \left[\boldsymbol{E}_t^{\mathrm{c}}\right]_{1,1} + \varepsilon_t(\mu_{0:t})\,. \tag{29}$$

B. Detailed Proof of Proposition 1

Proof (Sufficiency): Note that $\breve{e}_T \leqslant e_T(\mu_{0:T}) \leqslant \mathbb{V}\{\mathsf{x}_T^{(1)} \,|\, \mathbf{z}_{0:T}^{(1)}\} \leqslant \mathbb{V}\{\mathsf{x}_T^{(1)}\}$ for an arbitrary encoding strategy $\mu_{0:T}$. If $[1 \quad 0]^{\mathrm{T}}$ is orthogonal to the unobservable subspace of $(\boldsymbol{\Gamma}^{(1)}, \boldsymbol{A})$, then Lemma 1 shows that $\{\mathbb{V}\{\mathsf{x}_T^{(1)} \,|\, \mathbf{z}_{0:T}^{(1)}\}\}_{T\geqslant 0}$ is bounded. If $A^{(1)} < 0$, then $\{\mathbb{V}\{\mathsf{x}_T^{(1)}\}\}_{T\geqslant 0}$ is bounded. Therefore, Condition 1 in Proposition 1 ensures that $\{\breve{e}_T\}_{T\geqslant 0}$ is bounded. Moreover, for arbitrary encoding strategies $\mu_{0:T}$, it can be shown that $\{e_T(\mu_{0:T})\}_{T\geqslant 0}$ is also bounded.

Next, assume that Condition 2 holds instead. Consider an encoding strategy $\breve{\mu}_{0:T}$ with encoding function $\breve{\mu}_t$ at time t defined as

$$\breve{\mu}_t\left(\mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\right) := \breve{\alpha}_t\left(\mathsf{y}_t^{(1)} - \mathbb{E}\left\{\mathsf{y}_t^{(1)} \,\middle|\, \mathsf{r}_{0:t}\right\}\right)$$

where $\breve{\alpha}_t$ is a scalar such that

$$\mathbb{V}\left\{\breve{\alpha}_t\left(\mathsf{y}_t^{(1)} - \mathbb{E}\left\{\mathsf{y}_t^{(1)} \,\middle|\, \mathsf{r}_{0:t}\right\}\right)\right\} = P\,. \tag{30}$$

Note that $\breve{\mu}_{0:T}$ is a linear encoding strategy, i.e., $\breve{\mu}_t$ is a linear function for all $t \in [0, T]$. Let $\varepsilon_T(\breve{\mu}_{0:T})$ represent the MMSE for inferring $\mathsf{y}_T^{(1)}$ based on $\mathbf{z}_{0:T}^{(1)}$ and $\mathsf{r}_{0:T}$ if strategy $\breve{\mu}_{0:T}$ is employed. In addition, let $\breve{\varepsilon}_T(\breve{\mu}_{0:T})$ represent the MMSE for inferring $\mathsf{y}_T^{(1)}$ based only on $\mathsf{r}_{0:T}$ if $\breve{\mu}_{0:T}$ is employed. Then $\varepsilon_T(\breve{\mu}_{0:T}) \leqslant \breve{\varepsilon}_T(\breve{\mu}_{0:T})$. Viewing $\breve{\varepsilon}_T(\breve{\mu}_{0:T})$ as a function of T, we can show that this function solves the following initial value problem

$$\frac{d}{dT}\breve{\varepsilon}_T(\breve{\mu}_{0:T}) = \left(2A^{(1)} - \frac{P}{\kappa^2}\right)\breve{\varepsilon}_T(\breve{\mu}_{0:T}) + [1 \quad 0]\, \boldsymbol{E}_T^{\mathrm{c}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}\boldsymbol{E}_T^{\mathrm{c}}\begin{bmatrix}1\\0\end{bmatrix} \tag{31a}$$

$$\breve{\varepsilon}_0(\breve{\mu}_0) = \mathbb{V}\left\{\mathsf{y}_0^{(1)} \,\middle|\, \mathbf{z}_0^{(1)}\right\} \tag{31b}$$

where $\boldsymbol{E}_t^{\mathrm{c}}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Xi}$ are defined in (19), (6b), and (21a), respectively. To derive (31), we observe that $\mathsf{y}_t^{(1)}$ defined in (27) is the first component of $\mathbf{y}_t$ defined in (8b). Combining this with (24) gives

$$d\mathsf{y}_t^{(1)} = A^{(1)}\mathsf{y}_t^{(1)}\, dt + [1 \quad 0]\; \boldsymbol{E}_t^{\mathrm{c}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1/2} d\boldsymbol{\eta}_t \tag{32}$$

where $\boldsymbol{\eta}_t$ is defined in (23). If the encoding strategy $\breve{\mu}_{0:T}$ is employed, then the received messages $\mathsf{r}_{0:T}$ satisfy

$$d\mathsf{r}_t = \breve{\alpha}_t\left(\mathsf{y}_t^{(1)} - \mathbb{E}\left\{\mathsf{y}_t^{(1)} \,\middle|\, \mathsf{r}_{0:t}\right\}\right) dt + \kappa\, d\mathsf{w}_t\,. \tag{33}$$

The relationship (31) is then obtained by combining (30), (32), and (33). If Condition 2 of Proposition 1 holds, then the solution $\breve{\varepsilon}_T(\breve{\mu}_{0:T})$ to (31) is unique, and $\{\breve{\varepsilon}_T(\breve{\mu}_{0:T})\}_{T\geqslant 0}$ is bounded. Moreover, according to Lemma 1, $\{[\boldsymbol{E}_T^{\mathrm{c}}]_{1,1}\}_{T\geqslant 0}$ is bounded. Note that

$$\begin{aligned}\breve{e}_T \leqslant e_T(\breve{\mu}_{0:T}) &= \left[\boldsymbol{E}_T^{\mathrm{c}}\right]_{1,1} + \varepsilon_T(\breve{\mu}_{0:T})\\ &\leqslant \left[\boldsymbol{E}_T^{\mathrm{c}}\right]_{1,1} + \breve{\varepsilon}_T(\breve{\mu}_{0:T})\end{aligned} \tag{34}$$

where (29) is used for obtaining the equality in (34). Since both $\{[\boldsymbol{E}_T^{\mathrm{c}}]_{1,1}\}_{T\geqslant 0}$ and $\{\breve{\varepsilon}_T(\breve{\mu}_{0:T})\}_{T\geqslant 0}$ are bounded, $\{\breve{e}_T\}_{T\geqslant 0}$ is also bounded.

Next, it is proved that $\{e_T(\mu_{0:T}^{\mathrm{p}})\}_{T\geqslant 0}$ is bounded if Condition 2 of Proposition 1 holds. If the encoding strategy $\mu_{0:T}^{\mathrm{p}}$ is employed, then $\boldsymbol{Q}_T$ satisfies (26). Omitting the term after the second minus sign in (26), which is a positive semidefinite (PSD) matrix, we obtain

$$\frac{d\boldsymbol{Q}_t}{dt} \preccurlyeq \boldsymbol{A}\boldsymbol{Q}_t + \boldsymbol{Q}_t\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{E}_t^{\mathrm{c}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}\boldsymbol{E}_t^{\mathrm{c}} - \frac{P}{\kappa^2}\frac{1}{[\boldsymbol{Q}_t]_{1,1}}\boldsymbol{Q}_t\,\mathrm{diag}\{1, 0\}\,\boldsymbol{Q}_t\,. \tag{35}$$

Here, $\boldsymbol{X} \preccurlyeq \boldsymbol{Y}$ represents that $\boldsymbol{Y} - \boldsymbol{X}$ is PSD for symmetric matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$. Definitions (8) and (27) show that $\mathsf{y}_t^{(1)} = [1 \quad 0]\;\mathbf{y}_t$ and $\hat{\mathsf{x}}_t^{(1)} = [1 \quad 0]\;\hat{\mathbf{x}}_t$. Combining these with (25) and (28) give $\varepsilon_T(\mu_{0:T}^{\mathrm{p}}) = [1 \quad 0]\;\boldsymbol{Q}_T\,[1 \quad 0]^{\mathrm{T}}$. Therefore, left and right multiplying (35) by $[1 \quad 0]$ and $[1 \quad 0]^{\mathrm{T}}$, respectively, and combining the result with $\varepsilon_0(\mu_0^{\mathrm{p}})$, we obtain

$$\frac{d}{dT}\varepsilon_T(\mu_{0:T}^{\mathrm{p}}) \leqslant \left(2A^{(1)} - \frac{P}{\kappa^2}\right)\varepsilon_T(\mu_{0:T}^{\mathrm{p}}) + [1 \quad 0]\;\boldsymbol{E}_T^{\mathrm{c}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}\boldsymbol{E}_T^{\mathrm{c}}\begin{bmatrix}1\\0\end{bmatrix} \tag{36a}$$

$$\varepsilon_0(\mu_0^{\mathrm{p}}) = \mathbb{V}\{\mathsf{y}_0^{(1)} \,|\, \mathbf{z}_0^{(1)}\}\,. \tag{36b}$$

Comparing (31) with (36) and applying of [94, Ch. 3, Th. 4.1] gives $\varepsilon_T(\mu_{0:T}^{\mathrm{p}}) \leqslant \varepsilon_T(\breve{\mu}_{0:T})$. Combining this with (29) gives

$$e_T(\mu_{0:T}^{\mathrm{p}}) \leqslant [\boldsymbol{E}_T^{\mathrm{c}}]_{1,1} + \breve{\varepsilon}_T(\breve{\mu}_{0:T})\,.$$

Since both $\{[\boldsymbol{E}_T^{\mathrm{c}}]_{1,1}\}_{T\geqslant 0}$ and $\{\breve{\varepsilon}_T(\breve{\mu}_{0:T})\}_{T\geqslant 0}$ are bounded, $\{e_T(\mu_{0:T}^{\mathrm{p}})\}_{T\geqslant 0}$ is also bounded.

