

When Outdated Information Matters: Optimal Fusion with Memory for Multi-Sensor Networked Control With Random Delay

Jia Zhang^{1,2}, Wei Feng^{1,2,*}, Vincent Lau³, Chih-Chun Wang⁴

¹Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

²State Key Laboratory of Space Network and Communications, Tsinghua University, Beijing 100084, China

³Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

⁴School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA

* Corresponding author.

jia-zhang@tsinghua.edu.cn, fengwei@tsinghua.edu.cn, knlau@ust.hk, chihw@purdue.edu

Abstract— Internet-of-Things applications rely on fusing distributed sensor data for control, yet the common practice discards outdated information and uses only the freshest data from each sensor. This work studies Linear Quadratic Gaussian (LQG) feedback control systems with random delay and multiple sensors, and proves that such freshest-data-only fusion is strictly suboptimal for this system. We propose Cumulative State Fusion (CSF), a linear minimum mean square error (LMMSE) optimal scheme that exploits both the freshest and outdated information. By reformulating the fusion problem as a Kalman filter problem over a meta-state system, CSF automatically captures complex spatial-temporal correlations across sensors while explicitly accounting for quantization effects. Theoretically, a tight smallest achievable state norm is developed for this system. CSF achieves this theoretical performance limit for infinite-precision communication. Numerical experiments demonstrate 15% performance improvement over freshest-data-only fusion, with the gap widening under coarse quantization. These results establish that outdated information is not merely redundant, it is essential for optimal multi-sensor estimation.

Index Terms—LQG control system, Age of Information (AoI), LMMSE fusion with memory, quantization effect reduction, the smallest achievable state norm

I. INTRODUCTION

Many Internet-of-Things (IoT) applications - such as autonomous driving systems, smart grids, and emergency rescue - rely on a plant-sensor-controller feedback loop [1], [2], as illustrated in Fig. 1. In this loop, fusion design at the central controller is critical because it decides if the multi-faceted information from distributed sensors can be effectively synthesized to design control units. The majority of state-of-art covariance-based fusion solutions [3]–[6] are based on the *smart sensor* setting, where sensors generate local estimates $\hat{\mathbf{x}}_k(t)$ and send them to the controller. It admits superior communication efficiency when sensor-to-controller channels suffer traffic congestion and burst, hence we follow this setting.

One common fusion strategy, which is categorized as *fusion without memory* in [5], is to fuse only the freshest state estimate from each of the K sensors, discard outdated ones. [3] proposed the state-of-art LMMSE optimal fusion scheme

without memory. It has been widely adapted to systems with network communication effects, such as packet dropouts, cyber-attacks, and fading channels [6]. However, intuitively, discarding outdated information loses information. In this work, we generalize [3] for our random-delay channel setting, named the scheme Freshest State Fusion (FSF). Our numerical experiment shows that it is strictly suboptimal compared to our proposed fusion method with memory, named Cumulative State Fusion (CSF), for systems with $K \geq 2$ sensors. Besides, FSF is non-capable to model the quantization noises of local estimates, which becomes a non-negligible problem for scenarios where bandwidth is limited.

On the line of *fusion with memory*, where both the freshest and outdated information are utilized for fusion, some excellent works are published but no optimality is guaranteed [4], [5], [7]. [4] proposed the Information Matrix Fusion (IMF) method for multi-sensor vehicle motion detection. This method performs well when mainly two sensors participate in the fusion process per time interval. But its performance degrades when more sensors are involved simultaneously, as we will shown in later experiments. [5]’s Section 5.2 introduced the concept of LMMSE fusion, but no algorithmic computation method for the (cross-) covariance evolution is developed, and the formulation is limited to systems with only two sensors. [7] proposed a new structure combining the freshest local estimate fusion result and the globally fused process probabilistically, but no optimality is guaranteed. Therefore there is a gap between the state-of-art fusion with memory and the LMMSE optimal fusion with memory. It is also worth noting that, while large language model offers powerful semantic reasoning for heterogeneous information, covariance-based fusion remains preferable for low-latency and precise estimation tasks - such as multi-sensor tracking in dynamic environments. The prohibitive interference latency of LLMs renders them impractical in such scenarios [8]. [9] provide a more thorough summary for this field.

In order to evaluate the quality of a design, finding the smallest attainable L2 state norm, denoted as $D_{\min}$, of a multi-

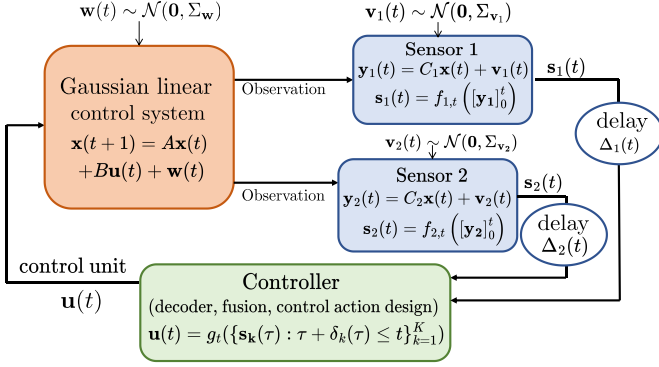


Fig. 1: A multi-sensor LQG system with random communication delays

sensor Gaussian linear control system is important. This is a sub-problem in the general framework of rate-cost tradeoff of LQG [10] systems. It has deepened roots in both control and information theory. The existing works focused on the rate-cost tradeoff of single-sensor LQG systems [11], [12]. Thus finding $D_{\min}$ for multi-sensor systems remains an open question.

