The finite-dimensional Witsenhausen counterexample

Pulkit Grover, Anant Sahai and Se Yong Park

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

{pulkit, sahai, separk}@eecs.berkeley.edu

Abstract

Recently, a vector version of Witsenhausen's counterexample was considered and it was shown that in that limit of infinite vector length, certain quantization-based strategies are provably within a constant factor of the optimal cost for all possible problem parameters. In this paper, finite vector lengths are considered with the vector length being viewed as an additional problem parameter. By applying the "sphere-packing" philosophy, a lower bound to the optimal cost for this finite-length problem is derived that uses appropriate shadows of the infinite-length bound. We also introduce lattice-based quantization strategies for any finite length. Using the new finite-length lower bound, we show that good lattice-based strategies achieve within a constant factor of the optimal cost uniformly over all possible problem parameters, including the vector length. For Witsenhausen's original problem — the scalar case — regular lattice-based strategies are observed to numerically attain within a factor of 8 of the optimal cost.

I. Introduction

Distributed control problems have long proved challenging for control engineers. In 1968, Witsenhausen [1] gave a counterexample showing that even a seemingly simple distributed control problem can be hard to solve. For the counterexample, Witsenhausen chose a two-stage distributed LQG system and provided a nonlinear control strategy that outperforms all linear laws. It is now clear that the non-classical information pattern of Witsenhausen's problem makes it quite challenging[1]; the optimal strategy and the optimal costs for the problem are still unknown — non-convexity of the problem makes the search for

[1] In words of Yu-Chi Ho [2], "the simplest problem becomes the hardest problem."

an optimal strategy hard [3]–[5]. Discrete approximations of the problem [6] are even NP-complete[2] [7].

In the absence of a solution, research on the counterexample has bifurcated into two different directions. Since the problem is non-convex, a body of literature (e.g. [4] [5] [8] and the references therein) is dedicated to finding optimal solution by searching over the space of possible control actions for a few choices of problem parameters. Work in this direction has yielded considerable insight into addressing non-convex problems in general.

In the other direction the emphasis is on understanding the role of *implicit communication* in the counterexample. In distributed control, control actions not only attempt to reduce the immediate control costs, they can also communicate relevant information to other controllers to help them reduce costs. Various modifications on the counterexample help understand if it is misalignment of these two goals of control and communication that makes the problems hard [3], [9]–[12] (see [13] for a survey of other such modifications). Of particular interest is the work of Rotkowitz and Lall [11] which shows that with extremely fast external channels, design of optimal controllers is computationally efficient. This suggests that allowing for an external channel between the two controllers in Witsenhausen's counterexample might simplify the problem. However, Martins [14] shows that finding optimal solutions can be hard even in the presence of an external channel[3]. To design good distributed control strategies, it is therefore imperative to develop good understanding of the implicit communication in the counterexample.

Witsenhausen [1, Section 6] and Mitter and Sahai [15] aim at developing systematic constructions based on implicit communication. Witsenhausen's two-point quantization strategy is motivated from the optimal strategy for two-point symmetric distributions of the initial state [1, Section 5] and it outperforms linear

[2]More precisely, results in [7] imply that discrete approximations are NP-complete if the assumption of Gaussianity of the primitive random variables is relaxed. Further, it is also shown in [7] that with this relaxation, a polynomial time solution to the original continuous problem would imply $P = NP$, and thus conceptually the relaxed continuous problem is NP-complete (or harder).

[3]Martins shows that nonlinear strategies that do not even use the external channel can outperform linear ones even at high SNR on the external channel. Indeed, the best a linear strategy can do is to communicate the initial state as well as possible on the external channel. But if the uncertainty in the initial state is large, the external channel is only of limited help and there may be substantial advantage in having the controllers talk through the plant.

strategies for certain parameter choices. Mitter and Sahai [15] propose multipoint-quantization strategies that, depending on the problem parameters, can outperform scalar strategies by an arbitrarily-large factor.

