

Implicit and explicit communication in decentralized control

Pulkit Grover and Anant Sahai

Department of EECS, University of California at Berkeley, CA-94720, USA

{pulkit, sahai}@eecs.berkeley.edu

Abstract—There has been substantial progress recently in understanding toy problems of purely *implicit* signaling. These are problems where the source and the channel are implicit — the message is generated endogenously by the system, and the plant itself is used as a channel. In this paper, we explore how implicit and explicit communication can be used synergistically to reduce control costs.

The setting is an extension of Witsenhausen’s counterexample where a rate-limited external channel connects the two controllers. Using a semi-deterministic version of the problem, we arrive at a binning-based strategy that can outperform the best known strategies by an arbitrarily large factor.

We also show that our binning-based strategy attains within a constant factor of the optimal cost for an asymptotically infinite-length version of the problem uniformly over all problem parameters and all rates on the external channel. For the scalar case, although our results yield approximate optimality for each fixed rate, we are unable to prove approximately-optimality uniformly over all rates.

I. INTRODUCTION

In his layered approach to design of decentralized control systems [1], Varaiya dedicates an entire layer for coordinating the actions of various agents. The question is: how can the agents build this coordination?

The most natural way to build coordination is through communication. To begin with, let us assume that the source and the channel have been specified explicitly. Even with this simplification, the general problem of multiterminal information theory has proven to be hard. The community therefore resorted to building a bottom-up theory that starts from Shannon’s toy problem of point-to-point communication [2]. The insights and tools obtained from this toy problem have helped immensely in the continuing development of multiterminal information theory.

A more accurate model of a dynamic control system is where the source can *evolve* with time, reflecting the impact of random perturbations and control actions. A counterpart of Shannon’s point-to-point toy problem that models evolution due to random perturbations is a problem of communicating an unstable Markov source across a channel. The problem is reasonably well understood [3]–[7], and again, building on the understanding for this toy problem, the community has begun exploring multicontroller problems [8], [9].

Do the above models encompass the possible ways of building coordination? Because these models are motivated by an architectural separation of estimation and control,

they do not model the impact of control actions in state evolution¹. Is this aspect important? Indeed, in decentralized control systems, it is often possible to modify what is to be communicated before communicating it. But at times, it is also often unclear what medium to use for communicating the message [12, Ch. 1]. That is, the sources and the channels may not be not as explicit as assumed in traditional communication models. To understand this issue, we informally define *implicit communication* to be one of the following two phenomena arising in decentralized control:

- Implicit message: the message itself is generated endogenously by the control system.
- Implicit channel: the system to be controlled is used as a channel to communicate.

The first phenomenon, that of implicit messages, poses an intellectual challenge to information theorists. How does one communicate a message that is endogenously generated, and hence can potentially be affected by the policy choice?

The second phenomenon, that of viewing the plant as an implicit communication channel, is challenging from a control theoretic standpoint. The control actions now perform a dual role — control of the system (*i.e.* minimizing immediate costs), and communication through the system (presumably to lower future costs).

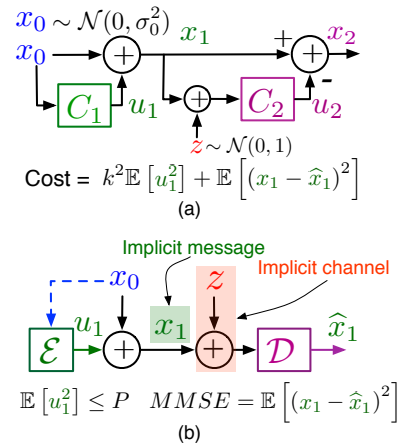


Fig. 1. The Witsenhausen counterexample, shown in (a) is the minimalist toy problem that exhibits the two notions of implicit communication, shown in (b), which is an equivalent representation [13].

¹Communication has also been used to build coordination by generating correlation between random variables [10], [11].

The counterpart of Shannon’s point-to-point problem in implicit communication is a decentralized two-controller problem called Witsenhausen’s counterexample [14] shown in Fig. 1. The message, state x_1 , is implicit, because it can be affected by the input u_1 of the first controller. The channel is implicit because the system state itself is used to communicate the message.

Despite substantial efforts of the community, the counterexample remains unsolved, and due to this the community could not build on the problem to address larger control networks of this nature. Recently, however, we showed that using the input to quantize the state (complemented by linear strategies) attains within a constant factor of the optimal cost uniformly over all problem parameters for the counterexample and its vector extensions [13], [15]. Building on this provable approximate-optimality we have been able to obtain similar results for many extensions to the counterexample² [12], [18]–[21].

When is it useful to communicate implicitly? To understand this, Ho and Chang [22] introduce the concept of partially-nested information structures. Their results can be interpreted in the following manner: when transmission delay across a noiseless, infinite-capacity external channel is smaller than the propagation delay of implicit communication, there is no advantage in communicating implicitly³. The system designer always has the engineering freedom to attach an external channel. Can this external channel obviate the need to consider implicit communication?

In practice, however, the channel is *never* perfect. In [12, Ch. 1], we compare problems of implicit and explicit communication where the respective channels are noisy. Assuming that the weights on quadratic costs on inputs and reconstruction are the same for implicit and explicit communication, we show that implicit communication can outperform various architectures of explicit communication by an arbitrarily large factor! The gain is due to implicit nature of the messages — the simplified source after actions of the controller can be communicated with much greater fidelity for the same power cost.

So an external channel should not be thought of as a substitute for implicit communication. But if an external channel is available, how should it be used in conjunction with implicit communication? To examine this, we consider an extension of Witsenhausen’s counterexample (shown in Fig. 2) where an external channel connects the two controllers. A special case when the channel is power constrained and has additive Gaussian noise has been considered by Shoarinejad *et al* [25] and Martins [26]. Shoarinejad *et al* observe that when the channel noise variance diverges to infinity, the problem approaches Witsenhausen’s counterexample, while

linear strategies are optimal in the limit of zero noise. Martins considers the case of finite noise variance and shows that in some cases, there exist nonlinear strategies that outperform all linear strategies.

In Section III, we provide an improvement over Martins’s strategy based on intuition obtained from a semi-deterministic version of the problem. In Section IV, we show that our strategy can outperform Martins’s strategy by an arbitrarily large factor. Because we interpret the problem as communication across two parallel channels — an implicit one and an explicit one — our strategy ensures that the information on implicit and explicit channels is essentially orthogonal. Without the implicit channel output, the message our strategy sends on the explicit channel would yield little information about the state. But the observations on the two channels jointly reveal a lot more about the state. This eliminates a redundancy in Martins’s strategies where the same message is duplicated over the implicit and explicit channels. In this sense, our results here also provide a justification for the utility of the concept of implicit communication.

For simplicity, we assume a fixed-rate noiseless external channel for most of the paper. In Section V-A, our binning strategy is proved to be approximately optimal for all problem parameters and all rates on the external channel for an *asymptotic vector version of the problem*. In Section V-B, using tools from large-deviation theory and KL-divergence, we obtain a lower bound on the costs for finite vector-lengths. Using this lower bound, we show that our improved strategy is within a constant factor of optimal for any fixed rate R_{ex} on the external channel for the *scalar case*. However, we do not yet have an approximately-optimal solution that is uniform over *external channel’s rate* — the ratio of upper and lower bounds diverges to infinity as $R_{ex} \rightarrow \infty$. We conclude in Section VI.

