Uniformly Bounded State Estimation Over Multiple Access Channels

Ghassen Zafzouf ©, Girish N. Nair ©, Fellow, IEEE, and Farhad Farokhi ©, Senior Member, IEEE

Abstract—This article addresses the problem of distributed state estimation via multiple access channels (MACs). We consider a scenario where two encoders are simultaneously communicating their measurements through a noisy channel. First, the zero-error capacity region of the general \( M \)-input, single-output MAC is characterized using tools from nonstochastic information theory. Next, we show that a tight condition to be able to achieve uniformly bounded state estimation errors can be given in terms of the channel zero-error capacity region. This criterion relates the channel properties to the plant dynamics. These results pave the way toward understanding information flows in networked control systems with multiple transmitters.

Index Terms—Distributed state estimation, multiple access channels, networked control systems, nonstochastic information, zero-error capacity region.

I. INTRODUCTION

O

VER the last few decades, the important progress in the fields of electronics, nanotechnology, and computing has led to the development of smaller and cheaper sensors that can perform real-time computation. The availability of such sensors to a wide range of users has contributed, among other factors, to a tremendous increase in the number of connected devices, i.e., large wireless sensor networks (WSNs) [1], [2], [3], [4], and to the emergence of new technological concepts, such as the Internet of Things and machine-to-machine communication [5].

Traditionally, classical control and estimation theory has assumed that communication between different components of a network occurs over point-to-point links. This standard assumption, however, cannot hold for recent emerging applications, where numerous subsystems are interacting with each other. In these applications, it may be impractical or costly to set up and maintain multiple physical point-to-point links between multiple subsystems. One solution is to use a shared medium, such as wireless, with frequency and/or time divided up into noninterfering slots, each for the dedicated use by a single transmitter [6]. In the case of time-slots, this leads to deterministic round-robin-like protocols for channel access [7]. However, such methods are generally suboptimal since they do not fully exploit channel resources in order to increase throughput. For real-time applications, the delays caused by a round-robin scheduling protocol may also degrade performance significantly. Event-based time scheduling is an alternative that allows sensors to transmit more quickly when their measurements exceed certain levels or increase rapidly [8]. This allows communication resources to be used where they are needed most. However, an additional network layer is typically required to make sure network access goes to the device that most needs it.

An alternative is to allow users to transmit their messages simultaneously over the shared channel, and use coding and modulation to mitigate interuser interference in addition to noise and other channel effects. Since resources, such as specific time slots or frequency bands are not preallocated to each user, this method potentially allows faster data flows to be achieved. In real-time applications with a large number of transmitters, it could also improve latency compared to a time-division approach. A simple model of simultaneous communication is the multiple access channel (MAC), consisting of different users who aim to each send an independent message reliably to a common receiver. This was initially introduced by Shannon in his seminal work [9]. In Section I-A, we present an example of a noisy MAC to familiarize the reader with this model.

In contrast to communication systems, where performance is studied in an average sense, many control and estimation systems are deployed for mission-critical purposes, with any error leading to potentially devastating consequences. This motivates system designs based on the worst-case scenario [10], [11], [12], [13]. With this in mind, the aim of this work is to understand the problem of remote estimation of dynamical systems over MACs with bounded noise rather than assuming a statistical distribution. In this context, we try to answer the following fundamental questions: Is it possible to provide a reliable estimation of the states of distinct plants observed by different sensors whose measurements are sent simultaneously over a wireless channel? If so, what is the connection between the intrinsic properties of the dynamical systems and the transmission data rates?

A. Multiple Access Channels and Zero-Error Capacity

As discussed above, allowing senders to transmit simultane-

ously can be advantageous in terms of improving throughput
and reducing latency. A channel model capturing the essence of this problem is the aforementioned MAC. A natural example is the Gaussian two-user MAC

\[ Y = X^1 + X^2 + Z \]  

(1)

where \( X^1 \) and \( X^2 \) are the inputs to the channel, \( Z \) denotes zero-mean Gaussian noise independent of \( X^1 \) and \( X^2 \), and \( Y \) is the channel output. In [14], it was shown that a reasonably good approximation of the Gaussian MAC can be obtained by approximating it as a deterministic two-user XOR MAC model carrying summation in \( \mathbb{Z}_2 \). A further example of MACs is the binary adder channel (BAC), which unlike the XOR model, performs the addition over \( \mathbb{Z} \) and will be discussed in greater detail in Section V-A.

In communications, the (ordinary) channel capacity \( C \) is the maximum rate at which information can be transferred with vanishingly small error probability across the channel [15]. Unlike point-to-channel problems, whose channel capacity \( C \) is a nonnegative scalar, in the context of multiuser channels, including the MAC in particular, we talk about a channel capacity region \( C \) that lies in an \( M \)-dimensional space, where \( M \) is the number of users in the system. For more details regarding this topic, see [16].

In the context of worst-case state estimation, it turns out that the notion of zero-error capacity \( C_0 \) is a more insightful figure of merit than classical channel capacity \( C \)[10],[11]. The zero-error capacity \( C_0 \) of a point-to-point channel is defined as the highest block-coding rate which yields exactly zero decoding errors at the receiver [17]. Intuitively, the block-codes with zero-error property are those leading to an output set that consists of distinct elements, i.e., no codewords result in the same output.

In a similar manner to \( C \), in a multiuser communication setup, the zero-error capacity region \( C_0 \) with \( M \) senders is contained in an \( M \)-dimensional space, rather than a single-axis like \( C_0 \). A formal definition of \( C_0 \) is introduced in Section III.

B. Literature Overview

The decisive role of information theory in digital communications arises largely from the fundamental coding bounds it provides. To approach these bounds, coding schemes with arbitrarily long block lengths are used, which result in vanishingly small probability of decoding errors. However, long block-lengths are ill-suited for real-time control systems, since they lead to long delays that degrade closed-loop performance significantly. Furthermore, in real-time control and estimation applications where there are safety requirements or mission-critical objectives, closed-loop performance is often quantified in a worst-case sense, rather than probabilistically.

An important step toward understanding communication requirements for worst-case state estimation was taken in [18], where it was shown that to achieve almost surely (a.s.) uniform state estimation of a linear, disturbed dynamical system over a stochastic discrete memoryless point-to-point channel, it is necessary and the open-loop topological entropy\(^1\) \( h \) does not exceed the channel zero-error capacity \( C_0 \). If \( C_0 > h \) strictly, then there exists a coding and estimation scheme that achieves a.s. uniform state estimation. By noting that \( C_0 \) depends on the combinatorial structure of the channel rather than its probabilistic nature, this result was rederived in [10] for surely bounded estimation, by introducing the framework of uncertain variables (uv’s) and nonstochastic information I.,. In a recent article [19], the authors introduced the notion of uncertain wiretap channel and studied the problem of secure state estimation of a noisy unstable dynamical system in a nonstochastic setup. Tools from this nonprobabilistic framework were also used to study the problem of bounded state estimation over point-to-point channels with finite memory [20],[21].

The main focus in previous works was on studying either state estimation or control problem in the context of point-to-point channels [22],[23],[24],[25]. Stochastic stabilization of networked control systems over multiple wireless channels has been extensively studied in, e.g., [26] and [27]. Furthermore, the problem of filtering for stochastic systems consisting of distributed sensors exchanging information over point-to-point links was investigated for both discrete, e.g., in [28] and [29], and continuous settings, e.g., in [30]. To the best of our knowledge, the problem of state estimation over an MAC with bounded noise has not been considered yet. An exception is the work of Zaidi et al. [31], where classical stochastic tools are used to present sufficient conditions ensuring the mean-square stabilization of two scalar linear time-invariant (LTI) systems over a noisy two-input, single-output MAC. The paper [32] also obtains necessary and sufficient conditions for stabilizing two scalar plants across a shared Gaussian MAC. In that paper, the authors distinguish between the case where encoders are entirely independent from each other, and where information sharing among them is allowed.

In contrast, we address the problem of worst-case state estimation over MAC with bounded noise without assuming a specific statistical distribution. Hence, with the aim of understanding distributed systems with multiple sensors sharing the same channel, the characterization of the zero-error capacity region \( C_0 \) of an \( M \)-input, single-output MAC with bounded noise becomes necessary. To date, no formula for such a region has been obtained and \( C_0 \) of MACs in general remains an open problem, in contrast to its ordinary capacity region studied in, e.g., [33] and [34]. The work in [35] characterized the zero-error capacity region for the special case of an MAC with two correlated transmitters. Another class of MACs has been studied in [36], namely, MAC with pairwise shared messages as well as a common message among all users. To derive an expression of the zero-error capacity region, we use nonstochastic information theory [10],[11],[12]. However, unlike the point-to-point zero-error capacity in [11], our definition requires the concept of conditional nonstochastic information.

C. Overview of Main Result

Our main contributions are hence the following: 1) We first generalize the characterization of the zero-error capacity region obtained in [35] for a two-user MAC to the \( M \)-user case with one common message using the framework of nonstochastic information (Theorem 2). To this end, we present a converse proof and establish the achievability argument by constructing a suitable coding scheme. 2) Next, the problem of state estimation over

\(^1\)The notion of topological entropy of a system is defined as the sum-log of its unstable pole magnitudes.
MACs with bounded noise is studied and a theorem (Theorem 3) linking between the intrinsic properties of the dynamical systems and the zero-error capacity of the communication channel is shown. A numerical example involving the BAC is then used to illustrate this theoretical result.

The first step toward understanding information flow in large networked systems with multiple sensors and estimators is to start with analyzing a simple network topology. Consider the networked systems with multiple sensors and estimators is to illustrate this theoretical result.

The operator int(·) denotes the interior of a given region.

state estimation errors can be constructed. This condition is considered tight as sufficiency and necessity differ only on the boundary, which is a set of Lebesgue measure zero.

Unlike our previous paper [35], where the two-user MAC model was studied, in this work we derive a channel coding theorem characterizing the zero-error capacity region of an $M$-user MAC with one common message and $M \geq 2$. In addition, we present here the detailed necessity proof of Theorem 3, which appeared as a brief communication in [37], using the notions of nonstochastic information as well as nonstochastic conditional information, and furthermore prove the sufficient condition, by constructing a suitable scheme. Finally, the setup (2a)–(2b) illustrated in Fig. 1 is shown here to be general for noiseless linear systems.

D. Structure of the Article

The rest of this article is organized as follows. In Section II, some basic definitions related to the nonprobabilistic framework are introduced and the MAC model along with the zero-error coding scheme are presented. Next, we extend an earlier result on the zero-error capacity region of a two-input, single-output MAC [35] to the more general case of $M$ inputs for any given block-length $n$ in Section III. Along with the results, both converse and achievability proofs are provided in this section. Then, the problem of uniform state estimation over two-user MAC is thoroughly investigated in Section IV. Furthermore, an example of state estimation over the BAC is studied in Section V. Finally, Section VI concludes this article.

