The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link
Part II: vector systems

Anant Sahai, Sanjoy Mitter

sahai@eecs.berkeley.edu, mitter@mit.edu

Abstract

In part I, we reviewed how Shannon's classical notion of capacity is not sufficient to characterize a noisy communication channel if we intend to use that channel as a part of a feedback loop to stabilize an unstable scalar linear system. While classical capacity is not enough, a parametric sense of capacity called "anytime capacity" was shown to be both necessary and sufficient for the stabilization of an unstable process over that channel. The rate required is the log of the open-loop system gain and the sense of reliability required comes from the desired sense of stability. This is sufficient even in cases with noisy observations and without any explicit feedback between the observer and the controller.

Here, in part II, the vector-state generalizations are established and it is the magnitudes of the unstable eigenvalues that play an essential role. To deal with such systems, we introduce the concept of the anytime rate-region which is the region of rates that the channel can support while still meeting potentially different anytime reliability targets for the parallel bitstreams. All the scalar results generalize on an eigenvalue by eigenvalue basis. For cases in which there is no explicit feedback of the noisy channel outputs, the intrinsic delay of the control system tells us what the feedback delay needs to be while evaluating the anytime-rate-region for the channel. We close with a numeric example involving a binary erasure channel that illustrates how differentiated service is required in any separation-based control architecture.

Index Terms

Real-time information theory, reliability functions, control over noisy channels, differentiated service, feedback, anytime decoding

A. Sahai is with the Department of Electrical Engineering and Computer Science, U.C. Berkeley. Some portions of this work appeared in his doctoral dissertation and at the Allerton Conference in 2000.

Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology. Support for S.K. Mitter was provided by the Army Research Office under the MURI Grant: Data Fusion in Large Arrays of Microsensors DAAD19-00-1-0466 and the Department of Defense MURI Grant: Complex Adaptive Networks for Cooperative Control Subaward #03-132 and the National Science Foundation Grant CCR-0325774.

The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link
Part II: vector systems

I. Introduction

One of Shannon's key contributions was the idea that bits could be used as a single universal currency for communication. For a vast class of point-to-point applications, the communication aspect of the problem can be reduced to transporting bits reliably from one point to another where the required sense of reliability does not depend on the application. The classical source/channel separation theorems justify a layered communication architecture with an interface that focuses primarily on the data rate. Data rate has the advantage of being additive in nature and so multiple applications can be supported over a single link by simple multiplexing of the data streams. The underlying noisy channel is thus abstracted away and considered only in terms of its capacity. This paradigm has been so successful in practice, that researchers often assume that it is always valid.

Interactive applications pose a challenge to this separation based paradigm because long delays are not allowed. In part I of this paper [1], we studied the requirements for a particular interactive application: stabilization of an unstable scalar linear system with feedback that must go through a noisy communication channel.[1] It turns out that data rate is not the only relevant parameter since the underlying noisy channel must also support enough anytime-reliability to meet the targeted sense of stability. However, the architectural implications of this result are unclear in the scalar case since there is only one data stream. In addition, scalar control does not provide a natural setting in which to explore interactions involving more than two parties.

To better understand the architectural implications of interactivity in a well defined mathematical setting, it is natural to consider the stabilization of linear systems with a vector-valued state. Prior work on communication-limited stabilization problems had also considered such vector problems, but primarily from a source coding perspective in that the communication constraint was expressed as a rate constraint. [2] showed that the minimum rate required is the sum of the logs of the magnitudes of the unstable eigenvalues and [3] extends the result to certain classes of unbounded driving disturbances. [4], [5] gave a bound on control performance in the vector case based on sequential rate distortion theory. These sequential rate-distortion theory bounds essentially calculated what control performance is possible using a noisy channel that is perfectly matched to the unstable open-loop system while being restricted to having a specified Shannon capacity. Thus, the prior necessary conditions on stabilization were only in terms of Shannon capacity and the prior sufficient conditions required noiseless channels.

In this paper, our goal is to generalize the tight necessary and sufficient conditions from the scalar case to the vector case, and to begin exploring the architectural implications of these results. The goal is to understand what happens in the point-to-point case with a single noisy link and thereby set the stage for understanding multiparty interactions.[2] Recently, Pradhan has investigated block-coding reliability regions for distributed channel coding without feedback.[8], [9] Our results in this paper indicate that reliability regions are important even in the point-to-point case.

After first briefly reviewing the main results from the scalar case in Section I-A, we introduce our model of vector valued linear control systems in Section I-B. The generalization of the scalar results to the vector case is done in stages. In Section II, we study the case of A matrices that are real and diagonal. This is essentially multiple copies of the scalar case and it shows what the requirements are going to be in the vector case with a minimum of technical difficulties. In Section III, we examine the generic case of vector systems that have diagonalizable A matrices. Here, the issues of controllability and observability become important, but can be addressed by treating time in the appropriate short blocks. There is also

[1] See [1] for more discussion of relevant prior work.

[2] The multiparty case has begun to be addressed in the control community [6], [7] under the assumption of noiseless channels.

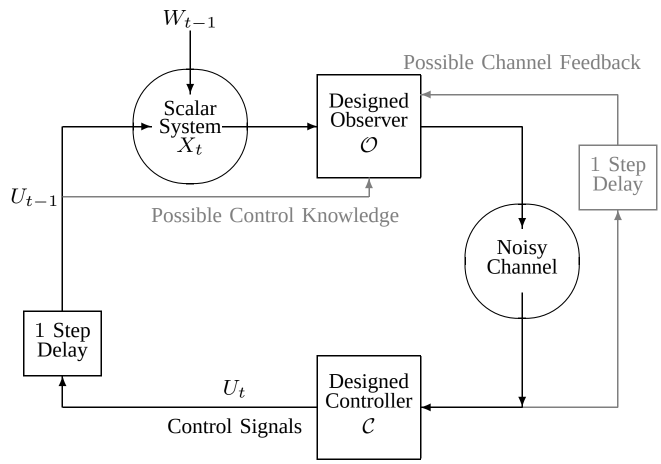

Fig. 1. Control over a noisy communication channel. The unstable system is persistently disturbed by W_t and must be kept stable in closed-loop through the actions of $\mathcal{O}, \mathcal{C}$.

a more subtle issue regarding intrinsic delays in control systems that is dealt with in Section IV. The remaining nongeneric case of nondiagonalizable matrices is addressed in Section V.

The significance of these results is demonstrated through a numeric example in Section VI involving stabilization of a vector-valued plant over a binary erasure channel. For this example, we point out that stabilization is impossible unless different bits are treated differently when it comes to transporting them across the noisy channel. These results establish that in interactive settings, a single "application" can fundamentally require different senses of reliability for its data streams. No single number can adequately summarize the channel and any layered architecture for reliable communication should allow applications to individually adjust the reliabilities on bitstreams.

There are many results in this paper and most of them involve straightforward generalizations of results from [1]. In order not to unduly burden the reader with details and unnecessarily lengthen this paper, we have adopted a discursive style in some of the proofs. The reader familiar with [1] should not have any difficulty in filling in the omitted details.

A. *Review of scalar system results from part I*

This section just restates the key results and ideas from [1] that we will build upon here. The reader is referred to [1] for complete proofs, motivations, and other discussion. The complete problem is illustrated in Figure 1 and the core unstable scalar system is modeled as:

$$X_{t+1} = \lambda X_t + U_t + W_t, \quad t \geq 0 \tag{1}$$

where $\{X_t\}$ is a $\mathbb{R}$-valued state process. $\{U_t\}$ is a $\mathbb{R}$-valued control process and $\{W_t\}$ is a bounded noise/disturbance process s.t. $|W_t| \leq \frac{\Omega}{2}$. This bound is assumed to hold with certainty. For convenience, we also assume a known initial condition $X_0 = 0$.

Our goal is to stabilize the system in closed loop:

Definition 1.1: (Definition 2.2 in [1]) A closed-loop dynamic system with state X_t is *η-stable* if there exists a constant K s.t. $E[\|X_t\|^\eta] \leq K$ for all $t \geq 0$.

Definition 1.2: (Definition 3.1 in [1]) As illustrated in figure 2, a rate R *communication system* over a noisy channel is an encoder $\mathcal{E}$ and decoder $\mathcal{D}$ pair such that:

- R-bit message M_i enters[3] the encoder at discrete time i

[3] In what follows, we will often consider messages as composed of bits for simplicity of exposition. The i-th bit arrives at the encoder at time $\frac{i}{R}$ and thus M_i is composed of the bits $S_{\lfloor (i-1)R \rfloor + 1}^{\lfloor iR \rfloor}$.

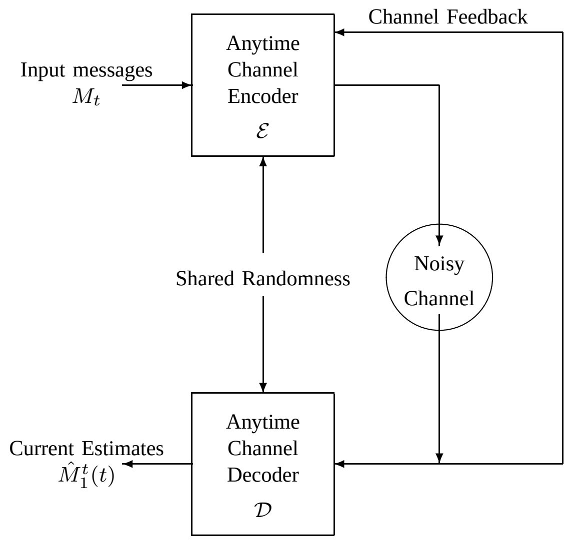

Fig. 2. The problem of communicating messages in an anytime fashion. The important feature is that both the encoder $\mathcal{E}$ and decoder $\mathcal{D}$ are forced to be causal maps and the decoder in principle provides updated estimates for *all* past messages. We further require that these estimates converge to the true message values appropriately rapidly with increasing delay.

- The encoder produces a channel input at integer times based on all information that it has seen so far. For encoders with access to feedback with delay $1+\theta$, this also includes the past channel outputs $B_1^{t-1-\theta}$.
- The decoder produces updated channel estimates $\hat{M}_i(t)$ for all $i \leq t$ based on all channel outputs observed till time t.

A rate R sequential communication system achieves *anytime reliability* α if there exists a constant K such that:

$$\mathcal{P}(\hat{M}_1^i(t) \neq M_1^i) \leq K 2^{-\alpha(t-i)} \tag{2}$$

holds for every i, t. The probability is taken over the channel noise, the R bit messages M_i, and all of the common randomness available in the system.

If (2) holds for every possible realization of the messages M, then we say that the system achieves *uniform anytime reliability* α.

Communication systems that achieve *anytime reliability* are called *anytime codes* and similarly for *uniform anytime codes*.

Definition 1.3: (Definition 3.2 in [1]) The α*-anytime capacity* $C_{\text{any}}(\alpha)$ of a channel is the least upper bound of the rates R (in bits) at which the channel can be used to construct a rate R communication system that achieves uniform anytime reliability α.

Feedback anytime capacity is used to refer to the anytime capacity when the encoder has access to noiseless feedback of the channel outputs with unit delay.

Theorem 1.4: (Theorem 3.3 in [1]) For a given noisy channel and $\eta > 0$, if there exists an observer $\mathcal{O}$ (with or without channel feedback or access to the controls) and controller $\mathcal{C}$ for the unstable scalar system that achieves $E[|X_t|^\eta] < K$ for all sequences of bounded driving noise $|W_t| \leq \frac{\Omega}{2}$, then $C_{\text{any}}(\eta \log_2 \lambda) \geq \log_2 \lambda$ bits per channel use for the noisy channel considered with the encoder having access to noiseless feedback.

This proof proceeded by a direct reduction of the anytime communication problem to a problem of stabilization. The key idea was to embed the messages into the open-loop uncontrolled state of the unstable plant by suitable choice of disturbances. The core equation was:

$$\check{X}_t = \gamma\lambda^t \sum_{k=0}^{\lfloor Rt \rfloor} (2+\epsilon_1)^{-k} S_k \tag{3}$$

where S_k is the k-th bit to be transmitted (as either ± 1),

$$\epsilon_1 = 2^{\frac{\log_2 \lambda}{R}} - 2 \tag{4}$$

and

$$\gamma = \frac{\Omega}{2\lambda^{1+\frac{1}{R}}} \tag{5}$$

was the scaling term to ensure we stayed within the bound for the disturbance. The key fact that enabled all this to work was that the minimum gap between the encoded state corresponding to two sequences of bits that first differ in bit position i is given by $\text{gap}_i(t) =$

$$\inf_{\bar{S}:\bar{S}_i \neq S_i} |\check{X}_t(S) - \check{X}_t(\bar{S})| > \begin{cases} \lambda^{t-\frac{i}{R}} \left(\frac{2\gamma\epsilon_1}{1+\epsilon_1}\right) & \text{if } i \leq \lfloor Rt \rfloor \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

This allowed us to robustly recover the data at the decoder using a variation on the traditional serial A/D algorithm.

On the sufficiency side, the core theorem is:

Theorem 1.5: (Theorem 4.2 in [1]) It is possible to control an unstable scalar process driven by a bounded disturbance over a noisy channel so that the η-moment of $|X_t|$ stays finite for all time if the channel with feedback has $C_{\text{any}}(\alpha) > \log_2 \lambda$ for some $\alpha > \eta \log_2 \lambda$ and the observer is allowed to observe the channel outputs and the state exactly.