The fact that nonlinear strategies can be arbitrarily better brings us to a question that has received little attention in the literature — how far are the proposed nonlinear strategies from the optimal? It is believed that the strategies of Lee, Lau and Ho [5] are close to optimal. In Section VI, we will see that these strategies can be viewed as an instance of the "dirty-paper coding" strategy in information theory, and quantify their advantage over pure quantization based strategies. Despite their improved performance, there is no guarantee that these strategies are indeed close to optimal[4]. Witsenhausen [1, Section 7] derived a lower bound on the costs that is loose in the interesting regimes of small k and large σ_0^2 [13], [16], and hence is insufficient to obtain any guarantee on the gap from optimality.

Towards obtaining such a guarantee, a strategic simplification of the problem was proposed in [13], [17] where we consider an asymptotically-long vector version of the problem. This problem is related to a toy communication problem that we call "Assisted Interference Suppression" (AIS) which is an extension of the dirty-paper coding (DPC) [18] model in information theory. There has been a burst of interest in extensions to DPC in information theory mainly along two lines of work — multi-antenna Gaussian channels, and the "cognitive-radio channel." For multi-antenna Gaussian channels, a problem of much theoretical and practical interest, DPC turns out to be the optimal strategy (see [19] and the references therein). The "cognitive radio channel" problem was formulated by Devroye *et al* [20]. This work has inspired many other works in asymmetric cooperation between nodes [21]–[25]. In our work [13], [17], we developed a new lower bound to the optimal performance of the vector Witsenhausen problem. Using this bound, we show that depending on the problem parameters, linear and vector-quantization based strategies attain within a factor of 4.45 of the optimal cost for all problem parameters in the limit of infinite vector length. Further, combinations of linear and DPC-based strategies attain within a factor 2 of the optimal

[4]The search in [5] is not exhaustive. The authors first find a good quantization-based solution. Inspired by piecewise linear strategies (from the neural networks based search of Baglietto *et al* [4]), each quantization step is broken into several small sub-steps to approximate a piecewise linear curve.

cost[5]. While a constant-factor result does not establish true optimality, such results are often helpful in the face of intractable problems like those that are otherwise NP-hard [27]. This constant-factor spirit has also been useful in understanding other stochastic control problems [28], [29] and in asymptotic analysis of problems in multiuser wireless communication [30], [31].

While the lower bound in [13] holds for all vector lengths, and hence for the scalar counterexample as well, the ratio of the costs attained by the strategies of [15] and the lower bound diverges in the limit $k \to 0$ and $\sigma_0 \to \infty$. This suggests that there is a significant finite-dimensional aspect of the problem that is being lost in the infinite-dimensional limit: either quantization-based strategies are bad, or the lower bound of [13] is very loose. This effect is elucidated in [16] by deriving a different lower bound that shows that quantization-based strategies indeed attain within a constant[6] factor of the optimal cost for Witsenhausen's original problem. The bound in [16] is in the spirit of Witsenhausen's original lower bound, but is more intricate. It captures the idea that observation noise can force a second-stage cost to be incurred unless the first stage cost is large.

In this paper, we revert to the line of attack based on the vector simplification of [13]. Building upon the vector lower bound, a new lower bound is derived in the spirit of information-theoretic bounds for finite-length communication problems (e.g. [32]–[35]). In particular, it extends the tools in [35] to a setting with unbounded distortion. The resulting lower bound (on numerical evaluation) shows that quantization-based strategies attain within a factor of 8 of the optimal cost for the scalar problem. To understand the significance of the result, consider the following. At $k = 0.01$ and $\sigma_0 = 500$, the cost attained by optimal linear scheme is close to 1. The cost attained by a quantization-based[7] scheme is 8.894×10^{-4}. Our new lower bound on the cost is 3.170×10^{-4}. Despite the small value of the lower bound, the ratio of the quantization-based upper bound and the lower bound for this choice of parameters is less than three!

[5]This factor was later improved to 1.3 in [26].

[6]The constant is large in [16], but as this paper shows, this is an artifact of the proof rather than reality.

[7]The quantization points are regularly spaced about 9.92 units apart. This results in a first stage cost of about 8.2×10^{-4} and a second stage cost of about 6.7×10^{-5}.