E. Notation

In this article, we use upper case letters, e.g., $X$, to denote uv’s, and lower case letters, e.g., $x$, for their realizations. Calligraphic letters, such as $\mathcal{X}$, stand for sets, and the cardinality of set $\mathcal{X}$ is represented by $|\mathcal{X}|$. Note that a pair of vertical bars, i.e., $|a|$, is also used to denote the absolute value of a scalar $a$. We represent the sequence $\{x_i\}_{i=m}^n$ by $x_{m:n}$. Furthermore, the max-norm of a vector $v$ is denoted by $\|v\|$. The expression $[m : n]$ represents the sequence of integers $\{m, m+1, \ldots, n-1, n\}$. Finally, the symbol $\lceil \cdot \rceil$ stands for the ceiling function, and the logarithms throughout this article are in base 2.
A. Uncertain Variables, Unrelatedness, and Markovianity

Consider the sample space $\Omega$, as shown in Fig. 2. A unv $X$ consists of a mapping from $\Omega$ to a set $\mathcal{X}$. Hence, each sample $\omega \in \Omega$ induces a particular realization $X(\omega) \in \mathcal{X}$. Given a pair of unv’s $X$ and $Y$, the marginal, joint, and conditional ranges are denoted as

$$[X] := \{X(\omega) : \omega \in \Omega\} \subseteq \mathcal{X}$$

$$[X, Y] := \{(X(\omega), Y(\omega)) : \omega \in \Omega\} \subseteq \mathcal{X} \times \mathcal{Y}$$

$$[X|y] := \{X(\omega) : Y(\omega) = y, \omega \in \Omega\} \subseteq \mathcal{X}$$

The dependence on $\Omega$ will normally be hidden, with most properties of interest expressed in terms of operations on these ranges. As a convention, unv’s are denoted by upper case letters, while their realizations are indicated in lower case. The family $\{[X|y] : y \in \mathcal{Y}\}$ of conditional ranges is denoted $[X|Y]$.

**Definition 1 (Unrelatedness [10]):** The unv’s $X_1, X_2, \ldots, X_n$ are said to be (mutually) unrelated if

$$[X_1, X_2, \ldots, X_n] = \{[X_1] \times [X_2] \times \ldots \times [X_n]\}. \quad (8)$$

We denote two unrelated unv’s $X_1$ and $X_2$ by $X_1 \perp X_2$.

**Remark 1:** Unrelatedness is closely related to the notion of qualitative independence [38] between discrete sets. It can be proven that unrelatedness is equivalent to the following conditional range property:

$$[X_k|x_{1:k-1}] = [X_k] \forall x_{1:k-1} \in [X_{1:k-1}], k \in [2:n]. \quad (9)$$

**Definition 2 (Conditional Unrelatedness [10]):** The unv’s $X_1, \ldots, X_n$ are said to be conditionally unrelated given $Y$ if

$$[X_1, \ldots, X_n|y] = [X_1|y] \times \ldots \times [X_n|y] \forall y \in \mathcal{Y}. \quad (10)$$

**Definition 3 (Markovianity [10]):** The unv’s $X, Y$, and $X_2$ form a Markov uncertainty chain denoted as $X_1 \leftrightarrow Y \leftrightarrow X_2$, if

$$[X_1|y, x_2] = [X_1|y] \forall (y, x_2) \in [Y, X_2]. \quad (11)$$

The Markov chain $X_1 \leftrightarrow Y \leftrightarrow X_2$ is equivalently denoted by $X_1 \perp X_2|Y$, i.e., $X_1$ and $X_2$ are unrelated given $Y$.

**Remark 2:** It can be shown that Definition 3 is equivalent to $X_1$ and $X_2$ being conditionally unrelated given $Y$, i.e.,

$$[X_1, X_2|y] = [X_1|y] \times [X_2|y] \forall y \in \mathcal{Y}. \quad (12)$$

By the symmetry of (12), we conclude that $X_1 \leftrightarrow Y \leftrightarrow X_2$ if $X_2 \leftrightarrow Y \leftrightarrow X_1$.

B. Preliminaries on Nonstochastic Information

Before presenting the notion of nonstochastic information, we first present some important background concepts. Throughout this subsection $X, Y, Z, Z'$, and $W$ denote unv’s.

**Definition 4 (Overlap Connectedness [10]):** Two points $x \in [X]$ and $x' \in [X]$ are said to be $[X|Y]$-overlap connected, denoted $x \leftrightarrow x'$, if there exists a finite sequence $\{X_{y_i}\}_{i=1}^m$ of conditional ranges such that $x \in [X_{y_1}]$, $x' \in [X_{y_m}]$ and $[X_{y_i}] \cap [X_{y_{i+1}}] \neq \emptyset$, for each $i \in \{2, \ldots, m\}$.

Obviously, the overlap connectedness is both transitive and symmetric, i.e., it is an equivalence relation. Thus, it results in equivalence classes that cover $[X]$ and form a unique partition.

We call this family of sets the $[X|Y]$-overlap partition, denoted by $[X|Y]$.

**Definition 5 (Nonstochastic Information [10]):** The nonstochastic information between $X$ and $Y$ is given by

$$I_s[X; Y] = \log_2 |[X|Y]|. \quad (13)$$

**Remark 3:** Note that the nonstochastic information is symmetric, i.e., $I_s[X; Y] = I_s[Y; X]$.

**Remark 4:** Similar to classical information theory, there is an analogous data processing inequality in nonstochastic information theory. For any Markov uncertainty chain $W \leftrightarrow X \leftrightarrow Y$, it holds that $I_s[W; Y] \leq I_s[X; Y]$ [10]. In other words, inner unv pairs in the chain share more information than outer ones.

**Example** Consider unv’s $X$ and $Y$ with conditional range family $[X|Y] = \{[X|y_1], [X|y_2]\}$. Fig. 3(a) illustrates an example of overlap disconnected points. Observe that $[X|y_1] \cap [X|y_2] = \emptyset$, and hence, the unique overlap partition $[X|Y]$ consists of two singleton sets. Thus, the nonstochastic information in this case is $I_s[X; Y] = \log_2 2 = 1$ bit. On the other hand, Fig. 3(b) shows the case where the points are connected in overlap sense. In this case, it is easy to see that $[X|y_1] \cap [X|y_2] \neq \emptyset$ and $[X|Y]$ does not consist of two singleton sets. Hence, $I_s[X; Y] = \log_2 1 = 0$ bit.

**Definition 6 (Common Variables [39], [40]):** A unv $Z$ is said to be a common variable (cv) for $X$ and $Y$ if there exist functions $f$ and $g$ such that $Z = f(X) = g(Y)$.

**Remark 5:** Note that no cv can take more distinct values than the maximal one. The concept of a maximal common variable was first presented by Shannon in the framework of random variables [39], to which he referred by the term “common information element” for a maximal cv.

The nonstochastic information $I_s[X; Y]$ is precisely the log-cardinality of the range of a maximal cv between $X$ and $Y$. This is because it can be shown that $\forall (x, y) \in [X, Y]$, the partition set in $[X|Y]$, that contains $x$ also uniquely specifies the set in $[Y|X]$, that contains $y$. Thus, these overlap partitions define a cv for $X$ and $Y$, with corresponding functions $f$ and $g$ given by the labeling. Furthermore, this cv can be proved to be maximal. See [11] for further details.

**Definition 7 (Conditional $I_s$ [11]):** The conditional nonstochastic information between $X$ and $Y$ given $W$ is

$$I_s[X; Y|W] := \min_{w \in [W]} \log_2 |[X|Y, w]|. \quad (14)$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
$R_j = \{ \text{achievable } Z \text{ that generates estimate } w \}$

$n \geq 1$, the messages are encoded into channel input sequences $X_{1,n}, X_{2,n}, \ldots, X_{M,n}$ as

$$X_{i,n} = \varnothing(W^0, W^j), \ j \in [1 : M]$$

where $\{ \varnothing \}_{j=1}^M$ are the coding laws at each transmitter. Note that the common message $W^0$ is seen by all encoders while the private messages $W^j$ are only available to their respective transmitters. The code rate for each message is defined as

$$R_i := (\log_2 w^i_{\text{max}})/n, \ i \in [0 : M].$$

This general system configuration, where a common message is seen by all encoders, allows us to incorporate a form of relatedness among the channel input sequences in the model. In the case where the common message can take only one value, so that $R^0 = 0$, each channel input is generated in isolation and is mutually unrelated with the others. At the other extreme, if the private messages can each take only one value so that $R^1 = R^2 = \ldots = R^M = 0$, then the channel inputs are generated in complete cooperation.

The encoded data sequences are then sent through a stationary memoryless MAC, as depicted in Fig. 5. The output $Y_k \in \mathcal{Y}$ of the MAC is given in terms of a fixed function $f : \mathcal{Y} \times \ldots \times \mathcal{Y} \times \mathcal{X} \rightarrow \mathcal{Y}$

$$Y_k = f(X^1_k, X^2_k, \ldots, X^M_k, Z_k) \in \mathcal{Y}, \ k \in \mathbb{Z}_{\geq 0}$$

where the channel noise is denoted by $Z_k$ and is mutually unrelated with all messages, and past channel noise, i.e., $Z_{k-1}, W^0, W^1, W^2, \ldots, W^M$. We further assume that the range $[Z_k] = \mathcal{Z}$ is constant.

The receiver consists of a decoder $\varnothing$ that generates estimates $\hat{W}^0, \hat{W}^1, \ldots, \hat{W}^M$ of the transmitted messages using the channel output sequence $Y_{1:n}$. In the context of zero-error communication, these estimates must always be exactly equal to the original messages, even with the existence of channel noise or interuser interference. This requirement means that for any $i \in [0 : M]$ the conditional range $[W^i|Y_{1:n}]$ consists of one element for any channel output sequence $Y_{1:n} \in [Y_{1:n}]$.

This communication system is an extension of the nonstochastic MAC operating with two or more users.

For a given code block-length $n$, the operational zero-error $n$-capacity region $\mathcal{C}_{0,n}$ of the MAC is defined as the set of rate tuples $R = (R^0, R^1, \ldots, R^M)$ for which zero-error communication is possible by suitable choice of coding and decoding functions. Note that this is well defined for finite block-lengths $n$, and so it is of interest in safety-critical low-latency applications. This is unlike the Shannon capacity region $\mathcal{C}$, which requires $n \rightarrow \infty$ so as to yield vanishingly small decoding error probabilities.

If we are allowed to use arbitrarily long blocks, i.e., $n \rightarrow \infty$, then the relevant zero-error capacity region $\mathcal{C}_0$ is given by the closed union

$$\mathcal{C}_0 = \bigcup_{n \geq 1} \mathcal{C}_{0,n}$$

and we define the notion of achievable rate as follows.

**Definition 8:** A rate tuple $R = (R^0, R^1, \ldots, R^M)$ is called achievable if there exists a sequence of $\left( \left[ \frac{2^n R^0}{n} \right], \left[ \frac{2^n R^1}{n} \right], \ldots, \left[ \frac{2^n R^M}{n} \right], n \right)$ zero-error codes, with
$n \in \mathbb{Z}_{\geq 1}$ and $R_n = (R_0^n, R_1^n, \ldots, R_M^n) \in \mathcal{C}_{0,n}$, converging to $R$.

We have then the following result.

**Theorem 1**: The zero-error capacity region $\mathcal{C}_0$ (18) is convex.

**Proof**: See Appendix A.

**B. Zero-Error Capacity of $M$-User MAC Via Nonstochastic Information**

At this stage, we use nonstochastic information to establish a multiletter characterization of the zero-error $n$-capacity region $\mathcal{C}_{0,n}$ for the $M$-user MAC.