There are a host of theorems that extend the previous result to situations with various types of limitations:

Theorem 1.6: (Theorem 4.4 in [1]) If for all $\Omega > 0$, it is possible to stabilize a particular unstable scalar system with gain λ^n and arbitrary disturbance signal bounded by Ω when we are allowed n uses of a particular channel between when the control-system evolves, then for any $\Omega > 0$ it is also possible to stabilize an unstable scalar system with gain λ that evolves on the same time scale as the channel using an observer restricted to only observe the system every n time steps.

If the channel input at time t is allowed to explicitly depend on the channel outputs at time $t - v$, this permitted feedback is the same in both cases and is not changed by n.

Theorem 1.7: (Theorem 4.5 in [1]) Theorem 1.5 continues to hold if the control signal U_t is required to depend only on the channel outputs up through time $t - v$ where $v \geq 0$. Only the constants grow larger.

Theorem 1.8: (Theorem 4.6 in [1]) If for all $\Omega > 0$, it is possible to $\eta-$stabilize a particular unstable scalar system with arbitrary disturbance signal bounded by Ω given the ability to apply precise control signals, then for all $\Gamma_c > 0$ and $\Omega > 0$, it remains possible to $\eta-$stabilize the same unstable scalar system with arbitrary disturbance signal bounded by Ω given the ability to apply only Γ_c-precise control signals.

Corollary 1.1: (Corollary 4.1 in [1]) It is possible to control an unstable scalar process driven by a bounded disturbance over a noisy channel so that the η-moment of $|X_t|$ stays finite for all time if the channel with noiseless feedback has $C_{\text{any}}(\alpha) > \log_2 \lambda$ for some $\alpha > \eta \log_2 \lambda$ and the observer is allowed to observe the channel outputs exactly and has a boundedly noisy view of the state.

Beyond the above limitations in terms of noisy observation or delayed/noisy actions, there are a few important results having to do with limitations on the feedback structure in which we deny the system observer direct access to the past channel outputs:

Theorem 1.9: (Theorem 5.2 in [1]) It is possible to control an unstable scalar process driven by a bounded disturbance over a noisy channel so that the η-moment of $|X_t|$ stays finite for all time if the channel *without feedback* has $C_{\text{any}}(\alpha) > \log_2 \lambda$ for some $\alpha > \eta \log_2 \lambda$ and the observer has only boundedly noisy access to the state process.

Specializing to the case of discrete memoryless channels (DMCs), we also gave a specific randomized construction for the observer that was nearly memoryless:

Theorem 1.10: (Theorem 5.3 in [1]) It is possible to control an unstable scalar process driven by a bounded disturbance over a DMC so that the η-moment of $|X_t|$ stays finite for all time if the channel without feedback has random coding error exponent $E_r(R) > \eta \log_2 \lambda$ for some $R > \log_2 \lambda$ and the observer is allowed boundedly noisy access to the state process.

Furthermore, there exists an $n > 0$ so this is possible by using an observer consisting of a time-varying random scalar quantizer that samples the state every n time steps and outputs a random label for the bin index. This random label is chosen iid from the channel input alphabet $\mathcal{A}^n$ according to the distribution that maximizes the random coding error exponent at R. The controller is assumed to have access to the randomness used to choose the random bin labels.

A simplified interpretation of the above theorem is given by:

Corollary 1.2: (Corollary 5.1 in [1]) If the observer is allowed boundedly noisy access to the plant state, and the noisy channel is a DMC with Shannon capacity $C > \log_2 \lambda$, then there exists some $\eta > 0$ and an observer/controller pair that stabilizes the system in closed loop so that the η-moment of $|X_t|$ stays finite for all time.

Theorem 1.11: (Corollary 5.3 in [1]) Given a noisy channel with a countable output alphabet, identify the channel output alphabet with the integers and suppose that there exist[4] $K > 0, \beta > 0$ so that the channel outputs B_t satisfy: $\mathcal{P}(|B_t| \geq i) \leq K e^{-\beta i}$ for all t.

Then, it is possible to control an unstable scalar process driven by a bounded disturbance over a that channel so that the η-moment of $|X_t|$ stays finite for all time if the channel with feedback has $C_{\text{any}}(\alpha) > \log_2 \lambda$ for some $\alpha > \eta \log_2 \lambda$ and the observer is allowed boundedly noisy access to the system state.

There are also analogous results for continuous time where nats are used instead of bits, and the rate is the unstable gain itself rather than the logarithm of the magnitude of the gain. Similarly, for the case of almost-sure stabilization[5], we had a short sequence of results beginning with a key lemma:

Lemma 1.1: (Lemma 4.1 in [1] If it is possible to η'-stabilize a persistently disturbed system from (1) with open-loop gain λ' when driven by any driving noise W' bounded by Ω, then there exists a time-varying observer with noiseless access to the state and a time-varying controller so that any undisturbed system (1) with initial condition $|X_0| \leq \frac{\Omega}{2}$, $W_t = 0$, and $0 < \lambda < \lambda'$ can be stabilized in the sense that there exists a K so that:

$$E[|X_t|^{\eta'}] \leq K(\frac{\lambda}{\lambda'})^{\eta' t} \tag{7}$$

and culminating in

Corollary 1.3: (Corollary 5.2 in [1]) If the observer is allowed perfect access to the plant state, and the noisy channel is a DMC with Shannon capacity $C > \log_2 \lambda$, then there exists an observer/controller pair that stabilizes the system (1) in closed loop so that:

$$\lim_{t \to \infty} X_t = 0 \text{ almost surely}$$

as long as the initial condition $|X_0| \leq \frac{\Omega}{2}$ and the disturbances $W_t = 0$.

[4]This condition is most naturally satisfied by a channel with a finite output alphabet. It is also satisfied naturally for certain channels with quantized outputs coming from inputs with finite dynamic range —- for example, a quantized AWGN channel with a hard input amplitude constraint.

[5]In which there was no disturbance and only uncertainty about the initial condition, the desired property was that the controlled state should go almost-surely to zero.

Furthermore, this is possible by using an observer consisting of a time-varying random scalar quantizer that samples the state every n time steps and outputs a random label for the bin index. This random label is chosen iid from the channel input alphabet $\mathcal{A}^n$ according to the distribution that maximizes the random coding error exponent at $\log_2 \lambda < R < C$. The controller is assumed to have access to the randomness used to choose the random bin labels.

B. Our model for vector-valued unstable linear systems

The vector-systems model follows (1) except that everything is vector-valued now:

$$\vec{X}_{t+1} = A\vec{X}_t + B_u\vec{U}_t + B_w\vec{W}_t, \quad t \geq 0 \tag{8}$$

where $\{\vec{X}_t\}$ is an $\mathbb{R}^n$-valued state process $\{\vec{U}_t\}$ is an $\mathbb{R}^{m_u}$-valued control process and $\{\vec{W}_t\}$ is a bounded noise/disturbance process taking values in $\mathbb{R}^{m_w}$ s.t. $\|\vec{W}_t\| \leq \frac{\Omega}{2}$ where we can use any finite-dimensional norm that we find convenient. For convenience, we also assume a known initial condition $\vec{X}_0 = \vec{0}$.

In addition, we may restrict the input to the observer/encoder to be a linear function of the state, rather than the state itself.

$$\vec{Y}_t = C_y\vec{X}_t \tag{9}$$

where $\vec{Y}_t$ is an m_y dimensional vector. The matrices A, B_u, B_w, C_y above are all of the appropriate dimensionality so that the equations (8) and (9) make sense.[6]

Three fundamental issues arise from possible mismatches in dimensionality.

- The control vector might be of lower dimensionality than the state.
- The observation vector available to the encoder might be of lower dimensionality than the state.

These first two can be addressed using the traditional linear systems machinery. We know that even without any communication constraint, for us to be able to stabilize the system in closed-loop, we generally require (A, B_u) to be a controllable[7] pair and (A, C_y) to be an observable[8] pair. Throughout this paper, we will not worry about the slight distinction between controllable and stabilizable here. The modes of the linear system that are already stable are not going to be causing us trouble. If any such uncontrollable or unobservable stable modes exist, let them.

In our context, there is another dimensionality mismatch which is not immediately apparent:

- The bounded disturbance vector might be of lower dimensionality than the state.

To deal with this, we will make an additional requirement that (A, B_w) is controllable when determining necessary conditions for stabilization. This is a reasonable assumption as it corresponds to requiring that the disturbance can persistently excite all the unstable modes of the system. Otherwise, in the context of the assumption that the initial conditions $\vec{X}_0 = \vec{0}$, those unstable modes may never get excited and might as well not exist. All of these issues are discussed in Section III, but we first consider the simplest possible case of a vector problem.

II. The real diagonal case

The simplest possible case is where the real A matrix is diagonal with eigenvalues λ_i for $i = 1 \ldots n$. The controllability and observability assumptions translate into B_u, B_w, C_y all being full-rank. As such,

[6]For example, when B_u is a single column and C_y is a single row, then the system with vector state is considered single-input single-output (SISO) since $U(t)$ is a scalar as is $Y(t)$. It is possible to model any autoregressive moving average (ARMA) scalar system using an appropriate (8) and (9).

[7]A pair of matrices (A, B) are controllable if the matrix $[B, AB, A^2B, \ldots, A^{n-1}B]$ is of full rank. This condition assures us that by appropriate choice of inputs, we can control the behavior of all the modes of the linear dynamical system.

[8]A pair of matrices (A, C) are observable if the matrix $[C, CA, CA^2, \ldots, CA^{n-1}]^T$ is of full rank. This condition assures us that by combining enough raw observations, we can see the behavior of all the modes of the linear dynamical system.

we can change variables slightly by substituting $\vec{U}_t = B_u^{-1}\vec{U}'_t$, $\vec{W}_t = B_w^{-1}\vec{W}'_t$, and using $X_t = C_y^{-1}Y_t$ directly as our observation. In this case the problem decomposes into L parallel scalar problems:

$$X_i(t+1) = \lambda_i X_i(t) + U'_i(t) + W'_i(t) \tag{10}$$

for $i = 1 \ldots n$ with the only connection between the different parallel problems i being the common restriction $\|B_w^{-1}\vec{W}'_t\| \leq \frac{\Omega}{2}$. However, since all finite-dimensional norms are equivalent, we are free to translate that restriction into a pair of restrictions:

- For the necessity part in which we need to generate $\vec{W}_t$ signals to carry data across the channel, we can use an inscribed hypercube within the B_w^{-1} mapped $\frac{\Omega}{2}$ sized ball to allow ourselves to pick appropriate disturbances to the parallel control systems without worrying about what that means for other ones. This can be done since $\exists \Omega'$ s.t. $\|\vec{W}'\|_\infty \leq \frac{\Omega'}{2}$ implies $\|B_w^{-1}\vec{W}'_t\| \leq \frac{\Omega}{2}$. So, for the necessity theorems, the scalar theorems apply directly with Ω' playing the part of the bound.
- For the sufficiency proofs, we require a bounded range that the disturbance can take on system i. Again, we can use the equivalency of the $\infty-$norm to all other finite-dimensional norms and inscribe the B_w^{-1} mapped $\frac{\Omega}{2}$ sized ball from the original norm inside a hypercube $\bar{\Omega}$ on each side.

This same equivalency between norms tells us that if we have a finite η-moment of the $\infty-$norm of $\vec{X}$, then we also have a finite η-moment of any other norm and vice versa.[9] Furthermore, we have following simple lemma, proved in Appendix I-A, showing that if a finite collection of random variables all have finite η-moments, then they have a finite η-moment when collected together into a vector and vice versa.

Lemma 2.1: Let $|X_i|$ be positive random variables. There exists a K such that $E[|X_i|^\eta] \leq K$ for each i if and only if there exists a K' such that $E[\|\vec{X}\|^\eta] \leq K'$.

Furthermore, if $E[\|\vec{X}\|^\eta] \leq K$, then for any matrix L there exists K'' so that $E[\|L\vec{X}\|^\eta] \leq K''$.

Using Lemma 2.1 and the discussion above, we can immediately translate Theorem 1.4 on a component-wise basis. Before stating the appropriate corollary, we introduce the notion of an anytime rate region that generalizes the notion of a single anytime capacity to a rate region corresponding to a vector of anytime-reliabilities specifying how fast the different bitstreams have their probabilities of error tending to zero with delay.

Definition 2.1: The *anytime rate region* $\mathcal{R}_{\text{any}}(\vec{\alpha})$ of a channel is the set of rate-tuples which the channel can support so that each has a probability of error that asymptotically decays at least exponentially with delay at a rate α_i. There has to exist a uniform constant K such that for each i, all d and all times t we have

$$\mathcal{P}\left(\hat{M}_{i,1}^{t-d}(t) \neq M_{i,1}^{t-d}(t)\right) < K 2^{-\alpha_i d}$$

where the $M_{i,j}$ corresponds to the j-th message sent in the i-th bitstream.

The *θ-feedback anytime rate region* refers to the region when noiseless channel output feedback is available to the encoder with a delay of $\theta + 1$ time units. If θ is omitted, it is assumed to be zero.

The idea of the anytime rate region here is to transport parallel bitstreams across the noisy channel, giving each one a potentially different level of reliability.

Corollary 2.1: For a given noisy channel, diagonal A with diagonal elements $\lambda_i > 1$, bound Ω, and $\eta > 0$, if there exists an observer $\mathcal{O}$ and controller $\mathcal{C}$ for the unstable vector system that achieves $E[\|\vec{X}_t\|^\eta] < K$ for all sequences of bounded driving noise $\|\vec{W}_t\| \leq \frac{\Omega}{2}$, then for every $\vec{\epsilon} > 0$ we know that $(\log_2 \vec{\lambda} - \vec{\epsilon}) \in \mathcal{R}_{\text{any}}(\eta \log_2 \vec{\lambda})$ for the noisy channel considered with the encoder having access to noiseless feedback.

