

# Non-Coherence in Estimation and Control

Gireeja Ranade and Anant Sahai

Wireless Foundations

Electrical Engineering and Computer Sciences

University of California, Berkeley

{gireeja, sahai}@eecs.berkeley.edu

**Abstract**—Control strategies for systems with information bottlenecks often follow an estimate-then-control paradigm. This paper presents a “non-coherent” system where this strategy cannot work and provides an alternative.

The paper considers the estimation and control of a discrete-time linear system with continuous random observation gain, i.e. through a non-coherent channel. It is shown that such an unstable system is not mean-squared observable regardless of the density of the random observation gain: the mean-squared estimation error for any estimator must go to infinity. This is surprising in the context of threshold results for rate-limited estimation.

In contrast to other results with rate-limited feedback, the paper shows that the system can be closed-loop mean-square stabilized in a certain parameter regime even though its open-loop counterpart is not mean-square observable. Finally, carry-free models (generalized deterministic models) provide an intuitive interpretation for the results.

## I. INTRODUCTION

Stringent performance requirements for high-speed decentralized control systems (e.g. self-driving cars) push for a deeper understanding of the interactions between observers and controllers. Communication and control have a history of playing off of each other to explore these interactions. This work explores non-coherence in control systems (i.e. unknown random multiplicative gains) as a parallel to non-coherent communication.

Non-coherent channels emerge when phase-noise, frequency hopping or fast-fading make it impossible for the receiver to perfectly track the channel state. Such communication has been studied as far back as 1969 when Richters conjectured that even for continuous unknown fading-distributions the optimal input distribution is discrete [1]. The conjecture was proved in 2001 [2], [3]. Since then we know that the capacity of channels with unknown Gaussian fading scales as  $\log \log(SNR)$  as opposed to the  $\log(SNR)$  scaling of channels with known fading [4], [5].

Non-coherent observations in control systems can similarly arise from synchronization and sampling errors in control systems [6]. Consider a nearly trivial continuous-time system:

$$\dot{X}(t) = a \cdot X(t).$$

The system collects samples at regular intervals  $t_0, 2t_0, 3t_0$  and so on. The system differential equation above implies  $X(t) = e^{at}X(0)$ , which gives the difference equation:

$$X[n+1] = e^{at_0} \cdot X[n].$$

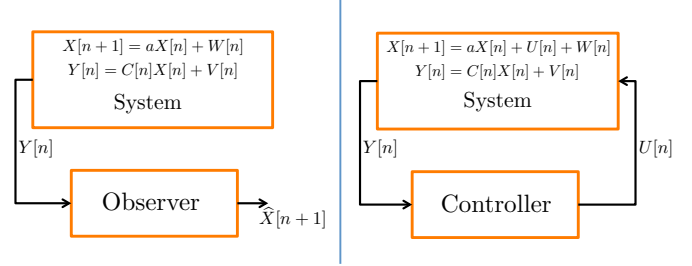


Fig. 1. This paper considers the estimation and control of a system with random multiplicative observation gain,  $C[n]$  and random initial state  $X[0]$ .  $V[n], W[n]$  are white noise and  $a$  is a scalar constant.

Say we are recording a sample  $Y[n]$  at time  $nt_0$ , but clock noise leads to the sample actually being collected at time  $nt_0 + \Delta t_0$ , where  $\Delta t_0$  is a continuous-valued random noise variable. Then we have:

$$Y[n] = e^{a(nt_0 + \Delta t_0)} X(0) = \beta[n] e^{a(nt_0)} X(0),$$

where  $\beta[n]$  is necessarily a continuous-valued random variable. We would have liked to observe  $Y[n] = e^{a(nt_0)} X(0)$ , but we must make do with the multiplicative random noise. Thus, with timing jitter, state information is received over a non-coherent channel. This paper shows that unstable linear systems with such random continuous multiplicative observation gains are in fact not mean-square observable in open-loop.

Systems with multiplicative noise were explored around the same time as questions regarding non-coherent transmission were being asked. In 1971, Rajasekaran et al. derived the **optimal linear filter for systems with unknown continuous observation gain** [7], but did not consider non-linear strategies. This optimal linear filter was further analyzed by Tugnait [8], who noted that error turns out to be stable only in situations when the system is *stable*. Kalman filtering with multiplicative noise was studied in [9], [10], [11], however these works were also limited to stable systems.

The interaction of information theory and control theory has highlighted the importance of non-linear strategies (e.g. [12], [13]), and it is natural to explore the more general setting for the problems above. The optimality of discrete input distributions for transmission over non-coherent channels suggests that optimal estimation and control strategies for non-coherent systems would also be non-linear. However, to the best of our knowledge, these have not been investigated.

In fact, the recently introduced MMSE dimension [14] can be interpreted as characterizing the gain achievable by non-linear estimation over linear estimation in the high-SNR regime. The MMSE dimension measures the scaling behavior of the MMSE relative to the  $SNR$  as  $SNR$  grows in the presence of *additive* noise. For purely continuous distributions, the MMSE-dimension is 1, which implies that non-linear estimation offers the same scaling as linear estimation. This leads to a conjecture for the case of multiplicative noise: **for systems with continuous distributions on the state, linear strategies should be essentially optimal.** On the other hand, the MMSE-dimension for discrete random variables is 0, i.e. non-linear estimation can offer significant gains. In a control problem control actions may or may not be able to discretize a continuous-valued random state. So can we use control to estimate better?

Unfortunately, the results from [14] cannot be directly applied to the problem at hand since the noise is multiplicative, and a naive strategy of taking logs of the system also does not help as is discussed in Sec. IV-B. Instead, **we reduce the estimation problem to a hypothesis-testing problem and use that to provide a lower-bound for the open-loop estimation error that itself must grow unboundedly with time.**

The estimation result in this paper contrasts with previous works in this vein that provide insightful thresholds to serve as design guidelines for systems. These thresholds provide a metric for the “uncertainty” in the system and provide regimes for observability and controllability. For instance, intermittent Kalman filtering results provide a threshold for the observation packet-drop probability parameter  $p$ , below which it is possible to track the system in a mean-square sense [15], [16], [17]. Neither the intermittent Kalman filtering setup nor the setup in this paper allow for coding over the observations, however, the discrete randomness in the Kalman filtering result allows for a threshold result.

Tatikonda and Mitter provide a joint threshold for observability and controllability with  $R$  bits of feedback with coding allowed [18].  $R > \log a$ , where  $a > 0$  is the gain of a scalar, linear system is necessary and sufficient for both estimation and control. The separation principle states that often the optimal control is a function purely of the optimal state estimate. This leads to the *estimate-then-control* paradigm. Tatikonda and Mitter in [18] observe: “If the state estimation error increases with time in an unbounded fashion there will come a point when we can no longer satisfy the control objective.”

This raises the natural question: can a system that cannot be tracked in open-loop be stabilized with feedback? The answer is: sometimes yes. The ability to interact with the system lets the controller use a simple linear strategy stabilize **if the uncertainty is small enough.** This is reminiscent of the **uncertainty threshold principle** [19], which gives stabilizability threshold in the presence of random multiplicative system

gains with perfect noiseless observations<sup>1</sup>.

The control problem here is similar to the larger set of rate-limited control and estimation problems such as [20], which looks at control over fading channels with erasures, or [21], [22] which provide a data-rate threshold for stabilization. Even though technically the feedback in the problem at hand may include infinitely many bits, the systems’s inability to choose and code these bits precludes the transmission of useful information. For further details [23], [24] provide extensive reviews of the work in rate-limited control and estimation.