Necessity: Assume that $\{\breve{e}_T\}_{T\geqslant 0}$ is bounded and Condition 1 of Proposition 1 does not hold. It will be shown that Condition 2 must hold. To see that $[1 \quad 0]^{\mathrm{T}}$ is orthogonal to the unobservable subspace of $(\boldsymbol{\Gamma}, \boldsymbol{A})$, note that $[\boldsymbol{E}_T^{\mathrm{c}}]_{1,1} \leqslant \breve{e}_T$ since the MSE of $\mathsf{x}_T^{(1)}$ achieved by the centralized MMSE estimator is no larger than that achieved by the distributed filter regardless of the employed encoding strategies. Since $\{\breve{e}_T\}_{T\geqslant 0}$ is bounded, $\{[\boldsymbol{E}_T^{\mathrm{c}}]_{1,1}\}_{T\geqslant 0}$ is also bounded. If $A^{(1)} > 0$, vector $[1 \quad 0]^{\mathrm{T}}$ must be orthogonal to the unobservable subspace of $(\boldsymbol{\Gamma}, \boldsymbol{A})$ to ensure the boundedness of $\{[\boldsymbol{E}_T^{\mathrm{c}}]_{1,1}\}_{T\geqslant 0}$.

It is next shown that $C > A^{(1)}$. If Condition 1 of Proposition 1 does not hold, then there exists a vector $\boldsymbol{u} \neq \boldsymbol{0}$ that satisfies the following equalities

$$\boldsymbol{\Gamma}^{(1)}\exp\{\boldsymbol{A}t\}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{0} \qquad \forall t \geqslant 0 \tag{37a}$$

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} = A^{(1)}\boldsymbol{u} \tag{37b}$$

where

$$\boldsymbol{B} := \mathrm{diag}\left\{\boldsymbol{B}^{(1)}, \boldsymbol{B}^{(2)}\right\}. \tag{38}$$

In fact, if $[1 \quad 0]^{\mathrm{T}}$ is not orthogonal to the unobservable subspace of $(\boldsymbol{\Gamma}^{(1)}, \boldsymbol{A})$, then either (i) $\mathcal{C}\big((\boldsymbol{\Gamma}^{(1)})^{\mathrm{T}}\big) \subseteq \mathcal{C}\big([0 \quad 1]^{\mathrm{T}}\big)$, or (ii) $\mathcal{C}\big((\boldsymbol{\Gamma}^{(1)})^{\mathrm{T}}\big) = \mathcal{C}\big([1 \quad a]^{\mathrm{T}}\big)$ for some $a \neq 0$, and $\boldsymbol{A} = A^{(1)}\boldsymbol{I}$. For situation (i), let $\boldsymbol{u} = [1 \quad 0]^{\mathrm{T}}$; for situation (ii), choose $\boldsymbol{u}$ satisfying $[1 \quad a]\ \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{u} = 0$. Define $e\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T;\mu_{0:T}\big)$ as the MMSE for inferring $\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T$ based on $\mathbf{z}^{(1)}_{0:T}$ and $\mathrm{r}_{0:T}$ if encoding strategy $\mu_{0:T}$ is employed. Moreover, define

$$\breve{e}\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T\big) := \inf_{\mu_{0:T}\in\mathcal{M}_T} e\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T;\mu_{0:T}\big) \tag{39}$$

where $\mathcal{M}_T$ represents the set of encoding strategies of horizon T. If $\{\breve{e}_T\}_{T\geqslant 0}$ is bounded, then $\{\breve{e}\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T\big)\}_{T\geqslant 0}$ can be shown to be also bounded. It will be shown in the following that $C > A^{(1)}$ must hold to ensure $\{\breve{e}\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T\big)\}_{T\geqslant 0}$ is bounded.

Note that node 1 computes the distributed filter by combining the received messages with its own sensor observations. This makes proving the necessity of Condition 2 challenging, as reasoning in existing literature where the receiver node does not obtain sensor observations cannot be applied directly. To address this challenge, consider the situation where node 2 can exploit observations obtained by node 1 in future time for generating transmitted messages. In particular, for a given horizon T, suppose node 2 can exploit the observations $\mathbf{z}^{(1)}_{0:T}$ for generating transmitted message at any $t \in [0, T]$. Therefore, the signal transmitted by node 2 at time t can be written as $\mu_{t,T}\big(\mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}, \mathrm{r}_{0:t}\big)$, where measurable function $\mu_{t,T}$ is referred to as a generalized encoding function. A collection of generalized encoding functions $\{\mu_{t,T}\}_{t\in[0,T]}$ is referred to as a generalized encoding strategy of horizon T if the power constraint (3) is satisfied with μ_t replaced by $\mu_{t,T}$. Define $\mu_{0:t,T} := \{\mu_{\tau,T}\}_{\tau\in[0,t]}$ for $t \in [0, T]$. If a generalized encoding strategy $\mu_{0:T,T}$ is employed, then the received message r_t is $\sigma\big(\mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}, \mathrm{w}_{0:t}\big)$-measurable. Define $e\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t;\mu_{0:t,T}\big)$ as the MMSE for inferring $\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t$ based on $\mathbf{z}^{(1)}_{0:T}$ and $\mathrm{r}_{0:t}$ if generalized encoding strategy $\mu_{0:T,T}$ is employed by node 2. Moreover, define $\breve{e}\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T;T\big)$ as

$$\breve{e}\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T;T\big) := \inf_{\mu_{0:T,T}\in\tilde{\mathcal{M}}_T} e\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T;\mu_{0:T,T}\big) \tag{40}$$

where $\tilde{\mathcal{M}}_T$ represents the set of generalized encoding strategies of horizon T. Since $\mathcal{M}_T \subseteq \tilde{\mathcal{M}}_T$, comparing (39) and (40) gives

$$\breve{e}\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T\big) \geqslant \breve{e}\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T;T\big)\,. \tag{41}$$

Define $\mathrm{y}_{t,T}$ as the MMSE estimator of $\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t$ based on $\mathbf{z}^{(1)}_{0:T}$ and $\mathbf{z}_{0:t}$, i.e.,

$$\mathrm{y}_{t,T} := \mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}\Big\} \qquad \text{for } t\in[0,T]\,. \tag{42}$$

In addition, define $e\big(\mathrm{y}_{t,T};\mu_{0:t,T}\big)$ as the MMSE for inferring $\mathrm{y}_{t,T}$ based on $\mathbf{z}^{(1)}_{0:T}$ and $\mathrm{r}_{0:t}$ if generalized encoding strategy $\mu_{0:T,T}$ is employed, i.e.,

$$e\big(\mathrm{y}_{t,T};\mu_{0:t,T}\big) := \mathbb{V}\Big\{\mathrm{y}_{t,T} - \mathbb{E}\Big\{\mathrm{y}_{t,T} \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\Big\}\Big\}\,.$$

It can be verified that $e\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t;\mu_{0:t,T}\big) \geqslant e\big(\mathrm{y}_{t,T};\mu_{0:t,T}\big)$. To see this, note that $\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t - \mathrm{y}_{t,T} \perp\!\!\!\perp \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}$ since the random vectors involved are jointly Gaussian. Moreover, $\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t - \mathrm{y}_{t,T} \perp\!\!\!\perp \mathrm{w}_{0:t} \,\big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}$ as $\mathrm{w}_{0:t} \perp\!\!\!\perp \mathbf{x}_t, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}$. Since $\mathrm{r}_{0:t}$ is $\sigma\big(\mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}, \mathrm{w}_{0:t}\big)$-measurable if $\mu_{0:T,T}$ is employed, Lemma 1 of [85] can be applied to conclude that $\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t - \mathrm{y}_{t,T} \perp\!\!\!\perp \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}, \mathrm{y}_{t,T}$. Therefore,

$$\begin{aligned} e\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t;\mu_{0:t,T}\big) &= \mathbb{E}\Big\{\Big(\mathrm{y}_{t,T} - \mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\Big\}\Big)^2\Big\} \\ &\quad + \mathbb{E}\Big\{\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t - \mathrm{y}_{t,T}\big)^2\Big\} \\ &\geqslant \mathbb{E}\Big\{\Big(\mathrm{y}_{t,T} - \mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\Big\}\Big)^2\Big\} \\ &\overset{\text{(a)}}{=} e\big(\mathrm{y}_{t,T};\mu_{0:t,T}\big) \end{aligned} \tag{43}$$

where equality (a) is because

$$\begin{aligned} &\mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\Big\} \\ &\quad = \mathbb{E}\Big\{\mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}, \mathrm{r}_{0:t}\Big\} \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\Big\} \\ &\quad = \mathbb{E}\Big\{\mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_t \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}\Big\} \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\Big\} \\ &\quad = \mathbb{E}\Big\{\mathrm{y}_{t,T} \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\Big\}\,. \end{aligned}$$