As a next step towards showing that approximately-optimal strategies can be found for all Witsenhausen-like problems, we consider the vector Witsenhausen problem with a finite vector length. The lower bounds derived here extend naturally to this case. For obtaining decent control strategies, we observe that the action of the first controller in the quantization-based strategy of [15] can be thought of as forcing the state to a point on a one-dimensional *lattice*. Extending this idea, we consider lattice-based strategies for finite dimensional spaces. We show that the class of lattice-based quantization strategies performs within a constant factor of optimal for any dimension. The approximation factor can be bounded by a constant uniformly over all choices of problem parameters, *including the dimension*.

The organization of the paper is as follows. In Section II, we define the vector Witsenhausen problem and introduce the notation. In Section III, lattice-based strategies for any vector length m are described. Lower bounds (that depend on m) on optimal costs are derived in Section IV. Section V shows that the ratio of the upper and the lower bounds is bounded uniformly over the dimension m and the other problem parameters. The conclusion in Section VI outlines directions of future research and speculates on the form of finite-dimensional strategies (following [13]) that we conjecture are optimal.

II. Notation and problem statement

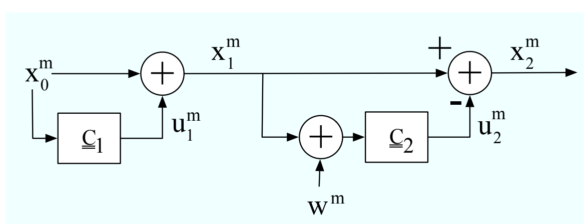

Fig. 1. Block-diagram for vector version of Witsenhausen's counterexample of length m.

Vectors are denoted in bold. Upper case tends to be used for random variables, while lower case symbols represent their realizations. $W(m, k^2, \sigma_0^2)$ denotes the vector version of Witsenhausen's problem of length m, defined as follows (shown in Fig. 1):

- The initial state $\mathbf{X}_0^m$ is Gaussian, distributed $\mathcal{N}(0, \sigma_0^2 \mathbb{I}_m)$, where $\mathbb{I}_m$ is the identity matrix of size

$m \times m$.

- The state transition functions describe the state evolution with time. The state transitions are linear:

$$\begin{aligned} \mathbf{X}_1^m &= \mathbf{X}_0^m + \mathbf{U}_1^m, \quad \text{and} \\ \mathbf{X}_2^m &= \mathbf{X}_1^m - \mathbf{U}_2^m. \end{aligned}$$

- The outputs observed by the controllers:

$$\begin{aligned} \mathbf{Y}_1^m &= \mathbf{X}_0^m, \quad \text{and} \\ \mathbf{Y}_2^m &= \mathbf{X}_1^m + \mathbf{Z}^m, \end{aligned} \tag{1}$$

where $\mathbf{Z}^m \sim \mathcal{N}(0, \sigma_Z^2 \mathbb{I}_m)$ is Gaussian distributed observation noise.

- The control objective is to minimize the expected cost, averaged over the random realizations of $\mathbf{X}_0^m$ and $\mathbf{Z}^m$. The total cost is a quadratic function of the state and the input given by the sum of two terms:

$$\begin{aligned} J_1(\mathbf{x}_1^m, \mathbf{u}_1^m) &= \frac{1}{m} k^2 \|\mathbf{u}_1^m\|^2, \text{ and} \\ J_2(\mathbf{x}_2^m, \mathbf{u}_2^m) &= \frac{1}{m} \|\mathbf{x}_2^m\|^2 \end{aligned}$$