**Theorem 2**: For a given block-length $n \geq 1$, let $\mathcal{R}(U, X_1^n, X_2^n, \ldots, X_M^n)$ be the set of rate tuples $(R_0, R_1, R_2, \ldots, R_M)$ such that

\[
R_0 \leq I(U; Y_1^n) \quad (19)
\]

\[
R_3 \leq I(X_1^n; Y_1^n | U) \quad \forall j \in [1 : M] \quad (20)
\]

where $2^{nR_0}, \in [0 : M]$, are positive integers, $X_1^n, j \in [1 : M]$, are sequences of inputs to the $M$-user MAC (17), $Y_1^n$ is the corresponding channel output sequence, and $U$ is an auxiliary uv.

Then, the operational zero-error $n$-capacity region $\mathcal{C}_{0,n}$ of the $M$-user MAC over $n$ channel uses coincides with the union of the regions $\mathcal{R}(U, X_1^n, X_2^n, \ldots, X_M^n)$ over all uv’s $U, X_1^n, X_2^n, \ldots, X_M^n$, such that

i) the sequences $X_1^n, j = 1, \ldots, M$, are conditionally unrelated given $U$;

ii) $U \rightarrow (X_1^n, X_2^n, \ldots, X_M^n) \rightarrow Y_1^n$ form a Markov uncertainty chain.

**Proof**: See Section III-C.

**Remark 7**: Under these circumstances, $\mathcal{R}(U, X_1^n, X_2^n, \ldots, X_M^n)$ becomes the set of the rate tuples $(R_1, R_2, \ldots, R_M)$ such that

\[
R_0 \leq I_i(X_1^n; Y_1^n | U), j \in [1 : M] \quad (21)
\]

with $2^{nR_0}$ being positive integers, $X_1^n$ being the input signals fed into the $M$-user MAC (17) and $Y_1^n$ corresponding to the channel output. The operational zero-error $n$-capacity region $\mathcal{C}_{0,n}$ of the $M$-user MAC over $n$ channel uses in this case coincides with the union of the regions $\mathcal{R}(X_1^n, X_2^n, \ldots, X_M^n)$ over all uv’s $X_1^n, X_2^n, \ldots, X_M^n$, such that the sequences $X_j^n, j \in [1 : M]$, are mutually unrelated.

Note that (21) is almost identical to the multiletter characterization of the ordinary capacity region of a 2-sender MAC [16, Th 4.1], apart from the use of $I_i$ instead of mutual information.

**Remark 8**: In the extreme case of setting $R_1 = \ldots = R_M = 0$ bits, whereas the rate $R_0$ is strictly positive, all users $\phi_1, \ldots, \phi_M$ transmit the same message, namely, the common message $W^0$.

Under these circumstances, (20) is trivially satisfied and by applying the data processing inequality [10] on the Markov uncertainty chain $U \rightarrow (X_1^n, X_2^n, \ldots, X_M^n) \rightarrow Y_1^n$, it can be seen that the maximum rate $R_0^*$ is achieved when $U = (X_1^n, X_2^n, \ldots, X_M^n)$.

**Remark 9**: In contrast to [35], where the two-transmitter MAC system with a common message was considered, this work provides a coding theorem for the general case of $M$-user MAC.

This result is a nonstochastic analogue of Shannon information-theoretic characterization of the ordinary capacity region [47]. Apart from its multiletter nature and the use of nonstochastic rather than Shannon information, there are several other notable differences. First, the inequalities are on each individual rate, and not the sum rate. Second, the conditioning argument in (20) involves only the auxiliary variable $U$, and not any of the channel input sequences. Finally, this characterization is valid for finite block-lengths $n$, not just as $n \rightarrow \infty$.

These differences originate from the zero-error requirement on our system, as well as the definition of conditional $I_i$. Note also that the lack of a sum-rate bound does not imply that senders can transmit with rates orthogonal to the others. This is because the zero-error $n$-capacity region is a union of the regions $\mathcal{R}(U, X_1^n, X_2^n, \ldots, X_M^n)$ over the set of all uv’s satisfying the conditions in Theorem 2. Though each of these regions is a hypercube aligned with the rate axes, their union may be more complicated than a hypercube.

**C. Proof of Theorem 2**

1) **Converse**: Consider the $M$-user MAC model defined in (17) and let $(R_0^i)_{i=0}^M (16)$ be the rates of some zero-error code (15) with block-length $n$. Furthermore, we set the uv $U = W^0$. By assumption, the messages $W^i, i \in [0 : M]$ are mutually unrelated and hence from (15) we conclude that the codewords satisfy

\[
\prod_{j=1}^{M} (X_1^n; Y_1^n | U) = [X_1^n, \ldots, X_M^n | U]. \quad (22)
\]

In addition, the unrelatedness of the channel noise $Z$ with the messages $W^i, i \in [0 : M]$, implies that $Z$ is also unrelated with the codewords $X_1^n, X_2^n, \ldots, X_M^n$. Thus, the Markov chain $Y_1^n \rightarrow (X_1^n, X_2^n, \ldots, X_M^n) \rightarrow U$ is satisfied.

As zero-error communication is assumed, the existence of a decoding function $\mathcal{D}$ such that

\[
W^0 = \mathcal{D}(Y_1^n) \quad (23)
\]

is then guaranteed. Moreover, since $U = W^0$ it can be directly deduced that $W^0$ is a common variable (cv) (Definition 6) between $U$ and $Y_1^n$. The maximal cv property of $I_i$, yields the following:

\[
nR_0 \equiv \log_2 |[W^0]| \leq I_i(U; Y_1^n). \quad (24)
\]

This proves expression (19) of Theorem 2. Next we show inequality (20) for $j \in [1 : M]$. First, note that given a specific realization $W^0 = w^0$ of the common message, there exists a unique message $w^j$ associated with the channel codeword $x_1^n$. This observation follows also from the zero-error property of the chosen code. In fact, if different realizations $W^j = w^j$ were mapped to the same codeword, then zero-error decoding would obviously be impossible and the assumption would have been violated. Thus, there certainly exists a function $g^j$ such that

\[
W^j = g^j(X_1^n, W^0), \quad j \in [1 : M]. \quad (25)
\]

Moreover, by the zero-error property there is indeed a decoding function $\mathcal{D}$ such that

\[
W^j = \mathcal{D}(Y_1^n) \quad \forall j \in [1 : M]. \quad (26)
\]
Hence, we conclude that the uv W \textsuperscript{j} is a cv between (X_{1:n} \textsuperscript{j}, W^0) and (Y_{1:n}, W^0). Recall that in the considered MAC model the private messages W are unrelated with U = W^0 for all j \in [1 : M]. Therefore, the interpretation of conditional I_n in terms of maximal cv’s results in

\[ nR^j \equiv \log_2 |[W^j]| \leq I_n[X_{1:n}^j; Y_{1:n}|W^0] \]

proving (20).

2) Achievability: The achievability proof is established by showing that if we have a set of uv’s U, X_{1:n}^j, \forall j \in [1 : M] for some n \geq 1 such that the outlined requirements in Theorem 2 are fulfilled, then it is possible to construct a zero-error coding scheme at rates achieving equality in (19) and (20).

a) Codebook generation: We first fix the rate R^0 such that R^0 \leq (I(U|Y_{1:n})/n. Next, select one point from each set of the family \{U|Y_{1:n}\}. We then denote the chosen points u(w^0) with w^0 = \{1,2,\ldots,2^{nR^j}\}.

Since \( nR^j = I_n[X_{1:n}^j; Y_{1:n}|U] \) for j \in [1 : M], (14) means that the following inequality holds:

\[ 2^{nR^j} \leq |X_{1:n}^j| \leq u(w^0), \forall j \in [1 : M] \] (28)

with w^0 \in [1 : 2^{nR^j}]. It is therefore possible to select 2^{nR^j} distinct codewords x_{1:n}^j from \{X_{1:n}^j|U = u(w^0)\} for any realization w^0 such that each nonempty set of the overlap partition \{X_{1:n}^j|Y_{1:n}, U = u(w^0)\}, contains exactly one element. Subsequently, these codewords denoted as \( \delta^j(w^0, j) \) for j \in [1 : 2^{nR^j}] correspond to the coding laws (15), where j \in [1 : M].

b) Zero-error decoding: At this stage of the proof, we show that it is possible to achieve zero decoding errors using the presented scheme.

First, recall that the uv’s U and \{X_{1:n}^j\}_{j=1}^M satisfy

\[ \prod_{j=1}^M |X_{1:n}^j| = |X_{1:n}, \ldots, X_M^j|U| \] (29)

Then, the n-tuples of the decoded codewords \( \delta^j(w^0, 1), \ldots, \delta^j(w^0, M) \) with w^0 \in [1 : 2^{nR^j}] and \forall j \in [1 : M] belong to the conditional joint range \{X_{1:n}, \ldots, X_M^j|U = u(w^0)\}. This means that any combination of w^0, w^1, \ldots, w^M is mapped to a valid point lying within \{X_{1:n}, \ldots, X_M^j|U\}. At the receiver, the decoding procedure consists of M + 1 stages.

1) First, the decoder determines the transmitted common message w^0. By construction, each of the 2^{nR^j} points u(w^0) is inside a separate set of the family \{U|Y_{1:n}\}. Furthermore, recall that the cv property of the overlap partition implies that each set in the family \{U|Y_{1:n}\}, containing u also uniquely specifies the matching set in \{Y_{1:n}|U\}, that contains y_{1:n}. Hence, the common message w^0 is decoded with zero error.

2) After having found w^0, it is now possible to determine which set of the conditional overlap partition \{Y_{1:n}, X_{1:n}^j, U = u(w^0)\}, contains the sequence y_{1:n}. In a similar way as step (1), this set uniquely determines the corresponding set of the family \{X_{1:n}^j|Y_{1:n}, U = u(w^0)\}, where the codeword \( \delta^j(m^0, m^1) \) lies. Since at most one codeword has been selected from each set of this family for each realization w^0, then the private message of user 1, namely, w^1, is uniquely decoded.

3) In the subsequent M − 1 stages, the decoder repeats step 2) with x_{1:n}^j for j \in [2 : M] and similarly recovers w^j with zero error.

Thus, the achievability of Theorem 2 is established.

Remark 10: Note that it suffices to take the cardinality of the auxiliary uv \{U\} \leq \min\{\{Y_{1:n}\}, \prod_{j=1}^M |X_{1:n}^j|\}. To see this, note from the proof above that U can be as taken as the common message W^0 in a zero-error code for the MAC, which is unambiguously determined by Y_{1:n}. Hence the constraint \{U\} \leq \{Y_{1:n}\}, may be imposed. Furthermore, W^0 is also uniquely determined by the tuple (X_{1:n}^1, \ldots, X_{1:n}^M) of channel input sequences; otherwise some valid combination of channel input sequences would be associated with multiple W^0, violating the zero-error property. Hence, we also need only consider U with \{U\} \leq \prod_{j=1}^M |X_{1:n}^j|. Putting these two constraints together yields the given bound.

Remark 11: Note that the achievability proof above is not intended to yield a practical way to find zero-error codes, but rather just prove their existence under the conditions of Theorem 2. We leave it as future work to explore how nonstochastic information ideas could be exploited to construct zero-error codes in practice.