So why are non-coherent observations different? One intuition is as follows: since the system is unstable, even a small multiplicative observation uncertainty leads to a large variation in the state estimate. The proof of the negative estimation result comes from first ignoring the additive noise in the state evolution and observations. Then the estimation problem is reduced to a hypothesis-testing problem with the help of a carefully constructed genie who sidesteps the prior on the initial state to provide side-information to the observer. With this reduction, the Chernoff-Stein lemma provides a lower-bound on the MMSE. The key technical lemmas for estimation are presented in Sec. IV-A to distill the components above, before the proof of the theorem is given in Sec. IV-B.

The last section Sec. VI uses carry-free models [25], [26] to provide alternative understanding of why the estimation error cannot be tracked and to illustrate the mechanics of a control strategy<sup>2</sup>. We start by setting up the problem below (Sec. II) and stating the main results (Sec. III).

## II. PROBLEM SETUP

Before we set up the problem, some notational conventions: scalars and constants are denoted by lower-case, random variables (r.v.s) are denoted by upper-case letters, realizations are lower-case. We use the  $Y_0^n$  to denote the sequence  $Y[0], Y[1], \dots, Y[n]$ , and so on. The density of  $X$  is denoted by  $f_X(\cdot)$ , and expectation by  $\mathbb{E}[X]$ .

### A. Estimation Setup

Consider the real-valued discrete-time scalar linear system (1) as shown in Fig. 1:

$$\begin{aligned} X[n+1] &= a \cdot X[n] + W[n], \\ Y[n] &= C[n] \cdot X[n] + V[n] \end{aligned} \quad (1)$$

$a \in \mathbb{R}, a > 1$  is a fixed and known scalar.  $W[n]$  and  $V[n]$  are i.i.d. white noise  $\mathcal{N}(0, 1)$  at each time  $n$ .  $C[n]$  are i.i.d. random variables. Let  $X[0] \sim \mathcal{N}(0, 1)$ . We focus on this case, but our results hold for more general densities over  $X[0]$ .

The observer has perfect recall and receives  $Y[n]$  at time  $n$ . Let  $\mathcal{F}_n$  be the  $\sigma$ -algebra defined by  $Y_0^n$ . The observer generates an estimate  $\hat{X}[n+1] \in \mathcal{F}_n$  of  $X[n+1]$  at time  $n$ ,

<sup>1</sup>The estimation results in this paper can also be thought of as the observability counterpart to the uncertainty threshold principle.

<sup>2</sup>As a side note, [14] implies that a finite number of bits of side information cannot change the scaling behavior of the MMSE, and carry-free models provide a visual interpretation for this.

based on all the past observations  $Y_0^n$ . Then the mean-squared estimation error is  $\mathbb{E}[Err[n]^2] = \mathbb{E}[|X[n+1] - \hat{X}[n+1]|^2]$ .

We consider the case where  $C[n] \sim \mathcal{N}(\mu, \sigma^2)$ .  $\mu, \sigma \in \mathbb{R}$  are known constants. We also briefly discuss the case where  $C[n] \sim \text{Unif}[\mu_1, \mu_2]$ .

### B. Control Setup

Consider control of the same system:

$$\begin{aligned} X[n+1] &= a \cdot X[n] + U[n] + W[n], \\ Y[n] &= C[n] \cdot X[n] + V[n] \end{aligned} \quad (2)$$

where  $C[n] \sim \mathcal{N}(\mu, \sigma^2)$ ,  $\mu \neq 0$ .  $X[0] \sim \mathcal{N}(0, 1)$  and  $U[n] \in \mathcal{F}_n$ .

*Definition 2.1:* The system (2) is mean-square stabilizable if there exists a causal controller that generates random variables  $U[n] \in \mathcal{F}_n$  such that  $\lim_{n \rightarrow \infty} \mathbb{E}[|X[n]|^2] < \infty$ .

## III. MAIN RESULTS

*Theorem 3.1:* For the system in Sec. II-A, eq. (1), with  $C[n] \sim \mathcal{N}(\mu, \sigma^2)$ ,  $X[0] \sim \mathcal{N}(0, 1)$ , and  $|a| > 1$  (i.e. the system is unstable),  $\lim_{n \rightarrow \infty} \mathbb{E}[Err[n]^2] = \infty$ .

*Theorem 3.2:* The system in Sec. II-B, eq. (2), with  $C[n] \sim \mathcal{N}(\mu, \sigma^2)$ ,  $X[0] \sim \mathcal{N}(0, 1)$ , and  $|a| > 1$ , is mean-square stabilizable if  $a^2 < 1 + \frac{\mu^2}{\sigma^2}$ .

The ideas from these results can be generalized to more general densities on  $C[n]$  and  $X[0]$ . While we focus on scalar systems here, the results also extend to vector systems<sup>3</sup>.

## IV. ESTIMATION

### A. Lemmas

Before we go into the proof of Thm. 3.1 we state a series of lemmas that we will use to bound the estimation problem using a binary-hypothesis-testing problem. Proofs of Lemmas 4.1, 4.2, 4.3 are discussed in Appendix A.

Lemma 4.1 states that we can lower-bound the MMSE of a Gaussian r.v. by the MMSE of a binary r.v.. While there are some details, the crux of the argument says that a genie giving the estimator more information about the the parameter  $\Theta$  can only help the scaling of the MMSE error. This is the first step to reduce the problem to a hypothesis-testing problem.

*Lemma 4.1:* Let  $Y[0], Y[1], \dots, Y[n] \sim \mathcal{P}_\Theta$ .  $\Theta$  is the parameter for the distribution and  $\Theta \sim \mathcal{N}(0, 1)$ . Let  $\mathcal{F}_n$  be the  $\sigma$ -algebra defined by  $Y[0], Y[1], \dots, Y[n]$ . Given fixed  $r_0, r_1, \epsilon$  such that  $r_1 > r_0$  and  $r_1 - r_0 > \epsilon > 0$ , there exists r.v.s  $R, R_1$  and a  $\sigma$ -algebra  $\mathcal{H}_n$ , such that

- $R_1 \sim \text{Unif}[r_0, r_1 - \epsilon]$
- $R$  is a discrete r.v. conditioned on  $R_1$ .  $R = R_1$  with probability  $\frac{1}{2}$  and  $R = R_1 + \epsilon$  with probability  $\frac{1}{2}$ .
- $\mathcal{H}_n = \mathcal{F}_n \cup \mathcal{F}_{R_1}$ , where  $\mathcal{F}_{R_1}$  is the  $\sigma$ -algebra generated by  $R_1$ .

<sup>3</sup>For estimation, consider a genie that provides the exact values of all but one component of the vector valued state to the observer. The arguments here show that even the last remaining component cannot be estimated. For control, standard extensions from the literature would establish a sufficient condition for control.

- $\min_{\Theta \in \mathcal{F}_n} \mathbb{E}|\Theta - \hat{\Theta}|^2 \geq \min_{\hat{R} \in \mathcal{H}_n} \gamma \mathbb{E}|R - \hat{R}|^2$ , where  $0 < \gamma < 1$  is a constant that depends on  $r_0, r_1$  but does not depend on  $\epsilon$ .

The second Lemma 4.2 shows that the MMSE for a binary r.v. can increase by at most a factor of four if the estimator is restricted to the two values in the range of the r.v..