II. NOTATION AND PROBLEM STATEMENT

Vectors are denoted in bold, with a superscript to denote their length (e.g. \mathbf{x}^m is a vector of length m). Upper case is used for random variables or random vectors (except when denoting power P), while lower case symbols represent their realizations. Hats ($\hat{\cdot}$) on the top of random variables denote the estimates of the random variables. The block-diagram for the extension of Witsenhausen’s counterexample considered in this paper is shown in Fig. 2. $\mathcal{S}^m(r)$ denotes a sphere of radius r centered at the origin in m -dimensional Euclidean space \mathbb{R}^m . $\text{Vol}(A)$ denotes volume of the set A in \mathbb{R}^m .

A control strategy is denoted by $\gamma = (\gamma_1, \gamma_2)$, where γ_i is the function that maps the observations at C_i to the control inputs. The first controller observes $\mathbf{y}_1^m = \mathbf{x}_0^m$ and generates a control input \mathbf{u}_1^m that affects the system state, and a message $W \in \{0, 1, \dots, 2^{mR} - 1\}$ (that can also be viewed as a control input) for the second controller that is sent across a parallel channel.

The second controller observes $\mathbf{y}_2^m = \mathbf{x}_1^m + \mathbf{z}^m$, where \mathbf{z}^m is the disturbance, or the noise at the input of the second

²Approximate-optimality results of this nature have proven useful in information theory as well — building on smaller problems [16], significant understanding has been gained about larger systems [17].

³The same conclusion is drawn in work of Rotkowitz and Lall [23] (as an application of quadratic-invariance) and that of Yüksel [24] in more general frameworks.

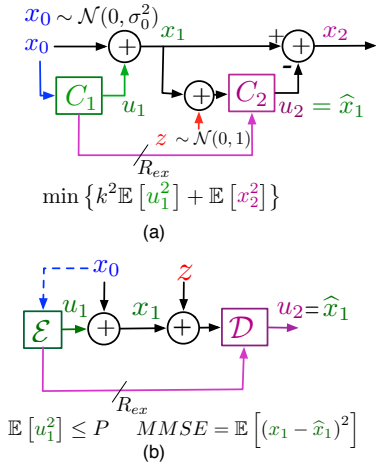


Fig. 2. The scalar version of the problem of implicit and explicit communication considered in this paper. An external channel connects the two controllers. In absence of implicit communication, the optimal strategy is linear. In absence explicit communication, an approximately-optimal strategy is quantization. Therefore, a natural strategy for this problem of implicit and explicit communication, proposed in [26], is to communicate linearly over the external channel, and use quantization over the implicit channel. Fig. 5 shows that our binning-based synergistic strategy can outperform this natural strategy by an arbitrarily large factor.

controller. It also observes perfectly the message W sent by the first controller. The total cost is a quadratic function of the state and the input given by:

$$J^{(\gamma)}(\mathbf{x}_0^m, \mathbf{z}^m) = \frac{1}{m} k^2 \|\mathbf{u}_1^m\|^2 + \frac{1}{m} \|\mathbf{x}_2^m\|^2, \quad (1)$$

where $\mathbf{u}_1^m = \gamma_1(\mathbf{x}_0^m)$, $\mathbf{x}_2^m = \mathbf{x}_0^m + \gamma_1(\mathbf{x}_0^m) - \mathbf{u}_1^m$ where $\mathbf{u}_2^m = \gamma_2(\mathbf{x}_0^m + \gamma_1(\mathbf{x}_0^m) + \mathbf{z}^m)$. The cost expression includes a division by the vector-length m to allow for natural comparisons between different vector-lengths.

Subscripts in expectation expressions denote the random variable being averaged over (e.g. $\mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}_G^m}[\cdot]$ denotes averaging over the initial state \mathbf{X}_0^m and the test noise \mathbf{Z}_G^m).

III. A SEMI-DETERMINISTIC MODEL

We extend the deterministic abstraction of Gaussian communication networks proposed in [17], [27] to a semi-deterministic model for our problem of Section II.

- Each system variable is represented in binary. For instance, in Fig. 3, the state is represented by $b_1 b_2 b_3 . b_4 b_5$, where b_1 is the highest order bit, and b_5 is the lowest.
- The location of the decimal point is determined by the signal-to-noise ratio (SNR), where signal refers to the state or input to which noise is added. It is given by $\lfloor \log_2(\text{SNR}) \rfloor - 1$. Noise can only affect the bit before the decimal point, and the bits following it that is, b_3 , b_4 and b_5 .
- The power of a random variable A , denoted by $\text{pow}(A)$ is defined as the highest order bit that is 1 among all the possible (binary-represented) values that A

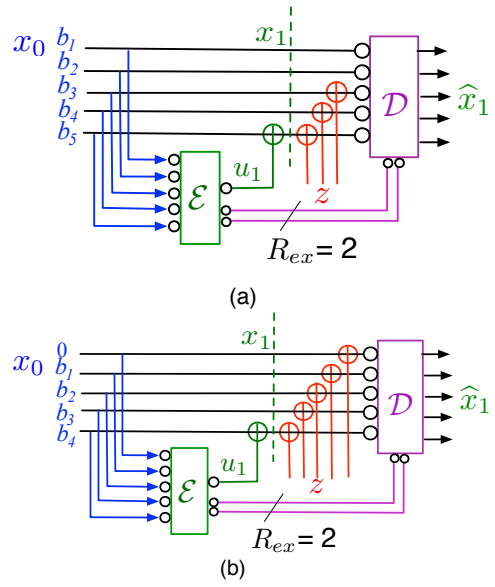


Fig. 3. A semi-deterministic model for the toy problem of implicit and explicit communication. An external channel (for this example, of capacity two bits) connects the two controllers. The case $\sigma_0^2 > 1$ is shown in (a), while $\sigma_0^2 < 1$ is shown in (b).

can take with nonzero probability⁴. For instance, if $A \in \{0.01, 0.11, 0.1, 0.001\}$, then A has the power $\text{pow}(A) = 0.1$.

- Additions/subtractions in the original model are replaced by bit-wise XORs. Noise is assumed to be iid Ber(0.5).
- The capacity of the external channel in the semi-deterministic version is the integer part (floor) of the capacity of the actual external channel.

We note here that unlike in the information-theoretic deterministic model of [17], the binary expansions in our model are valuable even after the decimal point (below noise level). Indeed, the model is not deterministic as random noise is modeled in the system⁵. This move from deterministic to semi-deterministic models is needed in decentralized control because one of the three roles of control actions is to improve the estimability of the state when observed noisily (the other two roles being control and communication). Since smart choices of control inputs can reduce the state uncertainty in the LQG model, a simplified model should allow for this possibility as well (the matter is discussed at length in [12]).

The semi-deterministic abstraction for our extension of Witsenhausen's counterexample is shown in Fig. 3. The original cost of $k^2 u_1^2 + x_2^2$ now becomes $k^2 \text{pow}(u_1) + \text{pow}(x_2)$. As in Fig. 2, the encoder for this semi-deterministic model observes x_0 noiselessly. Addition is represented by XORs, with the relative power of the terms to be added deciding which bits are affected. For instance, in Fig. 3, the power of the encoder input is sufficient to only affect the last bits of

⁴We note that our definition of $\text{pow}(A)$ is for clarity and convenience, and is far from unique in amongst good choices.