The $\log_2 \vec{\lambda}$ is shorthand for the vector consisting of the base two logarithms of all the components of $\vec{\lambda}$. Similarly, we can carry over the sufficiency results for the real diagonal case to get the basic vector corollary to Theorem 1.5:

[9] Whatever constant factor that bounds the norm from the ∞-norm will just be raised to the η-th power. This will change the finite expectation by at most that constant factor and hence finite expectations will remain finite.

Corollary 2.2: It is possible to control an unstable vector process with diagonal controllable dynamics and unstable eigenvalues $\vec{\lambda}$ driven by a bounded disturbance over a noisy channel so that the η-moment of $\|\vec{X}_t\|$ stays finite for all time if the channel with noiseless feedback has $(\log_2(\vec{\lambda}) + \vec{\epsilon}) \in \mathcal{R}_{\text{any}}(\eta \log_2 \vec{\lambda} + \vec{\epsilon})$ for some $\vec{\epsilon} > 0$ and the observer is allowed to observe the channel outputs perfectly.

The same carrying over can be done for Theorems 1.6, 1.7, 1.8, 1.9, and 1.11, as well as Corollary 1.1.

In the vector context, we can also provide a looser, but more easily checked, sufficient condition that assumes we use a feedback anytime-code that only deals with a single bitstream:

Corollary 2.3: It is possible to control an unstable vector process with diagonal controllable dynamics and unstable eigenvalues $\vec{\lambda}$ driven by a bounded disturbance over a noisy channel so that the η-moment of $\|\vec{X}_t\|$ stays finite for all time if the channel with noiseless feedback has $C_{\text{any}}(\alpha) > \sum_i \log_2 |\lambda_i|$ for some $\alpha > \eta \max_i \log_2 |\lambda_i|$ and the observer is allowed to observe the channel outputs perfectly and has boundedly noisy observations of the complete vector plant state.

Proof: Just multiplex together various scalar-stabilization bitstreams and send them over the feedback-anytime code. Since the anytime reliability α considers delays measured in time-units and not bit-units, they all will experience the same reliability α. The active constraint then comes from the largest unstable eigenvalue for which $\alpha > \eta \max_i \log_2 |\lambda_i|$ is clearly sufficient. □ The detailed example worked out in

Section VI shows that Corollary 2.3 is not tight and that we really do need to consider the full anytime rate regions in general.

The situation with generalizing Theorem 1.10 and Corollary 1.2 is slightly more interesting. Since all we want is to have some η for which the system is stable, we do not need to differentiate the service among the dimensions. We can operate at a rate larger than the sum of the logs of the unstable eigenvalues and then translate the resulting anytime reliability into a particular bound η.

Corollary 2.4: For diagonal controllable dynamics, if the observer is allowed boundedly noisy access to the complete vector plant state, and the noisy channel is a DMC with Shannon capacity $C > \sum_i \log_2 |\lambda_i|$, then there exists an observer/controller pair that stabilizes the system in closed loop so that the η-moment of $\|\vec{X}_t\|$ stays finite for all time for some $\eta > 0$ as long as the observer/controller are allowed to share common randomness.

Furthermore, this is possible by using an observer consisting of a time-varying random vector quantizer that samples the vector state every n' time steps and outputs a random label for the bin index. This random label is chosen iid from the $\mathcal{A}^{n'}$ according to the distribution that maximizes the random coding error exponent at $\sum_i \max(0, \log_2 |\lambda_i|) < R < C$. The controller is assumed to have access to the randomness used to choose the random bin labels.

Proof: The simple approach is to just apply Theorem 1.10 to each of the unstable plant states and then just multiplex the resulting sequence channel inputs across the noisy channel. The decoding and controller actions could proceed on a sub-system by sub-system basis and everything would work. However, this does not take advantage of the parallel channel coding advantage[10] and thereby gives up a lot of reliability in the process. In Appendix I-A, we give a better scheme. □

III. DIAGONAL JORDAN FORMS

We next show how we can extend the results for real diagonal A matrices to the case of linear systems that have real A matrices that have diagonal Jordan forms — i.e. those that have a full complement of eigenvectors. This represents a generic A matrix since the matrices that do not have a full complement of eigenvectors form a measure-zero set. The goal here is to show that the magnitudes of the unstable eigenvalues are all that matter.

The key idea is to use a coordinate transformation to diagonalize the dynamics of the system. This almost reduces the problem to the diagonal case of the previous section, except that we still face the potential dimensionality mismatch between $\vec{X}$ and $\vec{W}$. However, by examining time in blocks of at most n at a time, we can use the controllability of (A, B_w) to apply any desired bounded disturbance input that we want, fully recovering the diagonal case. This gives us:

Theorem 3.1: Assume that for a given noisy channel, system dynamics described by (8) with diagonalizable A and eigenvalues λ_i, and $\eta > 0$, that there exists an observer $\mathcal{O}$ and controller $\mathcal{C}$ for the unstable vector system that achieves $E[\|\vec{X}_t\|^\eta] < K$ for all sequences of bounded driving noise $\|\vec{W}_t\| \leq \frac{\Omega}{2}$. Furthermore, assume that the pair (A, B_w) is controllable.

Let $|\lambda_i| > 1$ for $i = 1 \ldots l$, and let $\vec{\lambda}$ be the l-dimensional vector consisting of only the exponentially unstable eigenvalues[10] of A. Then for every $\vec{\epsilon} > 0$ we know that $(\log_2 \vec{\lambda}_{||} - \vec{\epsilon}) \in \mathcal{R}_{\text{any}}(\eta \log_2 \vec{\lambda}_{||})$ for the noisy channel considered with the encoder having access to noiseless feedback. $\vec{\lambda}_{||}$ is our shorthand notation for the vector whose components are the magnitudes of the unstable eigenvalues.

The same ideas can be used to extend the sufficiency result from Corollary 2.2. Once again, a change of coordinates will diagonalize the system dynamics. In addition, we first examine time in groups of n and use the observability of (A, C_y) to noisily-estimate $\vec{X}$ from n consecutive observations of $\vec{Y}$. During this period, no controls are applied. In the next period of n times, we use the controllability of (A, B_u) to apply the desired control signal. This gives us:

Theorem 3.2: Assume that we have a noisy channel with a feedback anytime rate region for some $\vec{\epsilon} > 0$ that has $(\log_2(\vec{\lambda}_{||}) + \vec{\epsilon}) \in \mathcal{R}_{\text{any}}(\eta \log_2 \vec{\lambda}_{||} + \vec{\epsilon})$ where $\vec{\lambda}_{||}$ consists of the component-wise magnitudes of the $\vec{\lambda}$.

Then we can stabilize the linear system with dynamics described by (8) with diagonalizable A, unstable eigenvalues $\vec{\lambda}$, controllable (A, B_u), observable (A, C_y), bound Ω, by constructing an observer $\mathcal{O}$ and controller $\mathcal{C}$ for the unstable vector system that achieves $E[\|\vec{X}_t\|^\eta] < K$ for all sequences of bounded driving noise $\|\vec{W}_t\| \leq \frac{\Omega}{2}$ if the observer has perfect access to the channel outputs.

It is easy to see that the exact same arguments will work to generalize most of the sufficiency results from the scalar and diagonal case to the generic vector case. However, there is an interesting aspect to the generic generalization of Theorem 1.10 and Corollary 2.4. In the scalar and diagonal system case, it was possible to stabilize the unstable system for some $\eta > 0$ as long as $C_{\text{Shannon}} > \sum_i \log_2 \max(0, |\lambda_i|)$ by using nearly memoryless observers that sampled the state $\vec{X}$ every n' time steps. For generic vector systems, a single sample of the observation $\vec{Y}$ is not enough to tell us the current state of the system. Instead, we modify our sense of nearly memoryless encoders to instead take n consecutive samples of the output $\vec{Y}$, apply a linear transformation to them to recover an estimate for the current state $\vec{X}$, and then randomly vector quantize the result into $n' > 2n$ randomly chosen channel inputs. This allows us to get:

Corollary 3.1: For the linear system with dynamics described by (8) with diagonalizable A, unstable eigenvalues $\vec{\lambda}$, controllable (A, B_u), observable (A, C_y), and bounded driving disturbance $\vec{W}$, if the noisy channel is a DMC with Shannon capacity $C > \sum_i \log_2 |\lambda_i|$, then there exists an observer/controller pair that stabilizes the system in closed loop so that the η-moment of $\|\vec{X}_t\|$ stays finite for all time for some $\eta > 0$ as long as the observer/controller are allowed to share common randomness.

Furthermore, this is possible by using an observer consisting of a time-varying random vector quantizer that samples the observation $\vec{Y}$ in n consecutive time positions every $n' > 2n$ time steps, applies a possibly time-varying linear transformation to those n samples, and quantizes the result by outputting a random label for the bin index. This random label is chosen iid from the $\mathcal{A}^{n'}$ according to the distribution that maximizes the random coding error exponent at $\sum_i \max(0, \log_2 |\lambda_i|) < R < C$. The controller is assumed to have access to the randomness used to choose the random bin labels and only applies a nonzero control signal at times disjoint from the times during which output samples are used by the observer.

The case of using indirect feedback through the plant to compensate for having no direct channel feedback will be discussed in detail in the next section.

[10]Eigenvalues with multiplicity should appear multiple times in $\vec{\lambda}$

Finally, we state the obvious generalization of Lemma 1.1 to the vector case.

Lemma 3.1: If it is possible to η'-stabilize a persistently disturbed system from (8) and (9) with dynamics given by $(1+\xi)A$ when driven by any driving noise $\vec{W}'$ bounded by Ω, then there exists a time-varying observer with noiseless access to the observation $\vec{Y}$ and a time-varying controller so that any undisturbed system (8) with initial condition $|\vec{X}_0| \leq \frac{\Omega}{2}$, $\vec{W}_t = 0$, and dynamics A can be stabilized in the sense that there exists a K so that:

$$E[|\vec{X}_t|^{\eta'}] \leq K(\frac{1}{1+\xi})^{\eta' t} \tag{11}$$

Proof: All that is required is to scale up the observations of the undisturbed system with dynamics A by $(1+\xi)^t$ before feeding them to the observer and then scaling down the controls by $(1+\xi)^{-t}$ before applying them to the undisturbed system. Since the scaled system behaves like a system with dynamics $(1+\xi)A$ and is η'-stable, the actual undisturbed system will satisfy (11). □

Lemma 3.1 immediately gives almost-sure stabilization results in the style of Corollary 1.3 for vector cases.[11]

IV. Relaxing feedback to observers and intrinsic delay

By the strategy of the previous section of looking at the system in blocks of $2n$ — with controls being applied in the second half of the block while the first half of the block is used for state estimation — the problem of diagonalizable Jordan blocks reduces entirely to n parallel scalar problems, each one observed with noisy observations of the exact state. As such, all of the scalar results on sufficiency generalize to the vector case, including those that have no explicit access to the channel feedback at the observer.

However, the reduction to n scalar problems comes with a delay of as much as $2n$ in communication back from the controller to the observer through the plant. Since the channel feedback will also be delayed by $2n$, the sense of anytime reliability required for sufficiency is much stronger than what is required in Theorem 3.2. In this section, we give an example that shows that some added delay can be unavoidable and then define what the intrinsic delay is for a linear system. We then use this concept to give more refined necessity and sufficiency theorems that are tight and have additional delays in feedback.

Example 4.1: Consider the single input single output (SISO) system described by:

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

$$B_u = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$C_y = [1, 0, 0]$$

It controllable since $[B, AB_u, A^2B_u]$ is full rank and observable since (C, CA, CA^2) also span three dimensions. Since the eigenvalues $2, 3, 4$ are all distinct, it is certainly diagonalizable as well.

For such a system, initialized to 0, the impulse response is $0, 0, 1, 9, \ldots$. The input is not immediately visible at the output and takes 3 total time steps to show up at all. There is no possible way to communicate through the plant with only unit delay.

Example 4.1 motivates the following definition:

[11] In particular, Lemma 3.1 shows that a disturbance-free system can be almost-sure stabilized given a DMC with capacity strictly greater than the sum of the logs of the unstable eigenvalues. This is because an appropriately small $\xi > 0$ can be found so that $n \log_2(1+\xi) + \sum_i \log_2 |\lambda_i| < C$. Thus, Corollary 3.1 will be satisfied and there will exist some η for which the system with $(1+\xi)A$ dynamics can be η-stabilized. Lemma 3.1 tells us that the original system thus has an exponentially decaying η-th moment, and thus a finite sum of η-th moments through time. The only way that can happen is if the state is going to zero almost-surely.

Definition 4.1: The *intrinsic delay* $\Theta(A, B_u, C_y)$ of a linear system is the amount of time it takes the input to become visible at the output. It is the minimum integer[12] $i \geq 0$ for which $C_y A^i B_u \neq 0$.

For SISO systems, this is just the position of the first nonzero entry in the impulse response.

A. Refined necessity theorem

This definition of intrinsic delay allows us to state a more refined version of Theorem 3.1:

Theorem 4.2: Assume that for a given noisy channel, system dynamics described by (8) with diagonalizable A and eigenvalues λ_i, and $\eta > 0$, that there exists an observer $\mathcal{O}$ (without access to the control signals or channel outputs) and a controller $\mathcal{C}$ for the unstable vector system that achieves $E[\|\vec{X}_t\|^\eta] < K$ for all sequences of bounded driving noise $\|\vec{W}_t\| \leq \frac{\Omega}{2}$. Furthermore, assume that the pair (A, B_w) is controllable.

Then for every $\vec{\epsilon} > 0$ we know that $(\log_2 \vec{\lambda}_{||} - \vec{\epsilon}) \in \mathcal{R}_{\text{any}}(\eta \log_2 \vec{\lambda}_{||})$ for the noisy channel considered with the encoder having access to noiseless feedback delayed by $1 + \Theta(A, B_u, C_y)$.