IV. DISTRIBUTED STATE ESTIMATION OVER NONSTOCHASTIC MAC

A. Problem Formulation

We now consider three discrete LTI dynamical systems characterized by the following system equations for i \in \{0,1,2\}:

\[ X_i(k + 1) = A^iX_i(k) + V_i(k) \in \mathbb{R}^{d_i} \] (30a)

\[ Y_i(k) = C^iX_i(k) + W_i(k) \in \mathbb{R}^{h_i} \] (30b)

where the uv’s V_i(k) and W_i(k) denote process and measurement noise at time instant k \in \mathbb{Z}_{\geq 0}. Before being transmitted, the plant output sequences are first encoded into channel input signals S^i(k), S^2(k) via the coding functions \gamma^i as

\[ S^i(k) = \gamma^i(k, Y^0(0 : k), Y^1(0 : k)), l \in \{1,2\}. \] (31)

Note that the outputs of System 0 are available to both encoders, whereas Systems 1 and 2 are observed only by their respective users. The encoded data sequences are then sent through a stationary memoryless two-user MAC (17), as shown in Fig. 1. The received symbol Q(k) is the output of a fixed function f : \mathcal{A} \times \mathcal{B} \times \mathcal{Z} \rightarrow \mathcal{D}

\[ Q(k) = f(S^1(k), S^2(k), Z(k)) \in \mathcal{D}, k \in \mathbb{Z}_{\geq 0} \] (32)

where Z(k) is the channel noise at time k. At the receiver side, the symbols are used to generate an estimation \hat{X}(k) := (\hat{X}^0(k), \hat{X}^1(k), \hat{X}^2(k))^T of the original plant states X(k) :=
where $d = \sum_{i=0}^{2} d_i$ and, for $k = 0$, the initial estimate $\hat{X}(0) = 0$. The prediction error is denoted by the uv

$$E(k) := \begin{pmatrix} E^0(k) \\ E^1(k) \\ E^2(k) \end{pmatrix} = \begin{pmatrix} X^0(k) - \hat{X}^0(k) \\ X^1(k) - \hat{X}^1(k) \\ X^2(k) - \hat{X}^2(k) \end{pmatrix} \in \mathbb{R}^d. \quad (34)$$

Consider now the following definition.

Definition 9 (Uniformly Bounded Errors): For any noise ranges $[V^1(k)]$ and $[W^1(k)]$ with $\sup_{k \geq 0} ||V^1(k)|| < \infty$ and $\sup_{k \geq 0} ||W^1(k)|| < \infty$, $\exists \gamma > 0$ such that for any initial condition with range $[X(0)]$, that lies within the closed ball $B_t \subseteq \mathbb{R}^d$ centered at the origin and with radius $l$, it holds

$$\sup_{k \geq 0} ||E^i(k)|| = \sup_{k \geq 0} ||X^i(k) - \hat{X}^i(k)|| < \infty. \quad (35)$$

The aim is to design the coder-estimator tuple $(\gamma^1, \gamma^2, \delta)$ such that the resulting estimation error is uniformly bounded in the sense of Definition 9 with respect to the infinity norm. We impose the following assumptions $\forall i, j$ and $h \in \{0, 1, 2\}$ and $\forall k, t \in \mathbb{Z}_{\geq 0}$.

A1: Each matrix pair $(C^i, A^i)$ is observable.

A2: The noise terms $V^i(k)$, $W^i(k)$ are uniformly bounded, i.e., $\sup_{k \geq 0} ||V^i(k)||, \sup_{k \geq 0} ||W^i(k)|| < \infty$, where $V^i(k) := \begin{pmatrix} V^0(k) \\ V^1(k) \\ V^2(k) \end{pmatrix}$ and $W^i(k) := \begin{pmatrix} W^0(k) \\ W^1(k) \\ W^2(k) \end{pmatrix}$.

A3: The initial states $X^i(0)$ and the noise terms $V^i(k)$, $W^i(t)$ are all mutually unrelated, $\forall i, j, h \in \{0, 1, 2\}$ and $\forall k, t \geq 0$.

A4: The channel noise signal $Z(k)$ is unrelated with the combined initial state and noise signals $(X(0), V(0 : k - 1), W(0 : k))$.

A5: Each system matrix $A^i$ has only strictly unstable eigenvalues, i.e., $|\lambda^i_j| > 1$, $\ell = 1, \ldots, d_i$.

A6: The zero signal is a valid realization of measurement and process noise, i.e., $0 \in [V^i], [W^i]$.

Remark 12: Assumption (A5) is imposed mainly for the sake of conciseness. If $A^i$ were allowed to also have strictly stable eigenvalues, then it is straightforward to show that the state components associated with these eigenvalues are uniformly bounded over time. Hence the trivial zero estimator for these state components would yield estimation errors that are uniformly bounded. Thus, without loss of generality, strictly stable eigenvalues and their associated state components may be omitted from each $A^i$.

Eigenvalues with magnitude exactly equal to 1 need more careful treatment. This is because they lead to polynomial rather than exponential growth with time. As such, more delicate techniques than those used in this article are required. For reasons of space, we therefore also exclude such eigenvalues.

Remark 13: Notice that although the formulation with three decoupled subsystems may seem special, it will be shown later in Section IV-D that this setup is reasonably general for the case of noiseless linear systems.

B. Main Result

We present the main result of this section. First, for each subsystem matrix $A^i \in \mathbb{R}^{d_i \times d_i}$, let the topological entropy be

$$h_i := \sum_{1 \leq \ell \leq d_i, |\lambda^i_\ell| > 1} \log_2 |\lambda^i_\ell|. \quad (36)$$

Theorem 3: Consider the linear time-invariant systems in (30a)–(30b) whose outputs are coded (31) and estimated (33) via the two-input single output MAC (32). Suppose Assumptions (A1)–(A6) hold. If there exists a coder-estimator tuple $(\gamma^1, \gamma^2, \delta)$ yielding uniformly bounded estimation errors (Definition 9), then

$$h := (h_0, h_1, h_2)^T \in \mathcal{C}_0 \quad (37)$$

where $h$ is the vector of topological entropies of the corresponding systems and $\mathcal{C}_0$ refers to the zero-error capacity region (18) of the channel.

Furthermore, if $h \in \text{int} (\mathcal{C}_0)$, then a coder-estimator tuple that achieves uniformly bounded state estimation errors can be constructed.

Proof: See Section IV-C.

Remark 14: This result is an extension of [18] and [10], from centralized LTI systems with point-to-point channels, to distributed LTI systems estimated over an MAC.

Remark 15: The case where $h$ lies exactly on the boundary of the zero-error capacity region, i.e., $h \in \mathcal{C}_0 \setminus \text{int} (\mathcal{C}_0)$ introduces some technical issues and is not addressed here.

C. Proof of Theorem 3

1) Necessity Proof: Without loss of generality assume that for $i \in \{0, 1, 2\}$ the state matrix $A^i$ of Plant $i$ is in real Jordan canonical form, i.e., it consists of $m$ square blocks on its diagonal such that the $j$th block is denoted by $A^i_j \in \mathbb{R}^{d_i \times d_i}$ with $j \in [1 : m]$

$$A^i = \begin{pmatrix} A^i_1 & 0 & \ldots & 0 \\ 0 & A^i_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A^i_m \end{pmatrix} \in \mathbb{R}^{d_i \times d_i}. \quad (38)$$

In the following analysis, we consider only the unstable eigenvalues $\{\lambda^i_\ell\}_{\ell=1}^{d_i}$. In Appendix B, we establish the following inequality for $k \to \infty$ and $i \in \{0, 1, 2\}$:

$$I_k \{X^i(0 : k - 1); Q(0 : k - 1)\} \geq I_k \{X^i(0); Q(0 : k - 1)\} \geq \sum_{\ell=0}^{d_i} \log |\lambda^i_\ell|. \quad (39)$$
Before proceeding with the proof, we present the following lemma.

Lemma 1: Let $\Lambda$, $\Omega$ and $\Theta$ denote three uv’s such that $\Lambda$ and $\Theta$ are mutually unrelated. Then, the following relationship between $I_i[\Lambda;\Omega;\Theta]$ (13) and $I_i[\Lambda;\Omega]$ (14) holds:

$$I_i[\Lambda;\Omega;\Theta] \geq I_i[\Lambda;\Omega].$$

Proof: See Appendix C.

Since $X^0(0 : k − 1)$ and $X^l(0)$ are unrelated for $l \in \{1, 2\}$, it follows from Lemma 1:

$$I_i[X^l(0) ; Q(0 : k − 1)|X^0(0 : k − 1)] :I_i[X^0(0) ; Q(0 : k − 1)].$$

Note that $S^1(0 : k − 1) \leftrightarrow X^0(0 : k − 1) \leftrightarrow S^2(0 : k − 1)$, i.e., $S^1(0 : k − 1) \perp S^2(0 : k − 1)|X^0(0 : k − 1)$. Furthermore, the initial states $\{X^i(0)\}_{i=0}^1$ and additive noises $\{W^i(k), V^i(k)\}_{i=0}^1$ are mutually unrelated. Hence, the requirement (A3) results in the Markov chain $X^i(0) \leftrightarrow S^i(0 : k − 1) \leftrightarrow S^i(0 : k − 1)|X^0(0 : k − 1)$, for $l \in \{1, 2\}$. Thus, the conditional data processing inequality [11] yields

$$I_i[X^l(0) ; Q(0 : k − 1)|X^0(0 : k − 1)] \leq I_i[S^l(0 : k − 1) ; Q(0 : k − 1)|X^0(0 : k − 1)].$$

By combining this lower bound on $I_i[S^l(0 : k − 1) ; Q(0 : k − 1)|X^0(0 : k − 1)]$ with inequalities (39) and (41), we obtain

$$I_i[S^l(0 : k − 1) ; Q(0 : k − 1)|X^0(0 : k − 1)] > \sum_{\ell=0}^{d_t} \log |\Lambda^\ell_t|$$

for $k \to \infty$. From (39) and (43), we conclude that

$$I_i[X^0(0 : k − 1) ; Q(0 : k − 1)] \geq \sum_{\ell=0}^{d_t} \log |\Lambda^\ell_t|$$

for $l \in \{1, 2\}$. This completes the proof of necessity.

2) Sufficiency Proof: The sufficiency of (37) is now established using a separation structure between source and channel coding. To this end, we discuss in detail the structure of the encoder blocks $\gamma^1$ and $\gamma^2$, as depicted in Fig. 6.

First, a Luenberger observer $\delta^i$, with $i \in \{0, 1, 2\}$ generates the signals $\bar{X}^i(k)$ at time instant $k$. The state observer $\delta^i$ is defined by the following equation:

$$\bar{X}^i(k + 1) = A^i\bar{X}^i(k) + L^i(Y^i(k) − C^i\bar{X}^i(k))$$

where $L^i$ is a filter matrix of appropriate dimensions. The observability of $(A^i, C^i)$ for $i \in \{0, 1, 2\}$—Assumption (A1)—implies that there exists an observer (44) which guarantees asymptotic boundedness of estimation error. More details regarding linear state observers can be found in [48]. Note that the process noise corresponds to the innovations fed to the state observer, i.e.,

$$\bar{V}^i(k) = L^i(Y^i(k) − C^i\bar{X}^i(k)).$$

The $i^{th}$ observer’s output $\bar{X}^i(k + 1)$ are then down-sampled by a factor of $n$ to yield

$$\bar{X}^i(n(k + 1)) = (A^i)^n\bar{X}^i(nk + 1)$$

where it can be shown that the accumulated innovation noise $\bar{V}^i_r(k) := \sum_{\ell=1}^{n}(A^i)^{-\ell}\bar{V}^i(nk + \ell)$ is uniformly bounded $\forall r \in [0 : n − 1]$ over $k \in \mathbb{Z}_{\geq 0}$ when an appropriate filter $L^i$ is used.