⁵An erasure-based deterministic model for noise can instead be used. This model also has the same optimal strategies.

the state x_0 . The noise bits are assumed to be distributed iid $\text{Ber}(0.5)$.

A. Optimal strategies for the semi-deterministic abstraction

We characterize the optimal tradeoff between the input power $\text{pow}(u_1)$ and the power in the MMSE error $\text{pow}(x_2)$. The minimum total cost problem is a convex dual of this problem, and can be obtained easily. Let the power of x_0 , $\text{pow}(x_0)$ be σ_0^2 . The noise power is assumed to be 1.

Case 1: $\sigma_0^2 > 1$.

This case is shown in Fig. 3(b). The bits b_1, b_2 are communicated noiselessly to the decoder, so the encoder does not need to communicate them implicitly or explicitly. The external channel has a capacity of two bits, so it can be used to communicate two of b_3, b_4 and b_5 . It should be used to communicate the higher-order bits among those corrupted by noise, *i.e.*, bits b_3, b_4 . The control input u_1 should be used to modify the lower-order bits (bit b_5 in Fig. 3). In the example shown, if $P < 0.01$, $\text{MMSE} = 0.01$, else $\text{MMSE} = 0$.

In this case (shown in Fig. 3(b)), the signal power is smaller than noise power. All the bits are therefore corrupted by noise, and nothing can be communicated across the implicit channel. In order for the decoder to be able to decode any bit in the representation of x_1 , it must either a) know the bit in advance (for instance, encoder can force the bit to 0), or b) be communicated the bit on the external channel. Since the encoder should use minimum power, it is clear that the most significant bits of the state (bits b_1, b_2 in Fig. 3(b)) should be communicated on the external channel. The encoder, if it has sufficient power, can then force the lower order bits (b_3, b_4 in Fig. 3(b)) of x_1 to zero. In the example shown in Fig. 3(b), if $P < 0.001$, $\text{MMSE} = 0.001$, else $\text{MMSE} = 0$.

B. What scheme does the semi-deterministic model suggest over reals?

A linear communication scheme over the external channel would correspond to communicating the highest-order bits of the state. The scheme for the semi-deterministic abstraction (Section III) communicates instead the highest order bits *that are at or below the noise level*. This suggests that the external channel should not be used in a linear fashion — the higher order bits are already known at the decoder. Instead, the external channel should be used to communicate bits that are corrupted by noise — more refined information about the state that is not already implicitly communicated by the noisy state itself.

The resulting scheme for the problem over reals is illustrated in Fig. 4. The encoder forces lower order bits of the state to zero, thereby truncating the binary expansion, or effectively quantizing the state into bins. The higher order bits that are corrupted by noise (b_3, b_4 in Fig. 3(a)) are communicated via the external channel. These bits can be thought of as representing the color, *i.e.* the bin index, of quantization bins, where set of $2^{R_{ex}}$ consecutive quantization-bins are labelled with $2^{R_{ex}}$ colors with a fixed order (with zero, for

instance, colored blue). The bin-index associated with the color of the bin is sent across the external channel. The decoder finds the quantization point nearest to y_2 that has the same bin-index as that received across the external channel.

The scheme is very similar to the binning scheme used for Wyner-Ziv coding of a Gaussian source with side information [28], which is not surprising because of similarity of our problem with the Wyner-Ziv formulation.

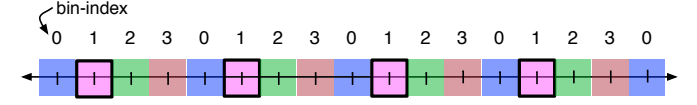


Fig. 4. The strategy intuited from the semi-deterministic model naturally yields a binning-based strategy for reals that leads to a synergistic use of implicit and explicit communication. The external channel get the decoder the bin-index (in this example, the index is 1). The more significant bits (coarse bin) is received from the implicit channel. Effectively, use of the external channel increases the distance between the ‘valid’ codewords by a factor of $2^{R_{ex}}$.

IV. GAUSSIAN EXTERNAL CHANNEL

A more realistic model of the external channel is a power constrained additive Gaussian noise channel, which was considered in [25], [26]. Without loss of generality, we assume that the noise in the external channel is also of variance 1.

At finite-lengths, an upper bound can be calculated using binning-based strategies. This binning-strategy turns out to outperform Martins’s strategy by a factor that diverges to infinity. The key is to choose the set of problems where the initial state variance and the power on the external channel, denoted by P_{ex} , are almost equal. In this case, a strategy that communicates the state on the external channel is not helpful — implicit channel can communicate the state at almost the same fidelity. Fig. 5 shows that fixing the relation $P_{ex} = \sigma_0^2$, as $\sigma_0^2 \rightarrow \infty$, the ratio of costs attained by the binning strategy to that attained by Martins’s strategy diverges to infinity.

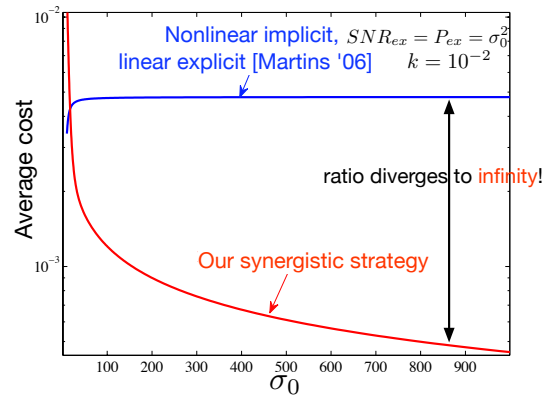


Fig. 5. If the SNR on the external channel is made to scale with SNR of the initial state, then our binning-based strategy outperforms strategy in [26] by a factor that diverges to infinity.

V. ASYMPTOTIC AND SCALAR VERSIONS OF THE PROBLEM

A. Asymptotic version

We now show that the binning strategy of Section III is approximately-optimal in the limit of infinite-lengths.

Theorem 1: For the extension of Witsenhausen's counterexample with an external channel connecting the two controllers,

$$\inf_{P \geq 0} k^2 P + \left(\left(\sqrt{\kappa_{new}} - \sqrt{P} \right)^+ \right)^2 \leq \bar{\mathcal{J}}_{opt} \leq \mu \inf_{P \geq 0} \left(\left(\sqrt{\kappa_{new}} - \sqrt{P} \right)^+ \right)^2,$$

where $\mu \leq 64$, $\kappa_{new} = \frac{\sigma_0^2 2^{-2R_{ex}}}{\bar{P}+1}$, where $\bar{P} = \left(\sigma_0 + \sqrt{P} \right)^2$ and the upper bound is achieved by binning-based quantization strategies. Numerical evaluation shows that $\mu < 8$.

Proof: Lower bound

We need the following lemma from [13, Lemma 3].

Lemma 1: For any three random vectors A , B and C ,

$$\sqrt{\mathbb{E}[\|B - C\|^2]} \geq \sqrt{\mathbb{E}[\|A - C\|^2]} - \sqrt{\mathbb{E}[\|A - B\|^2]}.$$

Proof: See [13]. ■

Substituting \mathbf{X}_0^m for A , \mathbf{X}_1^m for B , and \mathbf{U}_2^m for C in Lemma 1,

$$\begin{aligned} & \sqrt{\mathbb{E}[\|\mathbf{X}_1^m - \mathbf{U}_2^m\|^2]} \\ & \geq \sqrt{\mathbb{E}[\|\mathbf{X}_0^m - \mathbf{U}_2^m\|^2]} - \sqrt{\mathbb{E}[\|\mathbf{X}_0^m - \mathbf{X}_1^m\|^2]}. \end{aligned} \quad (2)$$

We wish to lower bound $\mathbb{E}[\|\mathbf{X}_1^m - \mathbf{U}_2^m\|]$. The second term on the RHS is smaller than \sqrt{mP} . Therefore, it suffices to lower bound the first term on the RHS of (2).