where $\|\cdot\|$ denotes the usual Euclidean 2-norm. The cost expressions are normalized by the vector-length m to allow for natural comparisons between different vector-lengths. A control strategy is denoted by $\gamma = (\gamma_1, \gamma_2)$, where γ_i is the function that maps the observation $\mathbf{y}_i^m$ at $\underline{\underline{C_i}}$ to the control input $\mathbf{u}_i^m$. For a fixed γ, $\mathbf{x}_1^m = \mathbf{x}_0^m + \gamma_1(\mathbf{x}_0^m)$ is a function of $\mathbf{x}_0^m$. Thus the first stage cost can instead be written as a function $J_1^{(\gamma)}(\mathbf{x}_0^m) = J_1(\mathbf{x}_0^m + \gamma_1(\mathbf{x}_0^m), \gamma_1(\mathbf{x}_0^m))$ and the second stage cost can be written as $J_2^{(\gamma)}(\mathbf{x}_0^m, \mathbf{z}^m) = J_2(\mathbf{x}_0^m + \gamma_1(\mathbf{x}_0^m) - \gamma_2(\mathbf{x}_0^m + \gamma_1(\mathbf{x}_0^m) + \mathbf{z}^m), \gamma_2(\mathbf{x}_0^m + \gamma_1(\mathbf{x}_0^m) + \mathbf{z}^m))$.
For given γ, the expected costs (averaged over $\mathbf{x}_0^m$ and $\mathbf{z}^m$) are denoted by $\bar{J}^{(\gamma)}(m, k^2, \sigma_0^2)$ and $\bar{J}_i^{(\gamma)}(m, k^2, \sigma_0^2)$ for $i = 1, 2$. We define $\bar{J}_{\min}^{(\gamma)}(m, k^2, \sigma_0^2)$ as follows

$$\bar{J}_{\min}(m, k^2, \sigma_0^2) := \inf_{\gamma} \bar{J}^{(\gamma)}(m, k^2, \sigma_0^2). \tag{2}$$

- The *information pattern* represents the information available to each controller at the time it takes an action (it has implicitly been specified above). Following Witsenhausen's notation in [36], the information pattern for the vector problem is

$$\mathcal{Y}_1 = \{\mathbf{y}_1^m\};\ \mathcal{U}_1 = \varnothing,$$
$$\mathcal{Y}_2 = \{\mathbf{y}_2^m\};\ \mathcal{U}_2 = \varnothing.$$

Here $\mathcal{Y}_i$ denotes the information about the outputs in (1) available at the controller $i \in \{1, 2\}$. Similarly, $\mathcal{U}_i$ denotes the information about the previously applied inputs available at the i-th controller. Note that the second controller does not have knowledge of the output observed or the input applied at the first stage. This makes the information pattern non-classical (and non-nested), and the problem distributed.

We note that for the scalar case of $m = 1$, the problem is Witsenhausen's original counterexample [1].

Observe that scaling σ_0 and σ_Z by the same factor essentially does not change the problem — the solution can also be scaled by the same factor. Thus, without loss of generality, we assume that the variance of the Gaussian observation noise is $\sigma_Z^2 = 1$ (as is also assumed in [1]). The pdf of noise $\mathbf{Z}^m$ is denoted by $f_Z(\cdot)$. In our proof techniques, we also consider a hypothetical observation noise $\mathbf{Z}_G^m \sim \mathcal{N}(0, \sigma_G^2)$ with the variance $\sigma_G^2 \geq 1$. The pdf of this test noise is denoted by $f_G(\cdot)$. We use $\psi(m, r)$ to denote $\Pr(\|\mathbf{Z}^m\| \geq r)$ for $\mathbf{Z}^m \sim \mathcal{N}(0, \mathbb{I})$.

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. LATTICE-BASED QUANTIZATION STRATEGIES

We introduce lattice-based quantization strategies as the natural generalizations of scalar quantization-based strategies [15]. An introduction to lattices can be found in [38], [39]. Relevant definitions are reviewed below. $\mathcal{B}$ denotes the unit ball in $\mathbb{R}^m$.