Next, each of the down-sampled sequences is processed by the respective adaptive quantizer $2^{\Psi}$ to generate the messages $M^i$ that are drawn from finite sets $\mathcal{M}^i$ with cardinalities $|\mathcal{M}^i|$ for $i \in \{0, 1, 2\}$. Finally, each pair of messages $(\{M^0, M^1\})_{i=0}^2$ is mapped into the codeword $S(nk : n(k + 1) − 1)$ with the block-length $n$ by the corresponding channel encoder $\psi^i$ at time instant $k$. The code rate for each message $M^i$ is then

$$R^i = (\log |\mathcal{M}^i|)/n,$$ for $i \in \{0, 1, 2\}.$

The receiver $\delta$ is a three-stage process that consists of the reverse operations performed by the encoding block (see Fig. 7). First, using the channel output $Q(nk : n(k + 1) − 1)$, the message estimates $\{M^i\}_{i=0}^2$ are produced by an appropriate channel decoder. Each of these estimates is then processed by an adaptive dequantizer $\Phi$ followed by an upsampling operation by $n$. This interpolation process consists in inserting $n − 1$ zeros between each couple of samples in $\bar{X}^i(nk)$ to generate the state estimations $\hat{X}^i(k)$ at time instant $k$ for $i \in \{0, 1, 2\}$.

As $n \in \mathcal{E}_0$ is, by definition, an open and non-empty set, there exists an open $l_\infty$-ball centered at $h$ and with arbitrarily small radius $\delta > 0$ denoted as $B_\delta(h) \subseteq \mathcal{E}_0$. Hence, for the given vector of topological entropies $h = (h_0, h_1, h_2)^T \in \int\mathcal{E}_0$, there exists a rate vector $R = (R^0, R^1, R^2)^T \in B_\delta(h)$, i.e., for $i \in \{0, 1, 2\}, R^i = h_i + \delta$. Let $B_\delta(R)$ be the $l_\infty$-ball of center $R$ and arbitrarily small radius $\epsilon > 0$, as depicted in Fig. 8. Since $R \in \mathcal{E}_0$, then there exists a rate vector $R_n \in \mathcal{E}_{0,n} \cap B_\delta(R)$ with
States of a noiseless (process and measurement) LTI system $J$ for $p$ $\Re$ that is observed by two different sensors $\tilde{\gamma}$ $\tilde{\gamma}_1 = \gamma$ is observable. In order to become $\Gamma$ $m$ such that system modes for $T + v[41]$ $\tilde{\gamma}_1 = 1 = (T + v)$ refer to the eigenvector corresponding to the eigenvalue $[0x0]$; and hence, it is guaranteed that there exists $\epsilon<\delta$ (48) may seem like a strict requirement. In this sub-

Consider a noiseless LTI system with a dynamic matrix $\tilde{A}$ $\tilde{A}$ that is observed by two different sensors $\gamma^1$ and $\gamma^2$, as depicted in Fig. 9. The system is described by

$$X(k + 1) = \tilde{A}X(k)$$

(51a)

$$Y^l(k) = \tilde{C}^lX(k), \text{ for } l \in \{1, 2\}$$

(51b)

where the joint system $((\tilde{C}^1, \tilde{C}^2), \tilde{A})$ is observable. In order to decouple the states of the observed system, it is possible to put the matrix $\tilde{A}$ into Jordan canonical form [41]. Note that we call system modes to refer to the states associated with a Jordan block. Supposing that $\tilde{A}$ has $p$ real eigenvalues $\{\lambda_k\}_{k=1}^p$ and $(n-p)/2$ conjugate complex eigenvalues $\{\lambda_k = \alpha_k + j\beta_k, (\lambda_k^* = \alpha_k - j\beta_k)\}_{k=p+1}^m$ with $m = p + (n-p)/2$, then there exists a transformation matrix $T$ such that

$$T = (v_1, \ldots, v_p, R(v_{p+1}), \mathcal{H}(v_{p+1}), \ldots, \mathcal{R}(v_m), \mathcal{F}(v_m))$$

(52)

where $v_k$ refer to the eigenvector corresponding to the eigenvalue $\lambda_k$. Using the matrix $T$, (51a)–(51b) become

$$X(k + 1) = T^{-1}\tilde{A}T^{-1}X(k)$$

(53a)

$$Y^l(k) = \tilde{C}^lT^{-1}X(k), \text{ for } l \in \{1, 2\}$$

(53b)

with the initial state $X(0) = T^{-1}\tilde{X}(0)$. The dynamic matrix $A$ has then the following form:

$$A = T^{-1}\tilde{A}T = \begin{pmatrix} J_1 & 0 \\ 0 & J_\Gamma \end{pmatrix}$$

(54)

with $\Gamma$ Jordan blocks such that the $i$th block is

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_i \end{pmatrix}$$

(55)

D. Generality of the Proposed Setup for Noiseless Linear Systems

The dynamic decoupling between the three subsystems in (30a)–(30b) may seem like a strict requirement. In this subsection, we show that any noiseless linear plant observed via two sensors can be put into this form, under a mild assumption on the output matrices. Subsystems 1 and 2 represent the plant modes that are observable only at sensors 1 and 2, respectively, while subsystem 0 represents the modes that are observable at each of the two sensors.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
for real eigenvalues, and

\[
J_i = \begin{pmatrix}
W_i & I_2 & 0 \\
\vdots & \ddots & \vdots \\
0 & W_i & I_2 \\
\end{pmatrix}, \quad W_i = \begin{pmatrix}
\alpha_i & \beta_i \\
-\beta_i & \alpha_i \\
\end{pmatrix},
\]

\[I_2 = \begin{pmatrix}
1 & 0 \\
0 & 1 \\
\end{pmatrix}
\]

for complex conjugate eigenvalues \(\alpha_i \pm j\beta_i\). For simplicity we suppose that each of the blocks \(\{J_i\}_{i=1}^l\) admits distinct eigenvalues. Note that the output matrix \(C^l\) can be expressed as a concatenation of submatrices \(C_i^l\) with \(i \in [1: \Gamma]\)

\[
C^l = \begin{pmatrix}
C_1^l & C_2^l & \ldots & C_{\Gamma}^l \\
\end{pmatrix} \in \mathbb{R}^{h \times d}
\]

where each submatrix \(C_i^l\) is a collection of column vectors such that

\[
C_i^l = \begin{pmatrix}
c_{1,i} & c_{2,i} & \ldots & c_{d_i,i} \\
\end{pmatrix} \in \mathbb{R}^{h \times d_i}.
\]

**Lemma 2:** Without loss of generality, consider the state matrix \(A\) to be in Jordan form \((54)\) and the output matrix \(C^l\) \((56)\) for \(l \in \{1, 2\}\). \(\Gamma\) denotes the number of Jordan blocks such that each block admits a distinct eigenvalue. Then the pair \((A, C^l)\) is observable if and only if the leading column of the block \(C_i^l\), namely, \(c_{1,i}\), is nonzero for all \(i \in [1: \Gamma]\).

**Proof:** See [42, Appendix D].

We now assume that the matrix \(C^l\) has the following structure: If there exists a nonzero coefficient \(c_{1,i}, i \in [1: d_1]\), associated with a Jordan block \(J_i\), then the leading column in \(C_i^l\) must be nonzero, i.e., \(c_{1,i} \neq 0\). For each of the systems \(\{(A, C^l)\}_{i=1}^l\), constrain the matrix \(A\) to the Jordan blocks that have nonzero contribution to the output \(y^l\). This results w.l.o.g. in a reduced system \((A_i^l, C_i^l)\) for \(l \in \{1, 2\}\). Based on the structure of \(C^l\) (and subsequently \(C_i^l\)) as well as Lemma 2, both systems are observable, and thus, the modes associated with the Jordan blocks are reconstructible. To clarify this structure for the reader, an example is provided.

**Example** In this example, we consider a system matrix \(A\) with two Jordan blocks as follows:

\[
A = \begin{pmatrix}
\lambda_1 & 1 & 0 & 0 \\
0 & \lambda_1 & 0 & 0 \\
0 & 0 & \lambda_2 & 1 \\
0 & 0 & 0 & \lambda_2 \\
\end{pmatrix} \in \mathbb{R}^{4 \times 4}
\]

and two output matrices such that \(C^1\) satisfies the assumed structure and \(C^2\) violates it

\[
C^1 = \begin{pmatrix}
c_{1,1}^1 & c_{2,1}^1 & 0 & 0 \\
\end{pmatrix} \in \mathbb{R}^{1 \times 4}
\]

\[
C^2 = \begin{pmatrix}
0 & c_{2,1}^2 & 0 & c_{2,2}^2 \\
\end{pmatrix} \in \mathbb{R}^{1 \times 4}
\]

with \(c_{1,1}^1, c_{2,1}^1, c_{2,1}^2,\) and \(c_{2,2}^2 \neq 0\). Then, the reduced system equations for \(l \in \{1, 2\}\) are, respectively,

\[
A^l = \begin{pmatrix}
\lambda_1 & 1 \\
0 & \lambda_1 \\
\end{pmatrix} \in \mathbb{R}^{2 \times 2}
\]

\[
C^l = \begin{pmatrix}
c_{1,1}^l & 0 \\
\end{pmatrix} \in \mathbb{R}^{1 \times 2}
\]

\[
A^2 = \begin{pmatrix}
\lambda_2 & 1 \\
0 & \lambda_2 \\
\end{pmatrix} \in \mathbb{R}^{2 \times 2}
\]

\[
C^2 = \begin{pmatrix}
0 & c_{2,2}^2 \\
\end{pmatrix} \in \mathbb{R}^{1 \times 2}.
\]

According to Lemma 2, the modes \((x_1, x_2)\) fall into the observability space of sensor 1, whereas \((x_3, x_4)\) are unobservable by both sensor 1 and 2.

W.l.o.g. we express \((53b)\) as

\[
Y^l(k) = \begin{pmatrix}
C^l \ast 0 \\
X_O^l(k) \\
\end{pmatrix}, \quad \text{for } l \in \{1, 2\}
\]

where \(X_O^1\) and \(X_O^2\) are the system states observable and unobservable at the \(l\)th sensor, respectively. Let \(X_O^1\) and \(X_O^2\) denote the subsets of observable and unobservable states at sensor \(l \in \{1, 2\}\). Then, by virtue of the discussion after Lemma 2, the system modes can be partitioned into three distinct classes as follows.

1) Modes common to both encoders \(\gamma^1\) and \(\gamma^2\), i.e., \(X^0 = X_O^1 \cap X_O^2\).
2) Modes observed only by \(\gamma^1\), i.e., \(X^1 = X_O^1 \cap X_O^2\).
3) Modes observed only by \(\gamma^2\), i.e., \(X^2 = X_O^1 \cap X_O^2\).