Towards the analytical gap of finding $D_{\min}$, and the practical gap of designing LMMSE fusion algorithm with memory, we made two contributions. First, we find an closed-form expression for $D_{\min}$. Second, we design CSF, which is LMMSE optimal when no bandwidth limitation is applied on the sensor-to-controller channels. Moreover, when bandwidth limitation exists and quantized information is sent out by the sensor, CSF successfully reduces the *quantization noise accumulation effect*, which otherwise severely degrades the performance.

Our numerical evaluation shows that in the regime of very small quantization error, CSF's performance is within 1% of the converse $D_{\min}$. This performance is 15% less than the average state norm achieved by FSF. Moreover, when the quantization cell is relatively coarser, the performance gap between CSF and FSF increases more significantly because of the quantization noise reduction benefit from CSF. Similar superiority of CSF compared to IMF also holds.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. LQG Systems With Random Communication Delay

Consider a Linear Quadratic Gaussian (LQG) system:

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + \mathbf{w}(t), \quad (1)$$

$$\mathbf{y}_k(t) = C_k \mathbf{x}(t) + \mathbf{v}_k(t), \quad \forall k \in [1, K], \quad (2)$$

where $\mathbf{x}(t)$, $\mathbf{u}(t)$, $\mathbf{w}(t)$, $\mathbf{y}_k(t)$, $\mathbf{v}_k(t)$ are vectors, and A, B, C_k are matrices. t is the integer time index and $t \geq 0$. Specifically, $\mathbf{x}(t)$ is the N -dimensional system state vector at time t ; $\mathbf{u}(t)$ is the control action vector; and $\mathbf{w}(t)$ is the state disturbance. A and B are matrices, where A is the $N \times N$ state evolution matrix and B is the $N \times N$ control matrix, and both are set as full-rank. It implies that (A, B) is controllable. There are K

sensors and the observation $\mathbf{y}_k(t)$ of sensor $k \in [1, K]$ is an M_k -dimensional vector; C_k is the $M_k \times N$ observation matrix; and $\mathbf{v}_k(t)$ is the observation noise. $\mathbf{w}(t)$ and $\mathbf{v}_k(t)$ are i.i.d. Gaussian with zero mean and covariance matrices being $\Sigma_{\mathbf{w}}$ and $\Sigma_{\mathbf{v}_k}$, respectively. See Fig. 1 for illustration. We assume (i) $\mathbf{x}(t) = \mathbf{u}(t) = \mathbf{0}$ for all $t \leq -1$; (ii) the overall system $(A, [C_1^T, \dots, C_K^T]^T)$ is *observable*; (iii) $\Sigma_{\mathbf{w}}$ and $\Sigma_{\mathbf{v}_k}$ are of full rank; and (iv) all quantities in (1) and (2) of real-valued.

At each time slot t , sensor k generates a binary bit string $\mathbf{s}_k(t)$ based on its past observations $\{\mathbf{y}_k(\tau) : \tau \leq t\}$, i.e.,

$$\mathbf{s}_k(t) = f_{k,t}(\{\mathbf{y}_k(\tau) : \tau \leq t\}) \quad (3)$$

where $f_{k,t}$ is the *encoding function* at time t , which includes the local estimation and quantization functionalities. We allow for *variable length coding*, i.e., the lengths of the string, denoted as $|\mathbf{s}_k(t)|$, vary depending on the random observations.

The bit string $\mathbf{s}_k(t)$ is then time-stamped and sent to the controller through a digital, noiseless but *randomly delayed* channel. That is, $\mathbf{s}_k(t)$ sent by sensor- k at time t will arrive at the controller at time $t + \Delta_k(t)$, where $\Delta_k(t)$ is i.i.d. with bounded support $\{0, \dots, \delta_{\max,k}\}$ for a sufficiently large $\delta_{\max,k} < \infty$. $\Delta_k(t)$'s distribution is described by its pmf $p_{k,\delta} = \text{Prob}(\Delta_k(t) = \delta)$ and can be different for each sensor k . We assume the pmfs of $\{\Delta_k(t) : \forall k\}$ are known *a priori*. Note that our i.i.d. delay model allows out-of-order delivery, i.e., the string $\mathbf{s}_k(t_1)$ sent at time $t_1 < t_2$ may arrive later than $\mathbf{s}_k(t_2)$ if $t_1 + \Delta_k(t_1) > t_2 + \Delta_k(t_2)$, which is a common phenomenon for networked communication channels [13].

At time t , the controller computes the action $\mathbf{u}(t)$ based on the $\mathbf{s}_k(\tau)$ it has received. That is,

$$\mathbf{u}(t) = g_t(\{\mathbf{s}_k(\tau) : k \in [1, K], \tau + \Delta_k(\tau) \leq t\}) \quad (4)$$

where $g_t(\cdot)$ is the control action function at time t . It includes the functionalities to decode $\mathbf{s}_k(\tau)$ s, to fuse the decoded local information, and to design $\mathbf{u}(t)$. The action $\mathbf{u}(t)$ is then applied instantaneously to $\mathbf{x}(t)$ via (1).

The converse is characterized by the following minimization problem:

$$D_{\min} \triangleq \min_{\{f_{k,t}, g_t\}} \sup_t E\{\mathbf{x}(t)^T \mathbf{x}(t)\}. \quad (5)$$

Namely, $D_{\min}$ is the smallest L2 state norm one can possibly achieve without any restriction on encoders $f_{k,t}(\cdot)$ and controller $g_t(\cdot)$. Next we will present two extensively in-use techniques, the Age of Information and Kalman filter.

