

Optimized data rate allocation for dynamic sensor fusion over resource constrained communication networks

Hyunho Jung¹ | Ali Reza Pedram¹ | Travis C. Cuvelier² | Takashi Tanaka³

¹Walker Department of Mechanical Engineering, University of Texas at Austin, Austin, Texas, USA

²Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, Texas, USA

³Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, Texas, USA

Correspondence

Hyunho Jung, Department of Mechanical Engineering, University of Texas at Austin, Austin, TX 78712, USA.

Email: jung.hyunho@utexas.edu

Funding information

Defense Advanced Research Projects Agency, Grant/Award Number: D19AP00078; Division of Electrical, Communications and Cyber Systems, Grant/Award Number: 1944318

Abstract

This article presents a new method to solve a dynamic sensor fusion problem. We consider a large number of remote sensors which measure a common Gauss–Markov process. Each sensor encodes and transmits its measurement to a data fusion center through a resource restricted communication network. The communication cost incurred by a given sensor is quantified as the expected bitrate from the sensor to the fusion center. We propose an approach that attempts to minimize a weighted sum of these communication costs subject to a constraint on the state estimation error at the fusion center. We formulate the problem as a difference-of-convex program and apply the convex-concave procedure (CCP) to obtain a heuristic solution. We consider a 1D heat transfer model and a model for 2D target tracking by a drone swarm for numerical studies. Through these simulations, we observe that our proposed approach has a tendency to assign zero data rate to unnecessary sensors indicating that our approach is sparsity-promoting, and an effective sensor selection heuristic.

KEYWORDS

control over communications, information theory and control, sensor fusion, sparsity-promoting, wireless sensor network

1 | INTRODUCTION

In this article, we develop techniques to efficiently allocate communication resources in dynamic sensor fusion over a resource-constrained network. Our principal motivation is networked autonomous systems where the communication between sensors and controllers is wireless. Networked autonomous systems play a major role in the heavy industry, transportation, and defense sectors. In such settings, wired communication may be impossible due to physical constraints (such as mobility). Wired communication may also be impractical due to the difficulty of routing and maintaining cables. This latter concern is not without consequence; a report from the US Navy indicates that over 1000 missions are aborted per year due to wiring faults.¹ Meanwhile, as control systems become more automated and perform more complex tasks, the number of sensor platforms tends to increase. For example, an Airbus A380-1000 model has about 10,000 sensors on its wings alone.² In wireless systems, communication resources are inherently constrained. Understanding the impact of control systems on wireless networks, and developing strategies to manage wireless resources efficiently subject to the particular requirements of control, will help enable the development of future autonomous systems.

We first give a high-level overview of our contributions before reviewing the literature.

1.1 | Our contribution

We consider a setup where several remote sensor platforms observe a common Gauss–Markov process and forward these observations to a fusion center over reliable point-to-point binary communication channels. The use of these *uplink* channels incurs a communication cost, quantified as the expected bitrate of a uniquely decodable source code. Given the observations it receives, the fusion center forms an estimate of the state. Using a *downlink* channel, which we assume incurs no cost, the fusion center feeds this estimate back to the sensor platforms. This formulation is applicable to a variety of physical layer resource allocation problems as detailed in (Sections 2.1 and 2.6). Our key contributions are presently summarized:

- We formulate a sensor rate allocation problem. Using entropy-coded dithered predictive quantization (cf. References 3 and 4) we show that the expected bitrate required to encode a quantized linear measurement with a specified reconstruction error can be approximated by a particular information measure.
- We propose to minimize a weighted average of these bitrates, over both sensors and a time horizon, subject to a constraint on the error covariance of an linear minimum mean square error estimator at the fusion center. This results in a nonconvex optimization.
- We convert the nonconvex optimization to difference-of-convex program⁵ and optimize via the heuristic convex-concave procedure (CCP). While there is no guarantee that the CCP procedure will produce the global minimum, it is guaranteed to produce a feasible local minimum.⁵
- We perform numerical experiments where we apply our approach to a time-invariant heat transfer system and a time-varying drone tracking system. We observed that as the estimator performance constraint is loosened, our algorithm tends to allocate zero data rate to more unneeded sensors. Our proposed approach thus tends to be sparsity-promoting, and may be used as a sensor selection heuristic.
- We provide analytical insight into the sparsity-promoting aspect of our approach via considering limiting examples in the scalar case.

1.2 | Literature review

A portion of this work appears in Reference 6. In this article, we motivate our problem formulation more concretely with an application to physical layer resource allocation in wireless communications (cf. Section 2.6). We also include new, more elucidating simulation results (cf. Section 5). Finally, we conclude with an additional discussion of the sparsity-promoting property of our approach (cf. Section 6).

The relevant prior work straddles several research areas, including optimal sensor selection and control with communication constraints, network information theory (in particular, multiterminal rate-distortion theory), and resource allocation in wireless communications. We highlight some relevant contributions from this areas and contrast them with our present work in the remainder of this section.

1.2.1 | Optimal sensor selection and control with communication constraints

Optimal sensor selection has been a well-studied problem for several decades. Generally speaking, such approaches aim to choose an optimal subset of sensors from some set. Information measures are often incorporated, via either the objective functions or the constraints, in the optimization over subsets. Early examples include heuristic subset optimizations that attempted to maximize the trace or determinant of the Fisher information matrix (FIM). For example, Reference 7 proposed a sensor selection method based on ranking sensors at different locations in terms of a surrogate for their contribution to the determinant of the FIM. By an iterative technique, the method removes insignificant sensors, resulting in a selected subset of sensors that tend to maximize the trace and determinant of the FIM. In Reference 8 sensor selection was performed via a genetic algorithm, with the determinant of the FIM as the figure of merit.

More relevant recent results include those that consider Bayesian estimation/tracking. In Reference 9, an entropy-based heuristic for sensor selection was introduced and applied to target localization. A semi-definite programming (SDP) relaxation of a sensor selection problem aiming to minimize the determinant of the estimation error

covariance matrix was proposed in Reference 10. In a target tracking problem,¹¹ used an estimate of the mutual information between a sensor's observation and the target state (computed via particle filter) to evaluate a notion of information gain. In Reference 12, an optimization method was proposed which maximizes Bayesian Fisher Information and mutual information while minimizing the number of selected sensors. The system model introduced in Reference 13 resembles the one in our present work. In Reference 13, a fusion center tracks a Gauss–Markov process with incoming observations from multiple remote sensors. A finite horizon optimization problem is proposed to identify a subset of sensors to be transmit their measurements at each time step. The problem is solved heuristically by a reweighted ℓ_1 relaxation. Sensor selection in aircraft engine health monitoring was studied in Reference 14. In particular, the selection was performed via evaluating entropy.

In this work we do not consider sensor selection, but rather *sensor rate allocation*. Rather than selecting a subset of sensors, we formulate an optimization problem to minimize a weighted average of data rates from the sensor platforms to the controller. Our present formulation offers additional flexibility; by choosing different weights, one can apply our present formulation to a variety of physical layer resource allocation problems (cf. Section 2.6). Notably, we also consider a system model where the sensors have strictly causal feedback access to the fusion center's best estimate. In Section 2.1, we argue that this is a reasonable assumption in several interesting and practical regimes.

Some of the prior art on control with communication constraints is also relevant to this work. The use of prefix-free coding and entropy-coded dithered quantization (ECDQ) for a control subject to data rate constraints was motivated by Silva et al.¹⁵ The work in Reference 15 was extended to the case of MIMO plants by Tanaka et al.⁴ In Reference 16, information-regularized control of a distributed system was studied with the objective of reducing intersubsystem communication while maintaining adequate control cost. The notion of communication cost in Reference 16 resembles the one of this present work. It was demonstrated that the optimal controller that jointly minimizes the control cost and the required data rate for intersubsystem communications has a sparse structure. While the communication cost in our present work is similar to the one in Reference 16, we consider a significantly different scenario—namely a distributed sensing paradigm where multiple independent sensor platforms independently convey their measurements to a fusion center. Our objective is to minimize bitrate while maintaining a constraint on the estimator mean squared error (MSE).

1.2.2 | Multiterminal rate-distortion

In this work, we address a problem that closely resembles the quadratic Gaussian CEO (chief executive/estimation officer) problem.¹⁷ In the CEO problem, the decoder, corresponding to the “executive” (analogous to the fusion center in this work), receives messages from a set of encoders or “agents”. Each agent observes a measurement of some random state variable that the executive would like to reconstruct. The tradeoff of interest is the sum rate of communication (usually measured in terms of fixed-length coding) versus the distortion in the CEO's estimate. Generally, it is assumed that the agents' measurements are conditionally independent, given the state, and that agents are not allowed to pool their data beforehand. In the “quadratic Gaussian” version of this problem (cf. Reference 18), agents observe a Gaussian process subject to independent additive Gaussian noise. The distortion metric is (block average) mean squared error. With an eye to distributed tracking,¹⁹ analyzed a causal version of the general CEO problem. A rate-distortion formulation is proposed where Massey's directed information (cf. Reference 20) is minimized subject to a constraint on the CEO's estimator distortion. This tradeoff is shown to be computable via a convex optimization in the Gaussian case. The rate loss due to the lack of communication between the sensors is also analyzed.

In this work, we deviate from the formulation of the CEO problem in that we minimize a weighted average of the data rates from sensor platforms to the fusion center, rather than just the sum rate. This is advantageous, for example, in an application to physical layer resource allocation (cf. Section 2.6). Most coding schemes for the CEO problem use fixed-length coding. We consider zero-delay uniquely decodable variable-length source coding, which may be advantageous in practice. Finally, we consider a setting where the sensors have feedback access to the decoder's estimate. Unfortunately, our lower bounds on codeword length are not completely general (they are with respect to a particular quantization scheme), and we have as-yet been unable to convexify the relevant optimization problem we derive.

It is worth mentioning that the CEO problem and the problem considered in this article are distinct from the standard vector Gaussian rate-distortion problem (which admits the well-known reverse water filling solution). Both the CEO problem and the problem considered in this article are inherently sequential; they consider tracking a dynamical system rather than and memoryless source. Even with a time-horizon of 1, the CEO problem and the scenario under investigation are fundamentally distributed source coding problems; in standard vector Gaussian rate-distortion, the compressor

has access to the entire source vector. In contrast, in the present work we assume that a set of independent compressors each has access to a scalar, linear, and potentially noisy measurement of the source vector.

1.2.3 | Resource allocation for control, estimation, and detection over wireless networks

Ultra-low latency reliable wireless communication (ULLRC) has been proposed as an enabler of next-generation networked autonomous systems—in particular for control systems using wireless communication.^{21,22} In the 5G cellular context, ULLRC is characterized by very short end-to-end latencies (on the order of a millisecond) and extremely small error probabilities (on the order of 10^{-5}).²² In Reference 23 a prediction-communication co-design approach was proposed with a network topology similar to the one considered in our present article. A nonconvex optimization was introduced to minimize the bandwidth required to decode a fixed-length message subject to constraints on block error probability and the probabilities of exceeding the latency or prediction error constraints. This approach was shown to effectively improve the tradeoff between reliability and performance (measured in terms of the aforementioned error rate metrics). In this work we consider a different notion of performance, namely estimator MSE. Our scheme is also able to dynamically allocate wireless resources by means of effective source coding; we optimize the actual number of bits conveyed from the sensors to the fusion center. In contrast to Reference 23, we do not include finite-blocklength channel coding in our analysis. In Reference 24, a general feedback control system using ULLRC is abstracted analogously to the CEO problem. The executive is assumed to operate on a fixed schedule and is required to produce its actions by a hard real-time deadline. Different agents make asynchronous linear/Gaussian measurements of a Gauss–Markov dynamical system, and transmit their measurements to the executive over a shared communication channel. The executive's estimator performance is measured in terms of MSE. Polling an observer incurs a communication cost (assumed known a priori), or *airtime*, and a branch and bound approach is applied to schedule the optimal sequence of observers that can be polled before the deadline. Our present work could be seen as optimizing the total (expected) airtime required to achieve some estimator performance (cf. Section 2.6).

The recent survey²⁵ on control systems incorporating wireless communication contains many relevant references, several practical case studies, and a detailed overview of a framework for control/communication co-design. It is argued that wireless systems for control applications should be dependable (measured in terms of control performance and stability), adaptive (reconfigurable), and efficient (with respect to the use of resources like time, bandwidth, and power). A control-guided communication approach was proposed where messages are transmitted according to a self-triggered protocol. The approach was shown to be effective through experiments in a cyber-physical testbed. In this work, we consider synchronous communication, which may be a necessity for control over congested networks. Our sensor rate allocation scheme is also inherently dynamic; in simulation, we demonstrate that different subsets of sensors are assigned nonzero bitrates over time. Furthermore, our choice to minimize bitrate allows us to draw a direct link between the utilization of physical layer resources and estimator performance, improving efficiency (at least on average).

Some prior work on sensor fusion for decentralized detection is particularly relevant to our present investigation. In Reference 26, a detection problem is considered where samples of a continuous-time signal are quantized at different rates over time. Given a sampling strategy and a constraint on total bitrate, different fixed-length quantizers are designed for each temporal sample using a criterion based on minimizing a notion of information loss. The approach in Reference 26 quantifies information loss in terms of general f -divergences. In Reference 27 several sensor platforms observe noisy scalar measurements of some deterministic unknown parameter. These measurements are conveyed to a fusion center whose goal is to test a hypothesis based on the deviation of the parameter from some known normal condition. A class of “smart” sensors transmits unquantized measurements over error free channels, meanwhile a class of “dumb” sensors transmits (fixed-length) quantized measurements over binary symmetric channels. Both a generalized likelihood ratio test (GLRT) and a (computationally simpler) Rao test for the aforementioned model are derived and analyzed, and an optimal quantizer design is pursued through nonconvex optimization. In Reference 28 a similar setup is considered with the measurement model extended to include (possibly) nonlinear, noisy, vector measurements of the deterministic unknown parameter. In Reference 28, each sensor can apply a linear precoder to its measurement before performing a fixed-length quantization. The quantized measurements are conveyed to the fusion center over a noisy binary symmetric channel. Again, both a GLRT and a (computationally simpler) Rao test are derived and analyzed. In particular, an asymptotic performance characterization lead to an optimal design for a one-bit quantizer (given some choice of precoder). The one-bit results are extended to multibit quantizers, and numerical study was performed to demonstrate that the Rao test achieves comparable performance to the GLRT.