*Lemma 4.2:* Let  $S$  be a discrete r.v. that takes values only on two points  $s_1 < s_2$ .  $f_S(s_1) > 0$ ,  $f_S(s_2) > 0$  and  $f_S(s) = 0 \forall s \neq s_1, s_2$ . Let  $\mathcal{H}_n$  be the  $\sigma$ -algebra of the observations. Then,

$$\min_{\hat{S} \in \mathcal{H}_n} \mathbb{E}[|S - \hat{S}|^2] \geq \frac{1}{4} \min_{\hat{S} \in \mathcal{H}_n, \hat{S} \in \{s_1, s_2\}} \mathbb{E}[|S - \hat{S}|^2]. \quad (3)$$

Lemma 4.3 uses the Chernoff-Stein Lemma to calculate the probability of type-2 error in a hypothesis-testing problem when the type-1 error is small and the two hypotheses are getting closer and closer together. If the KL-divergence between the hypothesis is small enough, the probability of error is constant.

*Lemma 4.3:* Let  $S$  be a r.v. taking values  $s_1$  and  $s_2$  with equal probability,  $P_S(s_1) = P_S(s_2) = 0.5$ , and let  $Y[1], Y[2], \dots, Y[n] \sim \mathcal{N}(\mu S, \sigma^2 S^2) = \mathcal{Q}$ .  $\sigma \neq 0$ . Let  $P_1$  be  $\mathcal{N}(\mu s_1, \sigma^2 s_1^2)$  and  $P_2$  be  $\mathcal{N}(\mu s_2, \sigma^2 s_2^2)$ . Consider the hypothesis test between  $H_1 : \mathcal{Q} = P_1$  and  $H_2 : \mathcal{Q} = P_2$ .  $H_1, H_2$  have priors  $P(H_1) = P(H_2) = 0.5$  and  $g(Y_0^n)$  is the decision rule used. Then define the probabilities of error as  $\alpha_n = P(g(Y_0^n) = s_2 | H_1 \text{ true})$  and  $\beta_n = P(g(Y_0^n) = s_1 | H_2 \text{ true})$ .

If  $s_2 - s_1 = \frac{1}{n}$ ,  $s_1 < s_2$ , and  $0 \leq \alpha_n < \frac{1}{n^{1+\zeta}} = \delta_n$ , for some  $\zeta > 0$ , then,  $\lim_{n \rightarrow \infty} \beta_n > e^{-\frac{1}{s_2}} 2^{\frac{1}{s_2}} = \kappa$ , a constant that does not depend on  $n$ .

*Remark 4.1:* The central limit theorem states that the error of the averaging estimator decays as  $\frac{1}{\sqrt{n}}$ , and it is natural to think that  $(s_2 - s_1)$  scaling as  $\frac{1}{\sqrt{n}}$  would be the right choice to get a constant probability of error. However, this turns out to be a little too slow, and in this case the probability of error can only be lower bounded by 0 using the techniques from this paper.

### B. Proof of Thm. 3.1

We focus on the Gaussian case with  $C[n] \sim \mathcal{N}(\mu, \sigma^2)$  and  $X[0] \sim \mathcal{N}(0, 1)$  since the proof is the most intuitive and elucidates the proof strategy. The following section remarks on extensions to other densities.

We first show that the estimation error cannot remain finite even for the simpler system<sup>4</sup>:

$$\begin{aligned} X[n+1] &= a \cdot X[n], \\ Y[n] &= C[n] \cdot X[n] \end{aligned} \quad (4)$$

<sup>4</sup>It is tempting to think that by removing the additive noise to get system (4), we have reduced the problem to a known one with additive observation noise as below. Indeed, if we assume the quantities are positive and take logs:

$$\log Y[n] = \log X[n] + \log C[n].$$

We reduce the tracking problem to a problem of estimating the initial state  $X[0]$  using  $Y_0^n$ . However, the estimation error on  $X[0]$  cannot be reduced fast enough to compensate for the rate at which the state is growing. The key idea of the proof is to bound the estimation error by a hypothesis-testing problem and use the Chernoff-Stein lower bound on the probability of error.

*Proof:* Consider a simpler system (4), without the noise,  $X[0] \sim \mathcal{N}(0, 1)$ ,  $C[n] \sim \mathcal{N}(\mu, \sigma^2)$  i.i.d.. Clearly, if the estimation error for system (4) is unbounded, then Thm. 3.1 must be true, since system (1) can be reduced to (4) by giving the noise variables  $V[n], W[n]$  as side information to the estimator.

For the rest of this section  $X[n], Y[n]$  are those associated with system (4). Now, consider

$$\begin{aligned} & \min_{\hat{X}[n+1] \in \mathcal{F}_n} \mathbb{E}|X[n+1] - \hat{X}[n+1]|^2 \\ &= \min_{\hat{X}[n+1] \in \mathcal{F}_n} \mathbb{E}|a^{n+1}X[0] - \hat{X}[n]|^2 \end{aligned} \quad (5)$$

$$= a^{n+1} \min_{\hat{X}_n[0] \in \mathcal{F}_n} \mathbb{E}|X[0] - \hat{X}_n[0]|^2. \quad (6)$$

where  $\hat{X}_n[0] = a^{n+1}\hat{X}[n]$ .

The only uncertainty in the system (4) comes from the initial state  $X[0]$ . For system (4),  $X[n] = a^{n+1}X[0]$  and estimating of  $X[n]$  is equivalent to estimating  $X[0]$ . Scaled appropriately,  $Y_0^n$  are observations drawn i.i.d. from  $\mathcal{P}_{X[0]}(\cdot)$ , where  $X[0]$  is a parameter.

$Y[n]|X[n] \sim \mathcal{N}(\mu X[n], \sigma^2 X[n]^2)$ , and we have  $\frac{Y[n]}{a^n}|X[0] \sim \mathcal{N}(\mu X[0], \sigma^2 X[0]^2) = \mathcal{P}_{X[0]}(\cdot)$ .

Recall  $\mathcal{F}_n$  is the  $\sigma$ -algebra generated by  $Y_0^n$ . We start with the term  $\min_{\hat{X}_n[0] \in \mathcal{F}_n} \mathbb{E}|X[0] - \hat{X}_n[0]|^2$ . We now use Lemma 4.1 to lower-bound the estimation error of  $X[0]$  by the estimation error that would result if  $X[0]$  were a binary random variable.

Choose  $r_0 = 1, r_1 = 2$  and  $\epsilon = \frac{1}{n}, n \geq 1$ . Lemma 4.1 implies that  $\exists R, R_1$  and  $0 < \gamma < 1$  such that  $R$  is a discrete r.v. that takes only two values  $r$  and  $r + \epsilon$  between  $r_0$  and  $r_1$ , and  $f_R(r) = 0.5, f_R(r + \epsilon) = 0.5$ . This gives (7).  $\mathcal{H}_n =$

One can use the average  $\frac{1}{n} \sum_i (Y[i] - i \log a)$  (with appropriate scaling to account for  $\log C[n]$ ) as an estimator for the initial state. The central limit theorem then says that the error on  $\log X[0]$  decays as  $\frac{1}{\sqrt{n}}$ . However, this is not sufficient for us to get a tight lower-bound on the MMSE error for  $X[0]$ : at best the logarithmic bound only helps bound the percentage error (ratio between  $X[0]$  and  $\hat{X}[0]$ ). For large values of the argument, the log function is flat and does not discriminate between its inputs, so a bound on  $|\log X[0] - \log \hat{X}[0]|$  cannot provide a tight bound on the error  $\mathbb{E}[Err[n]]$ . On the other hand, when  $X[0]$  is close to 0, the steep slope of log can allow for very accurate estimation of  $X[0]$ . This second situation is what prevents us from appropriating the log analysis for the lower-bound on the error. Similarly, results from [14] also cannot directly apply.