B. Age of Information

We define the Age-of-Information at time t by

$$\Theta_k(t) \triangleq t - \max_{\tau} \{\tau : \tau + \Delta_k(\tau) \leq t\} \quad (6)$$

for sensor- k data, which describes *how old is the most recently received packet $\mathbf{s}_k(\tau)$ from sensor k* . It describes the freshness of information at the controller's angle, which is different to but can be derived from $\Delta_k(t)$. It is provable that $\Theta_k(t)$ is a 1st order Markov process. See [12] for more derivation details. We further define the AoI vector for all K sensors by

$\vec{\Theta}(t) \triangleq (\Theta_1(t), \dots, \Theta_K(t))$, and its probability distribution is computed by

$$p_{\vec{\theta}}^{[AoI]} \triangleq \text{Prob}(\vec{\Theta}(t) = \vec{\theta}) \quad (7)$$

$$= \prod_{k=1}^K \left(\sum_{\delta=0}^{\theta_k} p_{k,\delta} \cdot \prod_{i=0}^{\theta_k-1} \left(\sum_{\delta=i+1}^{\delta_{\max}} p_{k,\delta} \right) \right). \quad (8)$$

C. Kalman Filtering

Kalman filter (KF) is originally derived under the *single-sensor* and *zero-delay* (1S-ZD) setting. Algorithm 1 summarizes the 1S-ZD KF algorithm.

Algorithm 1 1S-ZD Kalman Filter (KF) computation

- 1: **INPUT:** Matrices A , B , C , $\Sigma_{\mathbf{w}}$, and $\Sigma_{\mathbf{v}}$; and two sequences $\{\mathbf{y}(t) : t \geq 0\}$ and $\{\mathbf{u}(t) : t \geq 0\}$.
- 2: **OUTPUT:** A sequence of estimation $\{\hat{\mathbf{x}}(t) : t \geq 0\}$ and two sequences of matrices $\{P_t : t \geq 0\}$ and $\{\Gamma_t : t \geq 0\}$.
- 3: Initialize $\hat{\mathbf{x}}(-1) = 0$ and $P_{-1} = 0$.
- 4: **for** $t = -1$ to ∞ **do**

$$\Phi_{t+1} = AP_t A^T + \Sigma_{\mathbf{w}} \quad (9)$$

$$\Gamma_{t+1} = \Phi_{t+1} C^T (C \Phi_{t+1} C^T + \Sigma_{\mathbf{v}})^{-1} \quad (10)$$

$$\hat{\mathbf{w}}(t) = \Gamma_{t+1} (\mathbf{y}(t+1) - C(A\hat{\mathbf{x}}(t) + B\mathbf{u}(t))) \quad (11)$$

$$\hat{\mathbf{x}}(t+1) = A\hat{\mathbf{x}}(t) + B\mathbf{u}(t) + \hat{\mathbf{w}}(t) \quad (12)$$

$$P_{t+1} = \Phi_{t+1} - \Gamma_{t+1} C \Phi_{t+1} \quad (13)$$

5: **end for**

III. THE CONVERSE RESULT – $D_{\min}$

Since (5) places no restriction on $f_{k,t}(\cdot)$, when computing the converse $D_{\min}$ in (5) we can just assume

$$\mathbf{s}_k(t) = \{\mathbf{y}_k(\tau) : \tau \leq t\} \quad (14)$$

where “=” in (14) means that the string $\mathbf{s}_k(t)$ is representing the entire set of vectors $\{\mathbf{y}_k(\tau) : \tau \leq t\}$ with arbitrarily accurate (infinite) precision. Note that such a scheme is used only for converse analysis and is extremely wasteful in practice since (i) infinite precision is used, and (ii) old $\mathbf{y}_k(\tau)$ will be repeatedly transmitted for an unbounded number of times.

By (6) and (14), if the AoI vector at time t happens to be $\vec{\Theta}(t) = (\theta_1, \dots, \theta_K)$, then the controller will have perfect knowledge of the following set of sensor observations

$$\{\mathbf{y}_k(\tau) : k \in [1, K], \tau \in [0, t - \theta_k]\} \quad (15)$$

at time t . Looking back from time t , it is as if we are facing a deterministic delay setting, where the sensor- k -to-controller delay is always θ_k . The reason is that with the freshest packet from sensor k being $\mathbf{s}(t - \theta_k)$, observations $\mathbf{y}_k(\tau)$ are *continuously available* for $\tau = 0$ to $\tau = t - \theta_k$ by time t . Thus we can view each $\mathbf{y}_k(\tau)$ as received by the controller at $\tau + \theta_k$, which is a deterministic delay. Based on this critical observation, we have

Proposition 1. Define $D_{\min}^{(\vec{\theta})}$ as the minimal achievable L2 state norm under a deterministic delay setting with the sensor- k -to-controller delay being a constant θ_k . We then have

$$D_{\min} = \sum_{\{\vec{\theta}\}} p_{\vec{\theta}}^{[AoI]} \cdot D_{\min}^{(\vec{\theta})}. \quad (16)$$

We can compute $D_{\min}$ by averaging all possible $\vec{\Theta}(t)$ because of $\vec{\Theta}(t)$'s stationary property. We now discuss how to compute the value of $D_{\min}^{(\vec{\theta})}$ for any deterministic vector $\vec{\theta}$.