In contrast to References 26–28 the focus of our present article is on estimation and tracking of a dynamical system as opposed to detection. Furthermore, we consider variable-length source coding, in contrast to fixed-length quantization.

1.3 | Organization and notation

This article is organized as follows. Section 2 reviews ECDQ, defines the notion of communication cost, and connects this cost to the mutual information (MI) between two Gaussian random variables that arise in Kalman filtering. The sensor rate allocation (SRA) problem is proposed. We conclude the section with an application example motivated by physical layer resource allocation in a remote sensing scenario. In Section 3, we show that the communication cost can be written in terms of error covariance matrices from Kalman filtering. In these terms, we convert the SRA optimization to a difference-of-convex program. Section 4 defines an iterative heuristic algorithm based on the convex-concave procedure (CCP) to attack the resulting nonconvex optimization. Our numerical studies are introduced in Section 5. We provide insight into the sparsity-promoting nature of our approach in Section 6. We conclude in Section 7.

Lower case boldface symbols such as $\mathbf{x}$ denote random variables and use the notation $\mathbf{x}_{1:t} = (\mathbf{x}_1, \dots, \mathbf{x}_t)$ for subsets of random variables from some discrete-time random process. We use standard information-theoretic notation from Reference 29: the entropy of a discrete random variable $\mathbf{x}$ is denoted by $H(\mathbf{x})$, while the differential entropy of a continuous random variable $\mathbf{x}$ is denoted by $h(\mathbf{x})$. The mutual information between $\mathbf{x}$ and $\mathbf{y}$ is denoted by $I(\mathbf{x}; \mathbf{y})$, and the relative entropy is denoted by $D(\cdot \parallel \cdot)$. $\mathbb{S}^n$ represents the set of symmetric matrices of size $n \times n$. $X \in \mathbb{S}_+^n$ or $X \geq 0$ indicates that X is a positive semidefinite matrix, and $X \in \mathbb{S}_{++}^n$ or $X > 0$ represents that X is a positive definite matrix. We denote the set of natural numbers $\{1, 2, \dots, M\}$ as $1 : M$, and the set of positive natural numbers as $\mathbb{N}^+$.

2 | PROBLEM FORMULATION

We consider discrete-time estimation of a Gauss–Markov source and assume the source is an n -dimensional random process defined via

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + F\mathbf{w}_t, \quad \mathbf{w}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, I), \quad t = 1, 2, \dots, T \quad (1)$$

with initial condition $\mathbf{x}_1 \sim \mathcal{N}(0, P_{1|1})$, where $P_{1|1} \in \mathbb{S}_+^n$ and $A \in \mathbb{R}^{n \times n}$. We consider remote estimation over a star-like sensor network, as shown in Figure 1. Each of the M sensors makes a noisy, scalar, linear measurement of the state vector. At time t , sensor i observes a linear measurement of the plant with additive Gaussian noise. Denote the sensor's measurement matrix by $C_i \in \mathbb{C}^{1 \times n}$ and let $U_i \in \mathbb{S}_+^n$ and denote the measurement noise covariance. Let the measurement noise at the platform be given by $\mathbf{u}_{i,t} \sim \mathcal{N}(0, U_i)$ i.i.d over time. The i th sensor platform observes $\mathbf{y}_{i,t} = C_i\mathbf{x} + \mathbf{u}_{i,t}$. Concatenating the measurements from each sensor into $\mathbf{y}_t \in \mathbb{C}^M$, we have

$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{u}_t, \quad (2a)$$

where

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_{1,t} \\ \vdots \\ \mathbf{y}_{M,t} \end{bmatrix}, \quad C = \begin{bmatrix} C_1 \\ \vdots \\ C_M \end{bmatrix}, \quad \mathbf{u}_t = \begin{bmatrix} \mathbf{u}_{1,t} \\ \vdots \\ \mathbf{u}_{M,t} \end{bmatrix}. \quad (2b)$$

We assume that the noises at different sensors are independent, for example, we assume that for all t

$$U = \mathbb{E}(\mathbf{u}_t \mathbf{u}_t^T) = \begin{bmatrix} U_1 & 0 & \dots & 0 \\ 0 & U_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & U_M \end{bmatrix}. \quad (3)$$

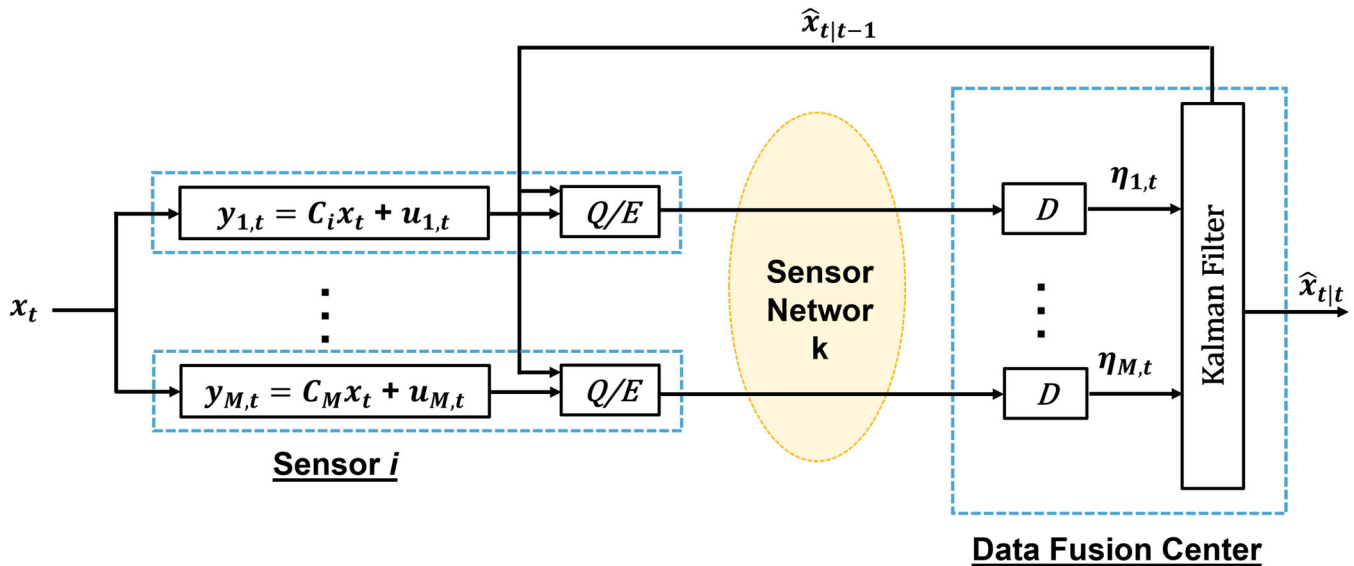


FIGURE 1 Data fusion center and distributed sensors. “Q/E” represents the quantizer/encoder module whereas “D” represents the decoder module

We assume that sensors must convey these measurements to the fusion center over a resource-constrained communication network, as detailed in the sequel. Throughout the article, we assume that the pair (A,C) is observable.

2.1 | Data fusion over resource-constrained network

We assume the platforms operate synchronously in discrete-time, analogous to a CAN-bus-like system. At time t , we assume that sensor i encodes $y_{i,t}$ into a binary codeword, say $z_{i,t}$, from a uniquely decodable variable-length code. The length of the codeword $z_{i,t}$, in bits, is denoted $\ell_{i,t}$. The expected length of $\ell_{i,t}$ measures the bitrate of uplink communication from the i^{th} sensor platform to the fusion center, and provides a notion of communication cost. In the CAN protocol, a packet frame consists of a header and a trailer in addition to the data field. For the sake of simplicity, we neglect the contributions from these components to the network’s bitrates. We assume that the uplink communication is *reliable* in the sense that the packet $z_{i,t}$ is received without error before time $t + 1$.

Having received packets from each of the M platforms, the fusion center decodes packets and computes the linear minimum mean-square error (LMMSE) estimate $\hat{x}_{t|t}$. The fusion center also calculates a step-ahead prediction $\hat{x}_{t+1|t}$ based on least mean-square-error estimate $\hat{x}_{t|t}$ and the source process (1). Once the prediction $\hat{x}_{t+1|t}$ is calculated, the fusion center transmits the prediction back to all the sensors as shown in Figure 1. Access to $\hat{x}_{t+1|t}$ at the sensor platforms allows them to apply *predictive quantization* at time step $t + 1$. Quantization and encoding the *innovation* $y_{i,t+1} - C_i \hat{x}_{t+1|t}$, rather than $y_{i,t+1}$ directly, helps to reduce the expected codeword length (cf. References 4 and 30).

We neglect the communication cost associated with the downlink communication and assume that the sensors receive $\hat{x}_{t+1|t}$ exactly. These assumptions are appropriate in a variety of circumstances. In one such scenario, the remote sensor platforms could be battery-powered meanwhile the fusion center could have direct access to power lines. If the communication between the sensors and the fusion center is wireless, the sensor platforms could have strictly constrained transmit powers meanwhile the fusion center could have no such constraint. In such a setting, the fusion center could encode an extremely finely quantized version of $\hat{x}_{t+1|t}$ and use an extremely high power to transmit this message reliably to all of the M sensor platforms. In a scenario with time-division duplexing, the uplink and downlink time-share the same frequency channel. As we discuss in Section 2.6, a reasonable notion of communication “cost” in terms of physical layer resources used is the amount of time the frequency channel is occupied by uplink and downlink transmissions. We discuss this in more detail in Section 2.6, however, at this stage it suffices to recognize that if the downlink transmit power is unconstrained, an arbitrary number of bits can be conveyed from the fusion center to the sensor platforms in a minimal amount of time.

It is also reasonable to neglect the downlink communication cost in a wireless communication scenario where each remote sensor platform can easily “overhear” (e.g., decode reliably) the transmissions from other platforms ostensibly directed to the fusion center. Formally, this is possible when, for each remote platform, the multicast capacity of the channel from the given platform to the others exceeds the channel capacity from the given platform to the fusion center. Essentially, this means that the channel capacity between any two platforms exceeds the capacity between any one platform and the fusion center. This occurs, for example, in a line-of-sight communication setting where the distance between the individual sensor platforms is small compared with the distance between the platforms and the fusion center. In this case, each user can form the estimate $\hat{\mathbf{x}}_{t+1|t}$ simply based on the transmissions it overhears on the uplink—no downlink communication is actually required.

2.2 | Entropy-coded dithered quantization

We assume that sensor platforms discretize, and encode their measurements into binary codewords, using *entropy-coded dithered quantization* (ECDQ).³¹ Define a uniform scalar quantizer with sensitivity Δ via

$$Q_{\Delta}(z) = k\Delta \quad \text{if} \quad \left(k - \frac{1}{2}\right)\Delta \leq z < \left(k + \frac{1}{2}\right)\Delta. \quad (4)$$

Essentially, the function Q_{Δ} rounds its input to the nearest multiple of Δ . *Dithering* introduces intentional randomness into the quantization process to make the quantization error tractable. Let $\mathbf{z}$ be a random variable that we would like to encode, and let $\xi \sim \text{unif}\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ independent of $\mathbf{z}$. Define the *quantization*

$$\mathbf{q} = Q_{\Delta}(\mathbf{z} + \xi). \quad (5)$$

In general, $\mathbf{q}$ is a discrete random variable with countable support. Define the *reconstruction*

$$\boldsymbol{\eta} = \mathbf{q} - \xi. \quad (6)$$

It can be shown (cf. References 3 and 4) that

$$\boldsymbol{\eta} = \mathbf{z} + \mathbf{v}, \quad (7)$$

where $\mathbf{v} \sim \text{unif}\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ and independent of $\mathbf{z}$. Note that the *same* realization of the dither signal ξ is used in the quantization and reconstruction steps.

Figure 2 illustrates the use of dithering in our communication network model. We assume that each sensor platform shares a common dither signal with fusion center. At time t , denote the i^{th} platform's dither signal $\xi_{i,t}$. The stochastic process (defined over time and platforms) $\{\xi_{i,t}\}_{i \in 1:M, t \in \mathbb{N}^+}$ consists of mutually independent uniform random variables

$$\xi_{i,t} \sim \text{unif}\left[-\frac{\Delta_{i,t}}{2}, \frac{\Delta_{i,t}}{2}\right], \quad (8)$$

where the sensitivities $\Delta_{i,t}$ will be designed in the sequel. Define the dither processes to be independent of $\{\mathbf{x}_t, \mathbf{w}_t, \mathbf{u}_{1:M,t}\}_{t \in \mathbb{N}^+}$. At time t , platform i computes the Kalman innovation corresponding to its measurement via

$$\theta_{i,t} = \mathbf{y}_{i,t} - \mathbf{C}_i \hat{\mathbf{x}}_{t|t-1}. \quad (9)$$

It then computes the dithered quantization

$$\mathbf{q}_{i,t} = Q_{\Delta_{i,t}}(\theta_{i,t} + \xi_{i,t}).$$

The sensor platform uses *lossless* variable-length *entropy coding* to encode $\mathbf{q}_{i,t}$ into a finite-length binary string. Denote this *codeword* $\mathbf{z}_{i,t}$. The length of $\mathbf{z}_{i,t}$ is a random variable, denote $\ell_{i,t}$. It can be shown that there exists a lossless source code, adapted to the conditional probability mass function of $\mathbf{q}_{i,t}$ given $\xi_{i,t}$, with expected codeword length satisfying [cf. Reference 29(theorem 5.5.1 and exercise 5.28)]

$$H(\mathbf{q}_{i,t}|\xi_{i,t}) \leq \mathbb{E}(\ell_{i,t}) < H(\mathbf{q}_{i,t}|\xi_{i,t}) + 1. \quad (10)$$