Proof: In order to generate the next simulated $\vec{Y}_{t+1}$ to feed to the observer $\mathcal{O}$, we only need the exact $\vec{U}$ values upto time $t - \Theta(A, B_u, C_y)$ since controls after that point have not become visible yet at the output. Thus, we can tolerate noiseless feedback that has $1+\Theta(A, B_u, C_y)$ delay rather than the unit delay assumed earlier in running the simulated control system at the anytime encoder. □

B. Refined sufficiency theorem

To close the gap entirely, we need to show how to communicate the exact channel output back to observer through the plant using only θ time steps. The key idea is illustrated in figure 3. As in the scalar case, the main applied control action will depend only on delayed channel outputs. Assuming that the observer has already recovered those channel outputs, it can remove their effect from its observations. The controller then applies an additional control input whose only purpose is to cause a movement in the observed $\vec{Y}$ that is too big to have been caused by the bounded disturbance $\vec{W}$. The observer sees these movements with a delay of $1 + \Theta$ and can thus recover the finite channel outputs with that delay. This enables the observer to operate the θ-feedback anytime code and gives us:

Theorem 4.3: Assume that we have a noisy finite-output-alphabet channel such that with access to noiseless feedback delayed by $1 + \theta$ time units, it has a θ-anytime rate region for some $\vec{\epsilon} > 0$ that has $(\log_2(\vec{\lambda}_{||}) + \vec{\epsilon}) \in \mathcal{R}_{\text{any}}(\eta \log_2 \vec{\lambda}_{||} + \vec{\epsilon})$ where $\vec{\lambda}_{||}$ consists of the component-wise magnitudes of the $\vec{\lambda}$.

Then we can η-stabilize the linear system with dynamics described by (8) with diagonalizable A, unstable eigenvalues $\vec{\lambda}$, controllable (A, B_u), observable (A, C_y), intrinsic delay $\Theta(A, B_u, C_y) = \theta$, by constructing an observer $\mathcal{O}$ and controller $\mathcal{C}$ for the unstable vector system that achieves $E[\|\vec{X}_t\|^\eta] < K$ for all sequences of bounded driving noise $\|\vec{W}_t\| \leq \frac{\Omega}{2}$ if the observer has access to the observations $\vec{Y}_t$ corrupted by bounded noise.

C. Comments

Theorems 4.2 and 4.3 show that in the vector case, there can be something fundamentally harder about stabilizing an unstable process when explicit noiseless unit-delay feedback of the channel outputs is not available. If the input-output behavior of the system has an intrinsic delay[13] associated with it, then we

[12] Technically speaking, this should be a theorem rather than part of the definition, but since it is an obvious consequence of linearity, we just put it in the definition itself for convenience.

[13] It is important to realize that it is only the intrinsic delay which introduces an additional reliability requirement. In particular, non-minimum phase zeros (zeros of the transfer function outside the unit circle) do not cause any fundamental challenge unless they are at ∞ and correspond to an intrinsic delay. This is in contrast to the results based on purely linear robust control approaches, which do encounter difficulties due to non-minimum phase zeros.[11]

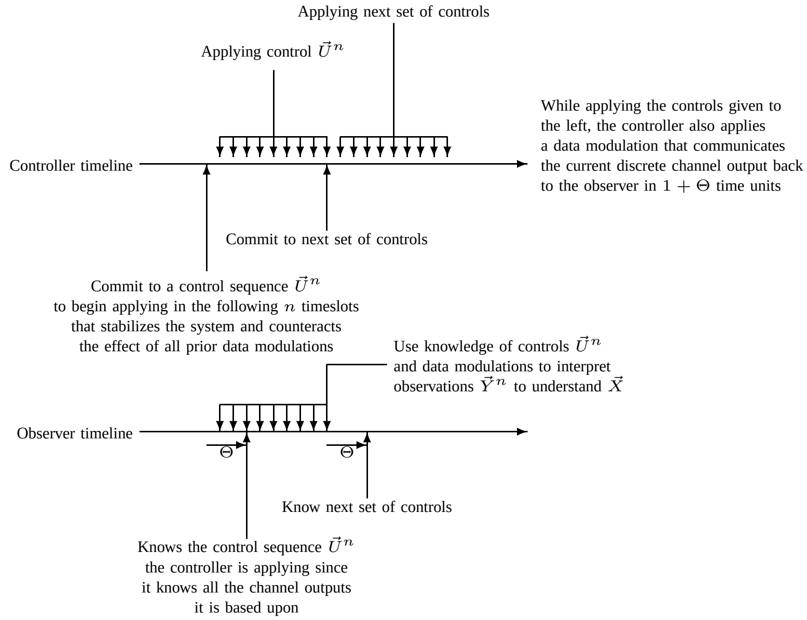

Fig. 3. When viewing time in blocks of n, we require the controller to commit to its primary controls 1 time step before actually putting them into effect. This way, by the time the observer can first see the effect of the control strategy, it already knows exactly what it is going to be since it knows all the channel outputs that it was based upon.

require that the noisy channel support enough anytime-reliability at the target rates even with feedback delayed by that intrinsic amount. Knowing the unstable eigenvalues and target moment η is not enough to evaluate a noisy channel.

Unfortunately, we know of no existing bounds on how much the anytime reliability is reduced by increasing the delay in the noiseless feedback path. However, it may be possible to derive such bounds by looking at the appropriate control problem. The impact of delay in the feedback path can be studied by constructing control systems with the appropriate intrinsic delay.

V. Non-diagonal Jordan blocks

The only remaining challenge concerns non-diagonal Jordan blocks. These arise for certain non-generic linear system structures and interfere with the straightforward diagonalization based arguments used earlier. Controllability and observability will give us the ability to independently control all the dimensions as well as observe all the state components with bounded noise. However, the diagonalization arguments will only be enough to separate the problem into ones involving parallel real-valued positive Jordan blocks. To address this challenge, it suffices to consider an n-dimensional square A matrix that represents a single real Jordon block.

$$A = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}$$

There are two key observations. The first is that the dynamics for $X_n(t)$ are the same as for the scalar case — $X_n(t+1) = \lambda X_n(t) + W_n(t) + U_n(t)$. The second is that the dynamics for all the other components are given by:

$$X_i(t+1) = \lambda X_i(t) + X_{i+1}(t) + W_i(t) + U_i(t) \tag{12}$$

For sufficiency with explicit or implicit feedback, recall that the sufficiency constructions were based on having a virtual controlled process that was assumed to be stabilized over a finite rate noiseless channel in a manner that held the virtual state to within a box of size Δ.[1] For the case of nondiagonal Jordan blocks, we group together $(X_{i+1}(t) + W_i(t))$ into a single disturbance term that remains bounded since both of its terms are bounded. Induction reduces the problem to a sequence of scalar problems and we have:

Theorem 5.1: Assume that we have a noisy channel with access to noiseless feedback with an anytime rate region for some $\vec{\epsilon} > 0$ that has $(\log_2(\vec{\lambda}_{||}) + \vec{\epsilon}) \in \mathcal{R}_{\text{any}}(\eta \log_2 \vec{\lambda}_{||} + \vec{\epsilon})$ where $\vec{\lambda}_{||}$ consists of the component-wise magnitudes of the λ_i.

Then we can stabilize the linear system with dynamics described by (8) with unstable eigenvalues λ_i, controllable (A, B_u), observable (A, C_y), by constructing an observer $\mathcal{O}$ and controller $\mathcal{C}$ for the unstable vector system that achieves $E[\|\vec{X}_t\|^\eta] < K$ for all sequences of bounded driving noise $\|\vec{W}_t\| \leq \frac{\Omega}{2}$.

It is clear that the exact same ideas can be used to generalize the finite-output-alphabet case of Theorem 4.3 as well. The necessity part is where there is a true question. Does each unstable eigenvalue add rate or is it only at the eigenvector level? The following theorem answers the question.

Theorem 5.2: Assume that for a given noisy channel, system dynamics described by (8) with eigenvalues λ_i and $\eta > 0$, that there exists an observer $\mathcal{O}$ and controller $\mathcal{C}$ for the unstable vector system that achieves $E[\|\vec{X}_t\|^\eta] < K$ for all sequences of bounded driving noise $\|\vec{W}_t\| \leq \frac{\Omega}{2}$.

Let $|\lambda_i| > 1$ for $i = 1 \ldots l$, and let $\vec{\lambda}$ be the l-dimensional vector consisting of only the exponentially unstable eigenvalues[14] of A. Then for every $\vec{\epsilon_1}, \vec{\epsilon_2} > 0$ we know that $(\log_2 \vec{\lambda}_{||} - \vec{\epsilon_1}) \in \mathcal{R}_{\text{any}}(\eta \log_2 \vec{\lambda}_{||} - \vec{\epsilon_2})$ for the noisy channel considered with the encoder having access to noiseless feedback delayed by $1 + \Theta(A, B_u, C_y)$.

This theorem tells us that each unstable eigenvalue, no matter whether it has its own eigenvector or not, induces a demand to reliably transport a bitstream. Here, we just describe the encoding and decoding algorithms. Since the only new feature is the nondiagonal Jordan block, we concentrate just on that one block. The remaining proof ideas not presented here are in Appendix I-D.

We encode n parallel bitstreams at rates $\log_2 \lambda > R = R_1 = R_2 = \cdots = R_n$ using the same bit-encoder as was used previously. At the decoder, notice that the last state is as before and only depends on its own bitstream. However, all the other states have a mixture of bitstreams inside of them since the later states enter as inputs into the earlier states. As a result, the decoding algorithm given in Section III.B.2 of [1] will not work on those states without modification.

The decoding strategy in the non-diagonal Jordon block case[15] will change to be successive-decoding in the style of decoding for the degraded broadcast channel.[12] Explicitly, the procedure goes as follows:

1) Set $i = n$. Set $D_j(t) = -\tilde{X}_j(t)$ for all j where $\tilde{X}_j$ represents the j-th component of the system in transformed coordinates driven only by the control inputs $\vec{U}'$, not the disturbances $\vec{W}'$. This is what is available at the decoder.
2) Decode the bits on the ith stream using the algorithm of Section III.B.2 of [1] applied to $D_i(t)$.
3) Subtract the impact of these decoded bits from the kth value of $D_k(t)$ for every $k < i$.
4) Decrement i and goto step 2.

Notice that if all the bits decoded up to this point are correct, then at the time we come to decode the bits on the ith stream using $D_k(t)$ as the input to the bit-extraction algorithm of Section III.B.2 of [1], the $D_k(t)$ will contain exactly what it would have contained had the A matrix been diagonal. As such, the error probability calculations done earlier would apply. However, this successive decoding strategy has the possibility of propagating errors between streams and so we need to get a better handle on how this error propagation can occur.

Lemma 5.1: Consider a Jordan block corresponding to λ. For every $\epsilon' > 0$, there exists a $K' > 0$ so that the maximum magnitude deviation of D_j due directly to decoding errors in stream $i > j$ occurring for bits corresponding to times after $t - d$ is $K' 2^{d(1+\epsilon')\log_2 \lambda}$

[14] If an eigenvalue has multiplicity, then it should appear in $\vec{\lambda}$ multiple times.

[15] Again, notice that all we use in this argument is that we have an upper-triangular block.

Lemma 5.1 shows that the error propagation can only cause a total deviation a little larger than λ^d. Since that is comparable to the error propagating up from our own stream's noise, the theorem holds.

VI. Differentiated Service Example

In this section, we give a very simple numeric example of a vector valued unstable plant with diagonal A matrix given by

$$A = \begin{bmatrix} 1.258 & 0 & 0 & 0 & 0 \\ 0 & 1.059 & 0 & 0 & 0 \\ 0 & 0 & 1.0585 & 0 & 0 \\ 0 & 0 & 0 & 1.058 & 0 \\ 0 & 0 & 0 & 0 & 1.0575 \end{bmatrix} \tag{13}$$

where the observer has noiseless access to both the state $\vec{X}$ and the applied control signals $\vec{U}$. The controller can apply any 5-dimensional input that it wishes. Assume that the bounded disturbance $\vec{W}$ is restricted to satisfy $\|\vec{W}_t\|_\infty \leq \frac{1}{2}$ for all times t. It is easy to see that this example essentially consists of five independent scalar systems that must share a single communication channel.

First, in Section VI-A, we show that it is possible to hold this system's state within a finite box over a noiseless channel using total rate $R = \frac{2}{3}$ consisting of one bitstream at rate $\frac{1}{3}$, and four other bitstreams, each of rate $\frac{1}{12}$. In Section VI-B, we consider a particular binary erasure channel and show that if we use it without distinguishing between the bitstreams, then we will not get mean-squared stability. In Section VI-C, we show how a simple priority based system can distinguish between the bitstreams in such a way as to allow us to get mean-squared stability while essentially using the observer/controller originally designed for the noiseless link. Finally, in Section VI-D, we discuss how this diagonal example can be transformed into a single-input single-output control problem that suffers from the same essential limitations.

A. Design for a noiseless channel

The system defined by (13) is essentially five independent systems and so falls under the treatment given in Section II. We notice that

$$\begin{aligned} \frac{1}{3} &> \log_2(1.258) \\ \frac{1}{12} &> \log_2(1.059) \end{aligned}$$

And so by Corollary 2.2, it is sufficient to use five parallel bitstreams of rates $R_1 = \frac{1}{3}$ (for the first sub-system) and $R_2 = R_3 = R_4 = R_5 = \frac{1}{12}$ (for the other four subsystems).