Step 1 of our computation is to define the “augmented” version of the system state $\mathbf{x}(t)$ as follows:

$$\bar{\mathbf{x}}(t) \triangleq (\mathbf{x}(t)^T, \mathbf{x}(t-1)^T, \dots, \mathbf{x}(t-\delta_{\max})^T)^T. \quad (17)$$

With deterministic delay θ_k , the controller will receive $\mathbf{y}_k(t - \theta_k)$ at time t , which can be viewed as a zero-delay (but noisy) version of $C_k \mathbf{x}(t - \theta_k)$, the output of the $\mathbf{x}(t - \theta_k)$ portion of the augmented state vector $\bar{\mathbf{x}}(t)$. *Namely, deterministically delayed observations in the original setting is equivalent to a zero-delay observation of part of the augmented state $\bar{\mathbf{x}}(t)$.*

Step 2 is to convert the multi-sensor setting to a single-sensor setting. This is done by vertically stacking the M_k dimensional observation $\mathbf{y}_k(t - \theta_k)$ of all K sensors to a single $\sum_{k=1}^K M_k$ dimensional vector $\bar{\mathbf{y}}(t) = (\mathbf{y}_1(t - \theta_1)^T, \dots, \mathbf{y}_K(t - \theta_K)^T)^T$.

Step 3: we apply the 1S-ZD KF in Algorithm 1 to the new single-sensor zero-delay augmented system. Denote P_{KF} as the $t \rightarrow \infty$ limit of the output P_t , which is a $(\delta_{\max} + 1)N$ -by- $(\delta_{\max} + 1)N$ matrix because of the state augmentation (17).

Lemma 1.

$$D_{\min}^{(\vec{\theta})} = \text{tr}(AP^{(\vec{\theta})}A^T) + \text{tr}(\Sigma_{\mathbf{w}}), \quad (18)$$

where matrix $P^{(\vec{\theta})}$ is the $N \times N$ sub-matrix in the top-left corner of P_{KF} .

Because of the page limit, we provide the general logic of proof instead. After applying Kalman filter to $\bar{\mathbf{x}}(t)$, we obtain $E\{\bar{\mathbf{x}}(t)|\bar{\mathbf{y}}^{(\vec{\theta})}(\tau), \tau \in [0, t]\} = E\{\bar{\mathbf{x}}(t)|\mathbf{y}_k^{(\theta_k)}(\tau), \tau \in [0, t], k \in [1, K]\}$. Thus according to the definition of $\bar{\mathbf{x}}(t)$, extracting the first N dimension of $E\{\bar{\mathbf{x}}(t)|\bar{\mathbf{y}}^{(\vec{\theta})}(\tau), \tau \in [0, t]\}$ gives $\hat{\mathbf{x}}^{(\vec{\theta})}(t) = E\{\mathbf{x}(t)|\bar{\mathbf{y}}^{(\vec{\theta})}(\tau), \tau \in [0, t]\}$, which is the MMSE estimate for $\mathbf{x}(t)$. Then design the optimal control action as $\mathbf{u}^*(t) = -B^{-1}A\hat{\mathbf{x}}^{(\vec{\theta})}(t)$ according to the classic certainty equivalence control law [14]. Applying $\mathbf{u}^*(t)$ to (2), we have $E\{\|\mathbf{x}(t+1)\|^2\} \geq E\{\|A\mathbf{x}(t) - A\hat{\mathbf{x}}^{(\vec{\theta})}(t)\|^2\} + \text{tr}(\Sigma_{\mathbf{w}})$ using the independence between $\mathbf{w}(t)$ and the estimation error. Therefore we have $E\{\|\mathbf{x}(t+1)\|^2\} \geq \text{tr}(AP^{(\vec{\theta})}A^T) + \text{tr}(\Sigma_{\mathbf{w}})$ after convergence.

IV. ACHIEVABILITY SCHEME

A. Smart-Sensor Achievability Schemes

The focus of this work is to design the LMMSE optimal fusion algorithm with memory. Thus we skip the details of other parts' designs in the feedback loop as shown in Fig. 1,

and provide a summary of it. A more-detailed description of the complete achievability scheme can be found in [15].

For each sensor k at time t , it will generate the local estimate $\hat{\mathbf{x}}_k(t)$ based on their observations according to the 1S-ZD KF scheme in Algorithm 1. Then it quantizes $\hat{\mathbf{x}}_k(t)$ to obtain $\hat{\mathbf{x}}_k^q(t)$, and denote the quantized output by $\hat{\mathbf{x}}_k^q(t)$. Define the quantization error at sensor k by

$$\mathbf{q}_k(t) \triangleq \hat{\mathbf{x}}_k^q(t) - \hat{\mathbf{x}}_k(t) \quad (19)$$

$$\Sigma_{\mathbf{q}_k}(t) \triangleq \mathbb{E}\{\mathbf{q}_k(t)(\mathbf{q}_k(t))^T\}. \quad (20)$$

Then $\hat{\mathbf{x}}_k^q(t)$ is encoded through a *variable length encoder* of average string length $R_k = \mathbb{E}\{|\mathbf{s}_k(t)|\}$ (bits) and sent to the controller. Obviously, the larger the average string length R_k is, the closer $\mathbf{x}_k(t)$ and $\hat{\mathbf{x}}_k^q(t)$, and the smaller $\Sigma_{\mathbf{q}_k}(t)$. We assume that $(R_k, \Sigma_{\mathbf{q}_k}(t))$ are known to the system designer.

The controller decodes received $\mathbf{s}_k(\tau)$ s and then performs state fusion on the quantized state estimates (q.st.est) it has received by time t :

$$\text{q.st.est}(t) \triangleq \{\hat{\mathbf{x}}_k^q(\tau) : k \in [1, K], \tau + \Delta_k(\tau) \leq t\} \quad (21)$$

$$\hat{\mathbf{x}}_f(t) = \hat{g}_t(\text{q.st.est}(t)). \quad (22)$$

$\hat{g}_t(\cdot)$ is the state fusion function that generates a single “fused estimate” $\hat{\mathbf{x}}_f(t)$. The controller then applies the *certainty-equivalence* control action by setting $\mathbf{u}(t) = -B^{-1}A \cdot \hat{\mathbf{x}}_f(t)$.