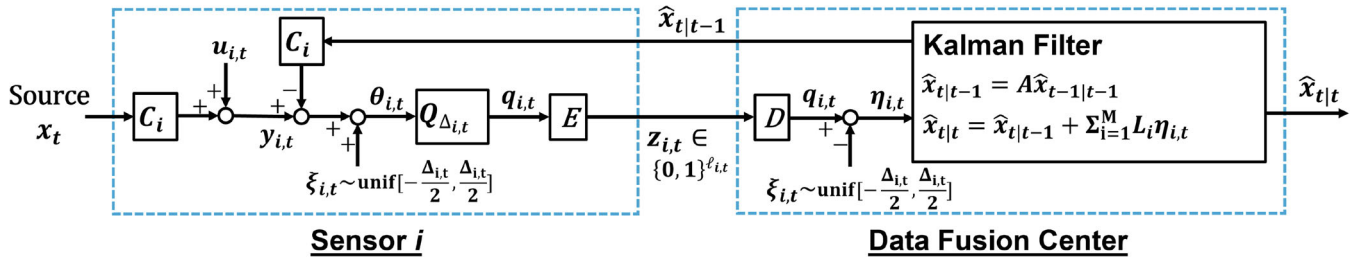


FIGURE 2 This figure gives an overview of the communication architecture between an individual sensor platform and the fusion center. Note that the dither signal, $\xi_{i,t}$ is assumed to be known at both the sensor platform and the fusion center. In practice, this “shared randomness” can be accomplished by using synchronized pseudorandom number generators between each sensor platform and the fusion center

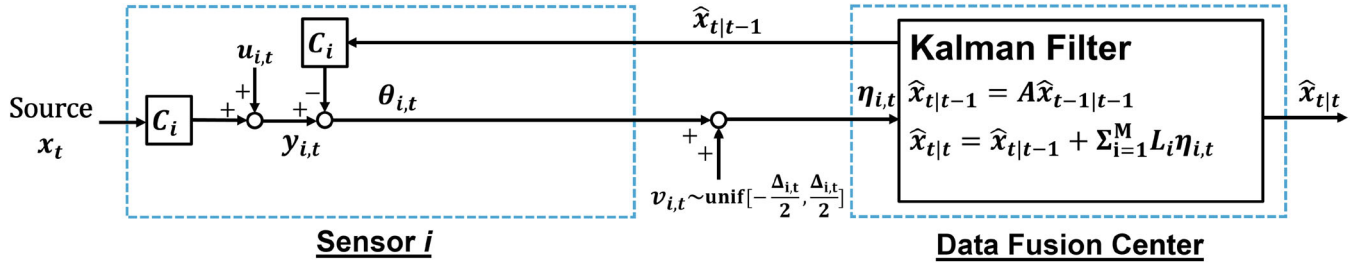


FIGURE 3 The net effect of the ECDQ scheme is that, from each sensor platform, the fusion center receives a linear measurement corresponding to the platform’s Kalman innovation corrupted by additive *uniform* noise

Alternative, more tractable bounds on codeword length are developed in the next section. Since the binary communication channel from the sensor platform to the fusion center is assumed to be reliable the sensor platform receives $z_{i,t}$ exactly. Since the entropy code is lossless, the decoder at the fusion center recovers $q_{i,t}$ exactly. The decoder then computes the dithered reconstruction

$$\eta_{i,t} = q_{i,t} - \xi_{i,t}. \quad (11)$$

Using (5,6,7) we have

$$\eta_{i,t} = \theta_{i,t} + \mathbf{v}_{i,t}, \text{ where } \mathbf{v}_{i,t} \sim \text{unif} \left[-\frac{\Delta_{i,t}}{2}, \frac{\Delta_{i,t}}{2} \right]. \quad (12)$$

It can be shown that the random variables $\{\mathbf{v}_{i,t}\}_{i \in 1:M, t \in \mathbb{N}^+}$ are mutually independent and independent of $\{\theta_{i,t}\}_{i \in 1:M, t \in \mathbb{N}^+}$.^{3,4} An end-to-end model of entropy-coded dithered quantization is shown in Figure 3. The joint distributions of the random variables in Figure 3 are identical to those in Figure 2.

2.3 | Communication cost approximation

We define the communication cost associated with the i th sensor over some time horizon T as the time-averaged expected codeword length, for example,

$$R_i = \frac{1}{T} \sum_{t=1}^T \mathbb{E}(\mathcal{L}_{i,t}). \quad (13)$$

The communication cost (13) can be interpreted as the time-average rate of reliable (e.g., error-free) communication required to convey the quantized measurements $\{q_{i,t}\}_{t=1}^T$ (defined in Section 2.2) from the i th sensor to the fusion center. Under several reasonable architectures from wireless communications, this cost is an effective surrogate for the (expected) amount of physical layer resources consumed by the i th platform on the uplink (cf. Section 2.6). Equation (10) gave a

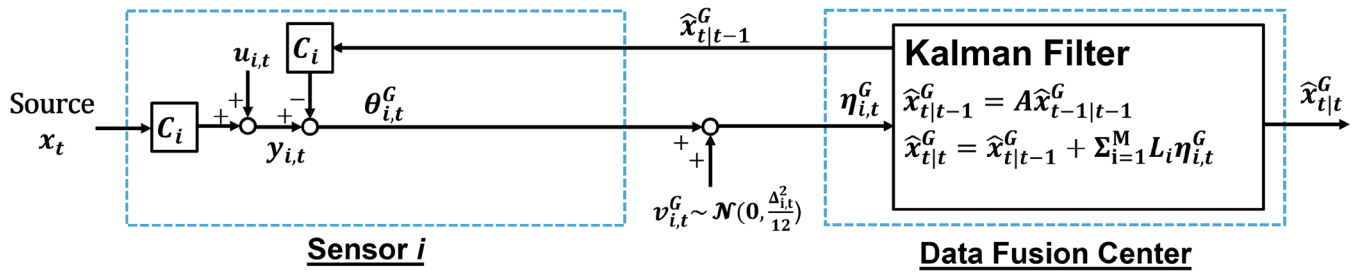


FIGURE 4 A “jointly Gaussian” version of the architecture shown in Figure 3. The uniform noise in Figure 3 is replaced with Gaussian noise with the same first and second-order statistics. The process defined by the random variables in this figure is jointly Gaussian and has identical first and second moments to the process defined by the random variables in Figure 3

bound on $\mathbb{E}(\mathcal{L}_{i,t})$ in terms of a conditional entropy. In the following lemma, we express these bounds in terms of a mutual information.

Lemma 1. *For every sensor i and time step t , the expected codeword length satisfies*

$$I(\theta_{i,t}; \eta_{i,t}) \leq \mathbb{E}(\mathcal{L}_{i,t}) < I(\theta_{i,t}; \eta_{i,t}) + 1,$$

where the mutual information is taken with respect to the joint distribution of $\theta_{i,t}$ and $\eta_{i,t}$ induced by the model in Figure 3.

Proof. See Appendix A. ■

Note that $\theta_{i,t}$ and $\eta_{i,t}$ are non-Gaussian. Directly evaluating the mutual information $I(\theta_{i,t}; \eta_{i,t})$ is rather difficult. In Reference 32, an approach to evaluating this mutual information was proposed via defining a “jointly Gaussian” version of the architecture in Figure 3, shown here in Figure 4. Define

$$V_{i,t} = \frac{\Delta_{i,t}^2}{12}. \quad (14)$$

Essentially, the uniform quantization noise $\mathbf{v}_{i,t}$ in Figure 3 is replaced with additive Gaussian noise, independent over sensors and time, with the same means and covariances, that is,

$$\mathbf{v}_{i,t}^G \sim \mathcal{N}(0, V_{i,t}), \text{ independent over } t, i, \text{ and independent of } \{\mathbf{x}_t\}_{t \in \mathbb{N}^+}. \quad (15)$$

Under this model, the measurements and reconstructions are jointly Gaussian, and denoted $\theta_{i,t}^G$ and $\eta_{i,t}^G$ respectively. Assume that the dynamical systems in Figures 3 and 4 have initial conditions with identical first and second moments. It follows from the linearity of the models in Figures 3 and 4, and in particular the fact that the Kalman filter recursions are identical, that the stochastic processes $\{\mathbf{x}_{t|t}^G, \mathbf{x}_{t|t-1}^G, \theta_t^G, \eta_t^G\}_{t=1,2,\dots,T}$ and $\{\mathbf{x}_{t|t}, \mathbf{x}_{t|t-1}, \theta_t, \eta_t\}_{t=1,2,\dots,T}$ have identical first and second moments.

The next lemma shows that $I(\theta_{i,t}; \eta_{i,t})$ can be bounded to within two bits of $I(\theta_{i,t}^G; \eta_{i,t}^G)$.

Lemma 2.

$$I(\theta_{i,t}^G; \eta_{i,t}^G) \leq I(\theta_{i,t}; \eta_{i,t}) < I(\theta_{i,t}^G; \eta_{i,t}^G) + \frac{1}{2} \log \frac{2\pi e}{12}.$$

Proof. See Appendix B. ■

Using Lemma 2, we can generate a new bound from Lemma 1 via

$$I(\theta_{i,t}^G; \eta_{i,t}^G) \leq \mathbb{E}(\mathcal{L}_{i,t}) < I(\theta_{i,t}^G; \eta_{i,t}^G) + \underbrace{1 + \frac{1}{2} \log \frac{2\pi e}{12}}_{\approx 1.254[\text{bits}]} \quad (16)$$

We will use the expression for $I(\theta_{i,t}^G; \eta_{i,t}^G)$ to approximate the communication cost in (13). In particular we define R_i^a as the *approximate* time-averaged expected bitrate for the i th sensor via

$$R_i^a = \frac{1}{T} \sum_{t=1}^T I(\theta_{i,t}^G; \eta_{i,t}^G) \quad (17)$$

Since $\mathbf{v}_{i,t}^G \sim \mathcal{N}(0, \Delta_{i,t}^2/12)$, the mutual information $I(\theta_{i,t}^G; \eta_{i,t}^G)$ is a function of the quantizer sensitivity $\Delta_{i,t}$. Tuning the set of sensitivities allows us to adjust the data rates allocated to different sensors over time. We demonstrate this explicitly in Section 3. In the remainder of this section, we will use some properties from Kalman filters to extend our notion of “second-order equivalence” between the random variables in Figure 3 and Figure 4 to the case of multiple sensors, and we define the sensor rate allocation problem.

2.4 | Linear minimum mean square error estimation

We consider sensor fusion in the setting that where the fusion center’s observations follow the models of either Figure 2 or Figure 4. In both cases, the fusion center uses a Kalman filter to recursively estimate $\mathbf{x}_t$. When all sensors conform to the non-Gaussian model of Figure 2, at time t , the Kalman filter computes

$$\hat{\mathbf{x}}_{t|t-1} = \text{the LMMSE estimate of } \mathbf{x}_t \text{ given } \{\eta_{i,k}\}_{i \in 1:M, k \in 1:t-1} \quad (18a)$$

and

$$\hat{\mathbf{x}}_{t|t} = \text{the LMMSE estimate of } \mathbf{x}_t \text{ given } \{\eta_{i,k}\}_{i \in 1:M, k \in 1:t}. \quad (18b)$$

Conversely, in the case where all sensors conform to the Gaussian model (cf. Figure 4), the LMMSE estimate computed by the Kalman filter corresponds to the minimum mean square error (MMSE) estimate, for example,

$$\hat{\mathbf{x}}_{t|t-1}^G = \mathbb{E}(\mathbf{x}_t | \eta_{1:t-1}^G) \quad (19a)$$

and

$$\hat{\mathbf{x}}_{t|t}^G = \mathbb{E}(\mathbf{x}_t | \eta_{1:t}^G). \quad (19b)$$

Analogously to the previous section, if the initial conditions in both the Gaussian and non-Gaussian systems have identical means and covariance matrices, the Kalman filter recursions, in particular the achieved mean squared errors, are identical. Thus, for any fixed choice of quantizer sensitivities $\Delta_{i,t}$ the estimator in Figure 2 (equivalently, Figure 3) achieves the same MSE performance as the estimator in Figure 4.

Let U be the diagonal matrix defined in (3) and let V_t be a diagonal matrix such with the $V_{i,t}$ s as diagonal entries that is,

$$V_t = \begin{bmatrix} \frac{\Delta_{1,t}^2}{12} & 0 & \dots & 0 \\ 0 & \frac{\Delta_{2,t}^2}{12} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\Delta_{M,t}^2}{12} \end{bmatrix}. \quad (20)$$

The Kalman gain at time t is given by the forward Riccati recursion via

$$L_t = P_{t|t-1} C^\top (C P_{t|t-1} C^\top + U + V_t)^{-1}. \quad (21)$$

Define $P_{t|t} \in \mathbb{S}_{++}^n$ and $P_{t+1|t} \in \mathbb{S}_{++}^n$ as the estimation error covariances $P_{t|t} := \text{Cov}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^G)$ and $P_{t+1|t} := \text{Cov}(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}^G)$. These matrices satisfy the recursion

$$P_{t|t}^{-1} = P_{t|t-1}^{-1} + C^\top (U + V_t)^{-1} C, \quad (22a)$$

$$P_{t+1|t} = A P_{t|t} A^\top + F F^\top. \quad (22b)$$

We reiterate that due to the “second-order equivalence” between the Gaussian and non-Gaussian systems, we have for all t that $\text{Cov}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^G) = P_{t|t}$ and $\text{Cov}(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}^G) = P_{t+1|t}$.