In order to estimate the different states, a dead-beat observer [43] is incorporated at each sensor. By collecting the measurements \((Y^l(n(k-1)), \ldots, Y^l(nk))\) for any \(k \in \mathbb{Z}_{\geq 1}\) over a period of time \(n\), the \(l\)th sensor recovers its observable states and generates the following measurements:

\[
Y^l(nk) = X_O^l(nk), \quad \text{for } l \in \{1, 2\}.
\]

Using the mode classes \(X^0, X^1,\) and \(X^2\), it is hence possible to decompose the original system into three subsystems with dynamic matrices \(A^0, A^1,\) and \(A^2\) and reform (53a) and \((53b)\) to obtain the setup depicted in Fig. 1 characterized by (30a) and (30b).

**V. NUMERICAL EXAMPLE: THE BINARY ADDER CHANNEL**

**A. Zero-Error Capacity Region of the BAC**

A well-known example of an MAC in literature is the BAC. This channel model consists of two independent users communicating with one receiver via a common discrete memoryless channel such that

\[
Y_k = X_k^1 + X_k^2 \in \{0, 1, 2\} \forall k \in \mathbb{Z}_{\geq 0}
\]

where \(X_k^1, X_k^2 \in \{0, 1\}\) are the inputs generated by users 1 and 2, respectively, and \(Y_k\) is the corresponding channel output. The \(i\)th user selects a particular codeword \(x_{i,n}^1\) from a finite set \(\mathcal{X}^i \subseteq \{0, 1\}^n\), where \(n \in \mathbb{Z}_{\geq 1}\) is the code block-length. At the receiver side, zero-error decoding is achieved when each element of the sumset

\[
\mathcal{Y} = \{x_{1,n}^1 + x_{2,n}^2 : \forall x_{1,n}^1 \in \mathcal{X}^1, x_{2,n}^2 \in \mathcal{X}^2\}
\]

appears exactly once, i.e., one-to-one correspondence, and hence unambiguous decoding is always possible. Note that the addition
here is componentwise over \( \mathbb{Z} \). No closed form of the BAC zero-error capacity region \( C_0 \) has been found yet. However, outer bounds [33], [34], [44] and inner bounds [45] are available. A zero-error code from [45] is used in the numerical example of the next section.

B. Application to the State Estimation Problem (Theorem 3)

We explore a setup where the states of two independent dynamical systems with given topological vector of entropies \( h = (h_1, h_2)^T \) are measured, encoded, and transmitted by users \( \mathcal{E}_1 \) and \( \mathcal{E}_2 \) through a BAC (67) and eventually recovered at the receiver end.

1) Simulation Details: We implement two scalar LTI systems described by the following equations:

\[
\begin{align*}
x^i(k + 1) &= a^i x^i(k) + v^i(k) \quad (69a) \\
y^i(k) &= x^i(k), \quad i \in \{1, 2\} \quad (69b)
\end{align*}
\]

where \( a^i = 2^{h_i} \) and \( v^i(k) \) denotes the process noise with range \([-1, 1]\). At time instant \( k = 0 \), the systems’ initial states \( \{x^i(0)\}_{i=1}^2 \) and noise term \( \{v^i(0)\}_{i=1}^2 \) are uniformly chosen from \([-1, 1]\). At each sampling time \( nk \) the encoder produces an estimate \( \hat{x}^i(nk) \) of the true state \( x^i(nk) \). Then, the error \( e^i(nk) := y^i(nk) - \hat{x}^i(nk) \) is processed by means of an adaptive uniform quantizer \( \mathcal{Q}^i \) as follows. First, the interval \([-1, 1]\) is partitioned into \( K_i \) subintervals of equal size whose respective midpoints are denoted by \( \omega(1), \ldots, \omega(K_i) \). Next, the error \( e^i(nk) \) is scaled and quantized via

\[
e^i_q(nk) := \mathcal{Q}^i(e^i(nk)/\ell^i_k) \in \{\omega(1), \ldots, \omega(K_i)\}.
\]

Both encoder and decoder agree on the initial value of the scaling factor \( \ell^i_0 \) and update it at each instant \( nk \) as follows:

\[
\ell^i_{k+1} := \frac{(a^i)^n}{K_i} \ell^i_k + \Psi^i_{\max}
\]

where the term \( \Psi^i_{\max} = \sum_{\xi=0}^{n-1} (a^i)^{n-1-\xi} \) forms an upper bound on the accumulated noise \( \Psi^i_{n-1}(k) \) during a time window of duration \( n \). The quantized error \( e^i_q(nk) \) is then encoded by \( \mathcal{E}_1 \) which employs a zero-error codebook with block-length \( n \) and a rate \( R^i = \log_2(K_i)/n \). The decoder \( \mathcal{D} \) uses the received codewords to generate the \( e^i_q(nk) \) with exactly zero decoding errors. Finally, the state estimate \( \hat{x}^i(n(k+1)) \) is updated by

\[
\hat{x}^i(n(k+1)) = (a^i)^n(\hat{x}^i(nk) + \ell^i_k e^i_q(nk)) \quad (72)
\]

Note that \( \hat{x}^i(k') \) for \( k' \in (nk, n(k+1)) \) is computed via

\[
\hat{x}^i(k') = (a^i)^{k-nk} \hat{x}^i(nk) \quad (73)
\]

2) Numerical Results and Discussion: The system (69a)–(69b) is simulated for different parameter settings, as outlined in Table I. Fig. 10 depicts a realization of the systems’ unstable states, i.e., \( X^1 = x^1 \) and \( X^2 = x^2 \), on both linear and logarithmic scales. Zero-error communication over the BAC is accomplished by means of the following binary codebooks [45] of block-length \( n = 6 \):

\[
\mathcal{D}^1 = \{0, 3, 6, 15, 17, 27, 36, 46, 48, 55, 60, 63\}
\]

\[
\mathcal{D}^2 = \{8, 9, 10, 13, 21, 22, 26, 28, 29, 30, 33, 34, 35, 37, 41, 42, 50, 53, 54, 55\}
\]

where \( \mathcal{D}^1 \) and \( \mathcal{D}^2 \) are expressed in the decimal basis. In this experiment, \( 10^5 \) realizations of system (69a)–(69b) with different initial conditions and noise values randomly chosen from the range \([-1, 1]\) are simulated. At each time step \( k \), the estimation error with the maximum norm over all realizations is selected and plotted in Figs. 11 and 12 for both scenarios 1 and 2. Note that the topological entropies in Scenario 1 are selected such that they lie within the inner region [45] of the BAC’s \( C_0 \). This guarantees that \( h \) is within \( C_0 \). On the other hand, for Scenario 2, the topological entropies are chosen from beyond the outer bound [44] on \( C_0 \). When the vector of topological entropies \( h \) lies within the BAC’s \( C_0 \), it is possible to find code rates such that \( R^i > h_i \) for \( i \in \{1, 2\} \). Thus, the estimation error is bounded in accordance with Theorem 3, as exhibited in Fig. 11. Scenario 2 on the other hand illustrates the case where \( h \notin \text{int}(C_0) \), and hence, it becomes impossible to reconstruct the state estimates at the receiver with zero error. One can see in Fig. 12 that the estimation error grows exponentially with time.

| Scenario | 1 | 2 |
|----------|---|---|
| \( h_1 \) | 0.15 | 0.85 |
| \( R^1 \) | 0.597 | 0.597 |
| \( h_2 \) | 0.18 | 0.95 |
| \( R^2 \) | 0.720 | 0.720 |
| \([V^i]_{i=1,2} \) | \([-1,1]\) | \([-1,1]\) |

Fig. 10. Example of state realizations \( x^1, x^2 \) and their corresponding estimations \( \hat{x}^1, \hat{x}^2 \) for unstable systems 1 (a) and 2 (b) on a logarithmic scale.
Empirical maximum error norms of systems 1 (a) and 2 (b) on a linear scale for Scenario 1. The topological entropy vector $h \in \text{int}(\mathcal{G}_0)$, and the code rates $R^1 > h_1$ and $R^2 > h_2$.

Fig. 11.

Empirical maximum error norms of systems 1 (a) and 2 (b) on a logarithmic scale for Scenario 2. The topological entropy vector $h \notin \text{int}(\mathcal{G}_0)$, and the zero-error code rates $R^1 < h_1$ and $R^2 < h_2$.

Fig. 12.

VI. CONCLUSION

This article represents a preliminary step in the direction of developing an in-depth understanding of the fundamental connection between the different components of a networked system in the context of distributed state estimation. Motivated by the relevance of zero-error capacity as figure of merit to assess system performance in worst-case scenarios, we derived a multilettter characterization of the zero-error capacity region of an $M$-user MAC with common message. Unlike classical information theory, the tools from its nonstochastic analog allowed us to obtain a result which is not only valid for asymptotically large coding block-lengths, but also true for finite ones. This MAC structure models a relevant scenario where $M$ sensors observe a different subset of the overall system’s dynamical modes, and the common message represents the latent correlation between their readings, i.e., the overlap between the observed subsets.

Next, the problem of distributed state estimation across a two-input, single-output MAC with bounded noise was studied. We provided tight necessity and sufficiency conditions to ensure uniformly bounded errors at the receiver end. It has been shown that in order to achieve this goal, the vector of the plants’ topological entropies must lie within the zero-error capacity region of the communication link. This result establishes a connection between the intrinsic properties of the linear systems and the channel characteristics.

Some of the remaining research questions that we need to address in future work include the following.

1) The problem we considered in this article is distributed state estimation. It is also interesting if the results obtained here are appropriately extended to cover the stabilizability of LTI systems across a shared MAC. To this end, a characterization of the zero-error capacity region of MAC with feedback in the nonstochastic framework has to be derived.

2) The coding scheme using Luenberger observers, which was constructed in this article, allowed us to establish the achievability argument to prove the main result. Nonetheless, in order to achieve a good estimation performance in practice, more sophisticated coding schemes, e.g., using zooming techniques [49], must be developed.

3) As outlined in the course of the article, an exact characterization for many MACs, e.g., BAC, remains unknown. Hence, it is also of interest to develop low-cost algorithms to compute nonstochastic information and to ultimately provide the specific zero-error capacity region of any desired MAC, or at least to reduce the gap between the upper and lower bounds characterizing the region.