Combining (40), (41), and (43) gives

$$\breve{e}\big(\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_T\big) \geqslant \inf_{\mu_{0:T,T}\in\tilde{\mathcal{M}}_T} e\big(\mathrm{y}_{T,T};\mu_{0:T,T}\big)\,. \tag{44}$$

A lower bound of the right-hand side (RHS) of (44) is derived as follows. Choose t and s such that $0 \leqslant t \leqslant s \leqslant T$. According to the chain rule of mutual information,

$$\begin{aligned} I\big(\mathrm{y}_{s,T};\mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:s}\big) &= I\big(\mathrm{y}_{s,T};\mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\big) \\ &\quad + I\big(\mathrm{y}_{s,T};\mathrm{r}_{t:s} \,\big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\big)\,. \end{aligned} \tag{45}$$

The second term on the RHS of (45) can be shown to satisfy

$$I\big(\mathrm{y}_{s,T};\mathrm{r}_{t:s} \,\big|\, \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t}\big) \leqslant \int_t^s C\,d\tau = C(s-t) \tag{46}$$

where C is the channel capacity given by (5). To investigate the first term on the RHS of (45), the following conditional independence will be used

$$\mathrm{y}_{s,T} \perp\!\!\!\perp \mathbf{z}^{(1)}_{0:T}, \mathrm{r}_{0:t} \,\big|\, \mathrm{y}_{t,T}\,. \tag{47}$$

To see this, define a scaled innovation process $\{\boldsymbol{\eta}_{\tau,T}\}_{\tau\in[0,T]}$ as

$$\begin{aligned} d\,\boldsymbol{\eta}_{\tau,T} &= \big(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\big)^{-1/2}\Big(d\mathbf{z}^{(2)}_{\tau} - \boldsymbol{\Gamma}^{(2)}_{\tau}\mathbb{E}\Big\{\mathbf{x}_\tau \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:\tau}\Big\}\,d\tau\Big) \\ \boldsymbol{\eta}_{0,T} &= \mathbf{z}^{(2)}_0\,. \end{aligned}$$

Process $\{\boldsymbol{\eta}_{\tau,T} - \boldsymbol{\eta}_{0,T}\}_{\tau\in[0,T]}$ can be shown to be a Brownian motion that satisfies

$$\{\boldsymbol{\eta}_{\tau,T} - \boldsymbol{\eta}_{0,T}\}_{\tau\in[0,T]} \perp\!\!\!\perp \mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0,T}\,. \tag{48}$$

Moreover,

$$\sigma\big(\mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:\tau}\big) = \sigma\big(\mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0:\tau,T}\big) \qquad \forall\tau\in[0,T] \tag{49}$$

where $\boldsymbol{\eta}_{0:\tau,T} := \{\boldsymbol{\eta}_{\tilde{\tau},T}\}_{\tilde{\tau}\in[0,\tau]}$. Define q_τ for $\tau\in[0,T]$ as

$$\begin{aligned}\mathsf{q}_\tau &:= \mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_\tau \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:\tau}\Big\} - \mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\mathbf{x}_0 \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0}\Big\} \\ &\quad - \int_0^\tau \mathbb{E}\Big\{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\mathbf{x}_{\tilde\tau} \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:\tilde\tau}\Big\}\, d\tilde\tau \qquad (50)\\ &= \mathsf{y}_{\tau,T} - \mathsf{y}_{0,T} - \int_0^\tau A^{(1)}\mathsf{y}_{\tilde\tau,T}\, d\tilde\tau \qquad (51)\end{aligned}$$

where (51) is obtained using (37b) and (42). Equation (50) shows that q_τ is $\sigma\big(\mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:\tau}\big)$-measurable, and it is thus also $\sigma\big(\mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0:\tau,T}\big)$-measurable. Moreover, the independence

$$\mathsf{q}_{t_2} - \mathsf{q}_{t_1} \perp\!\!\!\perp \sigma\big(\mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t_1}\big) = \sigma\big(\mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0:t_1,T}\big) \qquad (52)$$

holds for all $0 \leqslant t_1 \leqslant t_2 \leqslant T$. In particular, since q_{t_2}, q_{t_1}, $\mathbf{z}^{(1)}_{0:T}$, $\mathbf{z}^{(2)}_{0:t_1}$ are jointly Gaussian, (52) holds if

$$\mathbb{E}\big\{\mathbf{z}^{(1)}_\tau\big(\mathsf{q}_{t_2} - \mathsf{q}_{t_1}\big)\big\} = \mathbf{0} \qquad \forall\tau \in [0,T] \qquad (53a)$$

$$\mathbb{E}\big\{\mathbf{z}^{(2)}_\tau\big(\mathsf{q}_{t_2} - \mathsf{q}_{t_1}\big)\big\} = \mathbf{0} \qquad \forall\tau \in [0,t_1] \qquad (53b)$$

according to the monotone class lemma [95, Appendix A1]. Let $\boldsymbol{\varphi}$ represent an arbitrary random vector that is $\sigma\big(\mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t_1}\big)$-measurable. We can derive (54), shown at the bottom of the page, where $\mathbf{v}_{\tilde t} := \Big[\big(\mathbf{v}^{(1)}_{\tilde t}\big)^{\mathrm{T}} \quad \big(\mathbf{v}^{(2)}_{\tilde t}\big)^{\mathrm{T}}\Big]^{\mathrm{T}}$. In particular, (54a) is obtained using (50) and the following relationship

$$\mathbb{E}\Big\{\mathbb{E}\Big\{\mathbf{x}_{\tilde t} \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:\tilde t}\Big\} \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t_1}\Big\} = \mathbb{E}\Big\{\mathbf{x}_{\tilde t} \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t_1}\Big\} \quad \forall \tilde t \geqslant t_1 .$$

Setting $\boldsymbol{\varphi} = \mathbf{z}^{(1)}_\tau$ in (54), applying Fubini's Theorem, and using the independence between $\{\mathbf{n}^{(1)}_t\}_{t\geqslant 0}$ and $\{\mathbf{v}_t\}_{t\geqslant 0}$, we obtain

$$\mathbb{E}\Big\{\mathbf{z}^{(1)}_\tau(\mathsf{q}_{t_2} - \mathsf{q}_{t_1})\Big\} = \int_0^\tau \mathbb{E}\Big\{\boldsymbol{\Gamma}^{(1)}\mathbf{x}_{\tilde\tau}\Big(\int_{t_1}^{t_2} \boldsymbol{B}\, d\mathbf{v}_{\tilde t}\Big)^{\mathrm{T}}\boldsymbol{u}\Big\}\, d\tilde\tau . \qquad (55)$$

Combining (1) with (8a), we can write $\mathbf{x}_{\tilde\tau}$ as

$$\mathbf{x}_{\tilde\tau} = \exp\{\boldsymbol{A}\tilde\tau\}\mathbf{x}_0 + \int_0^{\tilde\tau} \exp\{\boldsymbol{A}(\tilde\tau - \tilde s)\}\boldsymbol{B}\, d\mathbf{v}_{\tilde s} .$$

Consequently,

$$\begin{aligned}&\mathbb{E}\Big\{\boldsymbol{\Gamma}^{(1)}\mathbf{x}_{\tilde\tau}\Big(\int_{t_1}^{t_2} \boldsymbol{B}\, d\mathbf{v}_{\tilde t}\Big)^{\mathrm{T}}\boldsymbol{u}\Big\} \\ &= \int_{t_1}^{\min\{\tilde\tau, t_2\}} \boldsymbol{\Gamma}^{(1)}\exp\{\boldsymbol{A}(\tilde\tau - \tilde t)\}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{u}\, d\tilde t = \mathbf{0} \qquad (56)\end{aligned}$$

where the first and second equality are due to the Itô isometry and (37a), respectively. Substituting (56) into (55) gives (53a). Setting $\boldsymbol{\varphi} = \mathbf{z}^{(2)}_\tau$ in (54) and using the relationship $\mathbf{z}^{(2)}_\tau \perp\!\!\!\perp \int_{t_1}^{t_2} \boldsymbol{B}\, d\mathbf{v}_{\tilde t}$ for all $\tau \in [0, t_1]$, we obtain (53b), and thus (52) holds. In particular, (52) shows that $\mathsf{q}_\tau \perp\!\!\!\perp \mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0,T}$ as $\mathsf{q}_0 = 0$ by definition. Since q_τ is $\sigma\big(\mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0:\tau,T}\big)$-measurable, there exists a function $\boldsymbol{g}_T : [0,T] \mapsto \mathbb{R}^{1\times k_2}$, where k_2 is the dimension of $\mathbf{z}^{(2)}_\tau$, such that [93, Ch. 3]

$$\mathsf{q}_\tau = \int_0^\tau \boldsymbol{g}_T(\tilde\tau)\, d\boldsymbol{\eta}_{\tilde\tau,T} \qquad \forall\tau \in [0,T] . \qquad (57)$$