Our goal is to design the state fusion function $\hat{g}_t(\cdot)$ in (22) such that: (a) For the infinite bit rate $R_k \rightarrow \infty$, the fused state is LMMSE optimal. That is,

$$\hat{\mathbf{x}}_f(t) = \mathbb{E}\{\mathbf{x}(t)|\text{q.st.est}(t)\}. \quad (23)$$

(b) For finite $R_k < \infty$, we would like to minimize

$$\mathbb{E}\{\|\mathbf{x}(t) - \hat{\mathbf{x}}_f(t)\|^2\}. \quad (24)$$

The main challenge of finding the optimal $\hat{\mathbf{x}}_f(t)$ in (22) is that an optimal fusion scheme must explicitly quantify the complicated spatial and temporal correlation of $\hat{\mathbf{x}}_k^q(\tau)$ across different sensors k and across the time axis τ . Our idea is to model this spatial-temporal information fusion problem as an *abstract control problem of the meta state*, on top of the physical control problem of real measurements.

Specifically, we define the *meta state* $\mathbb{X}(t)$ as a $(K+1)N$ -dimensional column vector:

$$\mathbb{X}(t) \triangleq (\mathbf{x}(t)^T, \hat{\mathbf{x}}_1(t)^T, \dots, \hat{\mathbf{x}}_K(t)^T)^T, \quad (25)$$

which combines the true system state and all K KF-estimates as a single vector. We then define the meta state perturbation vector $\mathbb{W}(t)$ by

$$\mathbb{W}(t) = (\mathbf{w}(t)^T, \mathbf{v}_1(t+1)^T, \dots, \mathbf{v}_K(t+1)^T)^T, \quad (26)$$

i.e., the collection of all randomness in the system. We then argue that $\mathbb{X}(t)$ and $\mathbb{W}(t)$ satisfy

$$\mathbb{X}(t+1) = \mathbb{A}(t)\mathbb{X}(t) + \mathbb{B}\mathbf{u}(t) + \mathbb{D}(t)\mathbb{W}(t), \quad (27)$$

where $\mathbb{A}(t)$ is the meta-state evolution matrix at time t , $\mathbb{B}$ is the meta-state control matrix; $\mathbf{u}(t)$ is the control action (the same as in (1)), and matrix $\mathbb{D}(t)$ specifies the impact of $\mathbb{W}(t)$.

We now explain why (27) is true and what are the values of $\mathbb{A}(t)$, $\mathbb{B}$, and $\mathbb{D}(t)$. By reformatting (1), we have

$$\begin{aligned} \mathbf{x}(t+1) &= [A, \mathbf{0}_{N \times (KN)}] \mathbb{X}(t) + \mathbf{I}_N \cdot \mathbf{u}(t) \\ &\quad + [\mathbf{I}_N, \mathbf{0}_{N \times (\sum_{k=1}^K M_k)}] \mathbb{W}(t) \end{aligned} \quad (28)$$

where $\mathbf{0}_{a \times b}$ denotes the $a \times b$ zero matrix and $\mathbf{I}_n$ is the $n \times n$ identity matrix. (28) gives the first N rows of (27).

We then notice that the iterative computation of the KF estimate $\hat{\mathbf{x}}_k(t)$ can be rewritten in a succinct way by first plugging (2) into (11) and then plugging (11) into (12). The end result is

$$\begin{aligned} \hat{\mathbf{x}}_k(t+1) &= \Gamma_{k,t+1} C_k A \mathbf{x}(t) + (A - \Gamma_{k,t+1} C_k A) \hat{\mathbf{x}}_k(t) \\ &\quad + \mathbf{u}(t) + \Gamma_{k,t+1} C_k \mathbf{w}(t) + \Gamma_{k,t+1} \mathbf{v}_k(t). \end{aligned} \quad (29)$$

By reformatting (29) into a matrix form, we have derived the matrix equation (27) for rows $(kN+1)$ to $(k+1)N$. Once we carefully set the values of $\mathbb{A}(t)$, $\mathbb{B}$, and $\mathbb{D}(t)$ using (28) and (29), we have derived (27). In (27), we see that all stochastic Gaussian noises are reordered in the $\mathbb{W}(t)$ random vector. This unifies the evolution of $\mathbf{x}(t)$ and $\hat{\mathbf{x}}_k(t)$ s, so that the cross-covariances evolve automatically along with the meta system.

Based on the received q.st.est(t_0) at time t_0 , the controller formulates a series of meta state observations $\mathbb{Y}_{t_0}(\tau)$ for $\tau = 0$ to $\tau = t_0$. That is,

$$\mathbb{Y}_{t_0}(\tau) = \mathbb{C}_{t_0}(\tau) \mathbb{X}(\tau) + \mathbb{E}_{t_0}(\tau) \mathbf{q}(\tau), \quad (30)$$

where $\mathbb{C}_{t_0}(\tau)$ is the meta state observation matrix at time τ , $\mathbb{E}_{t_0}(\tau)$ is the noise observation matrix at time τ , and $\mathbf{q}(\tau)$ is the quantization noise defined by

$$\mathbf{q}(\tau) = (\mathbf{q}_1(\tau)^T, \dots, \mathbf{q}_K(\tau)^T)^T, \quad (31)$$

where each $\mathbf{q}_k(\tau)$ is the quantization noise first specified in (19). The subscript “ t_0 ” in $\mathbb{Y}_{t_0}(\tau)$, $\mathbb{C}_{t_0}(\tau)$, and $\mathbb{E}_{t_0}(\tau)$ is to emphasize that they are constructed based on the quantized state estimates q.st.est(t_0) that are available at time t_0 and may change when a different t_0 is considered. In other words, $\mathbb{C}_{t_0}(\tau)$ and $\mathbb{E}_{t_0}(\tau)$ are indicator matrices which indicate if $\hat{\mathbf{x}}_k^q(\tau)$ has arrived or not by time t .