APPENDIX A

PROOF OF THEOREM 1

Select two achievable rate triples $R^\prime$, $R'' \in \mathcal{G}_0$ (18). For any $\epsilon > 0$, $\exists (n, R'_n)$ and $(m, R''_m)$ such that

$$||R^\prime - R'_n|| \leq \epsilon, \ ||R'' - R''_m|| \leq \epsilon$$

(74)

where $R'_n \in \mathcal{G}_{0,n}$ and $R''_m \in \mathcal{G}_{0,m}$ for sufficiently large $n, m \in \mathbb{Z}_{\geq 1}$ that denote the block-lengths of the zero-error codes operating at $R'_n$ and $R''_m$. First, we show that for any $\alpha \in (0, 1)$, we can construct an achievable rate triple

$$R = \alpha R^\prime + (1 - \alpha) R''.$$

(75)

We define the terms $e'_n$ and $e''_m$ as follows:

$$e'_n \ := \ R^\prime - R'_n, \ e''_m \ := \ R'' - R''_m.$$  

(76)

By transmitting $j$ blocks of length $n$ at rate $R'_n$, followed by the remaining $k$ blocks of length $m$ at rate $R''_m$, we obtain

$$R_{n,m} = \frac{j \cdot n}{j \cdot n + k \cdot m} (R^\prime - e'_n) + \frac{k \cdot m}{j \cdot n + k \cdot m} (R'' - e''_m).$$

(77)

For sufficiently small $\delta > 0$ and given integers $n, m \geq 1$, we seek integers $j, k$ such that

$$\left| \alpha - \frac{j \cdot n}{j \cdot n + k \cdot m} \right| \leq \delta.$$  

(78)
Thus, for a fixed $\alpha \in (0, 1)$, the ratio $(k/j) \in \mathbb{Q}$ must satisfy
$$\frac{n}{m} \left[ \frac{1}{\alpha + \delta} - 1 \right] \leq \frac{k}{j} \leq \frac{n}{m} \left[ \frac{1}{\alpha - \delta} - 1 \right].$$  
(79)

For arbitrarily small $\delta$ and $\alpha \in (0, 1)$, there exist $k, j \in \mathbb{Z}_{\geq 1}$ such that (79) holds. Then, we have
$$\| \hat{R}_{n,m} - R \| \leq \frac{j n}{j n + km} (R' - e'_{n})$$
$$+ \frac{km}{j n + km} (R'' - e''_{m}) - \alpha R'$$
$$- (1 - \alpha) R''$$
$$\leq \frac{j n}{j n + km} \| R' \| + \frac{j n}{j n + km} \| R'' \|$$
$$+ \alpha \| R'' \| + \frac{km}{j n + km} \| e'_{n} \| + \frac{km}{j n + km} \| e''_{m} \|$$
(80)

where (80) follows from the triangle inequality, and (81) holds by virtue of (78) and (74). Hence, by making the block-lengths $n$ and $m$ sufficiently large, therefore, for arbitrarily small $\epsilon, \delta$, the difference $\| \hat{R}_{n,m} - R \|$ can be made arbitrarily small, i.e., the rate $R$ is arbitrarily close to the zero-error code rate $\hat{R}_{n,m}$. We then deduce that $R \in \mathcal{C}_{0}$, which proves the desired property.

**APPENDIX B**

**PROOF OF (39)**

The following proof is similar to the technique used in [10] and [20] in the context of point-to-point channels but as the setup here is different, we include it for means of completeness. As mentioned in Section IV-C1, in the following proof we only consider the unstable system eigenvalues without loss of generality. Furthermore, it follows from Definition 9 that uniformly bounded errors are obtained for any uniformly bounded noise ranges $[V(k)]$ and $[W(k)]$ as well as any initial state range $[X(0)]$ contained in some l-ball $B_{l}$. Hence, for the upcoming analysis we set the noise terms to zero, i.e., $[V(k)] = [W(k)] = \{0\}$ and construct the range of initial states in the following manner. First, select $\epsilon \in (0, 1 - \max_{i,j,k} |\frac{1}{k} - \frac{1}{j}|)$. Note that this interval is nonempty by virtue of Assumption (A5). We then divide the interval $[-l, l]$ on the $t$th axis into $\kappa_{l}$ equal subintervals of length $2l/\kappa_{l}$ such that
$$\kappa_{l} := \frac{|(1 - \epsilon) \lambda_{l}|}{k}, \; l \in [1 : d_{l}].$$  
(82)

Let $P_{l}(s)$ denote the midpoints of the subintervals, for $s = [1 : \kappa_{l}]$ and construct an interval $I_{l}(s)$ centered at $P_{l}(s)$ such that its length is equal to $l/\kappa_{l}$. We define the family of hypercuboids $\mathcal{H}$ as follows:

$$\mathcal{H} = \left\{ \prod_{\ell=1}^{d_{l}} I_{\ell}(s) : s \in [1 : \kappa_{l}], \ell \in [1 : d_{l}] \right\}.$$  
(83)

Observe that any two hypercuboids from $\mathcal{H}$ are separated by a distance of $l/\kappa_{l}$ along the $t$th axis for each $\ell \in [1 : d_{l}]$.

In the following analysis, the superscript $i$ referring to the respective plant is omitted unless otherwise stated. The initial state range is set as $[X^{i}(0)] = \cup_{H \in \mathcal{H}} H \subset B_{l}(0)$. The operator $d_{m}(\cdot)$ is defined as the diameter of a set using $l_{\infty}$-norm. Then, as $[E_{j}(k)] \supseteq [E_{j}(k)] [q(0 : k - 1)]$,
$$\text{dm}([E_{j}(k)]) \geq \text{dm} ([E_{j}(k)] [q(0 : k - 1)])$$
(84)

$$= \text{dm} ([X_{j}(k) - \delta_{j}] (k, q(0 : k - 1)) [q(0 : k - 1)])$$
$$= \sup_{r, p \in [X_{j}(0)] [q(0 : k - 1)]} \| (A_{j})^{k} (r - p) \|_{2}$$
$$\geq \sup_{r, p \in [X_{j}(0)] [q(0 : k - 1)]} \sigma_{\min} \left( (A_{j})^{k} \right) \| r - p \|_{2}$$
$$\geq \sigma_{\min} \left( (A_{j})^{k} \right) \frac{\text{dm} ([X_{j}(0)] [q(0 : k - 1)])}{\sqrt{d_{l}}}$$
(85)

where $k \in \mathbb{Z}_{\geq 0}$, $q(0 : k - 1) \in [Q(0 : k - 1)]$, $\| \cdot \|_{2}$ denotes the Euclidean norm and $\sigma_{\min}(\cdot)$ refers to the smallest singular value. The inequality (84) holds because conditioning reduces the range, and (85) is valid since the diameter of a uv range is translation invariant. Now, note that the Yamamoto identity [50] states that
$$\lim_{k \to \infty} \left( \sigma_{\min} \left( (A_{j})^{k} \right) \right)^{1/k} = \lambda_{\min} \left( A_{j} \right)$$  
(86)

with $\lambda_{\min}$ being the eigenvalue with the smallest magnitude. Then, since $j \in [1 : m] < \infty$, i.e., there are finitely many real Jordan blocks $A_{j}$, there exists $k_{c} \in \mathbb{Z}_{\geq 0}$ such that
$$\sigma_{\min} \left( (A_{j})^{k} \right) \geq \left( 1 - \frac{\epsilon}{2} \right)^{k} \lambda_{\min} (A_{j})^{k}, \; k \geq k_{c}.$$  
(87)

In addition, by *uniform boundedness of errors* (35) there exists $\xi > 0$ such that
$$\xi \geq \sup \| [E_{j}(k)] \| \geq \frac{1}{2} \text{dm} \left( [E_{j}(k)] \right)$$
$$\geq \left( 1 - \frac{\epsilon}{2} \right) \lambda_{\min} (A_{j})^{k} \frac{\text{dm} ([X_{j}(0)] [q(0 : k - 1)])}{2 \sqrt{d_{l}}}.$$  
(88)

For large enough $k \in \mathbb{N}$, the hypercuboid family $\mathcal{H}$ in (83) is an $[X^{i}(0)] [Q(0 : k - 1)]$-overlap isolated partition of $[X^{i}(0)]$. To show this, we suppose in contradiction that $\exists H \in \mathcal{H}$ that is overlap connected in the family $[X^{i}(0)] [Q(0 : k - 1)]$ with another hypercuboid from $\mathcal{H}$. Therefore, there would exist a set $[X^{i}(0)] [q(0 : k - 1)]$ which contains a point $r_{j} \in H$ and a point $p_{j} \in H'$, with $H' \in \mathcal{H} \setminus H$ such that $r_{j}$ and $p_{j}$ are overlap
connected. This implies
\[
\|p_j - r_j\| \leq d_m \left( |X^i(0)| q(0 : k - 1) \right) \leq \frac{2\sqrt{\kappa_\xi}}{|(1 - \frac{\ell}{L}) \lambda_{\min}(A_j^i)|^k}
\]
for \( j \in [1 : m] \) and \( k \geq k_j \). Nonetheless, note that by construction the distance between any two hypercuboids in \( \mathcal{H} \) is equal to \( l/\kappa_\ell \) along the \( \ell \)th axis. Therefore,
\[
\|p_j - r_j\| \geq \frac{l}{\kappa_\ell} \leq \frac{l}{(1 - \epsilon)^k |\lambda_{\min}(A_j^i)|^k}.
\]
For sufficiently large \( k \), it is possible to obtain \( (1 - \epsilon^2)/(1 - \epsilon) > 2\sqrt{\kappa_\xi}/l \). Hence, the RHS of (89) would exceed the RHS of (90) resulting in a contradiction. Thus, when \( k \) is large enough, the family \( \mathcal{H} \) is \([X^i(0)](Q(0 : k - 1)]\)-overlap isolated partition of \([X^i(0)]\). As the cardinality of any \([X^i(0)](Q(0 : k - 1)]\)-overlap isolated partition is upper bounded by the maximin information \([X^i(0)](Q(0 : k - 1)]\), we obtain
\[
I_\epsilon[X^i(0); Q(0 : k - 1)] = \log \left( |X^i(0)| |Q(0 : k - 1)| \right)
\geq \log |\mathcal{H}| = \log \left( \prod_{\ell=1}^{d_i} \left( |(1 - \epsilon) \lambda_{\ell}^i \right)^k \right)
\geq \log \left( \prod_{\ell=1}^{d_i} \frac{0.5 |(1 - \epsilon) \lambda_{\ell}^i |^k}{k} \right)
= k \left( d_i \log(1 - \epsilon) - \frac{d_i}{k} + \sum_{\ell=0}^{d_i} \log |\lambda_{\ell}^i| \right).
\]
Hence, for \( i \in \{0, 1, 2\} \) it follows from (91)
\[
I_\epsilon[X^i(0); Q(0 : k - 1)] \geq \frac{d_i}{k} \log(1 - \epsilon) - \frac{d_i}{k} + \sum_{\ell=0}^{d_i} \log |\lambda_\ell^i|.
\]
Using the monotonicity of \( I_\epsilon \) and then letting \( k \to \infty \) and \( \epsilon \to 0 \), we obtain
\[
I_\epsilon[X^i(0); Q(0 : k - 1)] \geq \frac{d_i}{k} \log |\lambda_\ell^i|.
\]
In addition, the set \( \mathcal{Z} \) consists of all \( uv \)'s \( Z \) such that \( Z = \phi(\Lambda) = \psi(\Omega) \), and hence,
\[
I_\epsilon[\Lambda; \Omega] = \max_{Z \in \mathcal{Z}} \log |Z|.
\]
Since \( \Lambda \perp \Theta \), then \( \phi(\Lambda) \perp \Theta \) and subsequently \( Z \perp \Theta \). Therefore, \( \mathcal{Z} \subseteq \mathcal{Z}_\Theta \), and thus,
\[
\log |Z| \leq \log |Z_\Theta|.
\]
By maximizing both left-hand side and RHS of (94) over \( \mathcal{Z} \) and \( \mathcal{Z}_\Theta \), we obtain
\[
I_\epsilon[\Lambda; \Omega] \leq I_\epsilon[\Lambda; \Omega | \Theta].
\]

**References**

[1] S. Marano, V. Matta, and P. Willett, “Distributed estimation in large wireless sensor networks via a locally optimum approach,” *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 749–756, Feb. 2008.