Combining (51) with (57) shows that $\{\mathsf{y}_{\tau,T}\}_{\tau\in[0,T]}$ satisfies

$$d\mathsf{y}_{\tau,T} = A^{(1)}\mathsf{y}_{\tau,T}\, d\tau + \boldsymbol{g}_T(\tau)\, d\boldsymbol{\eta}_{\tau,T} \qquad (58)$$

and thus $\mathsf{y}_{s,T}$ can be written as

$$\mathsf{y}_{s,T} = \exp\{A^{(1)}(s-t)\}\mathsf{y}_{t,T} + \int_t^s \exp\{A^{(1)}(s-\tau)\}\boldsymbol{g}_T(\tau)\, d\boldsymbol{\eta}_{\tau,T} .$$

Since $\{\boldsymbol{\eta}_{\tau,T} - \boldsymbol{\eta}_{0,T}\}_{\tau\in[0,T]}$ is a Brownian motion and satisfies (48),

$$\int_t^s \exp\Big\{A^{(1)}(s-\tau)\Big\}\boldsymbol{g}_T(\tau)\, d\boldsymbol{\eta}_{\tau,T} \perp\!\!\!\perp \sigma\big(\mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0:t,T}\big) = \sigma\big(\mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0:t,T}, \mathsf{y}_{t,T}\big)$$

where the equality is obtained based on (42) and (49). Applying [85, Lemma 2] gives $\mathsf{y}_{s,T} \perp\!\!\!\perp \mathbf{z}^{(1)}_{0:T}, \boldsymbol{\eta}_{0:t,T} \,|\, \mathsf{y}_{t,T}$. Since $\mathsf{w}_{0:t} \perp\!\!\!\perp \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t}, \mathsf{y}_{t,T}, \mathsf{y}_{s,T}$, Equation (47) is obtained by applying [85, Lemma 1].

Using [85, Lemma 3], the first term on the RHS of (45) satisfies

$$\begin{aligned}&I\big(\mathsf{y}_{s,T}; \mathbf{z}^{(1)}_{0:T}, \mathsf{r}_{0:t}\big) \\ &\leqslant f\Big(I\big(\mathsf{y}_{t,T}; \mathbf{z}^{(1)}_{0:T}, \mathsf{r}_{0:t}\big); \mathbb{V}\big\{[\mathsf{y}_{s,T} \quad \mathsf{y}_{t,T}]^{\mathrm{T}}\big\}\Big) . \qquad (59)\end{aligned}$$

Combining (45), (46), and (59) gives

$$\begin{aligned}&I\big(\mathsf{y}_{s,T}; \mathbf{z}^{(1)}_{0:T}, \mathsf{r}_{0:s}\big) - I\big(\mathsf{y}_{t,T}; \mathbf{z}^{(1)}_{0:T}, \mathsf{r}_{0:t}\big) \\ &\leqslant \varphi_T(s-t, t) - \varphi_T(0, t)\end{aligned}$$

where function φ_T is defined as

$$\begin{aligned}\varphi_T(\Delta, t) &:= f\Big(I\big(\mathsf{y}_{t,T}; \mathbf{z}^{(1)}_{0:T}, \mathsf{r}_{0:t}\big); \mathbb{V}\big\{[\mathsf{y}_{t+\Delta,T} \quad \mathsf{y}_{t,T}]^{\mathrm{T}}\big\}\Big) \\ &\quad + C\Delta \qquad \text{for } t \in [0,T], \quad \Delta \in [0, T-t]\end{aligned}$$

with $\varphi_T(0,t) = I\big(\mathsf{y}_{t,T}; \mathbf{z}^{(1)}_{0:T}, \mathsf{r}_{0:t}\big)$ in particular. Therefore,

$$\begin{aligned}&\frac{d}{dt} I\big(\mathsf{y}_{t,T}; \mathbf{z}^{(1)}_{0:T}, \mathsf{r}_{0:t}\big) \\ &\leqslant \lim_{s\searrow t} \frac{1}{s-t}\big(\varphi_T(s-t,t) - \varphi_T(0,t)\big) \\ &= C - \frac{1}{2}\mathbb{V}\{\mathsf{y}_{t,T}\}^{-1}\Big(\exp\Big\{2I\big(\mathsf{y}_{t,T}; \mathbf{z}^{(1)}_{0:T}, \mathsf{r}_{0:t}\big)\Big\} \\ &\qquad -1\Big)\boldsymbol{g}_T(t)\boldsymbol{g}_T^{\mathrm{T}}(t) \qquad (60)\end{aligned}$$

$$\mathbb{E}\big\{\boldsymbol{\varphi}(\mathsf{q}_{t_2} - \mathsf{q}_{t_1})\big\} = \mathbb{E}\Big\{\mathbb{E}\Big\{\boldsymbol{\varphi}\boldsymbol{u}^{\mathrm{T}}\Big(\mathbf{x}_{t_2} - \mathbf{x}_{t_1} - \int_{t_1}^{t_2} \boldsymbol{A}\mathbf{x}_{\tilde t}\, d\tilde t\Big) \,\Big|\, \mathbf{z}^{(1)}_{0:T}, \mathbf{z}^{(2)}_{0:t_1}\Big\}\Big\} \qquad (54a)$$

$$= \mathbb{E}\Big\{\boldsymbol{\varphi}\boldsymbol{u}^{\mathrm{T}}\Big(\mathbf{x}_{t_2} - \mathbf{x}_{t_1} - \int_{t_1}^{t_2} \boldsymbol{A}\mathbf{x}_{\tilde t}\, d\tilde t\Big)\Big\} = \mathbb{E}\Big\{\boldsymbol{\varphi}\boldsymbol{u}^{\mathrm{T}}\int_{t_1}^{t_2} \boldsymbol{B}\, d\mathbf{v}_{\tilde t}\Big\} \qquad (54b)$$

where $s \searrow t$ represents that s approaches t from above. Define function $I_T(t)$ as the solution to the following initial value problem on $[0, T]$

$$\frac{d}{dt} I_T(t) = C - \frac{1}{2} \mathbb{V}\{\mathsf{y}_{t,T}\}^{-1} \Big(\exp\{2 I_T(t)\} - 1\Big) \boldsymbol{g}_T(t) \boldsymbol{g}_T^{\mathrm{T}}(t) \tag{61a}$$

$$I_T(0) = I\big(\mathsf{y}_{0,T}; \mathbf{z}_{0:T}^{(1)}\big) . \tag{61b}$$

Comparing (60) with (61) and applying [94, Ch. 3, Th. 4.1] gives $I\big(\mathsf{y}_{T,T}; \mathbf{z}_{0:T}^{(1)}, \mathsf{r}_{0:T}\big) \leqslant I_T(T)$. Applying [85, Lemma 4],

$$\begin{aligned} e\big(\mathsf{y}_{T,T}; \mu_{0:T,T}\big) &\geqslant \mathbb{V}\{\mathsf{y}_{T,T}\} \exp\Big\{-2I\big(\mathsf{y}_{T,T}; \mathbf{z}_{0:T}^{(1)}, \mathsf{r}_{0:T}\big)\Big\} \\ &\geqslant \mathbb{V}\{\mathsf{y}_{T,T}\} \exp\big\{-2I_T(T)\big\} . \end{aligned} \tag{62}$$

Moreover, using (58) and (61), $\mathbb{V}\{\mathsf{y}_{T,T}\} \exp\{-2I_T(T)\}$ can be written as

$$\begin{aligned} &\mathbb{V}\{\mathsf{y}_{T,T}\} \exp\big\{-2I_T(T)\big\} \\ &= \mathbb{V}\{\mathsf{y}_{0,T}\} \exp\Big\{-2I\big(\mathsf{y}_{0,T}; \mathbf{z}_{0:T}^{(1)}\big) + 2\big(A^{(1)} - C\big)T\Big\} \\ &\quad + \int_0^T \exp\big\{2\big(A^{(1)} - C\big)(T - \tau)\big\} \boldsymbol{g}_T(\tau) \boldsymbol{g}_T^{\mathrm{T}}(\tau)\, d\tau . \end{aligned} \tag{63}$$

Combining (62) and (63), if $C \leqslant A^{(1)}$, then

$$e\big(\mathsf{y}_{T,T}; \mu_{0:T,T}\big) \geqslant \int_0^T \boldsymbol{g}_T(\tau) \boldsymbol{g}_T^{\mathrm{T}}(\tau)\, d\tau . \tag{64}$$