With what distortion can \mathbf{x}_0^m be communicated to the decoder? The capacity of the parallel channel is the sum of the two capacities $C_{sum} = R_{ex} + C_{implicit}$. The capacity $C_{implicit}$ is upper bounded by $\frac{1}{2} \log_2(1 + \bar{P})$ where $\bar{P} := \left(\sigma_0 + \sqrt{P} \right)^2$. Using Lemma 1, the distortion in reconstructing \mathbf{x}_0^m is lower bounded by

$$\begin{aligned} D(C_{sum}) &= \sigma_0^2 2^{-2C_{sum}} = \sigma_0^2 2^{-2R_{ex} - 2C_{implicit}} \\ &\geq \frac{\sigma_0^2 2^{-2R_{ex}}}{\bar{P} + 1} = \kappa_{new}. \end{aligned}$$

Thus the distortion in reconstructing \mathbf{x}_1^m is lower bounded by

$$\left(\left(\sqrt{\kappa_{new}} - \sqrt{P} \right)^+ \right)^2.$$

This proves the lower bound in Theorem 1.

Upper bound

Quantization: This strategy is used for $\sigma_0^2 > 1$. Quantize \mathbf{x}_0^m at rate $C_{sum} = R_{ex} + C_{implicit}$. Bin the codewords randomly into $2^{nR_{ex}}$ bins, and send the bin index on the

external channel. On the implicit channel, send the codeword closest to the vector \mathbf{x}_0^m .

The decoder looks at the bin-index on the external channel, and keeps only the codewords that correspond to the bin index. This subset of the codebook, which now corresponds to the set of valid codewords, has rate $C_{implicit}$. The required power P (which is the same as the distortion introduced in the source \mathbf{x}_0^m) is thus given by

$$\frac{1}{2} \log_2 \left(\frac{\sigma_0^2}{P} \right) \leq R_{ex} + \frac{1}{2} \log_2 (1 + \sigma_0^2 - P),$$

which yields the solution $P = \frac{(1 + \sigma_0^2) - \sqrt{(1 + \sigma_0^2)^2 - 4\sigma_0^2 2^{-2R_{ex}}}}{2}$ which is smaller than 1. Thus,

$$\begin{aligned} P &= \frac{(1 + \sigma_0^2) - \sqrt{(1 + \sigma_0^2)^2 - 4\sigma_0^2 2^{-2R_{ex}}}}{2} \\ &= \frac{1}{2} (1 + \sigma_0^2) \left(1 - \sqrt{1 - 4 \frac{\sigma_0^2}{(1 + \sigma_0^2)^2} 2^{-2R_{ex}}} \right). \end{aligned}$$

Now note that $\frac{\sigma_0^2}{(1 + \sigma_0^2)^2}$ is a decreasing function of σ_0^2 for $\sigma_0^2 > 1$. Thus, $\frac{\sigma_0^2}{(1 + \sigma_0^2)^2} < \frac{1}{4}$ for $\sigma_0^2 > 1$, and $1 - 4 \frac{\sigma_0^2}{(1 + \sigma_0^2)^2} 2^{-2R_{ex}} > 0$. Because $0 < 1 - 4 \frac{\sigma_0^2}{(1 + \sigma_0^2)^2} 2^{-2R_{ex}} < 1$,

$$\sqrt{1 - 4 \frac{\sigma_0^2}{(1 + \sigma_0^2)^2} 2^{-2R_{ex}}} \geq 1 - 4 \frac{\sigma_0^2}{(1 + \sigma_0^2)^2} 2^{-2R_{ex}},$$

and therefore

$$\begin{aligned} P &\leq \frac{1}{2} (1 + \sigma_0^2) \left(1 - \left(1 - 4 \frac{\sigma_0^2}{(1 + \sigma_0^2)^2} 2^{-2R_{ex}} \right) \right) \\ &= \frac{1}{2} (1 + \sigma_0^2) \left(4 \frac{\sigma_0^2}{(1 + \sigma_0^2)^2} 2^{-2R_{ex}} \right) \\ &= \frac{2\sigma_0^2}{1 + \sigma_0^2} 2^{-2R_{ex}} \leq 2 \times 2^{-2R_{ex}}. \end{aligned}$$

The other strategies that complement this binning strategy are the analogs of zero-forcing and zero-input.

Analog of the zero-forcing strategy The state \mathbf{x}_0^m is quantized using a rate-distortion codebook of $2^{mR_{ex}}$ points. The encoder sends the bin-index of the nearest quantization-point on the external channel. Instead of forcing the state all the way to zero, the input is used to force the state to the nearest quantization point. The required power is given by the distortion $\sigma_0^2 2^{-2R_{ex}}$. The decoder knows exactly which quantization point was used, so the second stage cost is zero. The total cost is therefore $k^2 \sigma_0^2 2^{-2R_{ex}}$.

Analog of Zero-input strategy

Case 1: $\sigma_0^2 \leq 4$.

Quantize the space of initial state realizations using a random codebook of rate R_{ex} , with the codeword elements chosen i.i.d $\mathcal{N}(0, \sigma_0^2(1 - 2^{-2R_{ex}}))$. Send the index of the nearest codeword on the external channel, and ignore the implicit channel. The asymptotic achieved distortion is given by the distortion-rate function of the Gaussian source $\sigma_0^2 2^{-2R_{ex}}$.

Case 2: $R_{ex} \leq 2$. Do not use the external channel. Perform an MMSE operation at the decoder on the state \mathbf{x}_0^m . The resulting error is $\frac{\sigma_0^2}{\sigma_0^2+1}$.

Case 3: $\sigma_0^2 > 4, R_{ex} > 2$.

Our proofs in this part follow those in [29]. Let $R_{code} = R_{ex} + \frac{1}{2} \log_2 \left(\frac{\sigma_0^2}{3} \right) - \epsilon$. A codebook of rate R_{code} is designed as follows. Each codeword is chosen randomly and uniformly inside a sphere centered at the origin and of radius $m\sqrt{\sigma_0^2 - D}$, where $D = \sigma_0^2 2^{-2R_{code}} = 3 \times 2^{-2(R_{ex}-\epsilon)}$. This is the attained asymptotic distortion when the codebook is used to represent⁶ \mathbf{x}_0^m .

Distribute the $2^{mR_{code}}$ points randomly into $2^{mR_{ex}}$ bins that are indexed $\{1, 2, \dots, 2^{mR_{ex}}\}$. The encoder chooses the codeword \mathbf{x}_{code}^m that is closest to the initial state. It sends the bin-index (say i) of the codeword across the external channel.

Let $\mathbf{z}_{code}^m = \mathbf{x}_0^m - \mathbf{x}_{code}^m$. The received signal $\mathbf{y}_2^m = \mathbf{x}_0^m + \mathbf{z}^m = \mathbf{x}_{code}^m + \mathbf{z}_{code}^m + \mathbf{z}^m$, which can be thought of as receiving a noisy version of codeword \mathbf{x}_{code}^m with a total noise of variance $D + 1$, since $\mathbf{z}_{code}^m \perp \mathbf{z}^m$.