Further, the central limit theorem can only bound the one particular estimator at hand, and does not help us bound the other non-linear estimators that might perform better.

$\mathcal{F}_n \cup \mathcal{F}_{R_1}$ , where  $\mathcal{F}_{R_1}$  is the  $\sigma$ -algebra generated by  $R_1$ .

$$\begin{aligned} & \min_{\hat{X}_n[0] \in \mathcal{F}_n} \mathbb{E}|X[0] - \hat{X}_n[0]|^2 \\ & \geq \min_{\hat{R} \in \mathcal{H}_n} \gamma \cdot \mathbb{E}|R - \hat{R}|^2 \end{aligned} \quad (7)$$

$$\geq \min_{\hat{R} \in \mathcal{H}_n, \hat{R} \in \{r, r+\epsilon\}} \frac{\gamma}{4} \cdot \mathbb{E}|R - \hat{R}|^2 \quad (8)$$

$$= \frac{\gamma}{4} \cdot p_e(n, \epsilon) \cdot \epsilon^2 \quad (9)$$

$$\geq \frac{\gamma}{4} \cdot \frac{1}{2} \cdot (\alpha_n + \beta_n) \frac{1}{n^2}. \quad (10)$$

(8) states that the MMSE error considered can only be increased by a factor of 4 by reducing the domain of the estimator  $\hat{R}$  to  $\{r, r + \epsilon\}$  and follows from Lemma 4.2. To see (9), note that the probability of error  $p_e(n, \epsilon) = P(\hat{R} = r, R = r + \epsilon) + P(\hat{R} = r + \epsilon, R = r)$ , and hence,  $\mathbb{E}|R - \hat{R}|^2 = p_e(n, \epsilon) \cdot \epsilon^2$  if  $\hat{R} \in \{r, r + \epsilon\}$ .

For (10) define  $\alpha_n = P(\hat{R} = r | R = r + \epsilon)$  and  $\beta_n = P(\hat{R} = r + \epsilon | R = r)$ . Let  $\delta_n = \frac{1}{n^{1+\zeta}}, \zeta > 0$ . Let  $\beta_n^{\delta_n} = \min_{\alpha_n < \delta_n} \beta_n$ . Then,  $p_e(n, \epsilon) = \frac{1}{2}(\alpha_n + \beta_n)$ . The mean-squared error in both cases is symmetric and equal to  $\epsilon^2$ .  $\hat{R}$  is an estimator restricted to two values.

At this point, the estimation problem is effectively reduced to a binary-decision problem between the two hypotheses  $R = r$  and  $R = r + \epsilon$ . The Chernoff-Stein lemma characterizes the probability of error of such a binary-decision problem. This lemma is used to prove Lemma 4.3 which gives us the key step below. First, consider the case where  $\alpha_n > \delta_n$ ,

$$\begin{aligned} & \lim_{n \rightarrow \infty} a^{n+1} \min_{\hat{X}_n[0] \in \mathcal{F}_n} \mathbb{E}|X[0] - \hat{X}_n[0]|^2 \\ & \geq \lim_{n \rightarrow \infty} a^{n+1} \cdot \frac{\gamma}{4} \cdot \frac{1}{2} \cdot (\alpha_n + \beta_n) \cdot \frac{1}{n^2} \end{aligned} \quad (11)$$

$$\begin{aligned} & \geq \lim_{n \rightarrow \infty} a^{n+1} \cdot \frac{\gamma}{4} \cdot \frac{1}{2} \cdot \frac{1}{n^{1+\zeta}} \cdot \frac{1}{n^2} \\ & = \infty. \end{aligned} \quad (12)$$

On the other hand if  $\alpha_n < \delta_n$ , by Lemma 4.3 we have  $\lim_{n \rightarrow \infty} \beta_n^{\delta_n} > \kappa$

$$\begin{aligned} & \lim_{n \rightarrow \infty} a^{n+1} \min_{\hat{X}_n[0] \in \mathcal{F}_n} \mathbb{E}|X[0] - \hat{X}_n[0]|^2 \\ & \geq \lim_{n \rightarrow \infty} a^{n+1} \cdot \frac{\gamma}{4} \cdot \frac{1}{2} \cdot \beta_n^{\delta_n} \cdot \frac{1}{n^2} \end{aligned} \quad (14)$$

$$\begin{aligned} & > \kappa \lim_{n \rightarrow \infty} a^{n+1} \cdot \frac{\gamma}{4} \cdot \frac{1}{2} \cdot \frac{1}{n^2} \\ & = \infty. \end{aligned} \quad (15)$$

Since  $\kappa$  is a constant the exponential term dominates as  $n \rightarrow \infty$  which gives (16). ■

### C. Uniform $C[n]$ , General $X[0]$

Even in the case where  $C[n]$  are drawn from a uniform distribution and with  $X[0]$  drawn from the more general distribution as above, the estimation error must go to infinity. The proof above is based on the initial state  $X[0]$  being drawn from a normal distribution. We note that the same technique

can work for any continuous density on  $X[0]$ , since the crux of the proof lies in finding two points that are arbitrarily close to each other<sup>5</sup>. With some modifications we can also extend the proof to the case where  $C[n]$  has a uniform density. Details are omitted for space constraints and will be included in the full paper. The unbounded support of the Gaussian distribution is not the estimation holdup, nor are special properties of the Gaussian key for the result.

**Theorem 4.4:** For the system in Sec. II-A, eq. (1) with  $C[n] \sim \text{Unif}[\mu_1, \mu_2]$ ,  $|a| > 1$ ,  $X[0] \sim f_X(\cdot)$  where  $f_X(\cdot)$  is a continuous density,  $\lim_{n \rightarrow \infty} \mathbb{E}[\text{Err}[n]^2] = \infty$ .

Since the uniform distribution is the prototypical continuous density, using Thm. 4.4, we can use it to prove that the estimation error for a system (1) is unbounded with any continuous  $C[n]$ . The general density case reduces to the uniform case by considering a piecewise constant approximation to the density for  $C[n]$ , i.e. thinking of a general density as a mixture of tiny uniform densities. Then reduce the problem to the uniform case with a genie that reveals which interval  $C[n]$  was drawn from.

## V. CONTROL

This section gives the proof of Thm. 3.2. Appropriate feedback control can stabilize the system in a certain parameter range, even though it is not mean-square observable in open-loop.

*Proof:* Choose  $U[n] = -dY[n]$  for some  $d \in \mathbb{R}$ . Then

$$X[n+1] = aX[n] - dY[n] + W[n] \quad (17)$$

$$= (a - dC[n])X[n] - dV[n] + W[n]. \quad (18)$$

Because not zero-mean,  $\mathbb{E}[C(n)^2] \sim \text{Var}(C)$   
Hence,

$$\mathbb{E}[X[n+1]^2] \quad (19)$$

$$= \mathbb{E}[(a - dC[n])^2 X[n]^2] + \mathbb{E}[d^2 V[n]^2] + \mathbb{E}[W[n]^2] \quad (20)$$

$$= (a^2 - 2ad\mu + d^2(\mu^2 + \sigma^2))\mathbb{E}[X[n]^2] + d^2 + 1. \quad (21)$$

If  $K = a^2 - 2ad\mu + d^2(\mu^2 + \sigma^2) < 1$  the control can stabilize the system. Minimizing  $K$  gives  $d = \frac{a\mu}{\mu^2 + \sigma^2}$ , and for this choice,  $K < 1$  reduces to  $a^2 < \frac{\mu^2 + \sigma^2}{\sigma^2} = 1 + \frac{\mu^2}{\sigma^2}$ . ■

## VI. A BIT LEVEL INTERPRETATION

This section uses bit-level pictures to give an interpretation for the results. The model illustrates how the ability to modify the state (through control) lets the controller extract information from the system and stabilize it in closed-loop.