Using (57) and the Itô isometry, and then using (50),

$$\begin{aligned} &\int_0^T \boldsymbol{g}_T(\tau) \boldsymbol{g}_T^{\mathrm{T}}(\tau)\, d\tau \\ &\quad = \mathbb{E}\{(\mathsf{q}_T - \mathsf{q}_0)^2\} \\ &\quad = \mathbb{E}\bigg\{\Big(\mathsf{y}_{T,T} - \mathsf{y}_{0,T} - \int_0^T A^{(1)} \mathsf{y}_{\tau,T}\, d\tau\Big)^2\bigg\} . \end{aligned} \tag{65}$$

According to (58), the following holds

$$\mathbb{E}\{\mathsf{y}_{t,T}\ \mathsf{y}_{s,T}\} = \exp\{A^{(1)}(s - t)\} \mathbb{V}\{\mathsf{y}_{s,T}\} \tag{66}$$

for all t and s such that $0 \leqslant t \leqslant s \leqslant T$. Substituting (66) into (65),

$$\begin{aligned} \int_0^T \boldsymbol{g}_T(\tau) \boldsymbol{g}_T^{\mathrm{T}}(\tau)\, d\tau &= \mathbb{V}\{\mathsf{y}_{T,T}\} - \mathbb{V}\{\mathsf{y}_{0,T}\} \\ &\quad - 2A^{(1)} \int_0^T \mathbb{V}\{\mathsf{y}_{\tau,T}\}\, d\tau . \end{aligned} \tag{67}$$

Recalling the definition (42), and using the law of total covariance,

$$\mathbb{V}\{\mathsf{y}_{\tau,T}\} = \mathbb{V}\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_\tau\} - \mathbb{V}\Big\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_\tau \,\Big|\, \mathbf{z}_{0:T}^{(1)}, \mathbf{z}_{0:\tau}^{(2)}\Big\} \quad \forall \tau \in [0, T] . \tag{68}$$

In addition, combining (1) with (8a) and (37b),

$$\begin{aligned} \mathbb{V}\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_T\} &= 2A^{(1)} \int_0^T \mathbb{V}\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_\tau\}\, d\tau \\ &\quad + \boldsymbol{u}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{u} T + \mathbb{V}\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_0\} . \end{aligned} \tag{69}$$

Substituting (68) and (69) into (67) gives

$$\begin{aligned} &\int_0^T \boldsymbol{g}_T(\tau) \boldsymbol{g}_T^{\mathrm{T}}(\tau)\, d\tau \\ &\quad = \boldsymbol{u}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{u} T - \mathbb{V}\Big\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_T \,\Big|\, \mathbf{z}_{0:T}^{(1)}, \mathbf{z}_{0:T}^{(2)}\Big\} \\ &\quad\quad + \mathbb{V}\Big\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_0 \,\Big|\, \mathbf{z}_{0:T}^{(1)}, \mathbf{z}_0^{(2)}\Big\} \\ &\quad\quad + 2A^{(1)} \int_0^T \mathbb{V}\Big\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_\tau \,\Big|\, \mathbf{z}_{0:T}^{(1)}, \mathbf{z}_{0:\tau}^{(2)}\Big\}\, d\tau . \end{aligned} \tag{70}$$

Combining (44), (64), and (70)

$$\begin{aligned} \breve{e}\big(\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_T\big) &\geqslant \int_0^T \boldsymbol{g}_T(\tau) \boldsymbol{g}_T^{\mathrm{T}}(\tau)\, d\tau \\ &\geqslant \boldsymbol{u}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{u} T - \mathbb{V}\Big\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_T \,\Big|\, \mathbf{z}_{0:T}^{(1)}, \mathbf{z}_{0:T}^{(2)}\Big\} . \end{aligned} \tag{71}$$

The set $\big\{\mathbb{V}\{\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_T \,|\, \mathbf{z}_{0:T}^{(1)}, \mathbf{z}_{0:T}^{(2)}\}\big\}_{T \geqslant 0}$ can be verified to be bounded. Therefore, if $C \leqslant A^{(1)}$, then $\big\{\breve{e}\big(\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_T\big)\big\}_{T \geqslant 0}$ is unbounded according to (71). However, recall that if $\{\breve{e}_T\}_{T \geqslant 0}$ is bounded, then $\big\{\breve{e}\big(\boldsymbol{u}^{\mathrm{T}} \mathbf{x}_T\big)\big\}_{T \geqslant 0}$ must be bounded. This shows that if Condition 1 of Proposition 1 does not hold, then $C > A^{(1)}$ is necessary to ensure that $\{\breve{e}_T\}_{T \geqslant 0}$ is bounded. ☒

Appendix B
Proof of Proposition 2

A lemma to be used in the proof is presented first.

Lemma 2: Consider vector processes $\{\boldsymbol{\theta}_t\}_{t \geqslant 0}$ and $\{\boldsymbol{\xi}_t\}_{t \geqslant 0}$ described by the following SDE

$$d\boldsymbol{\theta}_t = \boldsymbol{A}_t \boldsymbol{\theta}_t\, dt + \boldsymbol{B}_t\, d\boldsymbol{\nu}_t \tag{72a}$$

$$d\boldsymbol{\xi}_t = \Big(\boldsymbol{G}_t \boldsymbol{\theta}_t + \boldsymbol{g}_t\big(\boldsymbol{\xi}_{0:t}\big)\Big)\, dt + \boldsymbol{F}_t\, d\boldsymbol{\nu}_t + \boldsymbol{L}_t\, d\boldsymbol{\omega}_t \tag{72b}$$

where $\boldsymbol{A}_t$, $\boldsymbol{B}_t$, $\boldsymbol{G}_t$, $\boldsymbol{F}_t$, and $\boldsymbol{L}_t$ are deterministic matrices such that $\boldsymbol{L}_t \boldsymbol{L}_t^{\mathrm{T}}$ is invertible for all $t \geqslant 0$; $\boldsymbol{g}_t$ is a linear function; $\{\boldsymbol{\nu}_t\}_{t \geqslant 0}$ and $\{\boldsymbol{\omega}_t\}_{t \geqslant 0}$ are independent Brownian motions. Moreover, $\boldsymbol{\theta}_0$ and $\boldsymbol{\xi}_0$ are jointly Gaussian and are independent of $\{\boldsymbol{\nu}_t\}_{t \geqslant 0}$ and $\{\boldsymbol{\omega}_t\}_{t \geqslant 0}$. Suppose regularity conditions for the existence of a strong solution to SDE (72) hold (see [84, Ch. 12]). Then processes $\big\{\hat{\boldsymbol{\theta}}_t\big\}_{t \geqslant 0}$ and $\big\{\tilde{\boldsymbol{\theta}}_t\big\}_{t \geqslant 0}$ are both Markov, where $\hat{\boldsymbol{\theta}}_t := \mathbb{E}\big\{\boldsymbol{\theta}_t \,\big|\, \boldsymbol{\xi}_{0:t}\big\}$ and $\tilde{\boldsymbol{\theta}}_t := \boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_t$.

Proof: According to [84, Th. 12.7], process $\big\{\hat{\boldsymbol{\theta}}_t\big\}_{t \geqslant 0}$ satisfies

$$d\hat{\boldsymbol{\theta}}_t = \boldsymbol{A}_t \hat{\boldsymbol{\theta}}_t\, dt + \boldsymbol{H}_t \Big(d\boldsymbol{\xi}_t - \Big(\boldsymbol{G}_t \hat{\boldsymbol{\theta}}_t + \boldsymbol{g}_t\big(\boldsymbol{\xi}_{0:t}\big)\Big)\, dt\Big) \tag{73}$$

where $\boldsymbol{H}_t$ is defined as

$$\boldsymbol{H}_t := \big(\boldsymbol{B}_t \boldsymbol{F}_t^{\mathrm{T}} + \mathbb{V}\{\boldsymbol{\theta}_t \,|\, \boldsymbol{\xi}_{0:t}\} \boldsymbol{G}_t^{\mathrm{T}}\big) \big(\boldsymbol{F}_t \boldsymbol{F}_t^{\mathrm{T}} + \boldsymbol{L}_t \boldsymbol{L}_t^{\mathrm{T}}\big)^{-1} .$$

Combination of (72) and (73) shows that $\big\{\tilde{\boldsymbol{\theta}}_t\big\}_{t \geqslant 0}$ satisfies the following SDE

$$d\tilde{\boldsymbol{\theta}}_t = \big(\boldsymbol{A}_t - \boldsymbol{H}_t \boldsymbol{G}_t\big) \tilde{\boldsymbol{\theta}}_t\, dt + \big(\boldsymbol{B}_t - \boldsymbol{H}_t \boldsymbol{F}_t\big)\, d\boldsymbol{\nu}_t - \boldsymbol{H}_t \boldsymbol{L}_t\, d\boldsymbol{\omega}_t .$$