The decoder receives the bin-index i on the external channel. Its goal is to find \mathbf{x}_{code}^m . It looks for a codeword from bin-index i in a sphere of radius $D + 1 + \epsilon$ around \mathbf{y}_2^m . We now show that it can find \mathbf{x}_{code}^m with probability converging to 1 as $m \rightarrow \infty$. A rigorous proof that MMSE also converges to zero can be obtained along the lines of proof in [13].

To prove that the error probability converges to zero, consider the total number of codewords that lie in the decoding sphere. This, on average, is bounded by

$$\begin{aligned} & \frac{2^{mR_{code}}}{Vol(S^m(m\sqrt{\sigma_0^2 - D + \epsilon}))} Vol(S^m(m\sqrt{D + 1 + \epsilon})) \\ &= \frac{2^{m(R_{ex} - \epsilon + \frac{1}{2} \log_2(\frac{\sigma_0^2}{3}))}}{Vol(S^m(m\sqrt{\sigma_0^2 - D + \epsilon}))} Vol(S^m(m\sqrt{D + 1 + \epsilon})) \\ &= \frac{2^{m(R_{ex} - \epsilon + \frac{1}{2} \log_2(\frac{\sigma_0^2}{3}))}}{(m\sqrt{\sigma_0^2 - D + \epsilon})^m} (m\sqrt{D + 1 + \epsilon})^m \\ &= 2^{m(R_{ex} - \epsilon)} 2^{\left(\frac{m}{2} \log_2 \left(\frac{\sigma_0^2(D+1+\epsilon)}{3(\sigma_0^2 - D + \epsilon)} \right)\right)}. \end{aligned}$$

Let us pick another codeword in the decoding sphere. Probability that this codeword has index i is $2^{-mR_{ex}}$. Using union bound, the probability that there exists another codeword in the decoding sphere of index i is bounded by

$$\begin{aligned} & 2^{-mR_{ex}} 2^{m(R_{ex} - \epsilon)} 2^{\left(\frac{m}{2} \log_2 \left(\frac{\sigma_0^2(D+1+\epsilon)}{3(\sigma_0^2 - D + \epsilon)} \right)\right)} \\ &= 2^{-m\epsilon} 2^{\left(\frac{m}{2} \log_2 \left(\frac{\sigma_0^2(D+1+\epsilon)}{3(\sigma_0^2 - D + \epsilon)} \right)\right)}. \end{aligned}$$

It now suffices to show that the second term converges to zero as $m \rightarrow \infty$. Since $D = 3 \times 2^{-2(R_{ex}-\epsilon)}$. Since $R_{ex} > 2$, $D < \frac{3}{4} \times 2^\epsilon < \frac{5}{6} - \epsilon$ for small enough ϵ . Since $\sigma_0^2 > 4$,

⁶In the limit of infinite block-lengths, average distortion attained by a uniform-distributed random-codebook and a Gaussian random-codebook of the same variance is the same [29].

$$D < \frac{5}{6} \frac{\sigma_0^2}{4} < \frac{\sigma_0^2}{4} + \epsilon,$$

$$\frac{\sigma_0^2(D+1+\epsilon)}{3(\sigma_0^2 - D + \epsilon)} < \frac{\sigma_0^2 \times (\frac{5}{6} + 1)}{3 \frac{3\sigma_0^2}{4}} = \frac{11}{6} = \frac{22}{27} < 1.$$

Thus the cost here is bounded by $3 \times 2^{-2(R_{ex}-\epsilon)}$ which is bounded by $4 \times 2^{-2R_{ex}}$ for small enough ϵ .

1) *Bounded ratios for the asymptotic problem:* The upper bound is the best of the vector-quantization bound, $2k^2 2^{-2R_{ex}}$, zero-forcing $k^2 \sigma_0^2 2^{-2R_{ex}}$, and zero-input bounds of $\sigma_0^2 2^{-2R_{ex}}$ and $4 \times 2^{-2R_{ex}}$.

Case 1: $P^* > \frac{2^{-2R_{ex}}}{16}$.

In this case, the lower bound is larger than $k^2 \frac{2^{-2R_{ex}}}{16}$. Using the upper bound of $4 \times 2^{-2R_{ex}}$, the ratio is smaller than 64.

Case 2: $P^* \leq \frac{2^{-2R_{ex}}}{16}, \sigma_0^2 \geq 1$.

Since $R_{ex} \geq 0$, $P^* \leq \frac{1}{16}$. Thus,

$$\kappa_{new} = \frac{\sigma_0^2 2^{-2R_{ex}}}{(\sigma_0 + \sqrt{P^*})^2 + 1} > \frac{1}{(1 + \frac{1}{4})^2 + 1} = \frac{16}{41} 2^{-2R_{ex}}.$$

Thus, the lower bound is greater than the MMSE which is larger than

$$\left(\sqrt{\frac{16}{41}} - \sqrt{\frac{1}{16}} \right)^2 2^{-2R_{ex}} \approx 0.14 \times 2^{-2R_{ex}}. \quad (3)$$

Using the upper bound of $4 \times 2^{-2R_{ex}}$, the ratio is smaller than 29.

Case 3: $P^* \leq \frac{2^{-2R_{ex}}}{16}, \sigma_0^2 < 1$.

If $P^* > \frac{\sigma_0^2 2^{-2R_{ex}}}{25}$, using the upper bound of $\sigma_0^2 2^{-2R_{ex}}$, the ratio is smaller than 25.

If $P^* \leq \frac{\sigma_0^2 2^{-2R_{ex}}}{25} < \frac{1}{25}$,

$$\begin{aligned} \kappa_{new} &= \frac{\sigma_0^2 2^{-2R_{ex}}}{(\sigma_0 + \sqrt{P^*})^2 + 1} \\ &\geq \frac{\sigma_0^2 2^{-2R_{ex}}}{(1 + \frac{1}{5})^2 + 1} \sigma_0^2 2^{-2R_{ex}} = \frac{25}{61} \sigma_0^2 2^{-2R_{ex}}. \end{aligned}$$

Thus, a lower bound on MMSE, and hence also on the total costs, is

$$\left(\sqrt{\frac{25}{61}} - \sqrt{\frac{1}{25}} \right)^2 \sigma_0^2 2^{-2R_{ex}} \approx 0.19 \sigma_0^2 2^{-2R_{ex}}.$$

Using the upper bound of $\sigma_0^2 2^{-2R_{ex}}$, the ratio is smaller than $\frac{1}{0.19} < 6$.

Numerical evaluations, shown in Fig. ??, show that the ratio is smaller than 8. ■

B. Scalar case

We first derive a lower bound for finite-vector lengths. The obtained bounds are tighter than those in Theorem 1 and depend explicitly on the vector length m .