Assume that the observer chooses to observe the first component of the plant only every three time steps. It is immediately clear that this has dynamics given by:

$$X_1(t+3) = (1.258)^3 X_1(t) + \sum_{k=0}^{2} (1.258)^k W_1(t+2-k) + \sum_{k=0}^{2} (1.258)^k U_1(t+2-k) \tag{14}$$

The $U_1(t)$ are known at both observer and controller and so all that matters is that the noise terms behave as though $\Omega_1 = 4.573429512$. Thus from [1], we know that by transmitting 1 bit every 3 time units, we can keep the uncertainty for the first part of the state within a box of width $\Delta_1 = \frac{4.573429512}{1-1.9908655122^{-1}} \approx 1002$. The controller for the first component also only needs to act at times that are divisible by three and applies a zero control at all other times.

Applying the same argument to the other four streams, and looking at time in multiples of twelve, we get for $i = 2$:

$$X_2(t+12) = (1.059)^{12} X_2(t) + \sum_{k=0}^{11} (1.059)^k W_2(t+11-k) + \sum_{k=0}^{11} (1.059)^k U_2(t+11-k) \tag{15}$$

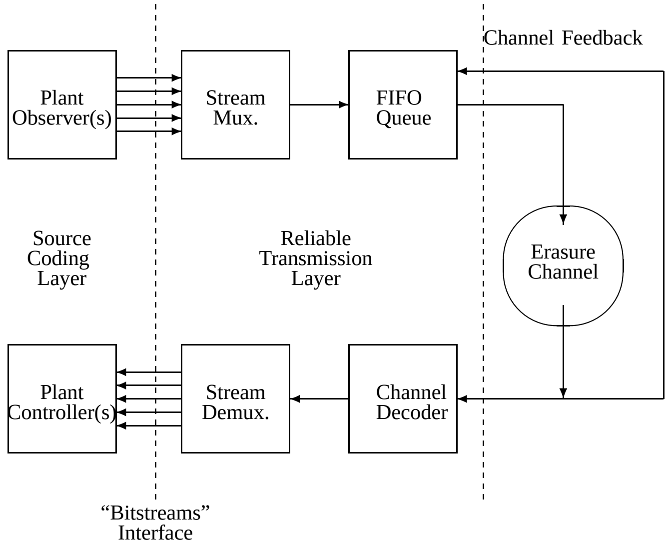

Fig. 4. Forcing all the bitstreams to get the same treatment for reliable transmission

and so effectively $\Omega_2 \approx 16.772$. Calculating Δ_2 gives ≈ 3205. Repeating for the others gives: $\Delta_3 = 1540.70, \Delta_4 = 1013.85, \Delta_5 = 755.38$. Once again, the controller for each of these four sub-systems only needs to act every twelve time steps.

Finally, over a noiseless $R = \frac{2}{3}$ channel, we need to specify the order in which the five-bitstreams are multiplexed together into a single bitstream. Pretty much anything reasonable will work, so we will specify the following repeating sequence $1, 2, 1, 3, 1, 4, 1, 5, 1, 2, \ldots$ for the bits, or with respect to time: $1, 2, 0, 1, 3, 0, 1, 4, 0, 1, 5, 0, 1, 2, 0 \ldots$ with the 0 representing the $\frac{1}{3}$ fraction of the time when we are not allowed to send any bits for a noiseless rate $\frac{2}{3}$ channel.

B. *Treating all bits alike*

Now, we will attempt to connect the observers and controllers from the previous section over a binary erasure channel with erasure probability $\delta = 0.27$ and noiseless feedback available to the encoder. Following the layering design paradigm, we want to use essentially the same observer/controller pair as before, making only the minimal set of changes required. There is clearly enough Shannon capacity since $1 - 0.27 = 0.73 > \frac{2}{3}$. To minimize latency per-bit, the natural choice of coding scheme is a FIFO queue in which bits are retransmitted until they get through correctly.

In this section, we will not discriminate between the various bits coming out of the observer and so there is only a single queue. When the queue is empty, we can just send a dummy bit across the channel. If that dummy bit is not erased, the receiver knows to ignore it since it knows that it is already caught up with the bitstream. Since the bits from the five bitstreams were multiplexed together in a deterministic fashion, and the FIFO queue does not change the order of bits, the receiver has no difficulty in dispatching the bit to whichever controller is waiting for it. The picture is illustrated in figure 4.

The only substantial change now is that the controller for sub-system i might receive some of the bits in a delayed fashion. If the controller would have applied control $U_i(t_o)$ if a particular data bit had arrived on-time, then it will apply $\lambda_i^d U_i(t_o)$ if that bit is delayed by d time-steps. This way, the controller is able to compensate for the known evolution of the plant due to the unstable dynamics.

It is clear that the controlled sub-state $|X_i(t)|$ can become big only when the bits in the i-th bitstream are delayed. In particular, there exists a constant K_i so that $\mathcal{P}(|X_i(t)| > K_i \lambda_i^d)$ must be bounded above by the probability that the bits at the receiver at time t are delayed by d or more. However, since the queue is FIFO, and the incoming rate is a deterministic $\frac{2}{3}$, a delay of d requires the queue to contain $\frac{2}{3}d$ bits in it.

To understand the asymptotic probability of having $\frac{2}{3}d$ bits waiting in the queue, we will group channel uses into blocks of three so that the number of bits awaiting transmission at the end of a block is the Markov state for the queuing system. This gives us:

$$\begin{aligned} p_{0,0} &= 3\delta(1-\delta)^2 + (1-\delta)^3 \\ p_{i,i+2} &= \delta^3 \\ p_{i,i+1} &= 3\delta^2(1-\delta) \\ p_{i,i} &= 3\delta(1-\delta)^2 \\ p_{i+1,i} &= (1-\delta)^3 \end{aligned}$$

The steady state distribution $\{\pi_i\}$ for the states can be calculated by looking at the local balance equations for the Markov chain. Assume $i \geq 2$ and take the cut separating the states $i-1$ and lower from the states i and higher. There is only one flow going across the cut in the negative direction: $(1-\delta)^3\pi_i$. There are three flows going across the cut in the positive direction: $\delta^3\pi_{i-2} + \delta^3\pi_{i-1} + 3\delta^2(1-\delta)\pi_{i-1}$. Setting them equal gives us a recurrence relation for π_i:

$$\pi_i = \frac{\delta^3\pi_{i-2} + (3\delta^2 - 2\delta^3)\pi_{i-1}}{(1-\delta)^3} \tag{16}$$

By simple application of the quadratic formula, we then have $\pi_i \propto \left(\frac{3\delta-2\delta^2+\sqrt{4\delta-3\delta^2}}{2(1-\delta)^3}\right)^i \approx (0.55256)^i$ for large i. The governing geometric for delay thus has the base $\approx (0.55256)^{\frac{2}{3}} \approx 0.673369$. Meanwhile, the first subsystem will grow as $(1.258)^d$ and thus the second moment will grow like $((1.258)^2)^d = (1.582564)^d$. Since $1.582564 * 0.673369 \approx 1.07 > 1$, the second moment of $|X(t)|$ will diverge for the first subsystem and hence for the plant as a whole. This particular observer/controller pair can not be successfully connected across this particular noisy channel with feedback in an application-blind way.

It turns out that this limitation is not restricted to the rate $\frac{2}{3}$ observer we are using here. If we multiplexed all the bits together into a single bitstream and used a single channel-code that did not differentiate among the substreams, then that code would give the same anytime reliability to all the constituent bits. We know that the minimum anytime reliability required is $\alpha^* = 2\log_2 1.258 \approx 0.6623$. For the binary erasure channel, we have an upper-bound[16] on the feedback anytime-capacity which is given by:[16]

$$C_{\text{any}}(\alpha) \leq \frac{\alpha}{\alpha + \log_2(\frac{1-\delta}{1-\delta 2^\alpha})} \tag{17}$$

Plugging in $\delta = 0.27$ and $\alpha = 2\log_2 1.258$ into (17) tells us that the channel can only carry $\approx 0.65418 \leq \log[2](1.258) + \log[2](1.059) + \log[2](1.0585) + \log[2](1.058) + \log[2](1.0575) \approx 0.65785$ bits/use with the required reliability. Thus we can not simultaneously attain the required rate/reliability pair by using an erasure channel code that treats all bits alike.

C. Differentiated service

The main difficulty we encountered in the previous section was that the most challenging reliability requirement came from the largest eigenvalue, while the total rate required involves all the eigenvalues. In this section, we explore the idea of differentiated service at the reliable transmission layer as illustrated in figure 5. In the context of the erasure channel, we use a simple priority-based scheme illustrated in figure 6 that gives extra reliability to the bitstream corresponding to the first subsystem at the expense of lower reliability for the other streams.

The priority scheme has a channel encoder that works as follows:

[16] This bound is actually achievable. A heuristic treatment of this bound is given in [13] and repeated in [14]. A more formal treatment is given in [15] and will also be given in [16].

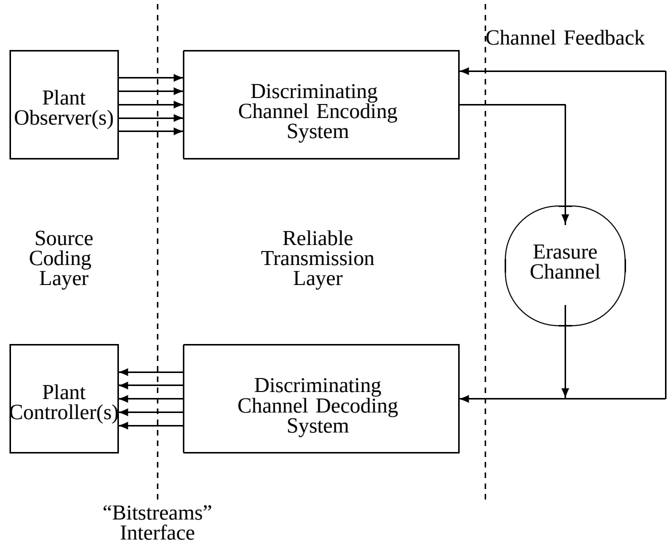

Fig. 5. Allowing the reliable transmission layer to discriminate between bitstreams

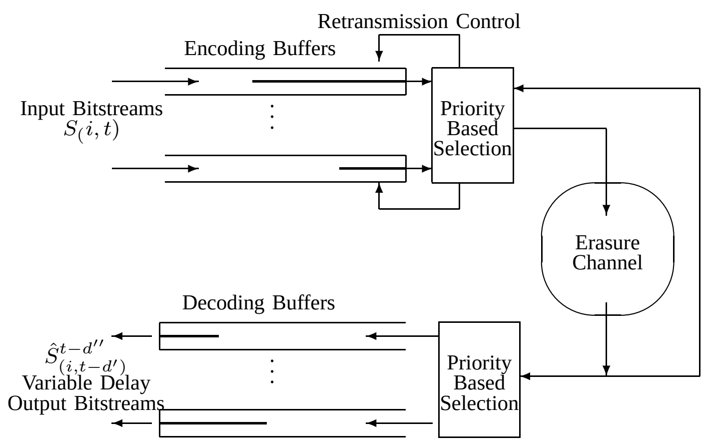

Fig. 6. The strict priority queuing strategy for discrimination between bitstreams. Lower priority buffers are served only if the higher priority ones are empty.

- Store the incoming bits from the different streams into prioritized FIFO buffers — one buffer for each distinct priority level.
- At every opportunity for channel use, transmit the oldest bit from the highest priority input buffer that is not empty.
- If the bit was received correctly, remove it from the appropriate input buffer.
- If there are no bits waiting in any buffer, then send a dummy bit across the channel.

We use exactly two priority levels. The higher one corresponds to the rate $R_1 = \frac{1}{3}$ bitstream coming from the first subsystem with eigenvalue 1.258. The lower one corresponds to the multiplexed stream at rate $R_2 + R_3 + R_4 + R_5 = \frac{1}{3}$ corresponding to the four subsystems with eigenvalues less than 1.059.

The decoder functions on a stream-by-stream basis. Since there is noiseless feedback and the encoder's incoming bitstreams are deterministic in their timing, the decoder can keep track of the encoder's buffer sizes. As a result, it knows which incoming bit belongs to which stream and can pass the received bit on to the appropriate subsystem's controller. The sub-system controllers are patched as in the previous section — they apply $\lambda_i^d U_i(t_o)$ if their bit arrives with a delay of d time-steps.

All that remains to do is to calculate the steady-state distribution of the delays by looking at the queue-length distributions and interpreting them appropriately.

1) The high priority stream: Since the highest priority stream preempts all lower priority streams, it effectively does not have to share the channel at all. As before, we get a simple Markov chain by grouping time into three time unit blocks. Then, the number of high-priority bits awaiting transmission at the end of a block is the Markov state and let $p_{i,j}$ represent the probability that the queue in state i will go next to state j.

$$\begin{aligned} p_{0,0} &= 3\delta^2(1-\delta) + 3\delta(1-\delta)^2 + (1-\delta)^3 \\ p_{i,i+1} &= \delta^3 \\ p_{i,i} &= 3\delta^2(1-\delta) \\ p_{i,i-1} &= \begin{cases} 3\delta(1-\delta)^2 + (1-\delta)^3 & \text{if } i = 1 \\ 3\delta(1-\delta)^2 & \text{if } i > 1 \end{cases} \\ p_{i,i-2} &= (1-\delta)^3 \end{aligned}$$

It is possible to calculate the steady state distribution π for this Markov chain. By examining a cut between states $i-2$ and state $i-$, we get the following local balance equation for $i \geq 2$

$$\delta^3 \pi_{i-2} = (1-\delta)^2(2\delta+1)\pi_{i-1} + (1-\delta)^3 \pi_i$$

which results in the following recurrence relation for π_i:

$$\pi_i = \frac{\delta^3 \pi_{i-2} - (1-\delta)^2(1+2\delta)\pi_{i-1}}{(1-\delta)^3} \tag{18}$$

Once again, a simple application of the quadratic formula gives us that $\pi_i \propto \left(\frac{2\delta^2-\delta-1+\sqrt{1+2\delta-3\delta^2}}{2(1-\delta)^2}\right)^i \approx (0.023718)^i$ for large i. The relevant rate here is just R_1 and so the governing geometric for delay has the base $\approx (0.023718)^{\frac{1}{3}} \approx 0.287314$. Recalling that the second moment of the first subsystem will grow with delay as $(1.582564)^d$ we check $1.582564 * 0.287314 = 0.454692 < 1$ and so the second moment of the first subsystem converges to a finite value.