Therefore, $\big\{\tilde{\boldsymbol{\theta}}_t\big\}_{t \geqslant 0}$ is a Markov process [82, Ch. 7]. Moreover, define a scaled innovation process $\{\boldsymbol{\eta}_t\}_{t \geqslant 0}$ as

$$d\boldsymbol{\eta}_t = \Big(\boldsymbol{F}_t \boldsymbol{F}_t^{\mathrm{T}} + \boldsymbol{L}_t \boldsymbol{L}_t^{\mathrm{T}}\Big)^{-1/2} \Big(d\boldsymbol{\xi}_t - \Big(\boldsymbol{G}_t \hat{\boldsymbol{\theta}}_t + \boldsymbol{g}_t\big(\boldsymbol{\xi}_{0:t}\big)\Big)\, dt\Big) \tag{74a}$$

$$\boldsymbol{\eta}_0 = \boldsymbol{\xi}_0 . \tag{74b}$$

Process $\{\boldsymbol{\eta}_t - \boldsymbol{\eta}_0\}_{t \geqslant 0}$ can be shown to be a Brownian motion. Combining (73) and (74), process $\big\{\hat{\boldsymbol{\theta}}_t\big\}_{t \geqslant 0}$ satisfies

$$d\hat{\boldsymbol{\theta}}_t = \boldsymbol{A}_t \hat{\boldsymbol{\theta}}_t\, dt + \boldsymbol{H}_t \Big(\boldsymbol{F}_t \boldsymbol{F}_t^{\mathrm{T}} + \boldsymbol{L}_t \boldsymbol{L}_t^{\mathrm{T}}\Big)^{1/2} d\boldsymbol{\eta}_t .$$

Therefore, $\{\hat{\boldsymbol{\theta}}_t\}_{t\geqslant 0}$ is also a Markov process. ☒

The three statements in Proposition 2 are proved next.

Proof:

Statement 1: Setting $\boldsymbol{\theta}_t = \mathbf{x}_t$ and $\boldsymbol{\xi}_t = \left[(\mathbf{z}_t^{(1)})^{\mathrm{T}} \quad (\mathbf{z}_t^{(2)})^{\mathrm{T}}\right]^{\mathrm{T}}$ in Lemma 2 shows that $\{\tilde{\mathbf{x}}_t\}_{t\geqslant 0}$ is a Markov process. Setting $i=1$ in (22b) and using (24), we obtain

$$d\mathbf{z}_t^{(1)} = \boldsymbol{\Gamma}^{(1)}\mathbf{y}_t\,dt + \left(\boldsymbol{\Xi}^{(1)}(\boldsymbol{\Xi}^{(1)})^{\mathrm{T}}\right)^{1/2} d\boldsymbol{\eta}_t^{(1)}. \tag{75}$$

Moreover, process $\{\mathsf{r}_t\}_{t\geqslant 0}$ satisfies

$$d\mathsf{r}_t = \left(\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{y}_t + g_t(\mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t})\right)dt + \kappa\, d\mathsf{w}_t. \tag{76}$$

Combining (24), (75), as well as (76), and applying Lemma 2, $\{\mathbb{E}\{\mathbf{y}_t \,|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\}\}_{t\geqslant 0}$ and $\{\mathbf{y}_t - \mathbb{E}\{\mathbf{y}_t \,|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\}\}_{t\geqslant 0}$ are shown to be Markov processes. Note that

$$\begin{aligned}\hat{\mathbf{x}}_t &= \mathbb{E}\left\{\mathbb{E}\left\{\mathbf{x}_t \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{r}_{0:t}\right\} \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\right\}\\ &\overset{(a)}{=} \mathbb{E}\left\{\mathbb{E}\left\{\mathbf{x}_t \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}\right\} \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\right\}\\ &= \mathbb{E}\left\{\mathbf{y}_t \,\middle|\, \mathbf{z}_{0:t}^{(1)}, \mathsf{r}_{0:t}\right\}\end{aligned} \tag{77}$$

where equality (a) is obtained using the relationship

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

which is proved in [85, Lemma 5]. Therefore, $\{\tilde{\mathbf{y}}_t\}_{t\geqslant 0}$ and $\{\hat{\mathbf{x}}_t\}_{t\geqslant 0}$ are Markov processes.

Statement 2: Random vectors $\tilde{\mathbf{x}}_t$, $\tilde{\mathbf{y}}_t$, and $\hat{\mathbf{x}}_t$ can be shown to be independent. Specifically, using the independence between the channel noise process and the state disturbance and sensor observation processes, we can show that

$$\mathbf{x}_t, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)} \perp\!\!\!\perp \mathsf{w}_{0:t}. \tag{78}$$

Moreover, (8b) and (11a) shows that $\tilde{\mathbf{x}}_t$ is $\sigma(\mathbf{x}_t, \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)})$-measurable. Combining this with (78), we obtain $\tilde{\mathbf{x}}_t \perp\!\!\!\perp \mathsf{w}_{0:t}$. In addition, (8b) and (11a) indicate that $\tilde{\mathbf{x}}_t \perp\!\!\!\perp \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}$. Therefore, $\tilde{\mathbf{x}}_t \perp\!\!\!\perp \mathbf{z}_{0:t}^{(1)}, \mathbf{z}_{0:t}^{(2)}, \mathsf{w}_{0:t}$ as the involved random quantities are jointly Gaussian. Since both $\tilde{\mathbf{y}}_t$ and $\hat{\mathbf{x}}_t$ are $\sigma(\mathbf{z}_t^{(1)}, \mathbf{z}_t^{(2)}, \mathsf{w}_{0:t})$-measurable, the relationship $\tilde{\mathbf{x}}_t \perp\!\!\!\perp \tilde{\mathbf{y}}_t, \hat{\mathbf{x}}_t$ holds. In addition, combining (11b) with (77) shows that $\tilde{\mathbf{y}}_t \perp\!\!\!\perp \hat{\mathbf{x}}_t$. Therefore, $\tilde{\mathbf{x}}_t$, $\tilde{\mathbf{y}}_t$, and $\hat{\mathbf{x}}_t$ are independent as they are jointly Gaussian. As a result, $\mathbb{V}\{\mathbf{s}_t\}^{-1}$ can be written as (79), which is shown at the bottom of the page. Therefore,

$$\begin{aligned}\frac{1}{2}\mathbf{s}_t^{\mathrm{T}}\mathbb{V}\{\mathbf{s}_\tau\}^{-1}\mathbf{s}_t = \frac{1}{2}\Big(&\tilde{\mathbf{x}}_t^{\mathrm{T}}\mathbb{V}\{\tilde{\mathbf{x}}_\tau\}^{-1}\tilde{\mathbf{x}}_t + \tilde{\mathbf{y}}_t^{\mathrm{T}}\mathbb{V}\{\tilde{\mathbf{y}}_\tau\}^{-1}\tilde{\mathbf{y}}_t\\ &+ \hat{\mathbf{x}}_t^{\mathrm{T}}\mathbb{V}\{\hat{\mathbf{x}}_\tau\}^{-1}\hat{\mathbf{x}}_t\Big).\end{aligned}$$

Letting τ in the above equation approach infinity and taking the expectation, the equation $E_t^{\mathbf{s}} = E_t^{\tilde{\mathbf{x}}} + E_t^{\tilde{\mathbf{y}}} + E_t^{\hat{\mathbf{x}}}$ is obtained.

Statement 3: Since $\tilde{\mathbf{x}}_t$, $\tilde{\mathbf{y}}_t$, and $\hat{\mathbf{x}}_t$ are independent, the following holds

$$\begin{aligned}S_t^{\mathbf{s}} = -D(P_{\mathbf{s}_t} \,\|\, \lambda_{\mathrm{L}}) &= -D(P_{[\tilde{\mathbf{x}}_t^{\mathrm{T}} \quad \tilde{\mathbf{y}}_t^{\mathrm{T}} \quad \hat{\mathbf{x}}_t^{\mathrm{T}}]^{\mathrm{T}}} \,\|\, \lambda_{\mathrm{L}})\\ &= S_t^{\tilde{\mathbf{x}}} + S_t^{\tilde{\mathbf{y}}} + S_t^{\hat{\mathbf{x}}}.\end{aligned}$$