Theorem 2 (Refined lower bound for finite-lengths):

For a finite-dimensional vector version of the problem, if for a strategy $\gamma(\cdot)$ the average power $\frac{1}{m} \mathbb{E}_{\mathbf{x}_0^m} [\|\mathbf{U}_1^m\|^2] = P$,

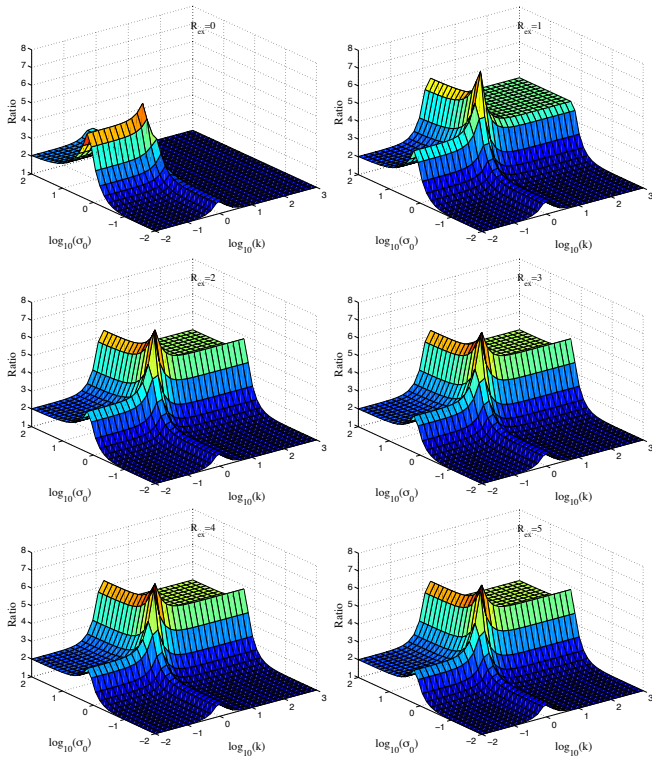


Fig. 6. The ratio of upper and lower bounds for the asymptotic problem are bounded by a factor of 8 for all k, σ_0 and R_{ex} .

the following lower bound holds on the second stage cost for any choice of $\sigma_G^2 \geq 1$ and $L > 0$

$$\bar{J}_2^{(\gamma)}(m, k^2, \sigma_0^2) \geq \eta(P, \sigma_0^2, \sigma_G^2, L).$$

where $\eta(P, \sigma_0^2, \sigma_G^2, L) =$

$$\frac{\sigma_G^m}{c_m(L)} \exp\left(-\frac{mL^2(\sigma_G^2 - 1)}{2}\right) \times \left(\left(\sqrt{\kappa_2(P, \sigma_0^2, \sigma_G^2, L)} - \sqrt{P}\right)^+\right)^2,$$

where $\kappa_2(P, \sigma_0^2, \sigma_G^2, L) :=$

$$\frac{\sigma_0^2 \sigma_G^2 2^{-2R_{ex}}}{c_m^{2/m}(L) e^{1-d_m(L)} \left((\sigma_0 + \sqrt{P})^2 + d_m(L) \sigma_G^2\right)},$$

$c_m(L) := \frac{1}{\Pr(\|\mathbf{Z}^m\|^2 \leq mL^2)} = (1 - \psi(m, L\sqrt{m}))^{-1}$,
 $d_m(L) := \frac{\Pr(\|\mathbf{Z}^{m+2}\|^2 \leq mL^2)}{\Pr(\|\mathbf{Z}^m\|^2 \leq mL^2)} = \frac{1 - \psi(m+2, L\sqrt{m})}{1 - \psi(m, L\sqrt{m})}$,
 $0 < d_m(L) < 1$, and $\psi(m, r) = \Pr(\|\mathbf{Z}^m\| \geq r)$. Thus the following lower bound holds on the total cost

$$\bar{J}_{\min}(m, k^2, \sigma_0^2) \geq \inf_{P \geq 0} k^2 P + \eta(P, \sigma_0^2, \sigma_G^2, L), \quad (4)$$

for any choice of $\sigma_G^2 \geq 1$ and $L > 0$ (the choice can depend on P). Further, these bounds are at least as tight as those of Theorem 1 for all values of k and σ_0^2 .

Proof: We remark that the only difference in this lower bound as compared to that in [15] is the term for R_{ex} in the

expression for κ_2 . The proof follows along the lines of that of [15, Theorem 3]. See Appendix I for the proof. ■

Theorem 3 (Upper bound for scalar case): An upper bound on costs for the scalar case is given by

$$\bar{J}_{opt} \leq \min\{\inf_{P \geq 2^{-2R_{ex}}} k^2 P + \psi(3, 2^{2R_{ex}} P), \\ ck^2 \sigma_0^2, c\sigma_0^2 2^{-2R_{ex}}, \\ a^2 2^{-2R_{ex}} + (1+a)^2 e^{-\frac{a^2}{2} + \frac{3}{2}(1+\ln(a^2))}\},$$

where⁷ $c \leq 2.72$ and $\psi(m, r)$ is defined in Theorem 2.

Proof: Just as for the asymptotic case, each term in the upper bound corresponds to a certain strategy.

Quantization

Divide the real line into uniform quantization bins of size \sqrt{P} . The quantization points are located at the center of these bins. Number consecutive bins $i \pmod{2^{R_{ex}}}$ starting with bin 0 which contains the origin. The encoder forces the initial state to the quantization point closest to the initial state, requiring a power of at most P . It also sends the index of the quantization bin on the external channel.

The decoder looks at the bin-index, and finds the nearest quantization point corresponding to the particular bin-index. The resulting MMSE error is given by $\mathbb{E}[z^2 \mathbb{1}_{\{|z| > 2^{R_{ex}} \sqrt{P}\}}]$. This is shown to equal $\psi(3, 2^{R_{ex}} \sqrt{P})$ in [15]. This yields the first term.

Analog of zero-forcing

Quantize the real-line using a quantization codebook of rate R_{ex} . The encoder forces x_0 to the nearest quantization point, and sends the index of the point to the decoder. The distortion is bounded by $2.72\sigma_0^2 2^{-2R_{ex}}$ [33]. The decoder has a perfect estimate of x_1 , thus the total cost is given by $k^2 c\sigma_0^2 2^{-2R_{ex}}$.

Analog of zero-input

As for the asymptotic case, we break this case into two strategies. For $\sigma_0^2 \leq 4$, we again use a quantization codebook of rate R_{ex} , but instead of zero-forcing the state, we take the distortion hit at the decoder. The resulting cost is $c\sigma_0^2 2^{-2R_{ex}}$.

For $\sigma_0^2 > 4$, we use a construct based on the idea of sending coarse information across the implicit channel, and fine information across the explicit channel. Divide the entire line into coarse quantization-bins of size $2a$. Divide each bin into $2^{R_{ex}}$ sub-bins, each of size $2a2^{-R_{ex}}$. Number each of the sub-bins in any sub-bin from $0, 1, \dots, 2^{R_{ex}}$.

The encoder send the index of the sub-bin in which x_0 lies across the external channel. The decoder decodes this sub-bin by finding the nearest sub-bin to the received output that has the same index as that received across the external channel.