2) The low priority streams: Rather than doing a similar calculation for the low priority streams, we just reuse a calculation already already performed in Section VI-B. We upper-bound the queue-length of the low-priority queue by the combined length of the two queues. This combined queue length behaves according to the Markov chain in Section VI-B since it behaves as a single rate $\frac{2}{3}$ stream entering a queue with geometric service time. Recall that this chain has a steady state distribution with $\pi_i \propto \left(\frac{3\delta-2\delta^2+\sqrt{4\delta-3\delta^2}}{2(1-\delta)^3}\right)^i \approx (0.55256)^i$ for large i. Therefore we have for the bits in the lower priority queue and large delays d:

$$\begin{aligned} P_{\text{error}}(\text{Delay} = d) &\leq P\left(\text{Combined Buffer State} > d(R_2+R_3+R_4+R_5)\right) \\ &\leq K\left(\frac{3\delta-2\delta^2+\sqrt{4\delta-3\delta^2}}{2(1-\delta)^3}\right)^{\frac{d}{3}} \end{aligned}$$

which for $\delta = 0.27$ results in a probability of delay that dies at least as fast[17] as $\approx (0.55256)^{\frac{1}{3}d} \approx (0.820591)^d$. Meanwhile, the second moment of the subsystems 2 through 4 can only grow at most as $(1.059)^{2d} \approx (1.121481)^d$. Since $1.121481 * 0.820591 \approx 0.92028 < 1$, it is clear that the second moments of all the subsystems converge and so the closed loop system is stable with differentiated service across the link.

[17]It actually dies faster than that since many of the bits in the combined queue belong to the higher priority bitstream.

D. Interpreting and extending the example

The diagonal system example given here is subject to two interpretations. First and most directly, it can be interpreted as a group of five applications each representing a physically distinct control system. For whatever reason, these distinct applications need to share a common bottleneck communication link. In that case, it represents an information-theoretic example of how different interactive applications sharing the same communication link can require differentiated service by the reliable communication layer even in the context of an asymptotic binary performance objective like stabilization.

Alternatively, this example can be interpreted as a single system with vector valued state. This vector-state valued system can be at the heart of even a single-input single output control system. Consider the transformation matrix[18]:

$$T = \begin{bmatrix} 2 & 0 & 2 & 1 & 2 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 1 & 0 \end{bmatrix} \tag{19}$$

We use this transformation to define $\tilde{A} = TAT^{-1}$ using (19) and (13). The B_w matrix remains the identity while B_u and C_y are given by:

$$B_u = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{20}$$

and

$$C_y = [1, 0, 0, 0, 1] \tag{21}$$

In words, we can only immediately control the first dimension of the vector state and our observations are limited to the sum of the first and last dimensions of the state. However, it is easy to verify that $[B_u, \tilde{A}B_u, \tilde{A}^2 B_u, \tilde{A}^3 B_u, \tilde{A}^4 B_u]$ is of full rank and thus the system is controllable. Similarly, the observability conditions are also satisfied. Thus this unstable scalar-output system with scalar-input falls within the parameters of Theorems 3.1 and 3.2 and the results of the previous few sections will apply to it. In order to stabilize its second moment across a binary erasure channel with probability of erasure $\delta = 0.27$ using a separation architecture[19], we require an observer that isolates the 1.258 eigenvalue and then assigns the corresponding bitstream a higher priority for transport across the noisy channel.

Our final comment on this example is regarding the choice of noisy channel. This example was crafted with the binary erasure channel in mind. Although we expect that similar examples exist for most nontrivial discrete memoryless channels, it is also important to point out that there are special channels for which such examples do not exist. In particular, the average power-constrained AWGN channel with feedback provides one such example. This is because we showed in [1] that the AWGN channel had a feedback anytime capacity equal to its Shannon capacity regardless of α. Examples illustrating the need for differentiated service only exist when we have a nontrivial tradeoff between rate and reliability.

Even so, the results in this paper are significant in the case of Gaussian channels. They effectively show that stabilization is possible over an adequate capacity AWGN channel with noiseless feedback even when there is a dimensionality mismatch between the channel and the plant. Prior results involving only linear

[18]This T matrix was generated randomly for this example.

[19]Without the requirement of separating reliable communication from the underlying application, it is hard to interpret priority in a meaningful way. For example, an observer could just operate by finding the direction in which current state uncertainty at the decoder is greatest and then use its bit to communicate which half the state lies in. The corresponding controller just keeps track of this same uncertainty and applies a control designed to bring the state uncertainty back to around the origin. Upon reflection, it becomes clear that this approach corresponds to reliable communication through a longest-queue-first transmission strategy that implicitly favors the higher-rate flows from a delay perspective.

control theoretic techniques could not reach the capacity bound for all cases in which the dimension of the plant was different than the dimension of the channel.[4]

VII. Conclusions

With Theorems 5.2 and 5.1, we know that the problem of stabilization of a linear vector plant over a noisy channel is intimately connected to the problem of reliable sequential communication over a noisy channel of parallel bitstreams in the anytime sense where the encoder has access to noiseless feedback of the channel outputs. The anytime-capacity region of a channel with feedback is the key to understanding whether or not it is possible to stabilize an unstable linear system over that noisy channel. The two problems are related through three parameters. The primary role is played by the magnitudes of the unstable eigenvalues since their logs determine the required rates. The target moment η multiplies these logs to give the required anytime reliabilities. Finally, the intrinsic delay $\Theta(A, B_u, C_y)$ tells us the noiseless feedback delay to use while evaluating the required anytime reliabilities when explicit channel feedback is not available.

It should be immediately clear that all the arguments given in [1] on continuous-time models also apply to vector-valued state spaces. Standard results on sampling linear systems tell us that in the continuous-time case, the role of the magnitude of the unstable eigenvalues will be played by the real part of the unstable eigenvalues. Similarly, all the results regarding the almost-sure sense of stabilization when there is no persistent disturbance also carry over directly.

In addition, because the results given here apply for general state-space models, we know that they apply to all equivalent linear models. In particular, they also apply to the case of control systems modeled using ARMA models or with rational open-loop transfer functions of any finite order. Assuming that there is no pole/zero cancellation, such results can be proven using standard linear techniques establishing the equivalence of such SISO models to the general state-space forms considered here. In those cases, the unstable eigenvalues of the state-space model will correspond to the unstable poles of the ARMA model. The intrinsic delay will correspond to the number of leading zeros in the impulse response, i.e. the multiplicity of the zero at $z = \infty$.

The primary limitation of the results so far is that they only cover the binary question of whether the plant is stabilizable in the $\eta-$moment sense or not. They do not address the issue of performance. In [17], we are able to cleanly approach the question of performance for the related scalar estimation problem using rate-distortion techniques. The linear systems techniques of this paper apply directly to the estimation problem there and can generalize the results naturally to the vector case.[20] For the estimation problem where the limit of large estimation delays does not inherently degrade performance, it turns out that we require l parallel bitstreams corresponding to each unstable eigenvalue, each of rate $> \log_2 |\lambda_i|$, together with one residual bitstream that is used to boost performance in the end-to-end distortion sense. The unstable streams all require anytime reliability in the sense of Theorem 5.2 while the residual stream just requires Shannon's traditional reliability. Since there are no control signals in the case of estimation, intrinsic delay plays no role there. In the control context, we already have results from sequential rate-distortion theory that give us a rate bound on the total bitrate required for a target level of performance.[4] However, there is no corresponding sufficiency result for control performance over general channels since there is no available slack-parameter that we can let get large in order to overcome possible mismatches between the channel and the system.

A second limitation of the results so far is that we have no good bounds on the anytime rate and reliability regions beyond the ones that we have already for the single-rate/reliability region.[16] However, even without such bounds, we have learned something nontrivial about the relative difficulty of different stabilization problems. For example, consider a scalar system with a single unstable eigenvalue of $\lambda = 8$ as compared to a vector-state system with three unstable eigenvalues, all of which are $\lambda_i = 2$. From

[20] In particular, it is straightforward to apply these techniques to completely solve all the nonstationary auto-regressive cases left open in [18].

a total rate perspective, the two appear identical requiring at least 3 bits per unit time. However, we can distinguish them based on what anytime-reliability they require for a given η-moment stability. The scalar case requires anytime-reliability $\alpha > 3\eta$ while the vector case can make do with any $\alpha > \eta$. Since the three eigenvalues are identical in the vector case, there is also no need to prioritize any one of them over the others and thus we can interpret the "vector-advantage" as being a factor reduction in the anytime-reliability required. Thus, in a very precise sense[21], we can say that vector-stabilization problems are easier than the scalar-stabilization problem having the same rate requirement.[22] It seems that spreading the potential growth of the process across many independent dimensions reduces the reliability requirements demanded from the noisy channel.

APPENDIX I
LONGER PROOFS

A. Proofs from Section II

Proof of Lemma 2.1: First we assume that $E[\|\vec{X}\|^\eta] \leq K'$. Then we know by the equivalencies of finite-dimensional norms that:

$$\begin{aligned} K' &\geq E[\|\vec{X}\|^\eta] \\ &\geq \kappa^\eta E[\|\vec{X}\|_\infty^\eta] \\ &= \kappa^\eta E[(\max_i |X_i|)^\eta] \\ &\geq \kappa^\eta E[|X_i|^\eta] \end{aligned}$$

So we have $E[|X_i|)^\eta] \leq \frac{K'}{\kappa^\eta}$. To see the other direction, first assume that $E[|X_i|^\eta] \leq K$.

$$\begin{aligned} E[\|\vec{X}\|^\eta] &\leq \kappa'^\eta E[\|\vec{X}\|_\infty^\eta] \\ &= \kappa'^\eta E[(\max_i |X_i|)^\eta] \\ &= \kappa'^\eta \int_0^\infty \mathcal{P}((\max_i |X_i|)^\eta \geq t) dt \\ &\leq \kappa'^\eta \int_0^\infty \sum_i \mathcal{P}(|X_i|)^\eta \geq t) dt \\ &= \kappa'^\eta \sum_i \int_0^\infty \mathcal{P}(|X_i|)^\eta \geq t) dt \\ &= \kappa'^\eta \sum_i E[|X_i|^\eta] \\ &\leq n\kappa'^\eta K \end{aligned}$$

For the second part, we use the ∞-norm again and given the previous part, we just need to consider a

[21] Using the ideas from Section VII of [1] to make it precise.

[22] This vector advantage in terms of required anytime reliability is even more surprising in light of the performance bounds in terms of rate only. [4] gives us explicit bounds on the squared-error performance using sequential distortion-rate theory. Suppose the $\lambda = 8$ scalar plant was driven by a standard iid Gaussian disturbance while the vector plant was diagonal and driven by three iid Gaussians each of variance $\frac{1}{3}$. For a given rate R (in bits), the sequential distortion-rate bound on $E[|X_t|^2]$ is $\frac{1}{1-4^{3-R}}$ for the scalar system while it is $\frac{1}{1-4^{1-\frac{R}{3}}}$ for the vector system. For a given rate, the second-moment performance of the vector system is *worse* than the scalar one. For example, at rate 4 the scalar one gets to ≈ 1.33 while the vector one is ≈ 2.70. At high rates, the two approach each other in terms of second-moment performance but the anytime-reliability requirements for the scalar system remain much higher.

single term $\sum_{i=1}^{n} l_i X_i$ and show that its η-moment is bounded.

$$\begin{aligned} E[|\sum_{i=1}^{n} l_i X_i|^\eta] &\leq E[(\sum_{i=1}^{n} |l_i X_i|)^\eta] \\ &\leq E[\left(\sum_{i=1}^{n} (\max_i |l_i|)(\max_i |X_i|)\right)^\eta] \\ &\leq (n\|\vec{l}\|_\infty)^\eta E[\|\vec{X}\|_\infty^\eta] \\ &\leq (n\|\vec{l}\|_\infty)^\eta K'' \end{aligned}$$

which proves the result. □

Improved Proof of Corollary 2.4: The idea is to follow the same strategy as the proof to Theorem 1.10 in [1], but naturally generalized to the vector case in a way that gives us the parallel channel coding advantage. We sketch only the modifications here since most of the details are essentially identical. Assume[23] all $|\lambda_i| \geq 1$.

- We partition the unstable sub-space into evenly-sized large boxes that are Δ on each side where Δ is suitably large.
- We augment the partition by adding an additional box centered on each of the vertexes of the original partition boxes. This adds at most 2^n additional boxes.
- Each box is given an iid sequence of n' channel inputs drawn from the relevant error-exponent maximizing distribution. These sequences are also iid across the different boxes[24] and at different time instants.
- Given suitably large Δ, in n' time units, the uncertainty in each box can grow to overlap with no more than $2^n|\det(A)|^{n'} = 2^n \prod_i |\lambda_i|^{n'}$ new boxes in the original augmented partition.
- Since the controller knows its past control signals, the channel outputs behave like they are coming from a random trellis code with rate $\frac{\log_2(2^n \prod_i |\lambda_i|^{n'})}{n'} = \frac{n}{n'} + \sum_i \log_2 |\lambda_i|$ per channel use which is less than C for large enough n'.