For example, consider a 2-sensor system and suppose the state estimate of sensor 1 at time $\tau = 6$ arrives at the controller at time 7 (i.e., delay $\Delta_1(6) = 7 - 6 = 1$) and the state estimate of sensor 2 at time $\tau = 6$ arrives at time 9 (i.e., $\Delta_2(6) = 9 - 6 = 3$). When the controller is making a decision at time $t_0 = 8$, since $\hat{\mathbf{x}}_2^q(6)$ has not arrived yet, we have

$$\mathbb{C}_8(6) = [\mathbf{0}_{N \times N}, \mathbf{I}_N, \mathbf{0}_{N \times N}], \quad (32)$$

$$\mathbb{E}_8(6) = [\mathbf{I}_N, \mathbf{0}_{N \times N}]. \quad (33)$$

That is, at time $t_0 = 8$, because the controller has $\hat{\mathbf{x}}_1^q(6)$ (that *actually* arrives at time 7), it can *effectively* assume $\hat{\mathbf{x}}_1^q(6)$ arrives at time $\tau = 6$ and set the corresponding $\mathbb{C}_8(6)$ accordingly by (32). Herein the $\mathbf{q}(6)$ and $\mathbb{E}_8(6)$ represent the quantization error since the controller receives only the quantized estimate $\hat{\mathbf{x}}_1^q(6)$, not the original KF estimate $\hat{\mathbf{x}}_1(6)$.

On the other hand, when the controller is making a decision at time $t_0 = 10$, since both $\hat{\mathbf{x}}_1^q(6)$ and $\hat{\mathbf{x}}_2^q(6)$ have arrived by time 10 (in our example the latter arrived at time 9), it is as if we are facing a new pair of $\mathbb{C}_{10}(6)$ and $\mathbb{E}_{10}(6)$:

$$\mathbb{C}_{10}(6) = [\mathbf{0}_{2N \times N}, \mathbf{I}_{2N}] \quad (34)$$

$$\mathbb{E}_{10}(6) = [\mathbf{I}_{2N}], \quad (35)$$

that “delivers” both $\hat{\mathbf{x}}_1^q(6)$ and $\hat{\mathbf{x}}_2^q(6)$ at time $\tau = 6$. In sum, the $\mathbb{C}_{t_0}(6)$ and $\mathbb{E}_{t_0}(6)$ change for different t_0 , depending on whether $\hat{\mathbf{x}}_1^q(6)$ and $\hat{\mathbf{x}}_2^q(6)$ have arrived by time t_0 or not.

We now describe how to compute the aggregated estimate $\hat{\mathbf{x}}_f(t_0)$ at time t_0 . Specifically, KF in Algorithm 1 is based on *stationary* LQG and one can easily modify it to handle any *time-varying* LQG system. We then apply the general time-varying KF formulas to the new time-varying linear meta state system in (27) and (30), and denote the output KF estimate of the meta data by:

$$\mathbb{X}_{\text{KF}}(t_0) \triangleq \mathbb{E}\{\mathbb{X}(t_0) | \{\mathbb{Y}_{t_0}(\tau) : \tau \in [0, t_0]\}\}. \quad (36)$$

Finally, we set fused state estimate $\hat{\mathbf{x}}_f(t_0)$ to be the first N coordinates of the $\mathbb{X}_{\text{KF}}(t_0)$, also see (25).

Remark 1: By the meta state linear control construction (27) and (30), we have captured the complicated spatial and temporal correlation between the KF estimates $\hat{\mathbf{x}}_k(\tau)$ while taking into account the discrepancy between the quantized estimates $\hat{\mathbf{x}}_k^q(\tau)$ and the original $\hat{\mathbf{x}}_k(\tau)$. The only “approximation” we used is that the quantization error $\mathbf{q}_k(t)$ is generally *not* i.i.d.. Therefore, the KF computation on the meta state system may not give the exact MMSE estimate. However, when the bit rate $R_k \rightarrow \infty$, $\mathbf{q}_k(t_0) \rightarrow \mathbf{0}$. Our KF computation becomes exact and MMSE optimal. We thus have

Lemma 2. *When $R_k \rightarrow \infty$ for all k , the proposed CSF algorithm satisfies the optimality condition (23).*

Remark 2: At different time t_0 , the KF computation for time $\tau = 0$ to t_0 needs to be recomputed since the $(\mathbb{C}_{t_0}(\tau), \mathbb{E}_{t_0}(\tau))$ may change when focusing on different t_0 . The complexity of finding $\hat{\mathbf{x}}_{\text{fused}}(t_0)$ thus grows linearly versus t_0 , a devastating drawback since t_0 may grow unboundedly. However, observing that from $t_0 - 1$ to t_0 , the arrivals with time stamps $\leq t_0 - 1 - \delta_{\max}$ does not change, we can reuse part of the KF computation from the previous estimate $\mathbb{X}_{\text{KF}}(t_0 - 1 - \delta_{\max})$ when computing $\hat{\mathbf{x}}_f(t_0)$. It therefore keeps the complexity to be $O(\delta_{\max})$ for every time instant, regardless how large t_0 is.