2.5 | The sensor rate allocation (SRA) problem

We now propose the optimization problem to be studied in the remainder of this article. We propose to minimize a weighted average of the approximate data rates R_i^a assigned to each sensor $i = 1, 2, \dots, M$ subject to the constraint that the MSE performance of the fusion center’s estimator does not exceed β . Define a set of positive weights α_i , $i \in 1 : M$ to represent the cost of transmitting a single bit from platform i to the fusion center. The total communication cost is $\sum_{i=1}^M \alpha_i R_i$, where R_i is defined in (13). The approximate cost, defined in terms of the approximate bitrates (cf. (17)), is defined as $\sum_{i=1}^M \alpha_i R_i^a$. Our decision variables are the sensor platforms’ quantizer sensitivities, for example, $\{\Delta_{i,t}\}_{i \in 1:M, t \in 1:T}$. Via (20), it can be seen that for any time t the set of feasible quantizer sensitivities are in one-to-one correspondence with the set of diagonal, positive semidefinite matrices. Let the set of positive semidefinite diagonal $M \times M$ matrices be denoted $\mathbb{D}_+^M$. For β the constraint on the fusion center’s MSE, we define the sensor rate allocation problem as

$$\min_{\substack{V_t \in \mathbb{D}_+^M \\ t \in 1:T}} \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \alpha_i I(\theta_{i,t}^G; \eta_{i,t}^G), \quad (23a)$$

$$\text{s.t.} \quad \frac{1}{T} \sum_{t=1}^T \mathbb{E} \|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}\|^2 \leq \beta. \quad (23b)$$

Say $\{V_t^*\}_{t=1}^T$ is a feasible solution of (23) and f^* is the attained value. Choosing the quantization sensitivities such that $\frac{\Delta_{i,t}^2}{12} = V_{i,t}^*$ guarantees a communication cost satisfying

$$\sum_{i=1}^M \alpha_i R_i \leq f^* + \left(1 + \frac{1}{2} \log \frac{2\pi e}{12}\right) \times \sum_{i=1}^M \alpha_i. \quad (24)$$

We also consider the infinite horizon problem where the quantizer sensitivities are time invariant, namely

$$\min_{\substack{V \in \mathbb{D}_+^M \\ t \in 1:T}} \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \alpha_i I(\theta_{i,t}^G; \eta_{i,t}^G), \quad (25a)$$

$$\text{s.t.} \quad \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}\|^2 \leq \beta. \quad (25b)$$

Essentially, the optimal solution to the problem (25) can be interpreted as the best time-invariant solution to the infinite-horizon SRA problem. The analysis of time-varying solutions to the SRA problem presents an opportunity for future work (cf. Section 7). In the following subsection, we describe an application of the sensor rate allocation optimization to resource management in remote sensing scenario.

2.6 | Application example: Physical layer resource allocation

The minimizations in (23) and (25) can be applied to a variety of resource allocation problems in wireless (and, in principle, wireline) communication systems. By choosing the weights α_i in (25) appropriately, we can design a data compression scheme that seeks to minimize the number of physical layer resources required to achieve the required estimator performance at the fusion center. As an example, we consider a setup similar to the one proposed in Reference 24, where several remote sensing platforms communicate their observations to a data fusion center over a shared wireless medium. We present a simple example to conclude this section.

Consider the case where several remote sensors, using a fixed uplink transmit power P_{UL} (in Watts), communicate their measurements to the fusion center via time-division multiple access (TDMA) over a channel of passband

bandwidth B (in Hz). In such a system, the platforms take turns transmitting their measurements over the same bandlimited channel—see Figure 5. We assume that the i th sensor platform is at a range of r_i from the fusion center, that all sensor platforms have identical transmission systems, and that the channel from each sensor to the fusion center is line-of-sight. The amount of power received from transmitter i depends on range, as well as a variety of factors, like the gains from the antenna systems and the losses from other components. We let L encapsulate the factors that do not depend on the range—in the present case L is assumed the same for all transmitters. The additive noise at the receiver is assumed to be white and Gaussian with power spectral density N_0 . Since only one sensor platform transmits at any given time, the effective channel between the i th transmitter and the fusion center is a point-to-point additive white Gaussian noise (AWGN) channel with a signal-to-noise ratio (SNR) of

$$\text{SNR}_i = L \frac{P_{\text{UL}}}{r_i^2 B N_0}. \quad (26)$$

The *capacity* of the wireless channel from receiver i to the fusion center is

$$C_i = B \log_2(1 + \text{SNR}_i) \text{ bits/second}, \quad (27)$$

which is the maximum reliable (e.g., error-free) data-rate that can be sustained from the i th transmitter to the receiver. The *airtime*, τ_i , in seconds, required to send b_i bits from the i th sensor platform to the fusion center is lower bounded via

$$\tau_i \geq \frac{b_i}{C_i}. \quad (28)$$

This bound is in general loose, as the zero-error channel capacity is achievable only asymptotically (cf. Reference 33). However, with modern error-correcting codes (including low-density parity-check and turbo codes) it is possible to communicate with very small error probability near the capacity for the point-to-point AWGN channel.³⁴ We thus make the assumption that

$$\tau_i \approx \frac{b_i}{C_i}. \quad (29)$$

In particular, the airtime is directly proportional to the energy expended on communication. Under the assumption that all the sensor platforms use the same transmit power, the energy expended to transmit b_i bits is $E_i = P_{\text{UL}} \tau_i$.

If the system uses time-division duplexing in addition to TDMA, the downlink communication also time-shares the same frequency channel shared between the uplink users. Let N_i denote the noise power spectral density of the channel from the fusion center to the i th user, and let $\theta_{\text{max}}^2 = \max_i N_i r_i^2$. Assuming line-of-sight communication, and that the fusion center feeds back a message containing b_{DL} bits to all of the remote platforms, the airtime used by the downlink can be approximated by

$$\tau_{\text{DL}} \approx \frac{b_{\text{DL}}}{B \log_2(1 + L P_{\text{DL}} / (B \theta_{\text{max}}^2))}. \quad (30)$$

Note that for a fixed downlink message size b_{DL} , τ_{DL} decreases to 0 as P_{DL} increases. Thus, we claim that in a setting where the uplink is subject to a strict power constraint (as appropriate, for example, if the users are battery-powered) and the downlink power constraint, P_{DL} , is sufficiently high it is reasonable to assume that the fusion center can feed back a finely quantized version $\hat{\mathbf{x}}_{i|t-1}$ to the users using a negligible airtime.

Minimizing the uplink airtime of a sensing task has a great deal of utility. It minimizes the amount of energy the network spends on uplink communication. Furthermore, completing a sensing task with the minimum airtime is useful in and of itself, as it may allow other users and systems the opportunity to access the channel without causing (or experiencing) additional interference. If we make the choice

$$\alpha_i = \frac{1}{C_i}, \quad (31)$$

the optimizations in (23) can be seen as minimizing the expected airtime (and, consequently, the energy expended on communication) required to complete the sensing task in the finite horizon. Note that the expected airtime expended

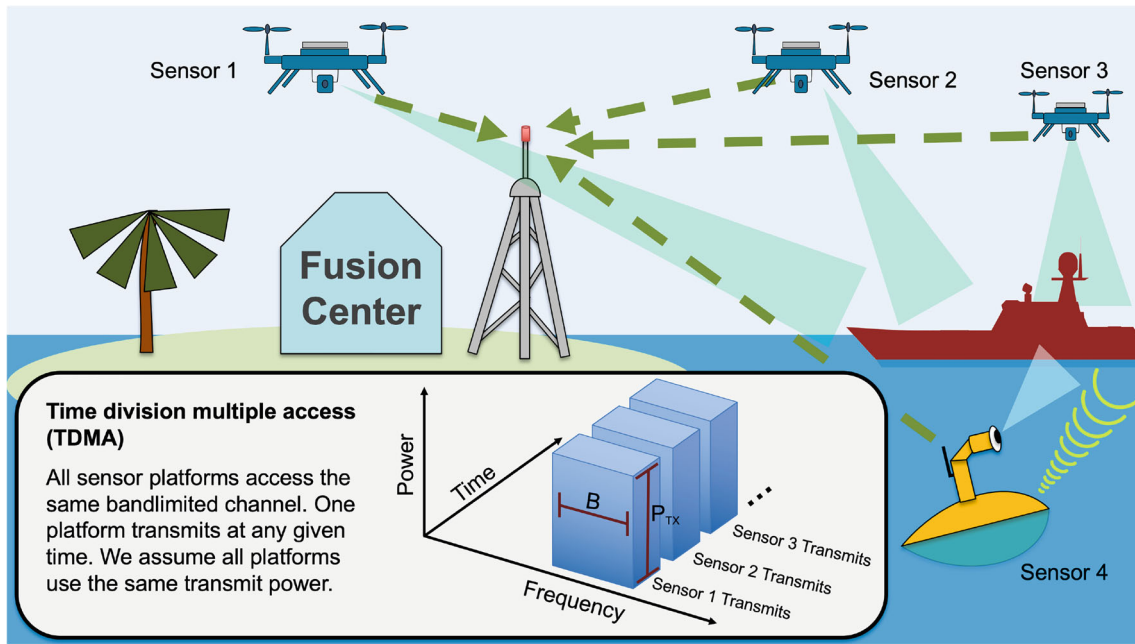


FIGURE 5 In this example, the different sensor platforms relay their measurements to the fusion center over a common wireless link using TDMA (see inset)

by each sensor can vary over sampling periods, since at sampling period t sensor i will send an average of about $[I(\mathbf{x}_t; \boldsymbol{\eta}_{i,t}^G | \boldsymbol{\eta}_{1:t-1}^G) + 1.254]$ bits, where the mutual information is computed under the probability measure imposed by the minimizing policy.

It should be noted that, due to our use of variable-length coding, the required airtime depends on the length of the codeword and is thus a random variable. In delay-tolerant applications, this may not be much of a problem. However, assume the sample rate of the sensors is fixed at T_s . The network will become congested if, during any sampling period, the required airtime exceeds T_s . In this setting, new measurements are being produced at a rate faster than the communication delay—more complicated scheduling algorithms may be required to ensure that the fusion center estimator achieves the desired performance by any real-time deadline (see, e.g., Reference 24). If the expected airtime, per sampling period, obtained from the optimization (23) is small compared to T_s , the Markov inequality can be used to derive a simple bound on the congestion probability. These ideas can be extended to more general types of communication channels (e.g., fading channels) and corresponding more general notions of capacity.

3 | CONVERSION TO DIFFERENCE-OF-CONVEX PROBLEM

In this section, we reformulate (23) and (25) as difference of convex (DC) optimization problems.

3.1 | Reformulation of mutual information

The mutual information (23a) can be expressed as:

$$\begin{aligned}
 I(\boldsymbol{\theta}_{i,t}^G; \boldsymbol{\eta}_{i,t}^G) &= h(\boldsymbol{\eta}_{i,t}^G) - h(\boldsymbol{\eta}_{i,t}^G | \boldsymbol{\theta}_{i,t}^G) \\
 &= h(C_i(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}^G) + \mathbf{u}_{i,t} + \mathbf{v}_{i,t}^G) - h(\boldsymbol{\theta}_{i,t}^G + \mathbf{v}_{i,t}^G | \boldsymbol{\theta}_{i,t}^G) \\
 &= \frac{1}{2} \log(C_i P_{t|t-1} C_i^\top + U_i + V_{i,t}) - \frac{1}{2} \log(V_{i,t}).
 \end{aligned} \tag{32}$$

Introducing the new variables $Q_{t|t-1} := P_{t|t-1}^{-1}$, $Q_{t|t} := P_{t|t}^{-1}$ and $\delta_{i,t} := (U_i + V_{i,t})^{-1}$, we can rewrite the expression derived for $I(\boldsymbol{\theta}_{i,t}^G; \boldsymbol{\eta}_{i,t}^G)$ in (32) as

$$\begin{aligned}
I(\theta_{i,t}^G; \eta_{i,t}^G) &= -\frac{1}{2} \log(\delta_{i,t}^{-1} - U_i) - \frac{1}{2} \log(\delta_{i,t}^{-1} + C_i Q_{t|t-1}^{-1} C_i^\top)^{-1} \\
&= \begin{cases} \min_{\psi_{i,t}, \gamma_{i,t}} & \frac{1}{2} \log \psi_{i,t} - \frac{1}{2} \log \gamma_{i,t} \\ \text{s.t.} & \psi_{i,t} \geq (\delta_{i,t}^{-1} - U_i)^{-1}, \\ & \gamma_{i,t} \leq (\delta_{i,t}^{-1} + C_i Q_{t|t-1}^{-1} C_i^\top)^{-1} \end{cases} \quad (33) \\
&= \begin{cases} \min_{\psi_{i,t}, \gamma_{i,t}} & \frac{1}{2} \log \psi_{i,t} - \frac{1}{2} \log \gamma_{i,t} \\ \text{s.t.} & \begin{bmatrix} \delta_{i,t} & \delta_{i,t} U_i & \delta_{i,t} \\ U_i \delta_{i,t} & U_i & 0 \\ \delta_{i,t} & 0 & \psi_{i,t} \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} \delta_{i,t} - \gamma_{i,t} & \delta_{i,t} C_i \\ C_i^\top \delta_{i,t} & Q_{t|t-1} + C_i^\top \delta_{i,t} C_i \end{bmatrix} \succeq 0. \end{cases} \quad (34)
\end{aligned}$$

The slack variables $\psi_{i,t}$ and $\gamma_{i,t}$ were introduced in (33). To obtain the first inequality constraint in (34), the Schur complement lemma is applied to the first constraint in (33). Likewise, the second inequality constraint in (34) follows from expanding the second inequality in (33) before applying the Schur complement lemma.

3.2 | Reformulation of mean square error

Using the variable $Q_{t|t}$, the fusion center's estimator MSE (23b) can be written as

$$\mathbb{E} \|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^G\|^2 = \text{Tr}(P_{t|t}) = \text{Tr}(Q_{t|t}^{-1}) = \begin{cases} \min_{S_t} & \text{Tr}(S_t) \\ \text{s.t.} & Q_{t|t}^{-1} \leq S_t \end{cases} = \begin{cases} \min_{S_t} & \text{Tr}(S_t) \\ \text{s.t.} & \begin{bmatrix} S_t & I \\ I & Q_{t|t} \end{bmatrix} \succeq 0 \end{cases}. \quad (35)$$

The final step of (35) follows from Schur complement formula.