If the decoder decodes the correct sub-bin, the error is bounded by $a^2 2^{-2R_{ex}}$. In the event when there is an error in decoding of the sub-bin, the error is bounded by $(|z| + a)^2$,

⁷This upper bound on c is the believed upper bound on the distortion-rate function $D_s(R) = c\sigma_0^2 2^{-2R}$ of a scalar Gaussian source. We have been unable to find a rigorous proof of this result, although the result is known to holds at high rates [31], and Lloyd's empirical results [32, Table VIII] suggest that the bound holds for all rates.

which averaged under the error event $|z| > a$ takes exactly the form of [15, Lemma 1]. Using that lemma, the MMSE in the error-event is bounded by

$$\begin{aligned} \mathbb{E}[(|z| + a)^2 \mathbb{1}_{\{|z| > a\}}] &\leq (\sqrt{\psi(3, a)} + a\sqrt{\psi(1, a)})^2 \\ &\stackrel{a > 1}{\leq} (1 + a)^2 e^{-\frac{a^2}{2} + \frac{3}{2}(1 + \ln(a^2))}. \end{aligned}$$

Thus the total MMSE is bounded by

$$MMSE \leq a^2 2^{-2R_{ex}} + (1 + a)^2 e^{-\frac{a^2}{2} + \frac{3}{2}(1 + \ln(a^2))}.$$

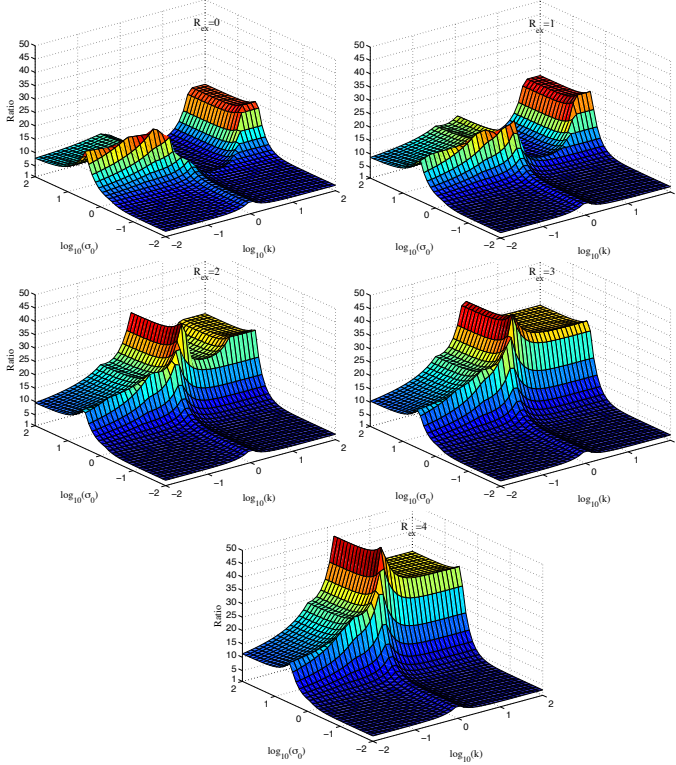


Fig. 7. Ratio of upper and lower bounds on the scalar problem for various values of R_{ex} . The ratio diverges to infinity as $R_{ex} \rightarrow \infty$.

Fig. 7 shows ratio of upper bound of Theorem 3 and lower bound of Theorem 2 in the (k, σ_0) -parameter space. Even though the ratio is bounded for each R_{ex} , it blows up as $R_{ex} \rightarrow \infty$.

VI. DISCUSSIONS AND CONCLUSIONS

The asymptotic result in Section V-A extends easily to an asymptotic version of problem with a Gaussian external channel (Section IV). This is because the error probability on the external channel converges to zero as the vector length $m \rightarrow \infty$ for any $R_{ex} < C_{ex}$, the capacity of the external channel, making it behave like a fixed rate external channel. Using large-deviation techniques, there is hope that the scalar problem with Gaussian external channel may also be solved approximately.

A rate-limited noiseless channel can be thought of as a model for limited-memory controllers. The problem of Fig. 2

can then be interpreted as a single controller system with finite memory. The problem considered here is also a toy problem that can design strategies for finite-memory controller problems.

ACKNOWLEDGMENTS

We would like to acknowledge most stimulating discussions with Tamer Başar while writing this paper. We also thank Aditya Mahajan for references, and Gireeja Ranade and Jiening Zhan for helpful discussions. Support of grant NSF CNS-0932410 is gratefully acknowledged.

APPENDIX I

PROOF OF LOWER BOUND FOR FINITE-LENGTH PROBLEM

Proof: From Theorem 1, for a given P , a lower bound on the average second stage cost is $\left(\left(\sqrt{\kappa_{new}} - \sqrt{P}\right)^+\right)^2$. We derive another lower bound that is equal to the expression for $\eta(P, \sigma_G^2, \sigma_L^2, L)$.

Define $S_L^G := \{\mathbf{z}^m : \|\mathbf{z}^m\|^2 \leq mL^2\sigma_G^2\}$ and use subscripts to denote which probability model is being used for the second stage observation noise. Z denotes white Gaussian of variance 1 while G denotes white Gaussian of variance $\sigma_G^2 \geq 1$.

$$\begin{aligned} &\mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}^m} \left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}^m) \right] \\ &= \int_{\mathbf{Z}^m} \int_{\mathbf{X}_0^m} J_2^{(\gamma)}(\mathbf{x}_0^m, \mathbf{z}^m) f_0(\mathbf{x}_0^m) f_Z(\mathbf{z}^m) d\mathbf{x}_0^m d\mathbf{z}^m \\ &\geq \int_{\mathbf{Z}^m \in S_L^G} \left(\int_{\mathbf{X}_0^m} J_2^{(\gamma)}(\mathbf{x}_0^m, \mathbf{z}^m) f_0(\mathbf{x}_0^m) d\mathbf{x}_0^m \right) f_Z(\mathbf{z}^m) d\mathbf{z}^m \\ &= \int_{\mathbf{Z}^m \in S_L^G} \left(\int_{\mathbf{X}_0^m} J_2^{(\gamma)}(\mathbf{x}_0^m, \mathbf{z}^m) f_0(\mathbf{x}_0^m) d\mathbf{x}_0^m \right) \frac{f_Z(\mathbf{z}^m)}{f_G(\mathbf{z}^m)} f_G(\mathbf{z}^m) d\mathbf{z}^m. \end{aligned} \quad (5)$$

The ratio of the two probability density functions is given by

$$\frac{f_Z(\mathbf{z}^m)}{f_G(\mathbf{z}^m)} = \frac{e^{-\frac{\|\mathbf{z}^m\|^2}{2}} \left(\sqrt{2\pi\sigma_G^2} \right)^m}{(\sqrt{2\pi})^m e^{-\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}}} = \sigma_G^m e^{-\frac{\|\mathbf{z}^m\|^2}{2} \left(1 - \frac{1}{\sigma_G^2} \right)}.$$

Observe that $\mathbf{z}^m \in S_L^G$, $\|\mathbf{z}^m\|^2 \leq mL^2\sigma_G^2$. Using $\sigma_G^2 \geq 1$, we obtain

$$\frac{f_Z(\mathbf{z}^m)}{f_G(\mathbf{z}^m)} \geq \sigma_G^m e^{-\frac{mL^2\sigma_G^2}{2} \left(1 - \frac{1}{\sigma_G^2} \right)} = \sigma_G^m e^{-\frac{mL^2(\sigma_G^2 - 1)}{2}}. \quad (6)$$