Remark 3: We have analytically traced the KF computation on the new meta state system $\mathbb{A}(\tau)$, $\mathbb{B}$, $\mathbb{D}(\tau)$, $\mathbb{C}_{t_0}(\tau)$, and $\mathbb{E}_{t_0}(\tau)$, and proved that

Lemma 3. *In a single-sensor system $K = 1$ with $R_1 \rightarrow \infty$, our CSF scheme will automatically discard any outdated observation and only use the freshest received state estimate $\hat{\mathbf{x}}_1(t_0 - \Theta_1(t_0))$ to form the final estimate $\hat{\mathbf{x}}_f(t)$.*

Lemmas 3 and 2 jointly show that outdated state estimates are useful *if and only if* in a multi-sensor $K \geq 2$ setting.

V. NUMERICAL EVALUATION

This section demonstrates the tightness of $D_{\min}$, the superiority of CSF to FSF, the quantization noise reduction benefit contributed from the meta system of CSF, and the superiority of CSF to a state-of-art fusion with memory algorithm, IMF.

A. Experiment 1: the tightness of $D_{\min}$ and the superiority of CSF compared to FSF

We simulate the following 5-dimension linear feedback control system with 3 sensors in Matlab using Monte Carlo simulation. $A = \begin{bmatrix} 1.3 & 0.5 & 0 & 0 & 0 \\ 0 & 1.1 & 0.5 & 0 & 0 \\ 0 & 0 & 0.9 & 0.5 & 0 \\ 0 & 0 & 0 & 1.6 & 0.5 \\ 0 & 0 & 0 & 0 & 1.5 \end{bmatrix}$, $B = \mathbf{I}_5$, $C_1 = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$, $C_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$, $C_3 = \begin{bmatrix} 1 & 0 & 0 & 0.5 & 0 \\ 0 & 1 & 1 & 0 & 0.8 \end{bmatrix}$. The system disturbance $\mathbf{w}(t)$ and the observation noises $\mathbf{v}_k(t)$ are i.i.d. Gaussian, $\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_5)$ and $\mathbf{v}_k(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{M_k})$. The maximal delays of each sensor are $(\delta_{\max,1}, \delta_{\max,2}, \delta_{\max,3}) = (5, 6, 7)$ and the minimal are $(\delta_{\min,1}, \delta_{\min,2}, \delta_{\min,3}) = (0, 0, 0)$ with $p_{\Delta_k(t)}$ s set as uniform distribution.

Figure 2 shows the performance curves of CSF and FSF. Each node on a curve is a simulation result for one quantization precision setting. Each simulation is the result of running a Monte Carlo simulation on System 1 for 40 time slots, when $\mathbf{x}(t)$ converges. For both CSF and FSF, $\mathbb{E}\{\|\mathbf{x}(t)\|^2\}$ increases monotonically when L decreases, equivalently when the quantization cells become coarser.

In Fig. 2, comparing the left-most nodes of the CSF and FSF performance curves, where L is relatively large, we see that CSF achieves $\mathbb{E}\{\|\mathbf{x}(t)\|^2\}$ 15% smaller than FSF does. Besides, when L decreases, the quantization error accumulates more severely over time, thus CSF outperforms FSF more significantly providing the same L . Notice that only CSF achieves $D_{\min}$, shown by the orange vertical line. This on the other hand shows the tightness of our analytical minimal attainable average state norm, $D_{\min}$.

B. Experiment 2: the quantization noise reduction benefit of CSF

Fig. 3 compares the simulation results of CSF to its counter-version, where $\mathbf{q}_k(t)$ s in (19) are forced to be zero, thus $\hat{\mathbf{x}}_k^q(t)$ are treated as $\hat{\mathbf{x}}_k(t)$ during fusion. The simulation result shows that when $\mathbf{q}_k(t)$ s are ignored (the green curve), the quantization error accumulates over time and causes substantial degradation to the original CSF’s performance.

C. Experiment 3: the superiority of CSF compared to IMF

In this experiment, we compare CSF to a state-of-art fusion with memory algorithm, IMF [4], with assuming the delays being deterministic instead, $(\Delta_1, \Delta_2, \Delta_3) = (\delta_{\max,1}, \delta_{\max,2}, \delta_{\max,3}) = (5, 1, 6)$, because otherwise IMF diverges on this system setting.

We generalize [4]’s equation (13) and (14) to be able to fuse multiple local estimates from different sensors simultaneously into the globally fused process $\hat{\mathbf{x}}_f(t|t-1)$. In Fig. 4, the experimental fusion error is computed by the Monte Carlo Simulation result $\sum_{i=1:S} \mathbb{E}\{\|\mathbf{x}^{(i)}(t) - \hat{\mathbf{x}}_f^{(i)}(t)\|^2\}$, where S denotes the number of samples, while the analytical fusion