[2] G. Chisci, H. Elsawy, A. Conti, M. -S. Alouini, and M. Z. Win, “Uncorrelated massive wireless networks: Spatiotemporal models and multiaccess strategies,” *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 918–931, Jun. 2019.

[3] A. Conti, S. Mazuelas, S. Bartoletti, W. C. Lindsey, and M. Z. Win, “Soft information for localization-of-things,” *Proc. IEEE*, vol. 107, no. 11, pp. 240–2264, Nov. 2019.

[4] F. Zabini and A. Conti, “Inhomogeneous poisson sampling of finite-energy signals with uncertainties in R^2,” *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4679–4694, Sep. 2016.

[5] M. Shirvanihaddam et al., “Short block-length codes for ultra-reliable low latency communications,” *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 130–137, Feb. 2019.

[6] W. W. L. Li, Y. J. Zhang, A. M.-C. So, and M. Z. Win, “Slow adaptive OFDMA systems through chance constrained programming,” *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3858–3869, Jul. 2010.

[7] M. Tabbara, D. Nesic, and A. R. Teel, “Stability of wireless and wireline networked control systems,” *IEEE Trans. Auto. Control*, vol. 52, no. 9, pp. 1615–1630, Sep. 2007.

[8] W. P. M. H. Heemels and N. van de Wouw, “Stability and stabilization of networked control systems,” *in Networked Control Systems (Lecture Notes in Control and Information Sciences)*, A. Bemporad, M. Heemels, and M. Johansson, (Eds.), vol. 406, Berlin, Germany: Springer, 2010.

[9] C. E. Shannon, “Two-way communication channels,” in *Proc. 4th Berkeley Symp. Math. Statist. Probability*, vol. 1, 1961, pp. 611–644.

[10] G. N. Nair, “A nonstochastic information theory for communication and state estimation,” *IEEE Trans. Auto. Control*, vol. 58, pp. 1497–1510, Jun. 2013.

[11] G. N. Nair, “Nonstochastic information concepts for estimation and control,” in *Proc. 54th IEEE Conf. Decis. Control.,* 2015, pp. 45–56.

[12] A. Rangi and M. Franceschetti, “Channel coding theorems in nonstochastic information theory,” in *Proc. IEEE Int. Symp. Inf. Theory, 2021*, pp. 1799–1795.

[13] M. Gagrani, Y. Ouyang, M. Rasouli, and A. Nayyar, “Worst-case guarantees for remote estimation of an uncertain source,” *IEEE Trans. Auto. Control*, vol. 66, no. 4, pp. 1794–1801, Apr. 2021.

[14] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Wireless network information flow: A deterministic approach,” *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Hoboken, NJ, USA: Wiley, 1991.

[16] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ., 2012.

[17] C. E. Shannon, “The zero-error capacity of a noisy channel,” *IRE Trans. Inf. Theory*, vol. 2, pp. 8–19, Sep. 1956.

[18] A. S. Matveev and A. V. Savkin, “Shannon zero-error capacity in the problems of state estimation and stabilization via noisy communication channels,” *Int. J. Control*, vol. 80, pp. 241–255, 2007.

[19] M. Wiese, T. J. Oechtering, K. H. Johansson, P. Papadimitratos, H. Sandberg, and M. Skoglund, “Secure estimation and zero-error secrecy capacity,” *IEEE Trans. Auto. Control*, vol. 64, no. 3, pp. 1047–1062, Mar. 2019.

[20] A. Saberi, G. N. Nair, and F. Farkhi, “Bounded estimation over finite-state channels: Relating topological entropy and zero-error capacity,” *IEEE Trans. Auto. Control*, vol. 67, no. 8, pp. 4029–4044, Aug. 2022.
[21] A. Saberi, F. Farokhi, and G. N. Nair, “Zero-error feedback capacity for bounded stabilization and finite-state additive noise channels,” IEEE Trans. Autom. Control, vol. 68, no. 10, pp. 6335–6355, Oct. 2022.

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

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

[24] V. S. Borkar, S. Mitter, and S. Tatikonda, “Markov control problems under communication constraints,” Commun. Inf. Syst., vol. 1, no. 1, pp. 15–32, 2001.

[25] Nigel J. Newton and Sanjoy K. Mitter, “Variational bayes and a problem of reliable communication I: Finite systems,” Commun. Inf. Syst., vol. 10 no. 3, 155–182, 2010.

[26] M. Rich and N. Elia, “Optimal mean-square performance for MIMO networked systems,” in Proc. Amer. Control Conf., 2015, pp. 6040–6045.

[27] E. Garone, B. Sinopoli, A. Goldsmith, and A. Casavola, “LQG control for MIMO systems over multiple erasure channels with perfect acknowledgment,” IEEE Trans. Autom. Control, vol. 57, no. 2, pp. 450–456, Feb. 2012.

[28] R. Olfati-Saber, “Distributed Kalman filtering for sensor networks,” in Proc. 46th IEEE Conf. Decis. Control. 2007, pp. 5492–5498.

[29] R. Olfati-Saber, “Kalman-consensus filter: Optimality, stability, and performance,” in Proc. 48th IEEE Conf. Decis. Control Held Jointly 28th Chin. Control Conf., 2009, pp. 7036–7042.

[30] A. Tanwani, “Suboptimal filtering over sensor networks with random communication,” IEEE Trans. Autom. Control, vol. 67, no. 10, pp. 5456–5463, Oct. 2022.

[31] A. A. Zaidi, T. J. Oechtering, and M. Skoglund, “Sufficient conditions for closed-loop control over multiple-access and broadcast channels,” in Proc. IEEE Conf. Dec. Cont., 2010, vol. 49, no. 9, pp. 4771–4776.

[32] J. Liu and V. Gupta, “Stabilizability conditions for linear time invariant systems across a gaussian MAC channel,” IEEE Trans. Autom. Control, vol. 64, no. 6, pp. 2310–2323, Jun. 2019.

[33] R. Ahktowe, “Two-way communication channels,” in Proc. IEEE Int. Symp. Inf. Theory, Tsahkador, Armenia, 1971, pp. 23–52.

[34] H. Liao, “Multiple-access channels,” Ph.D. dissertation, Univ. Hawaii, Honolulu, HI, USA, 1972.

[35] G. Zafzouf, G. N. Nair, and J. S. Evans, “Zero-error capacity of multiple access channels via nonstochastic information,” in Proc. IEEE Autom. Control Conf. 2019, pp. 1–5.

[36] G. Zafzouf and G. N. Nair, “Zero-error capacity region of a class of multiple access channels with inter-user correlation,” in Proc. IEEE Inf. Theory Workshop, 2021, pp. 1–5.

[37] G. Zafzouf and G. N. Nair, “Distributed state estimation with bounded errors over multiple access channels,” in Proc. IEEE Int. Symp. Inf. Theory, 2020, pp. 2861–2866.

[38] A. Rényi, Foundations of Probability, San Francisco, CA, USA, Holden-Day, 1970.

[39] C. E. Shannon, “The lattice theory of information,” Trans. IRE Professional Group Inf. Theory, vol. 1, pp. 105–107, Feb. 1953.

[40] S. Wolf and J. Wullschleger, “Zero-error information and applications in cryptography,” in Proc. Inf. Theory Workshop, 2004, pp. 1–6.

[41] R. A. Horn and C. R. Johnson, Matrix Analysis, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[42] G. Zafzouf, G. N. Nair, and F. Farokhi, “Uniformly bounded state estimation over multiple access channels,” 2022, arXiv:2002.03294.

[43] R. C. Dorf and R. H. Bishop, Modern Control Syst., 14th ed. Hoboken, NJ, USA: Prentice Hall, 2021.

[44] P. Austrin, P. Kasiki, M. Koivisto, and J. Nederlof, “Sharper upper bounds for unbalanced uniquely decodable code pairs,” IEEE Trans. Inf. Theory, vol. 64, no. 2, pp. 1368–1373, Feb. 2018.

[45] M. Mattas and P. R. J. Ostergärd, “A new bound for the zero-error capacity region of the two-user binary adder channel,” IEEE Trans. Inf. Theory, vol. 51, no. 9, pp. 3289–3291, Sep. 2005.

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

[47] D. Slepian and J. K. Wolf, “A coding theorem for multiple access channels with correlated sources,” Bell Syst. Tech. J., vol. 52, pp. 1037–1076, Sep. 1973.

[48] P. J. Antsaklis and A. Michel, Linear Systems, New York, NY, USA: McGraw-Hill, 1997.

[49] R. W. Brockett and D. Liberson, “Quantized feedback stabilization of linear systems,” IEEE Trans. Autom. Control, vol. 45, no. 7, pp. 1279–1289, Jul. 2000.

[50] T. Yamamoto, “On the extreme values of the roots of matrices,” J. Math. Soc. Jpn., vol. 19, no. 2, pp. 173–178, Apr. 1967.

Ghassen Zafzouf received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from the Technical University of Munich, Munich, Germany, in 2015 and 2018, respectively, the M.Eng. degree in electrical engineering from the University of Queensland, Brisbane, QLD, Australia, in 2017, and the Ph.D. degree in electrical engineering from the University of Melbourne, Parkville, VIC, Australia, in 2022.

His research interests include information theory, digital communication, and networked control systems.

Dr. Zafzouf was the recipient of several prizes and awards including Germany's National Scholarship in 2018 and the Melbourne Research Scholarship in 2019.

Girish N. Nair (Fellow, IEEE) was born in Malaysia. He received the B.E. (first class hons.) degree in electrical engineering, the B.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Melbourne, Parkville, VIC, Australia, in 1994, 1995, and 2000, respectively.

He is currently a Professor with the Department of Electrical and Electronic Engineering, University of Melbourne. From 2019 to 2024, he is the Principal Australian Investigator of an AUSMURI project. His research interests include information theory and control.

Dr. Nair was an ARC Future Fellow, from 2015 to 2019. He was the recipient of several prizes, including the IEEE CSS Axelby Outstanding Paper Award in 2014 and a SIAM Outstanding Paper Prize in 2006. He delivered a semipenial lecture at the 60th IEEE Conference on Decision and Control, in 2021.

Farhad Farokhi (Senior Member, IEEE) received the Ph.D. degree in automatic control from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2014.

He joined the University of Melbourne, Parkville, VIC, Australia, where he is currently a Senior Lecturer (equivalent to Assistant/Associate Professor in North America). From 2018 to 2020, he was also a Research Scientist with the CSIRO’s Data61, Canberra ACT, Australia. He has been involved in multiple projects on data privacy and cyber-security funded by the Australian Research Council, the Defence Science and Technology Group, the Department of the Prime Minister and Cabinet, the Department of Environment and Energy, and the CSIRO.

Dr. Farokhi was the recipient of the VESKI Victoria Fellowship from the Victoria State Government, Australia, and the McKenzie Fellowship, funded by the Australian Research Council, the Defence Science and Technology Group, the Department of the Prime Minister and Cabinet, the Department of Environment and Energy, and the CSIRO.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.