Using (5) and (6),

$$\begin{aligned} &\mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}^m} \left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}^m) \right] \\ &\geq \sigma_G^m e^{-\frac{mL^2(\sigma_G^2 - 1)}{2}} \times \\ &\int_{\mathbf{Z}^m \in S_L^G} \left(\int_{\mathbf{X}_0^m} J_2^{(\gamma)}(\mathbf{x}_0^m, \mathbf{z}^m) f_0(\mathbf{x}_0^m) d\mathbf{x}_0^m \right) f_G(\mathbf{z}^m) d\mathbf{z}^m \\ &= \sigma_G^m e^{-\frac{mL^2(\sigma_G^2 - 1)}{2}} \mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}_G^m} \left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}_G^m) \mathbb{1}_{\{\mathbf{Z}_G^m \in S_L^G\}} \right] \\ &= \sigma_G^m e^{-\frac{mL^2(\sigma_G^2 - 1)}{2}} \\ &\mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}_G^m} \left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}_G^m) | \mathbf{Z}_G^m \in S_L^G \right] \Pr(\mathbf{Z}_G^m \in S_L^G). \end{aligned} \quad (7)$$

It is shown in [15] that

$$\Pr(\mathbf{Z}_G^m \in \mathcal{S}_L^G) = \frac{1}{c_m(L)}. \quad (8)$$

From (7) and (8),

$$\begin{aligned} & \mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}_G^m} \left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}_G^m) \right] \\ & \geq \sigma_G^m e^{-\frac{mL^2(\sigma_G^2-1)}{2}} \mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}_G^m} \left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}_G^m) | \mathbf{Z}_G^m \in \mathcal{S}_L^G \right] \\ & \quad (1 - \psi(m, L\sqrt{m})) \\ & = \frac{\sigma_G^m e^{-\frac{mL^2(\sigma_G^2-1)}{2}}}{c_m(L)} \mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}_G^m} \left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}_G^m) | \mathbf{Z}_G^m \in \mathcal{S}_L^G \right] \quad (9) \end{aligned}$$

We now need the following lemma, which connects the new finite-length lower bound to the length-independent lower bound of Theorem 1.

Lemma 2:

$$\begin{aligned} & \mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}_G^m} \left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}_G^m) | \mathbf{Z}_G^m \in \mathcal{S}_L^G \right] \\ & \geq \left(\left(\sqrt{\kappa_2(P, \sigma_0^2, \sigma_G^2, L)} - \sqrt{P} \right)^+ \right)^2, \end{aligned}$$

for any $L > 0$.

Proof: This is a reworking of the proof for the asymptotic case to a channel which has a truncated Gaussian noise of (pre-truncation) variance σ_G^2 and a truncation for $|Z_G| \leq L$. Details are omitted due to space constraints. The derivation follows exactly the lines of [15, Lemma 2]. ■

The lower bound on the total average cost now follows from (9) and Lemma 2. ■

REFERENCES

- [1] P. Varaiya, "Towards a layered view of control," *Proceedings of the 36th IEEE Conference on Decision and Control (CDC)*, 1997.
- [2] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, Jul./Oct. 1948.
- [3] W. Wong and R. Brockett, "Systems with finite communication bandwidth constraints II: stabilization with limited information feedback," *IEEE Trans. Autom. Contr.*, pp. 1049–53, 1999.
- [4] V. Borkar and S. K. Mitter, "LQG control with communication constraints," in *Communications, Computation, Control, and Signal Processing: a Tribute to Thomas Kailath*. Norwell, MA: Kluwer Academic Publishers, 1997, pp. 365–373.
- [5] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2000.
- [6] A. Sahai, "Any-time information theory," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.
- [7] A. Matveev and A. Savkin, *Estimation and control over communication networks*. Springer, 2008.
- [8] S. Yüksel and T. Başar, "Communication constraints for stability in decentralized multi-sensor control systems," *IEEE Trans. Automat. Contr.*, submitted for publication.
- [9] —, "Optimal signaling policies for decentralized multicontroller stabilizability over communication channels," *Automatic Control, IEEE Transactions on*, vol. 52, no. 10, pp. 1969–1974, oct. 2007.
- [10] P. Cuff, "Communication requirements for generating correlated random variables," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*. IEEE, 2008, pp. 1393–1397.
- [11] P. Cuff, H. Permuter, and T. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.
- [12] P. Grover, "Control actions can speak louder than words," Ph.D. dissertation, UC Berkeley, Berkeley, CA, 2010.
- [13] P. Grover and A. Sahai, "Vector Witsenhausen counterexample as assisted interference suppression," *Special issue on Information Processing and Decision Making in Distributed Control Systems of the International Journal on Systems, Control and Communications (IJSCC)*, vol. 2, pp. 197–237, 2010.
- [14] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, Jan. 1968.
- [15] P. Grover, S. Park, and A. Sahai, "The finite-dimensional Witsenhausen counterexample," *Submitted to IEEE Transactions on Automatic Control*. Arxiv preprint arXiv:1003.0514, 2010.
- [16] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, Dec. 2008.
- [17] A. S. Avestimehr, "Wireless network information flow: A deterministic approach," Ph.D. dissertation, UC Berkeley, Berkeley, CA, 2008.
- [18] P. Grover, A. Wagner, and A. Sahai, "Information embedding meets distributed control," in *IEEE Information Theory Workshop (ITW)*, 2010, pp. 1–5.
- [19] P. Grover, S. Y. Park, and A. Sahai, "On the generalized Witsenhausen counterexample," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2009.
- [20] P. Grover and A. Sahai, "Distributed signal cancelation inspired by Witsenhausen's counterexample," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. Citeseer, 2010, pp. 151–155.
- [21] —, "Is Witsenhausen's counterexample a relevant toy?" *Proceedings of the 49th IEEE Conference on Decision and Control (CDC)*, Dec. 2010.
- [22] Y.-C. Ho and T. Chang, "Another look at the nonclassical information structure problem," *IEEE Trans. Automat. Contr.*, vol. 25, no. 3, pp. 537–540, 1980.
- [23] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control," *IEEE Trans. Automat. Contr.*, vol. 51, no. 2, pp. 1984–1996, Feb. 2006.
- [24] S. Yüksel, "Stochastic nestedness and the belief sharing information pattern," *IEEE Trans. Autom. Control*, pp. 2773–2786, 2009.
- [25] K. Shoarinejad, J. L. Speyer, and I. Kanellakopoulos, "A stochastic decentralized control problem with noisy communication," *SIAM Journal on Control and Optimization*, vol. 41, no. 3, pp. 975–990, 2002.
- [26] N. C. Martins, "Witsenhausen's counter example holds in the presence of side information," *Proceedings of the 45th IEEE Conference on Decision and Control (CDC)*, pp. 1111–1116, 2006.
- [27] A. S. Avestimehr, S. Diggavi, and D. N. C. Tse, "A deterministic approach to wireless relay networks," in *Proc. of the Allerton Conference on Communications, Control and Computing*, October 2007.
- [28] A. WYNER and J. ZIV, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, p. 1, 1976.
- [29] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge University Press, 2005.
- [30] P. Grover and A. Sahai, "A problem of implicit and explicit communication," in *Extended version of the paper to appear at the 48th Allerton Conference on Communication, Control, and Computing*, Oct. 2010. [Online]. Available: <http://www.eecs.berkeley.edu/~pulkit/files/Allerton10Online.pdf>
- [31] P. Panter and W. Dite, "Quantization distortion in pulse-count modulation with nonuniform spacing of levels," *Proceedings of the IRE*, vol. 39, no. 1, pp. 44–48, 1951.
- [32] S. Lloyd, "Least squares quantization in PCM," *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 129–137, mar. 1982.
- [33] R. Gray and D. Neuhoff, "Quantization," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2325–2383, oct. 1998.