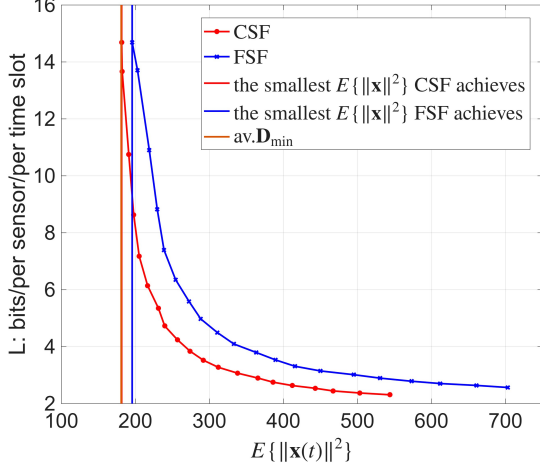


Fig. 2: CSF and FSF's performance and $D_{\min}$

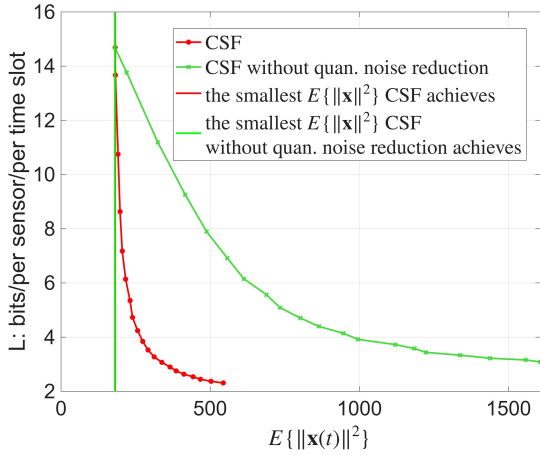


Fig. 3: Comparison of CSF and CSF without quantization noise reduction

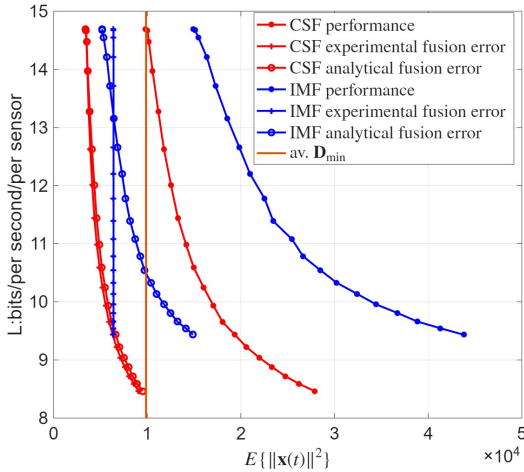


Fig. 4: CSF and IMF's performance and $D_{\min}$

errors are computed as the traces of Kalman filter error covariance and the IMF covariance matrices respectively. First, we notice that only CSF attains $D_{\min}$. Secondly, CSF overall performs remarkably better than IMF for different L levels.

The experimental and analytical fusion error curves of CSF align at all L levels because it successfully reduces the quantization noise accumulation effect, while IMF does not because it fails to take account of $\mathbf{q}_k(t)$ s.

This experiment shows CSF's superiority on both small L and large L regimes.

VI. CONCLUSION

This work considers multi-sensor LQG systems with random sensor-to-controller delay and limited bandwidth. We have characterized the smallest attainable L2 state norm $D_{\min}$, and designed a cumulative-state-fusion scheme, CSF, based on applying KF to a new metadata linear control system. Numerical comparison shows that outdated state estimates are still useful in a multi-sensor scenario with $K \geq 2$. Two interesting future research directions are 1) also considering packet dropouts on the sensor-to-controller channel, and 2) completing the rate-cost tradeoff of the multi-sensor scenario.

REFERENCES

- [1] P. Li, S. Sun, G. C. Lee, C. Chen, S. Zhu, X. Guan, and G. P. Fettweis, "State estimation and communication co-design in iiot systems: A brief survey," *Journal of Communications and Information Networks*, vol. 11, no. 1, pp. 1–21, 2026.
- [2] C. Lei, W. Feng, P. Wei, Y. Chen, N. Ge, and S. Mao, "Edge information hub: Orchestrating satellites, UAVs, MEC, sensing and communications for 6G closed-loop controls," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 1, pp. 5–20, 2025.
- [3] S.-L. Sun and Z.-L. Deng, "Multi-sensor optimal information fusion Kalman filter," *Automatica*, vol. 40, no. 6, pp. 1017–1023, 2004.
- [4] M. Aeberhard, S. Schlichtharle, N. Kaempchen, and T. Bertram, "Track-to-track fusion with asynchronous sensors using information matrix fusion for surround environment perception," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1717–1726, 2012.
- [5] X. Tian, T. Yuan, and Y. Bar-Shalom, "Track-to-track fusion in linear and nonlinear systems," *Advances in Estimation, Navigation, and Spacecraft Control*, pp. 21–41, 2015.
- [6] H. Song, S. Ye, P. Shi, W.-A. Zhang, and L. Yu, "Multi-rate sampled-data secure fusion estimation against malicious hybrid attacks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 11, pp. 401–412, 2025.
- [7] L. Lindenmaier, B. Czibere, S. Aradi, and T. Bécsei, "A robust and runtime-efficient track-to-track fusion for automotive perception systems," *IEEE Sensors Journal*, 2025.
- [8] H. Wang, X. Tu, H. Ke, H. Chai, D. Chen, and K. Han, "Im-meter: Unveiling runtime inference latency for on-device language models," in *ACM/IEEE Symposium on Edge Computing (SEC)*, 2025, pp. 1–17.
- [9] Y. Wang, S. Wu, C. Lei, J. Jiao, and Q. Zhang, "A review on wireless networked control system: The communication perspective," *IEEE Internet Things J.*, 2023.
- [10] X. Chen, "Lecture 6. linear quadratic gaussian (LQG) control," *Advanced control systems II, UC Berkeley*, 2010.
- [11] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4525–4540, 2019.
- [12] J. Zhang and C.-C. Wang, "On the rate-cost of gaussian linear control systems with random communication delays," in *IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2441–2445.
- [13] Z. Pang, C. Bai, G. Liu, Q. Han, and X. Zhang, "A novel networked predictive control method for systems with random communication constraints," *J. Syst. Sci. Complex.*, vol. 34, no. 4, pp. 1364–1378, 2021.
- [14] H. Van de Water and J. Willems, "The certainty equivalence property in stochastic control theory," in *IEEE Trans. Autom. Control*, vol. 26, no. 5, 1981, pp. 1080–1087.
- [15] J. Zhang, "On the rate-cost tradeoff of gaussian linear control systems with random communication delay," Ph.D. dissertation, Purdue University, 2022.