Carry-free models [25], [26] are a bit-level representation for communication that generalize ADT deterministic models [27]. The name “carry-free” is derived from the fact that the addition operation does not involve carry-overs. Multiplication between two bit-strings is convolution:  $y_m y_{m-1} \dots y_0 y_{-1} \dots = (c_m c_{m-1} \dots) \cdot (x_m x_{m-1} \dots)$ , then  $y_i = \sum_j c_j x_{i-j}$ . The effect of observation noise is captured

<sup>5</sup>In fact, all we really need is a distribution that contains an interval with a density.

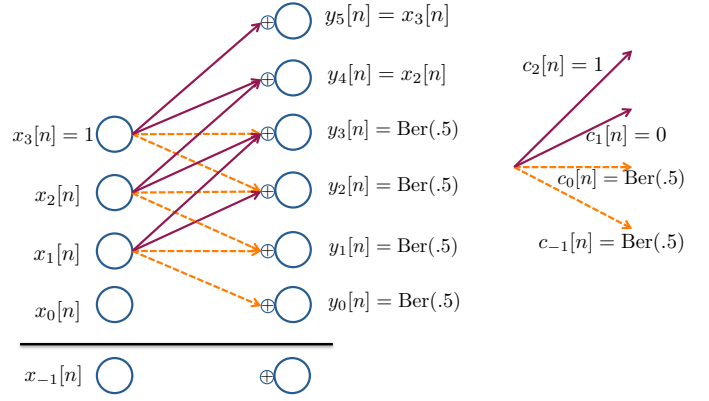


Fig. 2. Fig. 2 has  $m_n = 3$ . So  $y_5[n], y_4[n]$  are deterministic, and  $y_3[n]$  and below are random. Note that  $y_5[n] = c_{det} \cdot x_3[n] = x_3[n]$ , and  $y_4[n] = c_{det} \cdot x_2[n] + c_{det-1} \cdot x_2[n]$ .  $c_{det} = 2, c_{rand} = 0$ . We only show four levels for the gain  $c_2$  to  $c_{-1}$  for simplicity, but the gain levels may extend downwards indefinitely till  $-\infty$ .

by a truncation operation. For a polynomial  $a(z) = \sum_i a_i z^i$  define  $\lfloor a(z) \rfloor = \sum_{i \geq 0} a_i z^i$ .

We'll build a carry-free parallel to system (1), with the observations as transmissions over a channel with unknown fading. The estimator receives *uncoded* transmissions of the state. But before getting into the technical setup, we can see what is going on in Fig. 2. The  $x_i$ 's represent the state and  $y_i$ 's are the received signal. The gain  $c[n]$  has two deterministic bits  $c_2[n] = 1$  and  $c_1[n] = 0$ . Bits  $c_0[n], c_{-1}[n], \dots$  are unknown Bernoulli( $\frac{1}{2}$ ) r.v.s. Then  $y_5[n] = x_3[n]$ ,  $y_4[n] = x_2[n]$  are *clean* bits, but the bit-levels  $y_3[n]$  and lower are *contaminated* by the multiplicative noise. The *clean* levels of  $x[n]$ , i.e.  $x_3[n], x_2[n]$  can be perfectly estimated, but the observer can extract no information from the *contaminated* levels. So  $x_1[n]$  and lower are invisible. The level of contamination is increasing as the magnitude of the state (i.e. the maximum bit-level above noise) increases and hence the state estimation error grows unboundedly. However, a controller can knock-off the top clean bit-levels  $x_3[n], x_2[n]$  at each time step, thus keeping the state bounded. As long as the rate of increase of bit-levels is not higher than the ability to control, the state can be stabilized.

To formalize this, let the state as it evolves in time be represented by the bits  $x[n] = x_{m_n}[n]x_{m_n-1}[n] \dots x_1[n]x_0[n].x_{-1}[n]x_{-2}[n] \dots$ .  $n$  is the time index. The subscript denotes the bit-level, i.e.  $m_n$  is the highest non-zero bit-level occupied by the state at time  $n$ . In polynomial notation, we denote the state as  $x[n](z) = x_{m_n}[n]z^{m_n} + x_{m_n-1}[n]z^{m_n-1} + \dots + x_0[n] + x_{-1}z^{-1} \dots$ . We denote the carry-free state as  $\lfloor x[n](z) \rfloor$ . We will suppress the dummy polynomial variable  $z$  wherever obvious. The gain of the system is fixed as  $a(z) = 1 \cdot z^{g_a}$  and is not time varying.

Define the state evolution as:

$$x[n+1] = a \cdot x[n] \quad (22)$$

The initial state  $x[0](z)$  is a random binary polynomial of

degree  $m_0$ ,  $P(x_i[0] = 1) = 0.5$ ,  $\forall i \leq m_0$ . At each time step, the system magnitude increases by  $g_a$  bit-levels due to the gain  $a(z)$ . For simplicity we ignore the system evolution noise. The observation  $y[n](z)$  is:

$$y[n] = \lfloor c[n] \cdot x[n] \rfloor \quad (23)$$

We model the bits of the random gain  $c[n](z)$  as Bernoulli( $\frac{1}{2}$ ) variables.  $g_c$  is the highest bit-level of  $c[n](z)$ . The model in [26] focused on random gains where all the bits were random. However, this model does not accurately capture non-zero-mean random variables. We generalize that model to the one in eq. (25), which allows for both deterministic and random bits. The random bits are treated similar to noise, and above a certain level bits are deterministic.

$$c[n](z) = c_{g_c}[n] \cdot z^{g_c} + c_{g_c-1}[n] \cdot z^{g_c-1} + \dots \quad (24)$$

$$= 1 \cdot z^{c_{det}} + 0 \cdot z^{c_{det}-1} + 0 \cdot z^{c_{det}-2} + \dots \\ + c_{c_{rand}}[n] \cdot z^{c_{rand}} + c_{c_{rand}-1}[n] \cdot z^{c_{rand}-1} + \dots \quad (25)$$

$$= z^{c_{det}} + c_{rand}(z)[n]. \quad (26)$$

Define:

$$c_{rand} = \max\{i | c_i \sim \text{Bernoulli}(\frac{1}{2})\}. \quad (27)$$

$c_{rand}$  is the index of the highest random bit of  $c[n](z) \forall n$ .  $c_{rand}$  does not vary with  $n$ . On the other hand, the value of the bit at level  $c_{rand}$ , i.e.  $c_{c_{rand}}[n]$ , does. Further,  $c_i[n] = \text{Bernoulli}(\frac{1}{2})$ ,  $\forall i \leq c_{rand}$  and are unknown to the transmitter and the receiver. If  $c_{rand} < g_c$ , then we also define  $c_{det}$  as the highest deterministic bit-level of  $c$ .  $c_{c_{det}} = 1$ . All bits from  $c_{c_{det}-1}$  to  $c_{c_{rand}+1}$  are fixed to be 0 for simplicity.

$$c_{det} = \begin{cases} \max\{i | c_i = 1\} & \text{if } c_{rand} < g_c \\ c_{rand}, & \text{otherwise.} \end{cases} \quad (28)$$

By construction, we have  $c_{det} \geq c_{rand}$ .