3.3 | Reformulation of the original problem (23)

Using (22), (34), and (35), problem (23) can be reformulated as:

$$\min \quad \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \frac{\alpha_i}{2} (\log \psi_{i,t} - \log \gamma_{i,t}), \quad (36a)$$

$$\text{s.t.} \quad \begin{bmatrix} \delta_{i,t} & \delta_{i,t} U_i & \delta_{i,t} \\ U_i \delta_{i,t} & U_i & 0 \\ \delta_{i,t} & 0 & \psi_{i,t} \end{bmatrix} \succeq 0, \quad (36b)$$

$$\begin{bmatrix} \delta_{i,t} - \gamma_{i,t} & \delta_{i,t} C_i \\ C_i^\top \delta_{i,t} & Q_{t|t-1} + C_i^\top \delta_{i,t} C_i \end{bmatrix} \succeq 0, \quad (36c)$$

$$\begin{bmatrix} S_t & I \\ I & Q_{t|t} \end{bmatrix} \succeq 0, \quad \frac{1}{T} \sum_{t=1}^T \text{Tr}(S_t) \leq \beta, \quad (36d)$$

$$Q_{t|t} = Q_{t|t-1} + \sum_{i=1}^M \delta_{i,t} C_i^\top C_i, \quad (36e)$$

$$Q_{t|t-1}^{-1} = A Q_{t-1|t-1}^{-1} A^\top + F F^\top, \quad (36f)$$

where the decision variables are $[\psi_{i,t}, \delta_{i,t}, \gamma_{i,t}]$, for $i = 1, \dots, M$ and $t = 1, \dots, T$, S_t for $t = 1, \dots, T$, and $[Q_{t|t}, Q_{t|t-1}]$ for $t = 2, \dots, T$. The constraints (36e) and (36f) are imposed for $t = 2, \dots, T$, starting from $Q_{1|1} = P_{1|1}^{-1}$. All constraints in problem (36) are convex except (36f). In the following proposition, we show the nonconvex constraint (36f) can be replaced by a convex constraint without changing the nature of the problem. More precisely, we show that the equality in (36f) can be replaced by an inequality and we can solve (37), instead of (36), to obtain the optimal rate allocation.

$$\min \quad \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \frac{\alpha_i}{2} (\log \psi_{i,t} - \log \gamma_{i,t}), \quad (37a)$$

$$\text{s.t.} \quad \begin{bmatrix} \delta_{i,t} & \delta_{i,t} U_i & \delta_{i,t} \\ U_i \delta_{i,t} & U_i & 0 \\ \delta_{i,t} & 0 & \psi_{i,t} \end{bmatrix} \succeq 0, \quad (37b)$$

$$\begin{bmatrix} \delta_{i,t} - \gamma_{i,t} & \delta_{i,t} C_i \\ C_i^\top \delta_{i,t} & Q_{t|t-1} + C_i^\top \delta_{i,t} C_i \end{bmatrix} \succeq 0, \quad (37c)$$

$$\begin{bmatrix} S_t & I \\ I & Q_{t|t} \end{bmatrix} \succeq 0, \quad \frac{1}{T} \sum_{t=1}^T \text{Tr}(S_t) \leq \beta, \quad (37d)$$

$$Q_{t|t} = Q_{t|t-1} + \sum_{i=1}^M \delta_{i,t} C_i^\top C_i, \quad (37e)$$

$$Q_{t|t-1}^{-1} \succeq A Q_{t-1|t-1}^{-1} A^\top + F F^\top. \quad (37f)$$

Proposition 1. Let $[\psi_{i,t}^*, \delta_{i,t}^*, \gamma_{i,t}^*, S_t^*, Q_{t|t}^*, Q_{t|t-1}^*]$ be the optimal solution for (37). Then, $[\psi_{i,t}^*, \delta_{i,t}^*, \gamma_{i,t}^*, S_t^*, Q_{t|t}^{**}, Q_{t|t-1}^{**}]$ is the optimal solution for (36), where $Q_{t|t}^{**}$ and $Q_{t|t-1}^{**}$ are calculated recursively by

$$Q_{t|t-1}^{**} = A Q_{t-1|t-1}^{**} A^\top + F F^\top, \quad (38a)$$

$$Q_{t|t}^{**} = Q_{t|t-1}^{**} + \sum_{i=1}^M \delta_{i,t}^* C_i^\top C_i, \quad (38b)$$

starting from $Q_{1|1}^{**} = Q_{1|1}^*$. Moreover, if we denote the optimal values of (36) and (37) by J_1^* and J_2^* , respectively, we have $J_1^* = J_2^*$.

Proof. See Appendix C. ■

Constraint (37f) can be written as an equivalent linear matrix inequality (LMI)³⁵ condition. Therefore, problem (23) can be written as:

$$\min \quad \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \frac{\alpha_i}{2} (\log \psi_{i,t} - \log \gamma_{i,t}), \quad (39a)$$

$$\text{s.t.} \quad \begin{bmatrix} \delta_{i,t} & \delta_{i,t} U_i & \delta_{i,t} \\ U_i \delta_{i,t} & U_i & 0 \\ \delta_{i,t} & 0 & \psi_{i,t} \end{bmatrix} \succeq 0, \quad (39b)$$

$$\begin{bmatrix} \delta_{i,t} - \gamma_{i,t} & \delta_{i,t} C_i \\ C_i^\top \delta_{i,t} & Q_{t|t-1} + C_i^\top \delta_{i,t} C_i \end{bmatrix} \succeq 0, \quad (39c)$$

$$\begin{bmatrix} S_t & I \\ I & Q_{t|t} \end{bmatrix} \succeq 0, \quad \frac{1}{T} \sum_{t=1}^T \text{Tr}(S_t) \leq \beta, \quad (39d)$$

$$Q_{t|t} = Q_{t|t-1} + \sum_{i=1}^M \delta_{i,t} C_i^\top C_i. \quad (39e)$$

$$\begin{bmatrix} Q_{t|t-1} & Q_{t|t-1} A & Q_{t|t-1} F \\ A^\top Q_{t|t-1} & Q_{t-1|t-1} & 0 \\ F^\top Q_{t|t-1} & 0 & I \end{bmatrix} \succeq 0. \quad (39f)$$

It is simple to verify that (39) is the minimization of the difference of two convex functions subject to convex constraints and thus it is an instance of a difference-of-convex program. The infinite horizon (time-invariant) counterpart of (39) can be formulated as:

$$\min \sum_{i=1}^M \frac{\alpha_i}{2} (\log \psi_i - \log \gamma_i), \quad (40a)$$

$$\text{s.t.} \quad \begin{bmatrix} \delta_i & \delta_i U_i & \delta_i \\ U_i \delta_i & U_i & 0 \\ \delta_i & 0 & \psi_i \end{bmatrix} \succeq 0, \quad \forall i = 1, \dots, M, \quad (40b)$$

$$\begin{bmatrix} \delta_i - \gamma_i & \delta_i C_i \\ C_i^\top \delta_i & \hat{Q} + C_i^\top \delta_i C_i \end{bmatrix} \succeq 0, \quad \forall i = 1, \dots, M, \quad (40c)$$

$$Q = \hat{Q} + \sum_{i=1}^M \delta_i C_i^\top C_i, \quad \text{Tr}(S) \leq \beta, \quad (40d)$$

$$\begin{bmatrix} S & I \\ I & Q \end{bmatrix} \succeq 0, \quad \begin{bmatrix} \hat{Q} & \hat{Q}A & \hat{Q}F \\ A^\top \hat{Q} & Q & 0 \\ F^\top \hat{Q} & 0 & I \end{bmatrix} \succeq 0. \quad (40e)$$

4 | CONVEX-CONCAVE PROCEDURE (CCP)

In this section, we develop an iterative algorithm based on the convex-concave procedure (CCP)⁵ to find local minima of the time-varying and time-invariant sensor data-rate allocation problems (39) and (40). CCP starts by over-approximating the nonconvex terms of the DC program via linearization around a nominal point. The resulting convex problem can then be solved efficiently, and the algorithm iterates by convexifying the problem around the obtained solution. The process terminates when a local minimum is found. Starting from a feasible nominal point, as it is shown in Reference 5, all subsequent iterations will be feasible and the algorithm converges to a local minima.

We now show how the CCP algorithm operates for the time-invariant problem (40). For this problem, we linearize the concave term $\log \psi_i$ around the nominal point $\hat{\psi}_i$ which gives an upper bound as $\log \psi_i \leq \frac{1}{\hat{\psi}_i}(\psi_i - \hat{\psi}_i) + \log \hat{\psi}_i$. Thus, at each iteration of CCP algorithm we solve:

$$\min \sum_{i=1}^M \frac{\alpha_i}{2} (\psi_i / \hat{\psi}_i - 1 + \log \hat{\psi}_i - \log \gamma_i), \quad (41a)$$

$$\text{s.t.} \quad (40b) - (40e). \quad (41b)$$

If we denote the solution of (41) at iteration k by $(\psi_i^k, \delta_i^k, \gamma_i^k, S^k, Q^k, \hat{Q}^k)$, at iteration $k + 1$ (41) is solved for $\hat{\psi}_i = \psi_i^k$. Note that the optimal value of (41) is an upper bound for the optimal value of (40) with same decision variables $S, Q, \hat{Q}, \psi_i, \delta_i$, and

Algorithm 1. Convex-concave procedure (CCP)

Set *tolerance* sufficiently small;

Set initial value $\hat{\psi}_i \leftarrow 1$ for $i = 1, 2, \dots, M$;

for $k = 1, 2, \dots$ **do**

 solve (41).

$(\psi_i^k, \delta_i^k, \gamma_i^k, S^k, Q^k, \hat{Q}^k) \leftarrow$ Obtain optimal solution;

$f^k \leftarrow$ Optimal value of (41);

 Update $\hat{\psi}_i \leftarrow \psi_i^k$ for $i = 1 : M$;

if $f^{k-1} - f^k \leq \text{tolerance}$ **then**

 break;

end if

end for

γ_i for $i = 1, \dots, M$. The CCP algorithm we use to perform the time-invariant optimization (40) is outlined in Algorithm 1. A similar algorithm can be used to solve (39). Problem (41) is a log-det minimization and can be solved via SDP solvers in polynomial time.³⁶ As it will be demonstrated in Section 5, in our simulations Algorithm 1 tends to converge in a few tens of CCP iterations.

5 | NUMERICAL ANALYSIS

In this section, we present two numerical simulations showing the effectiveness of Algorithm 1. In both studies, we optimize the sum bitrate by setting $\alpha_i = 1$ for all sensors.

5.1 | Heat diffusion system

We first consider the problem of estimating the temperature distribution over a time-invariant 1D heat diffusion system.³⁷ We consider a slender concrete rod that is 15 m long. The length of the rod is equally divided by 60 nodes. The continuous-time heat diffusion system matrix A is given by

$$A_c = \frac{a}{h^2} \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}$$

where $a = 7.5 \times 10^{-7} \text{ m}^2/\text{s}$ is the thermal diffusivity and $h = 0.2459 \text{ m}$ is the length of each segment. We assume the system is driven by a random heat input at each node. Let $E_d = I - \frac{\Delta_t}{2} A_c$ and $A_d = I + \frac{\Delta_t}{2} A_c$. We consider tracking the temperature of the rod at each node. A discrete-time model for this system of the form (1) is found by setting

$$A = E_d^{-1} A_d, \text{ and} \quad (42a)$$

$$F = I. \quad (42b)$$

In these simulations, we set $\Delta_t = 1$. Our set of sensor platforms is a thermometer placed at each node; we assume that the thermometers can measure the node's temperature noiselessly. Let e_i for $i \in 1 : 60$ denote unit vectors from the standard basis. Formally, the measurement matrix for the i^{th} sensor is

$$C_i = e_i e_i^T. \quad (43)$$

The measurement data is transmitted to a data fusion center over a communication network as described in Section 2.

Figure 6 shows the performance of Algorithm 1 as a function of the number of CCP iterations assuming that the constraint on the fusion center's MSE is $\beta = 0.1$. The algorithm is initialized with an equal data rate (quantizer sensitivity) assigned to each sensor. Figure 6A plots the sum data rate obtained by the solution obtained after a given number of CCP iterations, while Figure 6B shows the data rates assigned to each sensor after each such iteration. In Figure 6B the assigned data rate is presented as color-coded block at each iteration. Figure 6A,B demonstrate that both the total data rate for the network, and the data rates assigned to each sensor platform, converge after a sufficient number of CCP iterations.

Figure 7 shows the data rate allocation obtained at convergence (i.e., after a sufficient number of CCP iterations) for three different values of β . We observe that a similar subset of sensors are given nonzero data rates for all three β values. As one might expect, the allocated data rates to individual sensors decrease as β increases. Increasing β corresponds to a less stringent demand on the performance of the fusion center's estimator MSE; a less demanding MSE constraint means that satisfactory performance can be achieved with less bitrate. Moreover, it is seen that more sensors are allocated zero data rate as β increases.

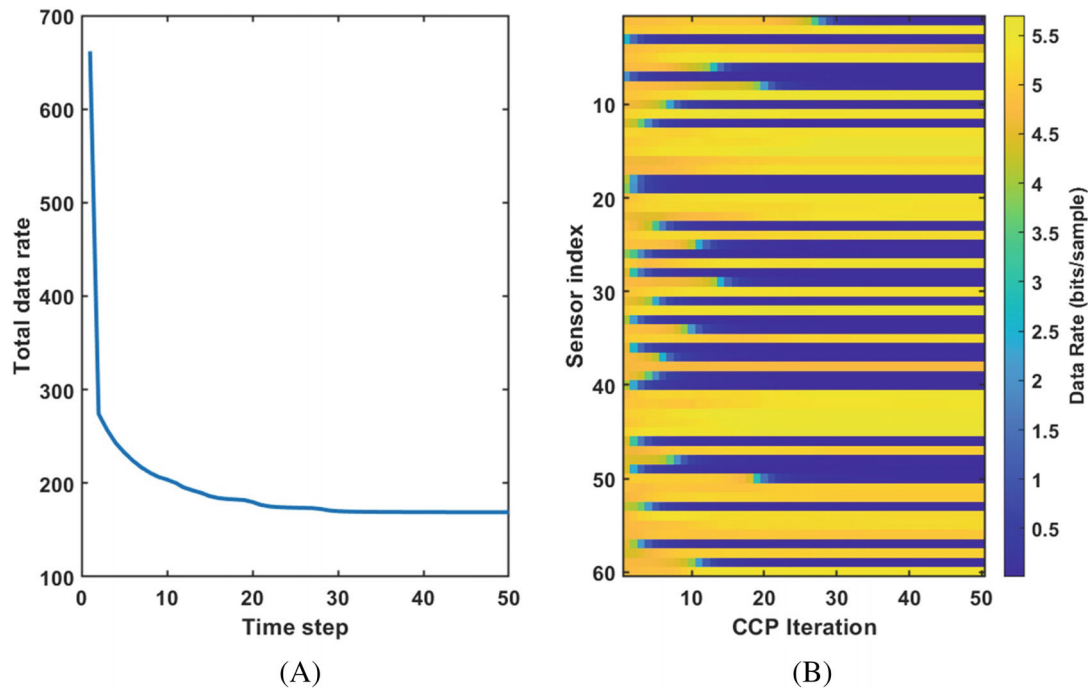


FIGURE 6 These plots show optimized data rates found after a given number of CCP iterations with a MSE constraint of $\beta = 0.1$. In plot (A), the total data rate is illustrated. Plot (B) illustrates the data rates allocated to individual sensors. The data rates are represented as color-coded blocks

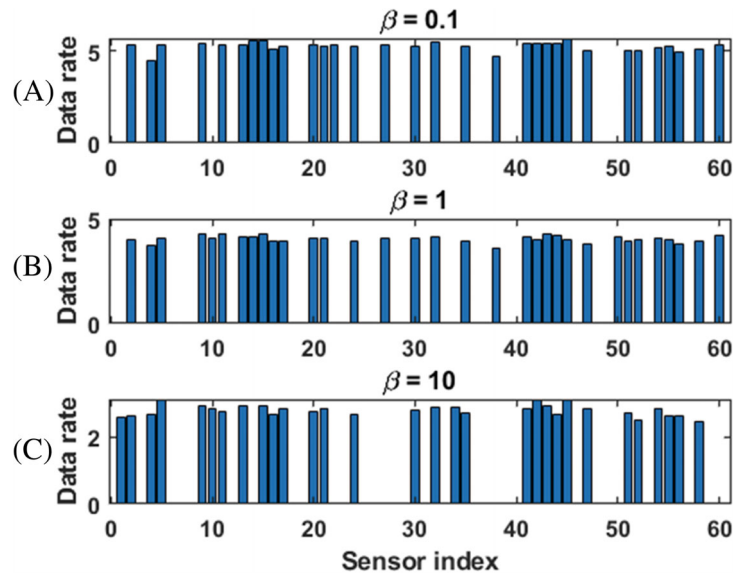


FIGURE 7 Data rate allocation obtained by CCP with (A) $\beta = 0.1$ (B) $\beta = 1$ (C) $\beta = 10$

To confirm this last observation, Figure 8 illustrates the number of sensors allocated nonzero data rate after a sufficient number of CCP iterations as β is varied from 1 to 180. The number of sensors tends to decrease, but is somewhat stagnant in the middle of the beta range. The relationship is not necessarily monotone. This plot demonstrates a sparsity-promoting property of the proposed data rate allocation method.