This is the desired result. ☒

Appendix C
Derivation of Energy Variation and Energy Flow Rates

Using Kalman–Bucy filtering results, we obtain

$$\frac{d}{dt}\mathbb{V}\{\mathbf{x}_t\} = \boldsymbol{A}\mathbb{V}\{\mathbf{x}_t\} + \mathbb{V}\{\mathbf{x}_t\}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} \tag{80a}$$

$$\begin{aligned}\frac{d}{dt}\mathbb{V}\{\tilde{\mathbf{x}}_t\} &= \boldsymbol{A}\mathbb{V}\{\tilde{\mathbf{x}}_t\} + \mathbb{V}\{\tilde{\mathbf{x}}_t\}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\\ &\quad - \mathbb{V}\{\tilde{\mathbf{x}}_t\}\boldsymbol{\Gamma}^{\mathrm{T}}(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}})^{-1}\boldsymbol{\Gamma}\mathbb{V}\{\tilde{\mathbf{x}}_t\}\end{aligned} \tag{80b}$$

$$\begin{aligned}\frac{d}{dt}\mathbb{V}\{\mathbf{y}_t\} &= \boldsymbol{A}\mathbb{V}\{\mathbf{y}_t\} + \mathbb{V}\{\mathbf{y}_t\}\boldsymbol{A}^{\mathrm{T}}\\ &\quad + \mathbb{V}\{\tilde{\mathbf{x}}_t\}\boldsymbol{\Gamma}^{\mathrm{T}}(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}})^{-1}\boldsymbol{\Gamma}\mathbb{V}\{\tilde{\mathbf{x}}_t\}\end{aligned} \tag{80c}$$

where

$$\boldsymbol{A} := \begin{bmatrix} A^{(1)} & \mathbf{0} \\ \mathbf{0} & A^{(2)} \end{bmatrix} \qquad \boldsymbol{B} := \begin{bmatrix} \boldsymbol{B}^{(1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}^{(2)} \end{bmatrix}$$
$$\boldsymbol{\Gamma} := \begin{bmatrix} \boldsymbol{\Gamma}^{(1)} \\ \boldsymbol{\Gamma}^{(2)} \end{bmatrix} \qquad \boldsymbol{\Xi} := \begin{bmatrix} \boldsymbol{\Xi}^{(1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Xi}^{(2)} \end{bmatrix}.$$

$$\mathbb{V}\{\mathbf{s}_t\}^{-1} = \begin{bmatrix} \mathbb{V}\{\tilde{\mathbf{x}}_t\}^{-1} & -\mathbb{V}\{\tilde{\mathbf{x}}_t\}^{-1} & \mathbf{0} \\ -\mathbb{V}\{\tilde{\mathbf{x}}_t\}^{-1} & \mathbb{V}\{\tilde{\mathbf{y}}_t\}^{-1} + \mathbb{V}\{\tilde{\mathbf{x}}_t\}^{-1} & -\mathbb{V}\{\tilde{\mathbf{y}}_t\}^{-1} \\ \mathbf{0} & -\mathbb{V}\{\tilde{\mathbf{y}}_t\}^{-1} & \mathbb{V}\{\hat{\mathbf{x}}_t\}^{-1} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}^{-1} \end{bmatrix}. \tag{79}$$

$$\frac{d}{dt}E_t^{\tilde{\mathbf{x}}} = \frac{1}{2}\mathrm{tr}\left\{\left[2\boldsymbol{A}\mathbb{V}\{\tilde{\mathbf{x}}_t\} + \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} - \mathbb{V}\{\tilde{\mathbf{x}}_t\}\boldsymbol{\Gamma}^{\mathrm{T}}(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}})^{-1}\boldsymbol{\Gamma}\mathbb{V}\{\tilde{\mathbf{x}}_t\}\right]\boldsymbol{\Sigma}_{\tilde{\mathbf{x}}}^{-1}\right\} \tag{82a}$$

$$\begin{aligned}\frac{d}{dt}E_t^{\tilde{\mathbf{y}}} = \frac{1}{2}\mathrm{tr}\Big\{\Big[&2\boldsymbol{A}\mathbb{V}\{\tilde{\mathbf{y}}_t\} + \mathbb{V}\{\tilde{\mathbf{x}}_t\}\boldsymbol{\Gamma}^{\mathrm{T}}(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}})^{-1}\boldsymbol{\Gamma}\mathbb{V}\{\tilde{\mathbf{x}}_t\} - \mathbb{V}\{\tilde{\mathbf{y}}_t\}\boldsymbol{\beta}_t\kappa^{-2}\boldsymbol{\beta}_t^{\mathrm{T}}\mathbb{V}\{\tilde{\mathbf{y}}_t\}\\ &- \left(\mathbb{V}\{\tilde{\mathbf{x}}_t\} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}\right)(\boldsymbol{\Gamma}^{(1)})^{\mathrm{T}}\left(\boldsymbol{\Xi}^{(1)}(\boldsymbol{\Xi}^{(1)})^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}^{(1)}\left(\mathbb{V}\{\tilde{\mathbf{x}}_t\} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}\right)\Big]\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}}^{-1}\Big\}\end{aligned} \tag{82b}$$

$$\begin{aligned}\frac{d}{dt}E_t^{\hat{\mathbf{x}}} = \frac{1}{2}\mathrm{tr}\Big\{\Big[&2\boldsymbol{A}\mathbb{V}\{\hat{\mathbf{x}}_t\} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}\boldsymbol{\beta}_t\kappa^{-2}\boldsymbol{\beta}_t^{\mathrm{T}}\mathbb{V}\{\tilde{\mathbf{y}}_t\}\\ &+ \left(\mathbb{V}\{\tilde{\mathbf{x}}_t\} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}\right)(\boldsymbol{\Gamma}^{(1)})^{\mathrm{T}}\left(\boldsymbol{\Xi}^{(1)}(\boldsymbol{\Xi}^{(1)})^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}^{(1)}\left(\mathbb{V}\{\tilde{\mathbf{x}}_t\} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}\right)\Big]\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}^{-1}\Big\}.\end{aligned} \tag{82c}$$

According to optimal filtering results for the system described by (24), (75), and (76),

$$\frac{d}{dt}\mathbb{V}\{\tilde{\mathbf{y}}_t\} = \boldsymbol{A}\mathbb{V}\{\tilde{\mathbf{y}}_t\} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}\boldsymbol{A}^{\mathrm{T}} - \mathbb{V}\{\tilde{\mathbf{y}}_t\}\boldsymbol{\beta}_t\kappa^{-2}\boldsymbol{\beta}_t^{\mathrm{T}}\mathbb{V}\{\tilde{\mathbf{y}}_t\} + \mathbb{V}\{\tilde{\mathbf{x}}_t\}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Gamma}\mathbb{V}\{\tilde{\mathbf{x}}_t\} - \left(\mathbb{V}\{\tilde{\mathbf{x}}_t\} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}\right)\left(\boldsymbol{\Gamma}^{(1)}\right)^{\mathrm{T}}\left(\boldsymbol{\Xi}^{(1)}\left(\boldsymbol{\Xi}^{(1)}\right)^{\mathrm{T}}\right)^{-1} \times \boldsymbol{\Gamma}^{(1)}\left(\mathbb{V}\{\tilde{\mathbf{x}}_t\} + \mathbb{V}\{\tilde{\mathbf{y}}_t\}\right) \tag{81a}$$

$$\frac{d}{dt}\mathbb{V}\{\hat{\mathbf{x}}_t\} = \frac{d}{dt}\mathbb{V}\{\mathbf{y}_t\} - \frac{d}{dt}\mathbb{V}\{\tilde{\mathbf{y}}_t\}. \tag{81b}$$

Combining (80b) and (81) with (12) gives (82), shown at the bottom of the previous page. The rates of energy flows are obtained by combining (14)–(18) with (80a), (80c), and (82). The results are given in of [92, Appendix A.3.3, eq. (A.91)].

REFERENCES

[1] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 146–181, Mar. 1974.

[2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[3] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for Online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

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

[5] 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, Nov. 2019.

[6] S. Gezici et al., "Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 70–84, Jul. 2005.

[7] 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.

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

[9] M. Chiani, A. Giorgetti, and E. Paolini, "Sensor radar for object tracking," *Proc. IEEE*, vol. 106, no. 6, pp. 1022–1041, Jun. 2018.

[10] J. Thomas, J. Welde, G. Loianno, K. Daniilidis, and V. Kumar, "Autonomous flight for detection, localization, and tracking of moving targets with a small quadrotor," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1762–1769, Jul. 2017.

[11] D. Wu, D. Chatzigeorgiou, K. Youcef-Toumi, and R. Ben-Mansour, "Node localization in robotic sensor networks for pipeline inspection," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 809–819, Apr. 2016.

[12] R. Karlsson and F. Gustafsson, "The future of automotive localization algorithms: Available, reliable, and scalable localization: Anywhere and anytime," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 60–69, Mar. 2017.

[13] G. Zhan and W. Shi, "LOBOT: Low-cost, self-contained localization of small-sized ground robotic vehicles," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 744–753, Apr. 2013.

[14] S. G. Nagarajan, P. Zhang, and I. Nevat, "Geo-spatial location estimation for Internet of Things (IoT) networks with one-way time-of-arrival via stochastic censoring," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 205–214, Feb. 2017.

[15] N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Data collection and wireless communication in Internet of Things (IoT) using economic analysis and pricing models: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2546–2590, 4th Quart., 2016.

[16] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient multi-sensor localization for the Internet of Things: Exploring a new class of scalable localization algorithms," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 153–167, Sep. 2018.

[17] S. Alletto et al., "An indoor location-aware system for an IoT-based smart museum," *IEEE Internet Things J.*, vol. 3, no. 2, pp. 244–253, Apr. 2016.

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

[19] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proc. IEEE*, vol. 107, no. 12, pp. 2327–2375, Dec. 2019.

[20] F. Morselli, S. M. Razavi, M. Z. Win, and A. Conti, "Soft information based localization for 5G networks and beyond," *IEEE Trans. Wireless Commun.*, early access, May 16, 2023, doi: 10.1109/TWC.2023.3275122.

[21] G. Torsoli, M. Z. Win, and A. Conti, "Blockage intelligence in complex environments for beyond 5G localization," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1688–1701, Jun. 2023.