**Definition 6.1:** Let  $\hat{x}[n](z)$  be any estimate of  $x[n](z)$  based on  $y[n]$ . Then the estimation error at time  $n$  is defined as  $e[n](z) = x[n](z) - \hat{x}[n](z)$

**Theorem 6.1:** Let  $g_e[n]$  be the highest non-zero bit-level of  $e[n](z)$ . Then  $\mathbb{E}[g_e[n]] \rightarrow \infty$ , i.e. the degree of the error polynomial is unbounded.

Clearly, the estimator learns the clean (the top  $c_{det} - c_{rand}$ ) bits of the state. However, these bits are unchanged as time increases and the same bits are learned by the estimator again and again. The second piece of information that is in the state is the highest non-zero level of  $x[n](z)$ ,  $m_n$ , which can be decoded from the highest non-zero level of  $y[n](z)$ . This is the part of the information that supposedly scales with the magnitude of the state, a la the  $\log \log SNR$  result. But according to our model,  $m_n$  is a deterministic function of  $m_{n-1}$ ,  $m_n = m_{n-1} + g_a$ , and again, the same information is re-learned over time. Since the state keeps growing and we can never learn more than a small finite number of bits ( $c_{det} - c_{rand} + \log m_0$ ), the error must grow unboundedly.

**Control Intuition:** The carry-free model also provides an intuition for why the condition  $a^2 < \frac{\mu^2 + \sigma^2}{\sigma^2}$  is sufficient to control system (2). Taking logs of this expression, we have  $\log a^2 < \log(\mu^2 + \sigma^2) - \log \sigma^2$ . The carry-free model suggests that if  $g_a < c_{det} - c_{rand}$  then the system can be stabilized, since at each step, we observe the top  $c_{det} - c_{rand}$  bits of  $x[n]$ . Through these, we can extract information and then cancel them with an appropriate control. This cancellation allows new bits to be learned at the next time step. Hence, the state remains bounded.  $(\mu^2 + \sigma^2)$  is the second-moment of  $C[n]$  and its log is the parallel to  $c_{det}$ , and  $\log \sigma^2$  captures the randomness in the system. This provides a direction to attack rate-limited control problems even when the pure estimation problem is fragile to multiplicative noise.

## VII. CONCLUSION

This problem highlights how the ability to interact with the system (through control) might allow an agent to extract information about the system state. This is reminiscent of other work in decentralized control and communication, where source simplification and modification are essential [13], [28]. This suggests future work to understand rate-limited non-coherent observations as well as non-coherent system gains.

## ACKNOWLEDGEMENTS

We acknowledge support of the NSF through the grant CNS-0932410 and a GRF. We would also like to thank Baris Nakiboglu for helpful discussions and Pulkit Grover, Abolfazl Motahari and Se Yong Park for feedback on the paper.

## APPENDIX A: ESTIMATION LEMMA PROOFS

### A. Proof of Lemma 4.1

Our proof shows that a we can reduce the estimation problem over the entire range of  $\Theta \sim \mathcal{N}(0, 1)$  to an estimation problem over just two points. We do this by generating  $\Theta$  as a mixture of two random variables  $S$  and  $T$  and using a genie argument.

Let  $R_1 \sim \text{Unif}[r_0, r_1 - \epsilon]$ . and  $R_1$  is independent of all problem parameters. Let

$$S = \begin{cases} R_1 & \text{with prob. } \frac{1}{2} \\ R_1 + \epsilon & \text{with prob. } \frac{1}{2} \end{cases} \quad (29)$$

for some fixed  $\epsilon < r_1 - \epsilon - r_0$ . The density of  $S$ ,  $f_S(s)$ , is given by:

$$f_S(s) = \begin{cases} \frac{1}{2(r_1 - \epsilon - r_0)}, & r_0 \leq s < r_0 + \epsilon \\ \frac{1}{(r_1 - \epsilon - r_0)}, & r_0 + \epsilon \leq s \leq r_1 - \epsilon \\ \frac{1}{2(r_1 - \epsilon - r_0)}, & r_1 - \epsilon < s \leq r_1 \\ 0, & \text{otherwise.} \end{cases} \quad (30)$$

The density of  $T$  is essentially a Gaussian density minus  $f_S(\cdot)$  with a scaling. Choose  $\gamma = \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r_1^2}$ . Note  $\gamma$  does



not depend on  $\epsilon$ .

$$f_T(t) = \begin{cases} \frac{1}{1-\gamma} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} - \gamma \frac{1}{2(r_1-\epsilon-r_0)} \right), & r_0 \leq t < r_0 + \epsilon \\ \frac{1}{1-\gamma} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} - \gamma \frac{1}{2(r_1-\epsilon-r_0)} \right), & r_0 + \epsilon \leq t \leq r_1 - \epsilon \\ \frac{1}{1-\gamma} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} - \gamma \frac{1}{2(r_1-\epsilon-r_0)} \right), & r_1 - \epsilon < t \leq r_1 \\ \frac{1}{1-\gamma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, & \text{otherwise} \end{cases} \quad (31)$$

The choice of  $\gamma$  ensures that  $f_T(t) \geq 0 \forall t$  and is a valid density. This mixture of  $S$  (with probability  $\gamma$ ) and  $T$  (with probability  $1-\gamma$ ) leads to  $\Theta \sim \mathcal{N}(0, 1)$ .

$H$  is a  $\gamma$ -biased coin which generates  $\Theta$  using  $S, T$ .

$$\Theta = \begin{cases} S & \text{if } H \text{ is heads (prob. } \gamma) \\ T & \text{if } H \text{ is tails (prob. } 1-\gamma) \end{cases} \quad (32)$$

Now consider the MMSE estimate of  $\Theta$ , given observations  $Y_0^n$ . This error could only decrease in the event that a genie were to reveal more information about  $\Theta$ . In particular, a genie observes the realization  $\Theta = \theta_1$  as well as  $S, R_1, T, H$ .

If  $H$  is tails then the exact value of  $\Theta$  is revealed, and  $\hat{\Theta} = \Theta$ . If  $H$  is heads, then the realization of the underlying r.v.  $R_1 = r$  is revealed. Let  $\mathcal{F}_{R_1}$  be the  $\sigma$ -algebra generated by  $R_1$ . Then,  $\hat{\Theta} \in \mathcal{H}_n$ , where  $\mathcal{H}_n = \mathcal{F}_n \cup \mathcal{F}_{R_1}$ .