5.2 | Target tracking by a drone swarm

Next, we consider the problem tracking multiple moving targets using a radar signal and multiple drones. As in Reference 38, we consider a scenario in which a base station, which also serves and the data fusion center, illuminates

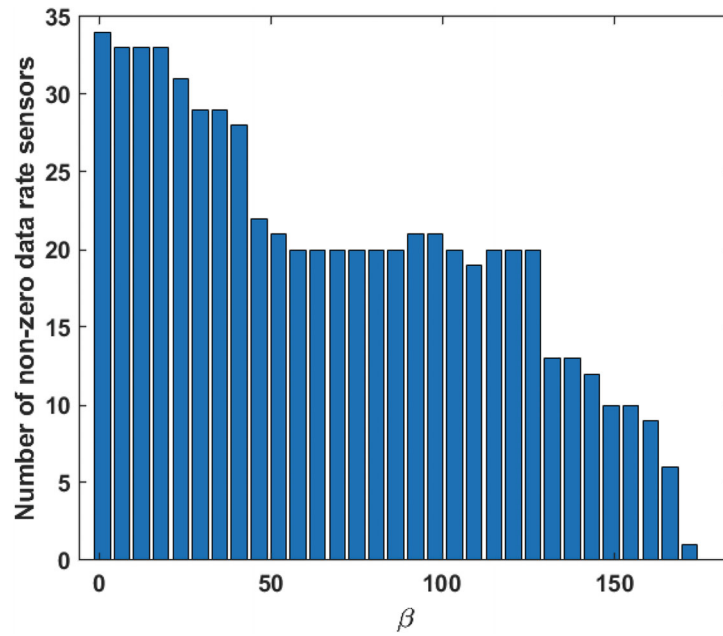


FIGURE 8 Number of sensors allocated nonzero data rate by CCP tested over the range $1 \leq \beta \leq 180$

the set of targets with a radar signal. A set of drones, nominally the sensor platforms, make time-delay and Doppler measurements of the radar signal reflected by the targets. The drones transmit these measurements back to the data fusion center, where the targets' positions are estimated by an extended Kalman filter (EKF). Unlike the previous example, the system “C” matrix in (2) is time-varying as the relative positions of the drones with respect to the targets change over time. Therefore, we consider an approach to recalculate the rate allocation at every time step by an repetitive executions of Algorithm 1. To improve computational efficiency, we adopt a “warm start” implementation of Algorithm 1 at every time step, that is, the optimal allocation from the previous time step is used as the initial condition for the CCP iteration in the next time step.

We assume there are five targets in the entire 2D environment. Each target is assumed to be a point mass driven by a random force which is modeled as an i.i.d. Gaussian process noise. The state vector to be estimated in this simulation study is therefore (1) with

$$x_t = \begin{bmatrix} x_t^1 & x_t^2 & x_t^3 & x_t^4 & x_t^5 \end{bmatrix}^T \in \mathbb{R}^{20},$$

where $x_t^i = \begin{bmatrix} p_{x,t}^i & p_{y,t}^i & v_{x,t}^i & v_{y,t}^i \end{bmatrix}^T$ are the position and the velocity of the i th target. Let

$$A_{\text{Target}} = \begin{bmatrix} 1 & 0 & \Delta_t & 0 \\ 0 & 1 & 0 & \Delta_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

where $\Delta_t = 1$ is the step size. The A matrix for the five target system is given by $A = \text{diag}(A_{\text{Target}}, A_{\text{Target}}, A_{\text{Target}}, A_{\text{Target}}, A_{\text{Target}})$. Likewise, let

$$F_{\text{Target}} = \begin{bmatrix} \sqrt{10} & 0 & 0 & 0 \\ 0 & \sqrt{10} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The F matrix for the five target system is defined as $F = \text{diag}(F_{\text{Target}}, F_{\text{Target}}, F_{\text{Target}}, F_{\text{Target}}, F_{\text{Target}})$.

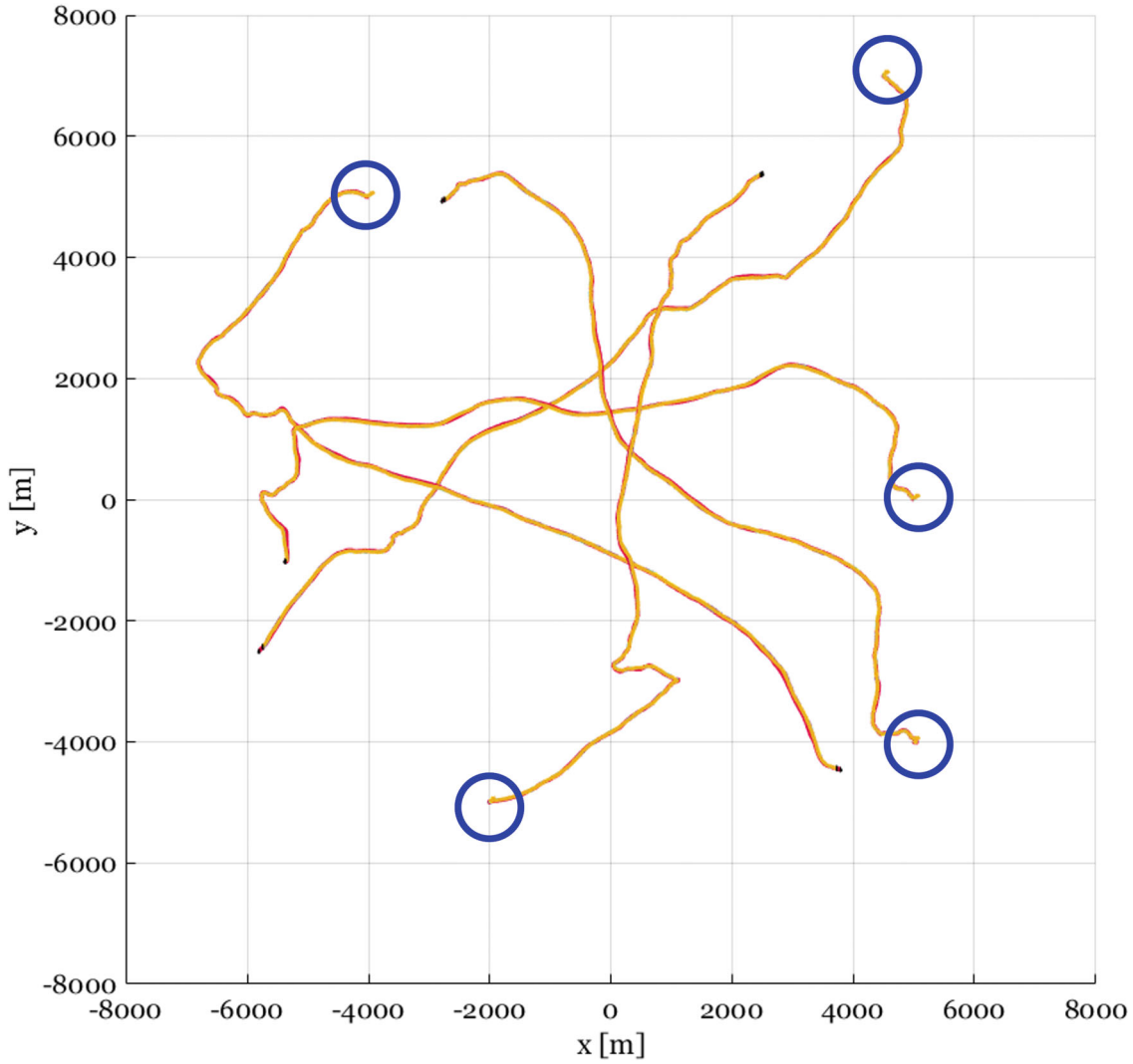


FIGURE 9 Target tracking simulation result in 5 separate regions. 1 target (red line) and 5 drones (other color lines) are in each region

As shown in Figure 9, we assume that five drones are available in the neighborhood of each target (so there are 25 drones in total). Each drone is modeled as a point mass in a 2D plane. Let $(p_{x,t}^{ij}, p_{y,t}^{ij})$ be the coordinate of the j th drone in the neighborhood of the i -th target, and $(v_{x,t}^{ij}, v_{y,t}^{ij})$ be its velocity. Drones are controlled by the data fusion center to track the target. In this simulation, we assume that a PD control with acceleration input is used for each drone:

$$\begin{aligned} a_{x,t}^{ij} &= K_P(\hat{p}_{x,t}^i - p_{x,t}^{ij}) + K_D(\hat{v}_{x,t}^i - v_{x,t}^{ij}) - \sum_{k \neq j} L_P(p_{x,t}^{i,k} - p_{x,t}^{ij}) - \sum_{k \neq j} L_D(v_{x,t}^{i,k} - v_{x,t}^{ij}), \\ a_{y,t}^{ij} &= K_P(\hat{p}_{y,t}^i - p_{y,t}^{ij}) + K_D(\hat{v}_{y,t}^i - v_{y,t}^{ij}) - \sum_{k \neq j} L_P(p_{y,t}^{i,k} - p_{y,t}^{ij}) - \sum_{k \neq j} L_D(v_{y,t}^{i,k} - v_{y,t}^{ij}). \end{aligned} \quad (44)$$

Here, $(\hat{p}_{x,t}^i, \hat{p}_{y,t}^i, \hat{v}_{x,t}^i, \hat{v}_{y,t}^i)$ is the estimated state of the target i computed by the data fusion center using the EKF described below. The gains K_P and K_D are tuned to keep the drones close to the target, whereas gains L_P and L_D are selected so that drones in the same neighborhood stay away from each other.

At every time step, the base station illuminates the targets with a radar signal. We assume that the reflected signal from the i th target is observed by the drones in the i th region. Each drone in the i th region obtains a two-dimensional measurement of the i -th target. The j th drone in the i th region

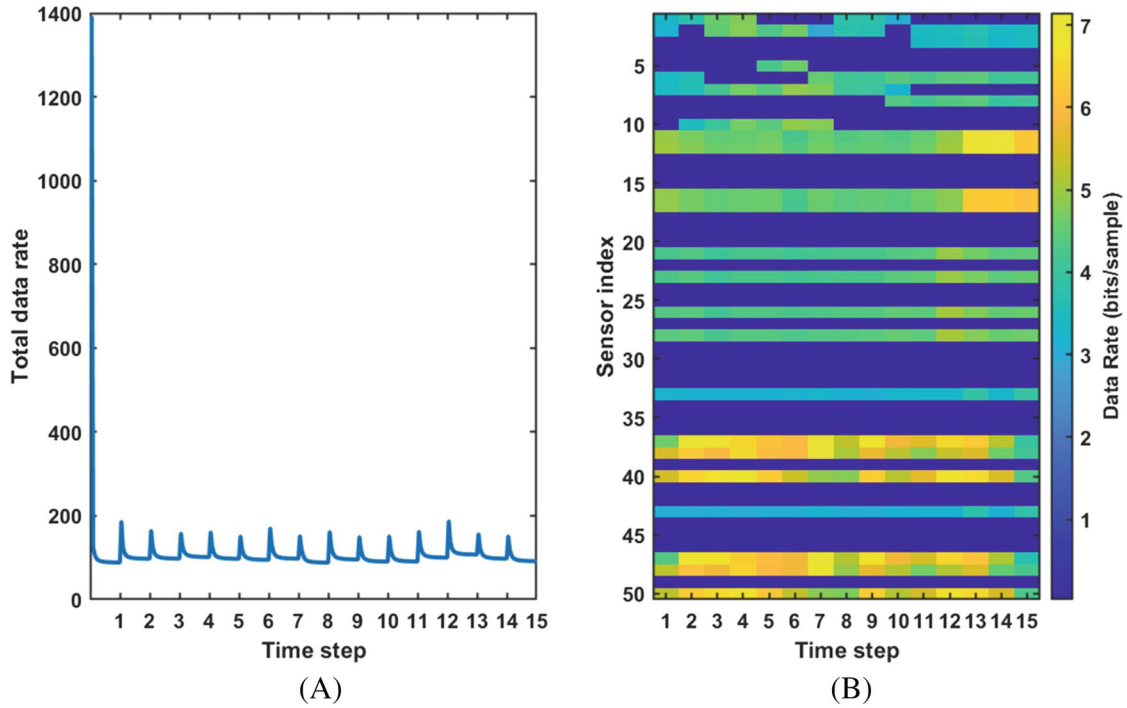


FIGURE 10 Optimized data rates over 15 time steps, with 15 iterations of the CCP algorithm performed per timestep. The MSE constraint is $\beta = 1$. (A) demonstrates the convergence of the CCP algorithm from the warm start solutions—when the time index on the horizontal axis updates from time $t - 1$ to time t , the optimization is restarted for the system matrices at time t . In between times t and $t + 1$, we plot the communication cost obtained for fifteen iterations of the CCP algorithm. (B) illustrates the data rates assigned to each sensor at the convergence of the CCP algorithm (nominally 15 iterations) at each time step. The assigned data rates are represented by color-coded blocks

receives Doppler f_t^{ij} and time delay τ_t^{ij} measurements of the reflected radar signal, which are given by Zhan et al.³⁸

$$f_t^{ij} = \frac{(v_{x,t}^i p_{x,t}^i + v_{y,t}^i p_{y,t}^i)}{\sqrt{(p_{x,t}^i)^2 + (p_{y,t}^i)^2}} + \frac{(v_{x,t}^i - v_{x,t}^{ij})(p_{x,t}^i - p_{x,t}^{ij}) + (v_{y,t}^i - v_{y,t}^{ij})(p_{y,t}^i - p_{y,t}^{ij})}{\sqrt{(p_{x,t}^i - p_{x,t}^{ij})^2 + (p_{y,t}^i - p_{y,t}^{ij})^2}}, \quad (45)$$

$$\tau_t^{ij} = \sqrt{(p_{x,t}^i)^2 + (p_{y,t}^i)^2} + \sqrt{(p_{x,t}^i - p_{x,t}^{ij})^2 + (p_{y,t}^i - p_{y,t}^{ij})^2}. \quad (46)$$

Since (45) and (46) are nonlinear observations of the state vector x_t , we linearize them around the current best estimate $(\hat{p}_{x,t}^i, \hat{p}_{y,t}^i, \hat{v}_{x,t}^i, \hat{v}_{y,t}^i)$, where these estimates are recursively computed using the EKF.