[22] M. Z. Win, W. Dai, Y. Shen, G. Chrisikos, and H. V. Poor, "Network operation strategies for efficient localization and navigation," *Proc. IEEE*, vol. 106, no. 7, pp. 1224–1254, Jul. 2018.

[23] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.

[24] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 89–103, Jan. 2005.

[25] Y. Shen, W. Dai, and M. Z. Win, "Power optimization for network localization," *IEEE/ACM Trans. Netw.*, vol. 22, no. 4, pp. 1337–1350, Aug. 2014.

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

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

[28] Y. Shen and M. Z. Win, "Fundamental limits of wideband localization–Part I: A general framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, Oct. 2010.

[29] Y. Shen, H. Wymeersch, and M. Z. Win, "Fundamental limits of wideband localization–Part II: Cooperative networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4981–5000, Oct. 2010.

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

[31] M. H. A. Davis, "Detection, mutual information and feedback encoding: Applications of stochastic calculus," in *Communication Systems and Random Process Theory*, J. Skwirzynski, Ed. Dordrecht, The Netherlands: Springer, 1978, pp. 705–720.

[32] T. T. Kadota, M. Zakai, and J. Ziv, "Mutual information of the white Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 17, no. 4, pp. 368–371, Jul. 1971.

[33] E. Mayer-Wolf and M. Zakai, "On a formula relating the Shannon information to the Fisher information for the filtering problem," in *Filtering and Control of Random Processes* (Lecture Notes in Control and Information Sciences), vol. 61. New York, NY, USA: Springer, 1984, pp. 164–171.

[34] Y. Shen, S. Mazuelas, and M. Z. Win, "Network navigation: Theory and interpretation," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 9, pp. 1823–1834, Oct. 2012.

[35] 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.

[36] K. Reif, S. Günther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter," *IEEE Trans. Autom. Control*, vol. 44, no. 4, pp. 714–728, Apr. 1999.

[37] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Trans. Autom. Control*, vol. 48, no. 2, pp. 246–258, Feb. 2003.

[38] S. Das and J. M. F. Moura, "Consensus+innovations distributed Kalman filter with optimized gains," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 467–481, Jan. 2017.

[39] S. Kar and J. M. F. Moura, "Consensus + innovations distributed inference over networks: Cooperation and sensing in networked systems," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 99–109, May 2013.

[40] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.

[41] U. A. Khan, S. Kar, and J. M. F. Moura, "Distributed sensor localization in random environments using minimal number of anchor nodes," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 2000–2016, May 2009.

[42] Z. Liu, W. Dai, and M. Z. Win, "Mercury: An infrastructure-free system for network localization and navigation," *IEEE Trans. Mobile Comput.*, vol. 17, no. 5, pp. 1119–1133, May 2018.

[43] B. Teague, Z. Liu, F. Meyer, A. Conti, and M. Z. Win, "Network localization and navigation with scalable inference and efficient operation," *IEEE Trans. Mobile Comput.*, vol. 21, no. 6, pp. 2072–2087, Jun. 2022.

[44] F. S. Cattivelli and A. H. Sayed, "Analysis of spatial and incremental LMS processing for distributed estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1465–1480, Apr. 2011.

[45] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2069–2084, Sep. 2010.

[46] A. A. Saucan and M. Z. Win, "Information-seeking sensor selection for ocean-of-things," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10072–10088, Oct. 2020.

[47] G. Papa, P. Braca, S. Horn, S. Marano, V. Matta, and P. Willett, "Multisensor adaptive Bayesian tracking under time-varying target detection probability," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2193–2209, Oct. 2016.

[48] P. Braca, P. Willett, K. LePage, S. Marano, and V. Matta, "Bayesian tracking in underwater wireless sensor networks with port-starboard ambiguity," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1864–1878, Apr. 2014.

[49] A. T. Ihler, J. W. Fisher III, R. L. Moses, and A. S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 809–819, Apr. 2005.

[50] J. Prieto, S. Mazuelas, and M. Z. Win, "Context-aided inertial navigation via belief condensation," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3250–3261, Jun. 2016.

[51] S. Mazuelas, Y. Shen, and M. Z. Win, "Belief condensation filtering," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4403–4415, Sep. 2013.

[52] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Eng.*, vol. 83, pp. 95–108, Mar. 1961.

[53] P. L. Falb, "Infinite-dimensional filtering: The Kalman–Bucy filter in Hilbert space," *Inf. Control*, vol. 11, nos. 1–2, pp. 102–137, Jul.-Aug. 1967.

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

[55] S. K. Mitter, "Control with limited information," *Eur. J. Control.*, vol. 7, nos. 2–3, pp. 122–131, 2001.

[56] V. S. Borkar and S. K. Mitter, "LQG control with communication constraints," in *Communications, Computation, Control, and Signal Processing: A Tribute to Thomas Kailath*, A. Paulraj, V. Roychowdhury, and C. D. Schaper, Eds. New York, NY, USA: Springer, 1997, pp. 365–373.

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

[58] S. Tatikonda, A. Sahai, and S. K. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1549–1561, Sep. 2004.

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

[60] A. Sahai and S. K. 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.

[61] N. C. Martins, M. A. Dahleh, and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link," *IEEE Trans. Autom. Control*, vol. 51, no. 3, pp. 438–447, Mar. 2006.

[62] 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.

[63] 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.

[64] G. N. Nair and R. J. Evans, "Exponential stabilisability of finite-dimensional linear systems with limited data rates," *Automatica*, vol. 39, no. 4, pp. 585–593, Apr. 2003.

[65] M. J. Khojasteh, M. Hedayatpour, J. Cortés, and M. Franceschetti, "Exploiting timing information in event-triggered stabilization of linear systems with disturbances," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 15–27, Mar. 2021.

[66] 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.

[67] T. Simsek, R. Jain, and P. Varaiya, "Scalar estimation and control with noisy binary observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1598–1603, Sep. 2004.

[68] 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.

[69] 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.

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

[71] 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.

[72] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints II: Stabilization with limited information feedback," *IEEE Trans. Autom. Control*, vol. 44, no. 5, pp. 1049–1053, May 1999.

[73] 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.

[74] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information feedback," *IEEE Trans. Autom. Control*, vol. 50, no. 6, pp. 910–915, Jun. 2005.

[75] R. W. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems," *IEEE Trans. Autom. Control*, vol. 45, no. 7, pp. 1279–1289, Jul. 2000.

[76] 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.

[77] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, and S. Sastry, "Optimal linear LQG control over lossy networks without packet acknowledgment," in *Proc. IEEE Conf. Decis. Control*, San Diego, CA, USA, Dec. 2006, pp. 392–397.

[78] 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.

[79] 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.

[80] 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.

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

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

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

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

[85] Z. Liu, A. Conti, S. K. Mitter, and M. Z. Win, "Distributed filtering in continuous-time scenarios: Encoding strategies and information theoretical interpretation," Massachusetts Inst. Technol., Lab. Inf. Decis. Syst. (LIDS), Cambridge, MA, USA, Internal Rep., 2023.

[86] P. A. Stavrou, T. Tanaka, and S. Tatikonda, "The time-invariant multidimensional Gaussian sequential rate-distortion problem revisited," *IEEE Trans. Autom. Control*, vol. 65, no. 5, pp. 2245–2249, May 2020.

[87] 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.

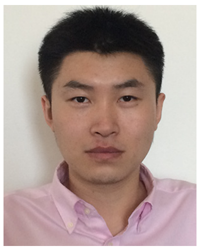
[88] T. Tanaka, K.-K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1896–1910, Apr. 2017.

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

[90] T. Weissman, Y.-H. Kim, and H. H. Permuter, "Directed information, causal estimation, and communication in continuous time," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1271–1287, Mar. 2012.

[91] J. L. Massey, "Causality, feedback and directed information," in *Proc. IEEE Int. Symp. Inf. Theory Its Appl.*, Waikiki, HI, USA, Nov. 1990, pp. 303–305.

[92] 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.
[93] M. H. Davis, *Linear Estimation and Stochastic Control*, 1st ed. London, U.K.: Chapman and Hall, 1977.
[94] P. Hartman, *Ordinary Differential Equations*, 2nd ed. Philadelphia, PA, USA: SIAM, 2002.
[95] J.-F. Le Gall, *Brownian Motion, Martingales, and Stochastic Calculus*. Cham, Switzerland: Springer, 2016.



Zhenyu Liu (Member, IEEE) 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, Beijing, China, in 2014 and 2011, respectively.

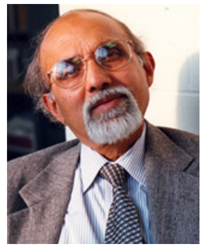
Since 2022, he has been a Post-Doctoral Associate in the wireless information and network sciences laboratory at MIT. His research interests include wireless communications, network localization, distributed inference, networked control, and quantum information science.

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 research 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, as well as 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 a 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 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. Recently, he 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, coauthored with students and colleagues, have received several awards. Other recognitions include 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). He is an ISI Highly Cited Researcher.