We are interested in the error that would be made due to confusion between  $r$  and  $r + \epsilon$  after this revelation. Then,

$$\min_{\hat{\Theta} \in \mathcal{F}_n} \mathbb{E}|\Theta - \hat{\Theta}|^2 \quad (33)$$

$$\geq \min_{\hat{\Theta} \in \mathcal{H}_n} \mathbb{E}|\Theta - \hat{\Theta}|^2 \quad (34)$$

$$= \min_{\hat{\Theta} \in \mathcal{H}_n} \gamma \mathbb{E}[|\Theta - \hat{\Theta}|^2 | H \text{ is heads}] + (1-\gamma) \mathbb{E}[|\Theta - \hat{\Theta}|^2 | H \text{ is tails}] \quad (35)$$

$$= \min_{\hat{\Theta} \in \mathcal{H}_n} \gamma \mathbb{E}[|\Theta - \hat{\Theta}|^2 | \Theta = S] + (1-\gamma) \mathbb{E}[|\Theta - \hat{\Theta}|^2 | \Theta = T] \quad (36)$$

$$\geq \min_{\hat{\Theta} \in \mathcal{H}_n} \gamma \mathbb{E}[|\Theta - \hat{\Theta}|^2 | \Theta = S] \quad (37)$$

$$= \min_{\hat{\Theta} \in \mathcal{H}_n} \gamma \mathbb{E}[|S - \hat{\Theta}|^2 | \Theta = S, R_1 = r] \quad (38)$$

$$= \min_{\hat{\Theta} \in \mathcal{H}_n} 0.5\gamma \left( \mathbb{E}[|S - \hat{\Theta}|^2 | S = r] + \mathbb{E}[|S - \hat{\Theta}|^2 | S = r + \epsilon] \right) \quad (39)$$

(38) follows since when  $\Theta = S$  the genie reveals  $R_1 = r$ . For (39) note that  $S|R_1 = r$  with probability 0.5 and  $S|R_1 = r + \epsilon$  with probability 0.5. Let  $R \doteq S|R_1 = r$ .

$$R = \begin{cases} r & \text{with prob. } \frac{1}{2} \\ r + \epsilon & \text{with prob. } \frac{1}{2} \end{cases} \quad (40)$$

If we substitute  $R$  into (39), we get (41), which gives the desired result.

$$= \min_{\hat{\Theta} \in \mathcal{H}_n} \gamma \mathbb{E}[|R - \hat{\Theta}|^2 | \Theta = R] \quad (41)$$

$$= \min_{\hat{R} \in \mathcal{H}_n} \gamma \mathbb{E}[|R - \hat{R}|^2] \quad (42)$$

## B. Proof of Lemma 4.2

Let the posterior probabilities on  $s_1$  and  $s_2$  by the data be  $q_n$  and  $1 - q_n$ . Then, the MMSE estimator  $\hat{S}_n$  and associated mean squared error  $Err[n] = \min_{\hat{S}_n \in \mathcal{H}_n} \mathbb{E}[|S - \hat{S}_n|^2]$  are

$$\hat{S}_n = q_n s_1 + (1 - q_n) s_2 \quad (43)$$

$$Err[n] = q_n (s_1 - \hat{S}_n)^2 + (1 - q_n) (s_2 - \hat{S}_n)^2 \quad (44)$$

Now consider a quantization estimator  $S_{quant,n} \in \{s_1, s_2\}$ .  $Err_{quant}[n]$  is the associated mean-squared error.

$$S_{quant,n} = \begin{cases} s_1, & \text{if } (s_1 - \hat{S}_n)^2 \leq (s_2 - \hat{S}_n)^2 \text{ i.e. } q \geq 0.5 \\ s_2, & \text{if } (s_1 - \hat{S}_n)^2 > (s_2 - \hat{S}_n)^2 \text{ i.e. } q < 0.5 \end{cases} \quad (45)$$

If  $S_{quant,n}$  is in error, the magnitude of the error is always  $(s_2 - s_1)^2$ . Let  $p_{e,quant}$  be the probability with which  $S_{quant,n}$  is in error. So

$$Err_{quant}[n] = p_{e,quant} (s_2 - s_1)^2 \quad (46)$$

Clearly,  $Err_{quant}[n] \geq Err[n]$ , by the definition of  $\hat{S}_n$ .

Consider the expression,  $Err_{quant,genie}[n]$ , which reduces the error by a factor of 4.

$$Err_{quant,genie}[n] = \frac{1}{4} Err_{quant}[n] = p_{e,quant} \left( \frac{s_1 - s_2}{2} \right)^2. \quad (47)$$

We will show that

$$Err[n] \geq Err_{quant,genie}[n]. \quad (48)$$

To get (48), compare  $Err[n]$  and  $Err_{quant,genie}[n]$  on a realization by realization basis. Let  $q_n > 0.5$ . Then  $S_{quant,n} = s_1$ . Also,  $p_{e,quant} = 1 - q_n$ , since this is the probability with which  $S = s_2$ . Since  $q_n > 0.5$ , we have  $s_1 \leq \hat{S}_n < \frac{s_1 + s_2}{2} < s_2$ . If  $S = s_1$ ,  $|Err_{quant,genie}[n]|^2 = 0$ , and  $|S - \hat{S}_n|^2 = (s_1 - \hat{S}_n)^2$ . So,  $|Err_{quant,genie}[n]|^2 \leq |S - \hat{S}_n|^2$  in this case. If  $S = s_2$ ,  $|Err_{quant,genie}[n]|^2 = \left( \frac{s_1 - s_2}{2} \right)^2$  and  $|S - \hat{S}_n|^2 = (s_1 - s_2)^2$ .  $|Err_{quant,genie}[n]|^2 \leq |S - \hat{S}_n|^2$  in this case as well. A similar argument holds if  $q \leq 0.5$ . Hence, we have  $Err[n] \geq Err_{quant,genie}[n]$ .

Consider now the optimal estimator for  $S$  over those estimators that only take values on  $\{s_1, s_2\}$

$$S_{opt,n} = \underset{\hat{S} \in \mathcal{H}_n, \hat{S} \in \{s_1, s_2\}}{\operatorname{argmin}} \mathbb{E}[|S - \hat{S}|^2] \quad (49)$$

$Err_{opt}[n]$  is the associated mean squared error, and  $Err_{opt,genie}[n]$  is the associated error for the optimal estimator with the help of the genie so that  $Err_{opt,genie}[n] = \frac{1}{4} Err_{opt}[n]$ , since the genie reduces the error by a factor of 4.  $S_{opt,n}$  will have better performance than  $S_{quant,n}$  and hence the probability of error  $p_{e,opt} \leq p_{e,quant}$ . Hence,

$$Err_{quant,genie}[n] \geq Err_{opt,genie}[n] = \frac{1}{4} Err_{opt}[n] \quad (50)$$

Hence, (48), (50) imply that  $Err[n] \geq \frac{1}{4} Err_{opt}[n]$  i.e.

$$\min_{\hat{S} \in \mathcal{H}_n} \mathbb{E}[|S - \hat{S}|^2] \geq \frac{1}{4} \min_{\hat{S} \in \mathcal{H}_n, \hat{S} \in \{s_1, s_2\}} \mathbb{E}[|S - \hat{S}|^2] \quad (51)$$

### C. Proof of Lemma 4.3

$s_2 - s_1 = \frac{1}{n}$ , thus  $\frac{s_1}{s_2} = 1 - \frac{1}{ns_2}$  and  $\left(\frac{s_1^2}{s_2^2} - 1\right) = \frac{1}{n} \left(\frac{1}{ns_2^2} - \frac{2}{s_2}\right)$ .

Consider any decision rule,  $g(Y_0^n)$ . Associated with the decision rule is an acceptance region for  $H_1$  based on  $Y_0^n$ , i.e. the set of values of  $Y_0^n$  where  $g(Y_0^n) = s_1$ . Let  $B_n$  denote this acceptance region. Then  $\alpha_n = P_1(B_n^c) < \delta_n$ , so  $P_1(B_n) > 1 - \delta_n$ . Since  $D(P_1||P_2) < \infty$ , the Chernoff-Stein lemma (Lem 11.8.1, Thm 11.8.3, [29], pg. 383) implies  $P_2(B_n) > (1 - 2\delta_n)2^{-nD(P_1||P_2) + \delta_n}$ . Recall,  $\delta_n = \frac{1}{n^{1+\zeta}}$ .