Figure 9 shows the scenario described above simulated for 1000 time steps. Blue circles represent neighborhood regions of each target, and there are five drones in each region. The target trajectory is presented as red line and drone trajectories in other colors. Based on the calculated position and velocity of target and drones, we apply the CCP iterations in each time step. Figure 10 shows the rate allocation obtained as the CCP iterations converge in each time step. In Figure 10A we plot the total data rate allocation, while in Figure 10B we illustrate the assigned data rates to individual sensors, at each time step, at the convergence of the CCP algorithm (nominally 15 iterations). Note that there are 50 possible sensors, since each of the 25 drones receives both range and Doppler measurements. The result changes over time, reflecting the time-varying nature of the considered problem.

To gain further insight, Figure 11 shows the data rate allocated to drones in region 1 and their spatial positions at time step $t = 70$. Figure 11A indicates that nonzero data rates (yellow color block) are given only to sensors 2, 3, 7, and 8, which correspond to the delay and Doppler measurements obtained from drones 2 and 3. Figure 11C indicates that drones 2 and 3 are the left-most and right-most vehicles in the formation.

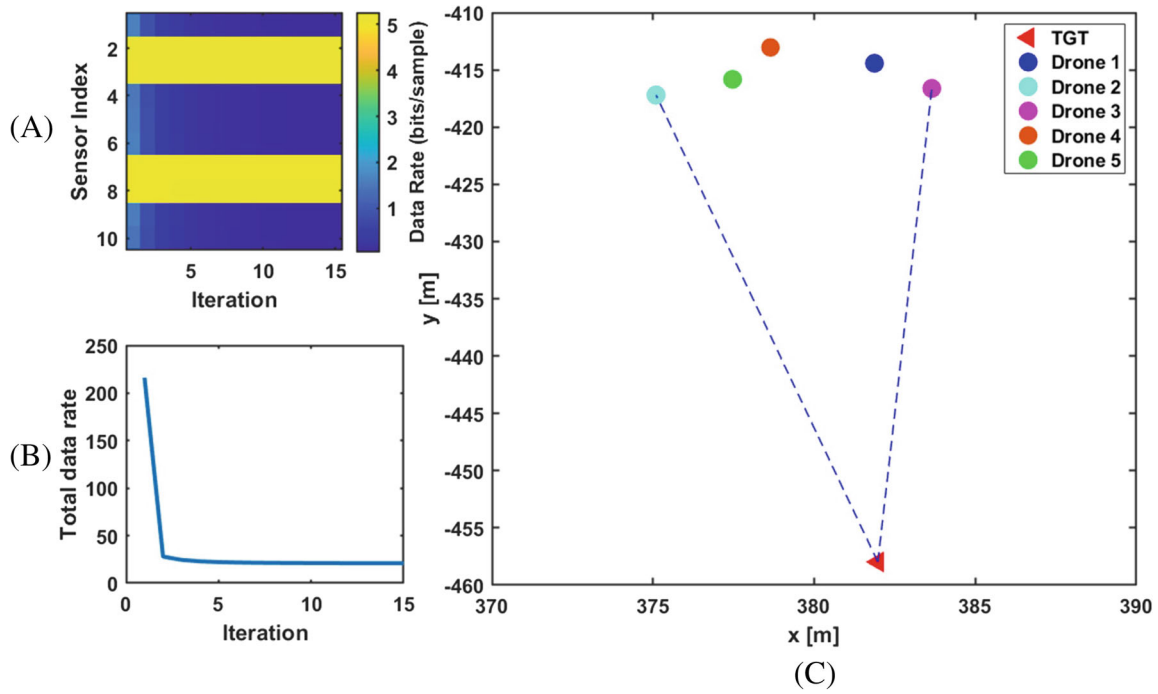


FIGURE 11 Data rate allocation in Region 1 at $t=70$ ($\beta = 1$). In (A), we show the data rate allocation to each sensor after a given number of CCP iterations. Indices 1–5 in (A) correspond to the time delay measurements $\tau_t^{1,j}$, $j = 1, \dots, 5$, while indices 6–10 correspond to the Doppler estimates $f_t^{1,j}$, $j = 1, \dots, 5$. In (B), we plot the total data rate allocation against the number of CCP iterations. In (C) we plot the physical position of the target (TGT) and drones. Dotted lines are drawn from drones allocated a nonzero data rate to the target. It is seen that the two drones allocated nonzero data rates have the largest possible angular spread, and, in a sense, the most “different” measurements, among pairs of drones

6 | DISCUSSION

As observed in Section 5, the SRA formulation (25) (similarly, (23)) admits sparse solutions, that is, for several sensors we have $\delta_i = 0$. In this section, we take a closer look at the mathematical structure of the SRA problem. We first remark that the sparsity-promoting nature of the solution can be attributed to the mathematical structure of the communication cost. The communication cost in (37a) is reminiscent of the *sum-of-logs* regularizer, a widely used heuristic to induce sparse solutions.³⁹

To develop further insight, we consider a special case of sensor rate allocation problem for which the closed-form solution is available. Consider a scalar system described by

$$\mathbf{x}_{t+1} = a\mathbf{x}_t + f\mathbf{w}_t, \quad \mathbf{w}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1),$$

and a single measurement

$$\boldsymbol{\eta}_t = \mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, V_t).$$

Since the system is observable, $P \triangleq \lim_{t \rightarrow \infty} P_{t|t}$ exists and is computed by the algebraic Riccati equation (ARE)

$$P^{-1} = (a^2P + f^2)^{-1} + V^{-1}, \quad (47)$$

where $V \triangleq \limsup_{t \rightarrow \infty} V_t$. The time-invariant rate allocation problem (25) for this system is

$$\begin{aligned} \min_{P, V \geq 0} \quad & \log \left(a^2 + \frac{f^2}{P} \right) \\ \text{s.t.} \quad & P \leq \beta \\ & P^{-1} = (a^2P + f^2)^{-1} + V^{-1}. \end{aligned} \quad (48)$$

Denote by $P = g(V)$ the unique positive solution to the ARE (47). It can be shown that $g(V)$ is a strictly increasing, and thus invertible, function of V with a bounded range $\left[g(0) = 0, g(\infty) = \frac{f^2}{1-a^2}\right]$. Therefore, problem (48) can be equivalently written as

$$\begin{aligned} \min_{P \geq 0} \quad & \log \left(a^2 + \frac{f^2}{P} \right) \\ \text{s.t.} \quad & P \leq \beta \\ & P \leq \frac{f^2}{1-a^2}. \end{aligned} \quad (49)$$

It is straightforward to verify that the minimizer of (49) is $P^* = \min \left\{ \beta, \frac{f^2}{1-a^2} \right\}$ and the optimal value of (49) is $\min \left\{ 0, \log \left(a^2 + \frac{f^2}{\beta} \right) \right\}$. The P^* is obtained by adopting $V^* = g^{-1}(P^*)$. The condition $P^* = \frac{f^2}{1-a^2}$, leading to $V^* = \infty$, corresponds to the condition when it is optimal to allot zero data rate to the sensor. This partly explains the sparse nature of the solutions we observed in Section 5. It is worth noting that problem (49) is equivalent to the scalar infinite-horizon sequential rate-distortion problem studied in References 40 and 41.

7 | CONCLUSION

In this article, we considered a sensor data rate allocation problem for dynamic sensor fusion over resource restricted communication networks. Using a system model motivated by practical remote sensing scenarios and well-understood methods of zero delay source coding, we proposed a data rate allocation problem between a group of remote sensors and a fusion center. We reformulated this problem using information-theoretic tools and ideas from Kalman filtering. Our proposed optimization took the form of a difference-of-convex program. We applied the CCP algorithm to find a heuristic solution. We conducted a numerical study on a 1D heat transfer model and on a 2D target tracking by drone swarm scenario. Our numerical results suggest that the proposed approach is sparsity-promoting. By considering a limiting case of the SRA problem that admitted an analytical solution, we gained insight into the sparsity-promoting property.

Possible directions for future research include analyzing the nonconvexity of the SRA problem and developing computationally efficient algorithms to perform the minimization. In particular, when considering the infinite horizon sensor resource allocation problem, we restricted our optimization to time-invariant rate allocations (i.e., time-invariant quantizer sensitivities). The existence (or lack thereof) of time-varying (e.g., periodic) solutions that outperform optimal time-invariant solutions is an interesting question that presents an opportunity for future work. In Reference 42, the optimality of time-invariant solutions is established for a related (but not equivalent) problem called the sequential rate-distortion (SRD) problem. The convexity of the multistage SRD problem allows one to effectively “single-letterize” the optimization. Unfortunately, in our current setting, the multistage SRA problem (e.g., Equation (36)) is not a convex program. This makes it difficult to apply the proof techniques from Reference 42 to the present setup.

Extensions to unreliable networks are attractive avenues for future research. Notably, our notion of airtime in Section 2.6 relied infinite blocklength expressions for channel capacity. It would be interesting to incorporate modern tools from finite blocklength information theory (cf. Reference 33) to better quantify “real world” communication costs and notions of reliability. Likewise, our numerical simulations in Section 5 assumed reliable communication channels. Incorporating realistic physical layer channel impairments into simulations like those in Section 5 presents another exciting opportunity for future investigations. Finally, it may also be possible to discover a more illuminating connection to the CEO problem.

ACKNOWLEDGMENT

This work was supported by the Defense Advanced Research Projects Agency (DARPA) under Grant D19AP00078 and National Science Foundation (NSF) under Award 1944318.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Hyunho Jung  <https://orcid.org/0000-0001-5491-3532>