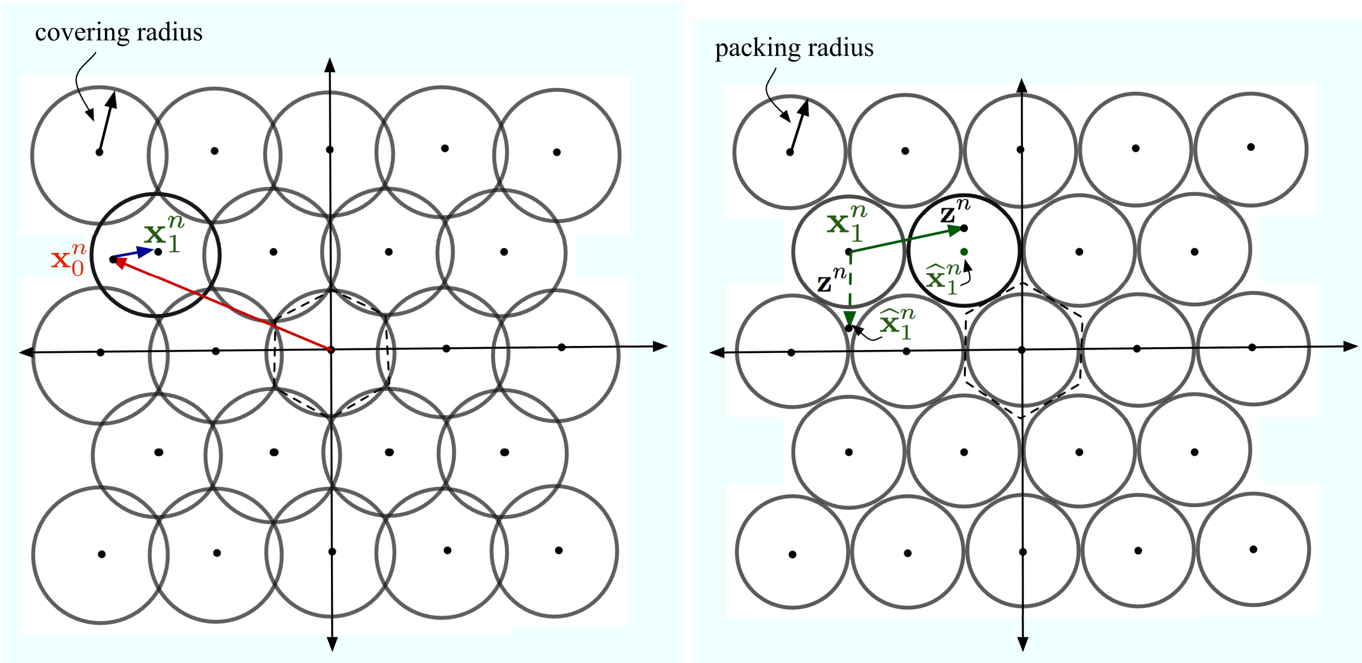

Fig. 2. Covering and packing for the 2-dimensional hexagonal lattice. The packing-covering ratio for this lattice is $\xi = \frac{2}{\sqrt{3}} \approx 1.15$ [37, Appendix C]. The first controller forces the initial state $\mathbf{x}_0^m$ to the lattice point nearest to it. The second controller estimates $\widehat{\mathbf{x}}_1^m$ to be a lattice point at the centre of the sphere if it falls in one of the packing spheres. Else it essentially gives up and estimates $\widehat{\mathbf{x}}_1^m = \mathbf{y}_2^m$, the received output itself. A hexagonal lattice-based scheme would perform better for the 2-D Witsenhausen problem than the square lattice (of $\xi = \sqrt{2} \approx 1.41$ [37, Appendix C]) because it has a smaller ξ.

Definition 1 (Lattice): An m-dimensional lattice Λ is a set of points in $\mathbb{R}^m$ such that if $\mathbf{x}^m, \mathbf{y}^m \in \Lambda$, then $\mathbf{x}^m + \mathbf{y}^m \in \Lambda$, and if $\mathbf{x}^m \in \Lambda$, then $-\mathbf{x}^m \in \Lambda$.

Definition 2 (Packing and packing radius): Given an m-dimensional lattice Λ and a radius r, the set $\Lambda + r\mathcal{B}$ is a *packing* of Euclidean m-space if for all points $\mathbf{x}^m, \mathbf{y}^m \in \Lambda$, $(\mathbf{x}^m + r\mathcal{B}) \bigcap (\mathbf{y}^m + r\mathcal{B}) = \emptyset$. The packing radius r_p is defined as $r_p := \sup\{r : \Lambda + r\mathcal{B} \text{ is a packing}\}$.