When  $\sigma \neq 0$ ,

$$\lim_{n \rightarrow \infty} P_2(B_n) > \lim_{n \rightarrow \infty} (1 - 2\delta_n)2^{-n(D(P_1||P_2) + \delta_n)} \quad (52)$$

$$= \lim_{n \rightarrow \infty} \left(1 - \frac{2}{n^{1+\zeta}}\right)2^{-\frac{n}{n^{1+\zeta}}} \lim_{n \rightarrow \infty} 2^{-nD(P_1||P_2)} \quad (53)$$

$$= 1 \cdot 2^0 \lim_{n \rightarrow \infty} 2^{-nD(P_1||P_2)} \quad (54)$$

$$D(P_1||P_2) = \log \frac{s_2}{s_1} + .5 \left( \frac{s_1^2}{s_2^2} \left( \frac{\mu^2 + \sigma^2}{\sigma^2} \right) - 1 \right) + \frac{\mu^2}{\sigma^2} \left( -\frac{s_1}{s_2} + \frac{1}{2} \right) \quad (55)$$

$$\lim_{n \rightarrow \infty} 2^{-nD(P_1||P_2)} = \lim_{n \rightarrow \infty} 2^{\left\{ -n \left( \log \frac{s_2}{s_1} + .5 \left( \frac{s_1^2}{s_2^2} \left( \frac{\mu^2 + \sigma^2}{\sigma^2} \right) - 1 \right) + \frac{\mu^2}{\sigma^2} \left( -\frac{s_1}{s_2} + \frac{1}{2} \right) \right) \right\}} \quad (56)$$

$$= \lim_{n \rightarrow \infty} \frac{s_1^n}{s_2^n} \cdot 2^{\left\{ -\frac{n}{2} \frac{1}{\sigma^2 s_2^2} (s_1^2 (\mu^2 + \sigma^2) - \sigma^2 s_2^2 - 2\mu^2 s_1 s_2 + \mu^2 s_2^2) \right\}} \quad (57)$$

$$= \lim_{n \rightarrow \infty} \frac{s_1^n}{s_2^n} \cdot 2^{\left\{ -\frac{n}{2} \frac{1}{\sigma^2 s_2^2} (\mu^2 (s_1^2 + s_2^2 - 2s_1 s_2) + \sigma^2 (s_1^2 - s_2^2)) \right\}} \quad (58)$$

$$= \lim_{n \rightarrow \infty} \frac{s_1^n}{s_2^n} \cdot 2^{\left\{ -\frac{n}{2} \frac{1}{\sigma^2 s_2^2} (\mu^2 (s_1 - s_2)^2 - \sigma^2 (s_2 - s_1)(s_1 + s_2)) \right\}} \quad (59)$$

$$= \lim_{n \rightarrow \infty} \left(1 - \frac{1}{ns_2}\right)^n \cdot 2^{\left\{ -\frac{n}{2} \frac{1}{\sigma^2 s_2^2} \left( \mu^2 \frac{1}{n^2} - \sigma^2 \frac{1}{n} (s_1 + s_2) \right) \right\}} \quad (60)$$

$$= e^{-\frac{1}{s_2}} 2^{\frac{1}{s_2}} \quad (61)$$

Thus,  $\lim_{n \rightarrow \infty} P_2(B_n) > \left(\frac{2}{e}\right)^{\frac{1}{s_2}} = \kappa$ . ■

### REFERENCES

- [1] J. Richters, "Communication over fading dispersive channels," DTIC Document, Tech. Rep., 1967.
- [2] Abou-Faycal, I.C. and Trott, M.D. and Shamai, S., "The capacity of discrete-time memoryless Rayleigh-fading channels," *Information Theory, IEEE Transactions on*, vol. 47, no. 4, pp. 1290–1301, 2001.
- [3] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2257–2270, 2004.
- [4] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electronics Letters*, vol. 33, no. 16, pp. 1368–1370, 1997.
- [5] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2426–2467, 2003.
- [6] H. Meyr, M. Moeneclay, and S. A. Fechtel, *Digital communication receivers: synchronization, channel estimation and signal processing*, 1998. John Wiley and Sons, New York, NY, USA.
- [7] P. Rajasekaran, N. Satyanarayana, and M. Srinath, "Optimum linear estimation of stochastic signals in the presence of multiplicative noise," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 3, pp. 462–468, 1971.
- [8] J. Tugnait, "Stability of optimum linear estimators of stochastic signals in white multiplicative noise," *Automatic Control, IEEE Transactions on*, vol. 26, no. 3, pp. 757–761, 1981.
- [9] L. Xie, Y. C. Soh, and C. E. de Souza, "Robust kalman filtering for uncertain discrete-time systems," *Automatic Control, IEEE Transactions on*, vol. 39, no. 6, pp. 1310–1314, 1994.
- [10] E. Gershon, D. Limebeer, U. Shaked, and I. Yaesh, "Robust H- $\infty$  filtering of stationary continuous-time linear systems with stochastic uncertainties," *Automatic Control, IEEE Transactions on*, vol. 46, no. 11, pp. 1788–1793, 2001.
- [11] E. Gershon, U. Shaked, and I. Yaesh, "Robust H- $\infty$  estimation of stationary discrete-time linear processes with stochastic uncertainties," *Systems & control letters*, vol. 45, no. 4, pp. 257–269, 2002.
- [12] H. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, p. 131, 1968.
- [13] P. Grover and A. Sahai, "The vector Witsenhausen problem as assisted interference cancellation," *International Journal on Systems, Control and Communications (IJSCC)*, 2008.
- [14] Y. Wu and S. Verdú, "MMSE dimension," *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 4857–4879, 2011.
- [15] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1453–1464, 2004.
- [16] Y. Mo and B. Sinopoli, "A characterization of the critical value for Kalman filtering with intermittent observations," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*. IEEE, 2008, pp. 2692–2697.
- [17] S. Park and A. Sahai, "Intermittent Kalman filtering: Eigenvalue cycles and nonuniform sampling," in *American Control Conference (ACC), 2011*. IEEE, 2011, pp. 3692–3697.
- [18] S. Tatikonda and S. Mitter, "Control under communication constraints," *Automatic Control, IEEE Transactions on*, vol. 49, no. 7, pp. 1056–1068, 2004.
- [19] M. Athans, R. Ku, and S. Gershwin, "The uncertainty threshold principle: Some fundamental limitations of optimal decision making under dynamic uncertainty," *Automatic Control, IEEE Transactions on*, vol. 22, no. 3, pp. 491–495, 1977.
- [20] N. Elia, "Remote stabilization over fading channels," *Systems & Control Letters*, vol. 54, no. 3, pp. 237–249, 2005.
- [21] G. Nair and R. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 413–436, 2004.
- [22] P. Minero, M. Franceschetti, S. Dey, and G. N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *Automatic Control, IEEE Transactions on*, vol. 54, no. 2, pp. 243–255, 2009.
- [23] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108–137, 2007.
- [24] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.
- [25] U. Niesen and M. Maddah-Ali, "Interference alignment: From degrees-of-freedom to constant-gap capacity approximations," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 2077–2081.
- [26] S. Park, G. Ranade, and A. Sahai, "Carry-free models and beyond," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 1927–1931.
- [27] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *Information Theory, IEEE Transactions on*, vol. 57, no. 4, pp. 1872–1905, 2011.
- [28] P. Grover and A. Sahai, "The "source-simplification" aspect of signaling," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 2432–2436.
- [29] T. Cover and J. Thomas, *Elements of information theory*, 2nd ed. Wiley, 2006.