Everything else is as it was in the scalar case and this achieves an internal reliability of the random coding error exponent evaluated at the sum of the log of the unstable eigenvalues. To get the η-sense of stability that is achieved, just divide that exponent by the log of the largest unstable eigenvalue. □

B. Proofs from Section III

Proof of Theorem 3.1: We want to turn the problem into one for which Corollary 2.1, the necessity theorem for the case of diagonal real matrices, applies.

Recall that at the encoder side, the disturbance input $\vec{W}$ is used to embed our data bits into the uncontrolled state. Since (A, B_w) is a controllable pair, we know that within at most n time steps, it is possible to take the system starting at $\vec{X}_t = \vec{0}$ and drive it to any $\vec{X}_{t+n+1} = \vec{w}'$ through the choice of disturbance inputs $\vec{W}_t^{t+n}$. Since the system is linear, given a bound on $\|\vec{W}\|_\infty$, we can determine an Ω' so that as long as $\|\vec{w}'\|_\infty < \frac{\Omega'}{2}$ we can choose an acceptable sequence of n disturbance vectors $\vec{W}$ to reach it.

As such, we can use linearity and rewrite the system dynamics (8) using blocks of time of size n as:

$$\vec{X}_{nk} = A^n \vec{X}_{n(k-1)} + \vec{U}_k' + \vec{W}_k', \quad t \geq 0 \tag{22}$$

where $\vec{U}_k'$ just combines the net effect of the controls over the past n time-steps and $\vec{W}_k'$ is now an arbitrary n-vector disturbance with $\|\vec{W}_k'\|_\infty \leq \frac{\Omega'}{2}$.

[23]Basically, ignore the stable subspace since it is not going to be involved at all.

[24]This independence across all boxes is what is important to get the parallel channel advantage in reliability.

Since A is diagonalizable, we know that there exists a nonsingular matrix V so that $VAV^{-1} = \Lambda$, a diagonal matrix with the first l diagonal elements corresponding to the exponentially unstable eigenvalues. Thus $A^n = V^{-1}\Lambda^n V$. This handles all the real eigenvalues and it is well known that we can change coordinates using V to get diagonal dynamics.

To avoid any complications arising from the complex eigenvalues, we rely on the real Jordan normal form.[19]. This tells us that there exists a nonsingular real matrix V so that VAV^{-1} is a diagonal sum of either traditional real Jordan blocks corresponding to the real eigenvalues and special "rotating" Jordan blocks corresponding to each pair of complex-conjugate eigenvalues. The rotating block for the pair $\lambda = \lambda_r + \lambda_j\sqrt{-1}$ and its conjugate is a two-by-two matrix:

$$\begin{aligned} & \begin{bmatrix} \lambda_r & \lambda_j \\ -\lambda_j & \lambda_r \end{bmatrix} \\ = & \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos(\angle\lambda) & \sin(\angle\lambda) \\ -\sin(\angle\lambda) & \cos(\angle\lambda) \end{bmatrix} \end{aligned}$$

which is clearly a product of a scaling matrix and a rotation matrix. Group these two-by-two blocks into a block-diagonal unitary matrix R. So we are guaranteed to have $VAV^{-1} = \Lambda R$ where Λ is now the real diagonal matrix consisting of the magnitudes of the eigenvalues.

The key is to take the rotating parts and move them into a rotating coordinate frame which makes the system dynamics real and diagonal. Transform to $\vec{X}'_{kn} = (R^{-kn}V)\vec{X}_{kn}$ using $R^{-kn}V$ as the real time-varying coordinate transformation. Notice that

$$\begin{aligned} \vec{X}'_{k+1} &= R^{-(k+1)}V\vec{X}_{k+1} \\ &= R^{-(k+1)}VA\vec{X}_k + \cdots \\ &= R^{-(k+1)}VAV^{-1}R^k\vec{X}'_k + \cdots \\ &= R^{-(k+1)}\Lambda R R^k\vec{X}'_k + \cdots \\ &= R^{-(k+1)}R^{k+1}\Lambda\vec{X}'_k + \cdots \\ &= \Lambda\vec{X}'_k + \cdots \end{aligned}$$

since the Λ diagonal matrix commutes with the unitary block-diagonal matrix R. Thus, this time-varying change of coordinates transforms (22) into

$$\vec{X}'_{nk} = \Lambda^n\vec{X}'_{n(k-1)} + \vec{U}''_k + \vec{W}''_k, \quad t \geq 0 \tag{23}$$

where Λ^n is a real diagonal matrix with the ith entry being $|\lambda_i|^n$. The $\vec{U}''$ and $\vec{W}''$ are essentially the same as they were in (22) except possibly with a smaller constant bound $\Omega'' > 0$ for $\vec{W}''$ due to the rotating change of basis. Notice that the time-varying nature of the transformation is due to taking powers of the unitary matrix R and so the Euclidean norm is not time-varying.

Now the second part of Lemma 2.1 can deal with the change of coordinates and we can apply Corollary 2.1 to the unstable subspace of (23). We observe that $\log_2|\lambda_i|^n = n\log_2|\lambda_i|$ and by expressing both the delay in unit time rather than in steps of n and the bitrates per unit time rather than in steps of n, we have the desired result. □ *Proof of Theorem 3.2:* First,

we consider the case where we can observe $\vec{X}_t$ directly at the encoder. Using the same ideas as the proof of Theorem 3.1, we can setup a rotating coordinate frame in which the plant behaves as though it were real and diagonal. However, there is still the issue of choosing control signals. Once again, we can use controllability, this time of (A, B_u) and treat time in blocks of n at a time. Since there is no constraint on the $\vec{U}_t$, we know that the controller can act in a way that completely subtracts the effect of $\hat{X}_{t-n}$ from X_t. So we have reduced the problem to that addressed in Corollary 2.2.

Now, consider the case where $\vec{X}_t$ is not observed directly by the encoder and only $\vec{Y}_t$ is available. We use the observability of (A, C_y). This tells us that with the n observations[25] $(Y_{nk}^{n(k+1)-1}, \vec{U}_{nk}^{n(k+1)-1})$ we can determine $\vec{X}_{nk}$, except for the linear effect of the unknown disturbances $\vec{W}_{nk}^{n(k+1)-1}$. Since the disturbance is bounded, so is the effect of the linear combination of observations used to extract $\vec{X}_{nk}$. Consider this a bounded observation noise and use Corollary 1.1 as straightforwardly generalized to the diagonal case. □

Proof of Corollary 3.1: Since the corresponding control signals are zero and the system is observable, applying an appropriate linear transformation to the n consecutive samples $\vec{Y}_{kn'}^{kn'+n-1}$ gives the value of $\vec{X}_{kn'}$ plus a linear combination of the bounded disturbances $\vec{W}_{kn'}^{kn'+n-2}$. This is just a boundedly noisy estimate of $\vec{X}_{kn'}$. If we quantize $\vec{X}_{kn'} + \vec{W}'_{kn}$ to a large enough box Δ' on each side, then we know that there exists another large Δ sized box guaranteed to contain the diagonalized $\vec{X}'_{kn'}$ state obtained under the possibly rotating coordinate transformation that diagonalizes (8). Thus, the uncertainty propagates after n' time steps exactly as it did earlier. Furthermore, controllability of (A, B_u) guarantees that the controller can apply whatever correction that it wants to apply in n time steps. Since $n' > 2n$, it is possible to do this in a manner disjoint from the observations used by the observer to estimate the state. Everything else is as before. Thus, the proof for Corollary 2.4 generalizes to the case of generic A matrices given large enough n'. □

C. Proofs from Section IV

Proof of Theorem 4.3: We show how to patch together the proofs of Theorem 3.2 and Theorem 1.11 for this case.

In the proof of Theorem 3.2, the observer's knowledge of the channel outputs was used for two purposes. First, it was used in the operation of the feedback-anytime code. Second, it was used to compute the control signal so that the observer could take it into account while observing $\vec{X}_{nk}$. Here, we need the channel outputs with delay $1 + \theta$ for running the anytime code. Since $\theta \leq n$, by the time the observer has seen enough $\vec{Y}$ to observe $\vec{X}_{nk}$, it will also have access to the channel outputs that were used to compute the controls applied during that period. So, it can still subtract the effect of the controls while interpreting its observations.

To actually communicate the channel outputs from the controller to the observer, we follow the strategy from the proof of Theorem 1.11 with the following modifi cations:

- The observer will use some component $Y_i(t+1)$ of $\vec{Y}_{t+1}$ that is potentially impacted by $\vec{U}_{t-\theta}$ to read off the value of $B_{t-\theta}$.
- By the time $t+1$ (by induction), the observer knows exactly what the values were for all the $B_1^{t-\theta-1}$ and so it already knows all the control signals so far except that part of $\vec{U}_{t-\theta}$ that will be used to communicate $B_{t-\theta}$. By linearity, it can therefore remove their entire impact from the output $Y_i(t+1)$.
- The $\vec{Y}_1^{t-1}$ and known controls so far give an estimate of $\vec{X}_{t-n}$ that differs from the true $\vec{X}_{t-n}$ by no more than some bound Γ' in any component. Through the linear operations of the observer, this translates into some constant Γ'' of prior uncertainty on $Y_i(t+1)$ that comes from the disturbances.
- Γ'' can thus play the role of Ω in the lattice encoding Fig. 16 from [1]. The encoder can encode the value of $B_{t-\theta}$ by modulating this data on top of $\vec{U}_{t-\theta}$ so that it causes a shift of an integer multiple of Γ'' guaranteed to be visible in $Y_i(t+1)$ at the observer.
- At the time the controller commits to a primary control sequence $\vec{U}_{kn+1}^{kn+n}$ for the next n control signals, it can also clean up for all the past data modulations that it has done to communicate past channel outputs. This way, the closed-loop system will stay stable since the deviations in the controls used to communicate will not have an impact that is allowed to propagate unstably through the system.

[25]From the noiseless feedback of channel outputs, we can reconstruct the control signals as well.

Everything else is as before. □

D. Proofs from Section V

Proof of Theorem 5.1: This is only a minor variation on Theorem 3.2. The state components corresponding to diagonal Jordan blocks or stable dynamics are dealt with as before. For each non-diagonal Jordan block, the final state is also dealt with as before since it is effectively a scalar system. For the other states, we have to modify the construction in Section IV.B of [1] slightly.

To enable induction, assume that we have managed to stabilize the virtual processes corresponding to states $X_{i+1}, X_{i+2}, \ldots, X_n$ and can hold those virtual processes to within boxes of size $\Delta_{i+1}, \Delta_{i+2}, \ldots, \Delta_n$. Since the state update equation (12) for state $X_i(t+1)$ has state $X_{i+1}(t)$ in it as well[26] as $W_i(t)$, we just add the Δ_{i+1} to the Ω to get a new disturbance $W_i'(t)$ that is bounded by:

$$\Omega_i = \Omega + \Delta_{i+1}$$

Since that just makes the constant Ω effectively bigger, it does not change the asymptotic rate or the sense of reliability required. It just results in a larger Δ_i for any given rate R_i. By induction, since n is finite, there is a maximum such Δ over all the state components. Thus, all the sufficiency proofs from generic case extend to the nongeneric case. □

Proof of Lemma 5.1: Assume all the rates $R_i = R$ for simplicity. We first write the expression corresponding to equation (3) for the states $i < n$. By (4), we know $(2 + \epsilon_i) = \lambda^{\frac{1}{R_i}}$ and so we can write $\check{X}_i(t) =$

$$\begin{aligned} &\gamma\lambda^t [\left(\sum_{k=0}^{\lfloor Rt \rfloor} \lambda^{-\frac{k}{R}} S_i(k)\right) \\ + &\left(\sum_{k=0}^{\lfloor Rt \rfloor} \lambda^{-\frac{k}{R}} p_1(\lfloor Rt \rfloor - k) S_{i-1}(k)\right) \\ + &\cdots + \left(\sum_{k=0}^{\lfloor Rt \rfloor} \lambda^{-\frac{k}{R}} p_{n-i}(\lfloor Rt \rfloor - k) S_n(k)\right) \end{aligned} \tag{24}$$

where the p_k represent polynomials. The key feature of polynomials is that for every ϵ, we can choose a K_i so that $p_k(\tau) \leq K_i 2^{\epsilon\tau}$. We will bound the maximum deviation possible by considering the case in which we make an error on all the bits after a certain point $t - d$ since the worst case is when every bit that could be wrong is wrong.

[26] It is easy to see that this argument easily generalizes to any real-normal form that was purely upper triangular.

In that worst case, the magnitude of the deviation in D_j due directly to decoding errors is given by:

$$\begin{aligned}
& \gamma\lambda^t \sum_{k=\lceil R(t-d)\rceil}^{\lfloor Rt\rfloor} \lambda^{-\frac{k}{R}} p_{i-j}(\lfloor Rt\rfloor - k)2 \\
\leq \;& 2K\gamma\lambda^t \sum_{k=\lceil R(t-d)\rceil}^{\lfloor Rt\rfloor} \lambda^{-\frac{k}{R}} 2^{\epsilon(\lfloor Rt\rfloor - k)} \\
\leq \;& 2K\gamma 2^{(R\epsilon+\log_2\lambda)t} \sum_{k=\lceil R(t-d)\rceil}^{\lfloor Rt\rfloor} 2^{-k(\epsilon+\frac{\log_2\lambda}{R})} \\
\leq \;& 2K\gamma 2^{(R\epsilon+\log_2\lambda)t} 2^{-R(t-d)(\epsilon+\frac{\log_2\lambda}{R})} \sum_{k=0}^{\infty} 2^{-k(\epsilon+\frac{\log_2\lambda}{R})} \\
= \;& K' 2^{(Rt\epsilon+t\log_2\lambda)-(Rt-Rd)(\epsilon+\frac{\log_2\lambda}{R})} \\
= \;& K' 2^{d(\epsilon R+\log_2\lambda)} \\
= \;& K' 2^{d(1+\frac{\epsilon R}{\log_2\lambda})\log_2\lambda}
\end{aligned}$$

Since ϵ was arbitrary, we can achieve this with any $\epsilon' = \frac{\epsilon R}{\log_2\lambda}$. □

Proof of Theorem 5.2: It suffices to consider a single nondiagonal Jordan block since the time-varying coordinate transformation techniques of Theorem 3.1 will decompose any problem into a set of such blocks. Similarly, the simulated control system being run at the anytime encoder only needs the channel outputs with delay $1+\theta$ if the only feedback is through the plant.