Ali Reza Pedram  <https://orcid.org/0000-0003-4619-2231>

REFERENCES

- Collins J. The challenges facing U.S. Navy aircraft electrical wiring systems. Proceedings of the 9th Annual Aging Aircraft Conference; 2006.
- Marr B. That's data science: airbus puts 10,000 sensors in every single wing; 2015 [Online]. <https://www.datasciencecentral.com/profiles/blogs/that-s-data-science-airbus-puts-10-000-sensors-in-every-single>
- Zamir R, Feder M. On universal quantization by randomized uniform/lattice quantizers. *IEEE Trans Inf Theory*. 1992;38(2):428-436. doi:10.1109/18.119699
- Tanaka T, Johansson KH, Oechtering T, Sandberg H, Skoglund M. Rate of prefix-free codes in LQG control systems. Proceedings of the IEEE International Symposium on Information Theory; 2016:2399-2403. doi:10.1109/ISIT.2016.7541729
- Lipp T, Boyd S. Variations and extension of the convex concave procedure. *Optim Eng*. 2016;17(2):263-287. doi:10.1007/s11081-015-9294-x
- Jung H, Tanaka T. Optimal data rate allocation for dynamic sensor fusion over resource constrained communication networks. Proceeding of the 2021 American Control Conference (ACC); 2021:5145-5150. doi:10.23919/ACC50511.2021.9482775
- Kammer D. Sensor placement for on-orbit modal identification and correlation of large space structures. Proceeding of the American Control Conference; 1990:2984-2990. doi:10.23919/ACC.1990.4791265
- Yao L, Sethares W, Kammer D. Sensor placement for on-orbit modal identification of large space structure via a genetic algorithm. Proceedings of the IEEE International Conference on Systems Engineering; 1992:332-335. doi:10.1109/ICSYSSE.1992.236889
- Wang H, Yao K, Pottie G, Estrin D. Entropy-based sensor selection heuristic for target localization. Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks; 2004:36-45. doi:10.1145/984622.984628
- Joshi S, Boyd S. Sensor selection via convex optimization. *IEEE Trans Signal Process*. 2009;57(2):451-462. doi:10.1109/TSP.2008.2007095
- Hoffmann GM, Tomlin CJ. Mobile sensor network control using mutual information methods and particle filters. *IEEE Trans Automat Contr*. 2010;55(1):32-47. doi:10.1109/TAC.2009.2034206
- Yan Q, Chen J. Sensor selection via maximizing hybrid Bayesian fisher information and mutual information in unreliable sensor networks. *Electronics*. 2020;9(2):1-15. doi:10.3390/electronics9020283
- Mo Y, Ambrosino R, Sinopoli B. Sensor selection strategies for state estimation in energy constrained wireless sensor networks. *Automatica*. 2011;47(7):1330-1338. doi:10.1016/j.automatica.2011.02.001
- Liu L, Liu D, Zhang Y, Peng Y. Effective sensor selection and data anomaly detection for condition monitoring of aircraft engines. *Sensors*. 2016;16(5):1-16. doi:10.3390/s16050623
- Silva EI, Derpich MS, Ostergaard J. A framework for control system design subject to average data-rate constraints. *IEEE Trans Automat Contr*. 2011;56(8):1886-1899. doi:10.1109/TAC.2010.2098070
- Stefan J, Tanaka T. Sparse LQR synthesis via information regularization. Proceedings of the 58th IEEE Conference on Decision and Control (CDC); 2019:7345-7351. doi:10.1109/CDC40024.2019.9029415
- Berger T, Zhang Z, Viswanathan H. The CEO problem [multiterminal source coding]. *IEEE Trans Inf Theory*. 1996;42(3):887-902. doi:10.1109/18.490552
- Viswanathan H, Berger T. The quadratic Gaussian CEO problem. *IEEE Trans Inf Theory*. 1997;43(5):1549-1559. doi:10.1109/18.623151
- Kostina V, Hassibi B. Fundamental limits of distributed tracking. Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT); 2020: 2438-2443. doi:10.1109/ISIT44484.2020.9174006
- Massey JL. Causality, feedback and directed information. Proceeding of International Symposium of Information Theory and its Applications (ISITA-90); 1990:303-305.
- Yilmaz ON, Wang YPE, Johansson NA, Brahmi N, Ashraf SA, Sachs J. Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case. Proceedings of the 2015 IEEE International Conference on Communication Workshop (ICCW); 2015:1190-1195. doi:10.1109/ICCW.2015.7247339
- Durisi G, Koch T, Popovski P. Toward massive, ultrareliable, and low-latency wireless communication with short packets. *Proc IEEE*. 2016;104(9):1711-1726. doi:10.1109/JPROC.2016.2537298
- Hou Z, She C, Li Y, Zhuo L, Vucetic B. Prediction and communication co-design for ultra-reliable and low-latency communications. *IEEE Trans Wirel Commun*. 2020;19(2):1196-1209. doi:10.1109/TWC.2019.2951660
- Jurdi R, Andrews JG, Heath RW. Scheduling observers over a shared channel with hard delivery deadlines. *IEEE Trans Commun*. 2021;69(1):133-148. doi:10.1109/TCOMM.2020.3032172
- Baumann D, Mager F, Wetzker U, Thiele L, Zimmerling M, Trimpe S. Wireless control for smart manufacturing: recent approaches and open challenges. *Proc IEEE*. 2021;109(4):441-467. doi:10.1109/JPROC.2020.3032633
- Yu CT, Varshney P. Bit allocation for discrete signal detection. *IEEE Trans Commun*. 1998;46(2):173-175. doi:10.1109/26.659474
- Cheng X, Ciuonzo D, Rossi PS. Multibit decentralized detection through fusing smart and dumb sensors based on rao test. *IEEE Trans Aersp Electron Syst*. 2020;56(2):1391-1405. doi:10.1109/TAES.2019.2936777

28. Ciunzio D, Javadi SH, Mohammadi A, Rossi PS. Bandwidth-constrained decentralized detection of an unknown vector signal via multisensor fusion. *IEEE Trans Signal Inf Process Netw*. 2020;6:744-758. doi:10.1109/TSIPN.2020.3037832
29. Cover T, Thomas J. *Elements of Information Theory*. 2nd ed. Wiley-Interscience; 2006.
30. Yüksel S. Jointly optimal LQG quantization and control policies for multi-dimensional systems. *IEEE Trans Automat Contr*. 2014;59(6):1612-1617. doi:10.1109/TAC.2013.2293414
31. Derpich MS, Ostergaard J. Improved upper bounds to the causal quadratic rate-distortion function for gaussian stationary sources. *IEEE Trans Inf Theory*. 2012;58(5):3131-3152. doi:10.1109/TIT.2012.2184669
32. Silva E. *A Unified Framework for the Analysis and Design of Networked Control Systems*. PhD thesis. School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, Australia; 2009.
33. Polyanskiy Y, Poor HV, Verdu S. Channel coding rate in the finite blocklength regime. *IEEE Trans Inf Theory*. 2010;56(5):2307-2359. doi:10.1109/TIT.2010.2043769
34. Costello DJ, Forney GD. Channel coding: the road to channel capacity. *Proc IEEE*. 2007;95(6):1150-1177. doi:10.1109/JPROC.2007.895188
35. Boyd S, Ghaoui L, Feron E, Balakrishnan V. *Linear Matrix Inequalities in System and Control Theory*. 1st ed. SIAM; 1994:7-12.
36. Ben-Tal A, Nemirovski A. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM; 2001.
37. Chahlaoui Y. A collection of benchmark examples for model reduction of linear time invariant dynamical systems. SLICOT Working Note 2002-2, Subroutine Library in Systems and Control Theory; 2002. <http://slicot.org/20-site/126-benchmark-examples-for-model-reduction>
38. Zhan P, Casbeer D, Swindlehurst A. A centralized control algorithm for target tracking with UAVs. Proceedings of the Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers; 2005:1148-1152. doi:10.1109/ACSSC.2005.1599940
39. Candes EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted L1 minimization. *J Fourier Anal Appl*. 2008;14(5-6):877-905. doi:10.1007/s00041-008-9045-x
40. Tanaka T, Kim KKK, Parrilo PA, Mitter SK. Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs. *IEEE Trans Automat Contr*. 2016;62(4):1896-1910. doi:10.1109/TAC.2016.2601148
41. Tatikonda S, Sahai A, Mitter S. Stochastic linear control over a communication channel. *IEEE Trans Automat Contr*. 2004;49(9):1549-1561. doi:10.1109/TAC.2004.834430
42. Tanaka T. Semidefinite representation of sequential rate-distortion function for stationary Gauss-Markov processes. Proceedings of the IEEE Conference on Control Applications; 2015:1217-1222. doi:10.1109/CCA.2015.7320778

How to cite this article: Jung H, Pedram AR, Cuvelier TC, Tanaka T. Optimized data rate allocation for dynamic sensor fusion over resource constrained communication networks. *Int J Robust Nonlinear Control*. 2023;33(1):237-263. doi: 10.1002/rnc.6076

APPENDIX A. PROOF OF LEMMA 1

Recall that $\eta_{i,t} = \mathbf{q}_{i,t} - \xi_{i,t}$. Thus

$$H(\mathbf{q}_{i,t} | \xi_{i,t}) = H(\eta_{i,t} | \xi_{i,t}). \quad (\text{A1})$$

Thus, by (10), we have

$$H(\eta_{i,t} | \xi_{i,t}) \leq \mathbb{E}(\mathcal{E}_{i,t}) < H(\eta_{i,t} | \xi_{i,t}) + 1. \quad (\text{A2})$$

Under the joint distribution of $\theta_{i,t}$ and $\eta_{i,t}$ induced by the model in Figure 2 it can be shown that [(cf. References 3 and Reference 4(lemma 1)]

$$H(\eta_{i,t} | \xi_{i,t}) = I(\theta_{i,t}; \eta_{i,t}). \quad (\text{A3})$$

The results of the previous section show that the joint distribution of $\theta_{i,t}$ and $\eta_{i,t}$ induced by the model in Figure 3 is equivalent to the joint distribution induced by Figure 2. Substituting (A3) into (A2) completes the proof.

APPENDIX B. PROOF OF LEMMA 2

To see the upper bound, first expand the mutual information in terms of differential entropy

$$I(\theta_{i,t}; \eta_{i,t}) = h(\eta_{i,t}) - h(\eta_{i,t} | \theta_{i,t}) \quad (\text{B1})$$

$$= h(\eta_{i,t}) - h(\eta_{i,t} - \theta_{i,t} | \theta_{i,t}) \quad (\text{B2})$$

$$= h(\boldsymbol{\eta}_{i,t}) - h(\mathbf{v}_{i,t}), \quad (\text{B3})$$

where (B2) follows from the definition of $\boldsymbol{\eta}_{i,t}$ and (B3) follows since $\mathbf{v}_{i,t}$ and $\boldsymbol{\theta}_{i,t}$ are independent (cf. below (12)). Likewise

$$I(\boldsymbol{\theta}_{i,t}^G; \boldsymbol{\eta}_{i,t}^G) = h(\boldsymbol{\eta}_{i,t}^G) - h(\mathbf{v}_{i,t}^G). \quad (\text{B4})$$

Gaussian random variables have the maximum entropy among all random variables with the same variance [cf. Reference 29(theorem 8.6.5)]. Thus, we have

$$h(\boldsymbol{\eta}_{i,t}) \leq h(\boldsymbol{\eta}_{i,t}^G). \quad (\text{B5})$$

Furthermore, by applying the definition of differential entropy, expanding the PDF of $\mathbf{v}_{i,t}^G$, and recalling that $\mathbf{v}_{i,t}^G$ and $\mathbf{v}_{i,t}$ have the same variance, we have

$$h(\mathbf{v}_{i,t}) = h(\mathbf{v}_{i,t}^G) - D(\mathbf{v}_{i,t} \| \mathbf{v}_{i,t}^G) \quad (\text{B6})$$

$$= h(\mathbf{v}_{i,t}^G) - \frac{1}{2} \log \frac{2\pi e}{12}. \quad (\text{B7})$$

Thus, recalling both (B4) and (B1), and subtracting both sides of (B6) from the respective sides of (B5) gives

$$\begin{aligned} I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t}) &\leq h(\boldsymbol{\eta}_{i,t}^G) - h(\mathbf{v}_{i,t}^G) + \frac{1}{2} \log \frac{2\pi e}{12} \\ &= I(\boldsymbol{\theta}_{i,t}^G; \boldsymbol{\eta}_{i,t}^G) + \frac{1}{2} \log \frac{2\pi e}{12}. \end{aligned}$$

On the other hand, the lower bound follows from Reference 32(Lemma C.1).

APPENDIX C. PROOF OF PROPOSITION 1

As the first step of the proof, we show by induction that for $t = 1, \dots, T$, we have

$$Q_{t|t}^{**} \geq Q_{t|t}^*, \quad (\text{C1a})$$

$$Q_{t+1|t}^{**} \geq Q_{t+1|t}^*. \quad (\text{C1b})$$

For the initial step, (C1a) trivially holds as $Q_{1|1}^{**} = Q_{1|1}^*$ by construction. Note that $Q_{2|1}^{*-1} = A Q_{1|1}^{*-1} A^\top + F F^\top$ by construction, and $Q_{2|1}^{*-1} \geq A Q_{1|1}^{*-1} A^\top + F F^\top$ since $Q_{1|1}^*$ and $Q_{2|1}^*$ are feasible solution to (36f). Thus $Q_{2|1}^{*-1} \leq Q_{2|1}^{*-1}$ or equivalently $Q_{2|1}^{**} \geq Q_{2|1}^*$. Therefore, relation (C1) holds for $t = 1$. For the induction step, we assume (C1) holds for $t = k(\geq 1)$:

$$Q_{k|k}^{**} \geq Q_{k|k}^*, \quad (\text{C2a})$$

$$Q_{k+1|k}^{**} \geq Q_{k+1|k}^*, \quad (\text{C2b})$$

and we show that (C1) holds for $t = k + 1$. From (36e) and (37e), we have

$$\begin{aligned} Q_{k+1|k+1}^{**} &= Q_{k+1|k}^{**} + \sum_{i=1}^M \delta_{i,k+1}^* C_i^\top C_i \quad \text{and} \\ Q_{k+1|k+1}^* &= Q_{k+1|k}^* + \sum_{i=1}^M \delta_{i,k+1}^* C_i^\top C_i. \end{aligned}$$

Thus, (C2b) directly implies

$$Q_{k+1|k+1}^{**} \geq Q_{k+1|k+1}^*. \quad (\text{C3})$$

Substituting $Q_{k+1|k+1}^{**}$ into (36e), we have

$$Q_{k+2|k+1}^{*-1} = A Q_{k+1|k+1}^{*-1} A^\top + F F^\top \quad (\text{C4a})$$

$$\leq A Q_{k+1|k+1}^{*-1} A^\top + F F^\top \quad (\text{C4b})$$

$$\leq Q_{k+2|k+1}^{*-1}, \quad (\text{C4c})$$

where (C4b) is due to (C3). The last inequality holds since $(Q_{k+1|k+1}^*, Q_{k+2|k+1}^*)$ is a feasible solution satisfying (37e). (C4) implies

$$Q_{k+2|k+1}^{**} \geq Q_{k+2|k+1}^*. \quad (\text{C5})$$

Inequalities (C3) and (C5) establish the relation (C1) for $t = k + 1$ and thus for any $t = 1 : T$. This completes the first step of the proof.

The proof can be completed as follows. Let J_2^* be the optimal value of (37) attained by $[\psi_{i,t}^*, \delta_{i,t}^*, \gamma_{i,t}^*, S_t^*, Q_{t|t}^*, Q_{t|t-1}^*]$. From (C1), we have

$$0 \leq \begin{bmatrix} \delta_{i,t}^* - \gamma_{i,t}^* & \delta_{i,t}^* C_i \\ C_i^\top \delta_{i,t}^* & Q_{t|t-1}^* + C_i^\top \delta_{i,t}^* C_i \end{bmatrix} \leq \begin{bmatrix} \delta_{i,t}^* - \gamma_{i,t}^* & \delta_{i,t}^* C_i \\ C_i^\top \delta_{i,t}^* & Q_{t|t-1}^{**} + C_i^\top \delta_{i,t}^* C_i \end{bmatrix}$$

and

$$0 \leq \begin{bmatrix} S_t^* & I \\ I & Q_{t|t}^* \end{bmatrix} \leq \begin{bmatrix} S_t^* & I \\ I & Q_{t|t}^{**} \end{bmatrix}.$$

Moreover, $(Q_{t|t}^{**}, Q_{t|t-1}^{**})$ satisfies the equality constraints (38) by construction. This means that $[\psi_{i,t}^*, \delta_{i,t}^*, \gamma_{i,t}^*, S_t^*, Q_{t|t}^{**}, Q_{t|t-1}^{**}]$ is a feasible solution of equation (36) and it obtains the value J_2^* . Therefore, $J_1^* \leq J_2^*$. However, (36) has more strict constraint than (37) which immediately implies $J_1^* \geq J_2^*$. Therefore, $J_1^* = J_2^*$.