The goal of the $\vec{\epsilon_2}$ is to allow us to give an slightly lower sense of reliability to the early streams within a Jordon block. From (6), we know how much of a deviation in $D_i(t)$ we can tolerate and still have no error in decoding bits from before d time steps ago. To get a bound on the probability of error, we will allocate that into $n-i+1$ equally-sized pieces. Each allocated margin will be of size $\frac{\gamma\epsilon_1}{(n-i+1)(1+\epsilon_1)} 2^{d\log_2\lambda}$. The first n of them will correspond to an allowance for propagating errors from later streams. The final piece will correspond to what we allow from the controlled state.

Using Lemma 5.1 and setting d' to the point in the past corresponding to the first error in the other stream, we can set the two terms equal to each other

$$\begin{aligned}
\frac{\gamma\epsilon_1}{(n-i+1)(1+\epsilon_1)} 2^{d\log_2\lambda} &= K' 2^{d'(1+\epsilon')\log_2\lambda} \\
\frac{\gamma\epsilon_1}{K'(n-i+1)(1+\epsilon_1)} 2^{d\log_2\lambda} &= 2^{d'(1+\epsilon')\log_2\lambda} \\
\frac{\log_2(\frac{\gamma\epsilon_1}{K'(n-i+1)(1+\epsilon_1)})}{(1+\epsilon')\log_2\lambda} + d\frac{1}{1+\epsilon'} &= d' \\
K'' + d\frac{1}{1+\epsilon'} &= d'
\end{aligned}$$

The key point to notice is that the tolerated delay d' on the other streams is a constant K'' plus a term that is almost equal to d.

As such, the probability of error on stream i on bits at delay d or earlier is bounded by

$$\sum_{j=i+1}^{n} \mathcal{P}(\text{Stream } j \text{ has an error at position } K''+d\frac{1}{1+\epsilon'} \text{ or earlier}) + P\left(|X_i(t)| \geq \frac{\gamma\epsilon_1}{(n-i+1)(1+\epsilon_1)} 2^{d\log_2\lambda}\right)$$

We can immediately prove what we need by induction. The base case, $i = n$ is obvious since it is just the scalar case by itself. Now assume that for every $j > i$, we have

$$\mathcal{P}(\text{Stream } j \text{ has an error at position } d \text{ or earlier}) \leq K_j''' 2^{-d\frac{\eta\log_2\lambda}{(1+\epsilon')^{n-j}}}$$

With the induction hypothesis and base case in hand, consider i:

$$
\begin{aligned}
& \mathcal{P}(\text{Stream } i \text{ has an error at position } d \text{ or earlier}) \\
\leq\ & \mathcal{P}(|X_i(t)| \geq \frac{\gamma\epsilon_1}{(n-i+1)(1+\epsilon_1)} 2^{d \log_2 \lambda}) + \sum_{j=i+1}^{n} K_j''' 2^{-(K''+d\frac{1}{1+\epsilon'})\frac{\eta \log_2 \lambda}{(1+\epsilon')^{n-j}}} \\
=\ & \mathcal{P}(|X_i(t)| \geq \frac{\gamma\epsilon_1}{(n-i+1)(1+\epsilon_1)} 2^{d \log_2 \lambda}) + \sum_{j=i+1}^{n} K_j''' 2^{-K''\frac{\eta \log_2 \lambda}{(1+\epsilon')^{n-j}}} 2^{-d\frac{\eta \log_2 \lambda}{(1+\epsilon')^{n-j+1}}} \\
\leq\ & K'''' 2^{-d\eta \log_2 \lambda} + K_i''' 2^{-d\frac{\eta \log_2 \lambda}{(1+\epsilon')^{n-i}}} \\
\leq\ & K_i''''' 2^{-d\frac{\eta \log_2 \lambda}{(1+\epsilon')^{n-i}}}
\end{aligned}
$$

where we used the induction hypothesis, the proof of Theorem 1.4 and that the sum of exponentials is bounded by a constant times the slowest exponential. Since ϵ' was arbitrary, we have proved the theorem since we can get as close as we want to $\alpha = \eta \log_2 \lambda$ in anytime reliability. □

APPENDIX II
NOTATION

We use the shorthand $a_1^t = (a_1, a_2, \ldots, a_{t-1}, a_t)$ to refer to a sequence. Vector valued variables will usually be marked as $\vec{X}$. The i-th component will be X_i. $X_i(t)$ refers to the i-th component of $\vec{X}_t$. We will also apply scalar functions to vectors on a component by component basis so $\log_2 \vec{\lambda}$ is simply shorthand for the vector whose components are $\log_2 \lambda_i$. Similarly $\vec{\lambda}_{||}$ is shorthand for the vector whose components are $|\lambda_i|$.

A. General Notation

Notation used throughout the paper is given here and symbols used only in particular sections will be listed in subsequent subsections for easy reference.

- $\mathcal{A}$ Channel input alphabet
- A The unstable system dynamics — $n \times n$ dimensional square matrix.
- B_t The random variable denoting the channel output at time t.
- B_u, B_w The input transforming matrices corresponding to the control and disturbance inputs correspondingly
- $\mathcal{C}$ The plant controller: takes the channel outputs so far and produces a control signal to apply to the plant.
- C The Shannon capacity of the channel
- $C_{\text{any}}(\alpha)$ The maximum rate at which the channel supports anytime reliability α from definition 1.3
- C_y The observation matrix — tells what linear combinations of the underlying state $\vec{X}$ are actually available at the observer $\mathcal{O}$.
- $\mathcal{D}$ The anytime decoder: produces updated estimates of all messages $\hat{M}_1^t(t)$ at time t.
- d The delay of interest. Measured in terms of channel uses whenever we are in discrete time.
- $\mathcal{E}$ The anytime encoder: takes all messages and feedback information it has access to and produces a channel input at each time step.
- i, j, k Integer valued indices used in various ways.
- K Some constant that we do not care about. Does not depend on delay d or time t.
- $\mathcal{O}$ The plant observer: takes observations of the plant and any other information it has access to and produces a channel input at each time step.
- $\mathcal{P}$ The probability function. Unless specified, all probabilities are taken over all sources of randomness in the system.

n The dimensionality of $\vec{X}$. Because of controllability and observability issues, it is also used to consider time in blocks.

$m_{\{u|w|y\}}$ The dimensionalities of the control input $\vec{U}$, bounded external disturbance$\vec{W}$, and the observation $\vec{Y}$.

M_t The message to be communicated that has arrived at time t. It consists of all bits that have arrived so far and not already batched into M_{t-1} or earlier messages.

$\vec{U}_t$ The vector control applied to the vector system at the end of time t.

$\vec{U}'_t$ The transformed vector control. It acts on each component separately.

$\vec{W}_t$ The bounded vector disturbance entering the vector system at the end of time t.

$\vec{W}'_t$ The transformed vector disturbance. It acts on each component separately.

$\vec{X}_t$ The vector plant state at time t.

$\vec{X}'_t$ The transformed vector state so that each component has independent dynamics.

R_i For multi-stream encoding, the rate of the i-th parallel bitstream.

$\mathcal{R}_{\text{any}}(\vec{\alpha})$ The rate region supported by a channel when the individual bitstreams must satisfy the anytime reliabilities α_i.

S_i The i-th bit-stream to be communicated. It is a function of time taking values ± 1.

$\vec{\alpha}$ A vector of anytime reliabilities α_i for different bitstreams being carried over a single noisy channel.

η The real moment of $|X|$ that is desired to be kept finite by the control system. $\eta > 0$

$\lambda_i, \vec{\lambda}$ The i-th unstable eigenvalue and the vector of all unstable eigenvalues.

Λ A diagonal matrix consisting only of the magnitudes of the eigenvalues of A.

Ω_i The effective bound on the disturbance that is used when trying to encode the i-th component of the transformed state corresponding to a non-diagonal Jordan block.

Ω' The inscribed bound for the $\vec{W}'$

$\bar{\Omega}$ The circumscribed bound for the $\vec{W}'$

B. Notation specific to Section IV

Θ The intrinsic delay of the specified linear system (A, B_u, C_y). It measures how long it takes the output to reflect an input. It is an integer ranging from 0 to $n-1$ for an n-dimensional system.

C. Notation specific to Section V

$D_k(t)$ The stripped version of $\tilde{X}_k(t)$ in which the impact of all bits estimated from streams $i > k$ has been removed. This is meant to approximate what would happen if the system had a purely diagonal Jordan block.

p_k The polynomial that captures the non-exponential part of the dependence of $\check{X}_i(t)$ on prior data bits in bitstreams $i + k$. This comes from the non-diagonal nature of the Jordan block.

$\tilde{X}_j(t)$ An open loop version of the j-th component of the system in transformed coordinates that is driven only by the control signals $\vec{U}$, but not the disturbance $\vec{W}$.

$\check{X}_i(t)$ An open loop version of the j-th component of the system that is driven only by the disturbance $\vec{W}$, but not the controls $\vec{U}$.

γ A suitably small scaling constant chosen to meet the disturbance bound while encoding data bits into real numbers.

Δ_i The size of the box that the observer attempts to keep the i-th component of the state in.

ϵ_i The slack used in encoding bits into a real number on stream i. $(2+\epsilon_i)$ is the "base" of the Cantor-set encoding for bit stream i. It is defined by $(2+\epsilon_i) = \lambda^{\frac{1}{R_i}}$.

$\vec{\epsilon_1}$ The amount we back off from $\log_2 \lambda$ in the target data rate. Because different backoffs are possible for different bitstreams, it is considered a vector.

$\vec{\epsilon_2}$ The amount we back off from $\eta \log_2 \lambda$ in the claimed anytime reliability. Because different backoffs are possible for different bitstreams, it is considered a vector.

D. Notation specific to Section VI

δ The probability of erasure in an erasure channel.

References

[1] A. Sahai and S. K. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link. part I: scalar systems," *IEEE Trans. Inform. Theory*, submitted for publication. [Online]. Available: http://www.eecs.berkeley.edu/~sahai/Papers/control-part-I.pdf

[2] S. Tatikonda and S. K. Mitter, "Control under communication constraints," *IEEE Trans. Automat. Contr.*, vol. 49, no. 7, pp. 1056–1068, July 2004.

[3] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 413–436, July 2004.

[4] S. Tatikonda, A. Sahai, and S. K. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Automat. Contr.*, vol. 49, no. 9, pp. 1549–1561, Sept. 2004.

[5] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[6] G. N. Nair, R. J. Evans, and P. E. Caines, "Stabilising decentralised linear systems under data rate constraints," in *Proceedings of the 43rd IEEE Conference on Decision and control*, Paradise Island, Bahamas, Dec. 2004, pp. 3992–3997.

[7] S. Yuksel and T. Basar, "Communication constraints for stability in decentralized multi-sensor control systems," *IEEE Trans. Automat. Contr.*, submitted for publication.

[8] L. Weng, S. Pradhan, and A.Anastasopoulos, "Error exponent region for Gaussian broadcast channels," in *Proceedings of the 2004 Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2004. [Online]. Available: http://www.eecs.umich.edu/~pradhanv/paper/ciss04_1.pdf

[9] ——, "Error exponent region for Gaussian multiple-access channels," in *Proceedings of the 2004 IEEE International Symposium on Information Theory*, Chicago, IL, June 2004. [Online]. Available: http://www.eecs.umich.edu/~pradhanv/paper/isit04_2.pdf

[10] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: John Wiley, 1971.

[11] N. Elia, "Remote stabilization over fading channels," *Systems and Control Letters*, vol. 54, no. 3, pp. 239–249, Mar. 2005.

[12] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY: John Wiley, 1991.

[13] A. Sahai, "Evaluating channels for control: Capacity reconsidered," in *Proceedings of the 2000 American Control Conference*, Chicago, CA, June 2000, pp. 2358–2362.

[14] ——, "Any-time information theory," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.

[15] T. Simsek, "Anytime channel coding with feedback," Ph.D. dissertation, University of California, Berkeley, 2004.

[16] A. Sahai, T. Simsek, and P. Varaiya, "Why block length is not the same as delay," *IEEE Trans. Inform. Theory*, In preparation.

[17] A. Sahai and S. K. Mitter, "Source coding and channel requirements for unstable processes," *IEEE Trans. Inform. Theory*, To be submitted. [Online]. Available: http://www.eecs.berkeley.edu/~sahai/Papers/anytime.pdf

[18] R. Gray, "Information rates of autoregressive processes," *IEEE Trans. Inform. Theory*, vol. 16, no. 4, pp. 412–421, July 1970.

[19] W. Kahan. (2000, Dec.) Math h110 notes: Jordan's normal form. [Online]. Available: http://www.cs.berkeley.edu/~wkahan/MathH110/jordan.pdf