Definition 3 (Covering and covering radius): Given an m-dimensional lattice Λ and a radius r, the set $\Lambda + r\mathcal{B}$ is a *covering* of Euclidean m-space if $\mathbb{R}^m \subset \Lambda + r\mathcal{B}$. The covering radius r_c is defined as $r_c := \inf\{r : \Lambda + r\mathcal{B} \text{ is a covering}\}$.

Definition 4 (Packing-covering ratio): The *packing-covering ratio* (denoted by ξ) of a lattice Λ is the ratio of its covering radius to its packing radius, $\xi = \frac{r_c}{r_p}$.

Because it creates no ambiguity, we do not include the dimension m and the choice of lattice Λ in the notation of r_c, r_p and ξ, though these quantities depend on m and Λ.

For a given dimension m, a natural control strategy that uses a lattice Λ of covering radius r_c and packing radius r_p is as follows. The first controller uses the input $\mathbf{u}_1^m$ to force the state $\mathbf{x}_0^m$ to the quantization point nearest to $\mathbf{x}_0^m$. The second controller estimates $\mathbf{x}_1^m$ to be the quantization point nearest to $\mathbf{y}_2^m$. For analytical ease, we instead consider an inferior strategy where the second controller estimates $\mathbf{x}_1^m$ to be a lattice point only if the lattice point lies in a sphere of radius r_p around $\mathbf{y}_2^m$. If no lattice point exists in the sphere, the second controller estimates $\mathbf{x}_1^m$ to be $\mathbf{y}_2^m$, the received sequence itself. The actions $\gamma_1(\cdot)$ of $\underline{\underline{C_1}}$ and $\gamma_2(\cdot)$ of $\underline{\underline{C_2}}$ are therefore given by

$$\begin{aligned}\gamma_1(\mathbf{x}_0^m) &= -\mathbf{x}_0^m + \arg\min_{\mathbf{x}_1^m \in \Lambda} \|\mathbf{x}_1^m - \mathbf{x}_0^m\|^2, \\ \gamma_2(\mathbf{y}_2^m) &= \begin{cases} \widetilde{\mathbf{x}}_1^m & \text{if } \exists\, \widetilde{\mathbf{x}}_1^m \in \Lambda \text{ s.t. } \|\mathbf{y}_2^m - \widetilde{\mathbf{x}}_1^m\|^2 < r_p^2 \\ \mathbf{y}_2^m & \text{otherwise} \end{cases}.\end{aligned}$$

The event where there exists no such $\widetilde{\mathbf{x}}_1^m \in \Lambda$ is referred to as *decoding failure*. In the following, we denote $\gamma_2(\mathbf{y}_2^m)$ by $\widehat{\mathbf{x}}_1^m$, the estimate of $\mathbf{x}_1^m$.

Theorem 1: Using a lattice-based strategy (as described above) for $W(m, k^2, \sigma_0^2)$ with r_c and r_p the covering and the packing radius for the lattice, the total average cost is upper bounded by

$$\bar{J}^{(\gamma)}(m, k^2, \sigma_0^2) \le \inf_{P \ge 0} k^2 P + \left(\sqrt{\psi(m+2, r_p)} + \sqrt{\frac{P}{\xi^2}}\sqrt{\psi(m, r_p)}\right)^2,$$

where $\xi = \frac{r_c}{r_p}$ is the packing-covering ratio for the lattice, and $\psi(m, r) = \Pr(\|\mathbf{Z}^m\| \ge r)$. The following looser bound also holds

$$\bar{J}^{(\gamma)}(m, k^2, \sigma_0^2) \le \inf_{P > \xi^2} k^2 P + \left(1 + \sqrt{\frac{P}{\xi^2}}\right)^2 e^{-\frac{mP}{2\xi^2} + \frac{m+2}{2}\left(1 + \ln\left(\frac{P}{\xi^2}\right)\right)}.$$

Remark: The latter loose bound is useful for analytical manipulations when deriving bounds on the ratio of the upper and lower bounds in Section V.

Proof: Note that because Λ has a covering radius of r_c, $\|\mathbf{x}_1^m - \mathbf{x}_0^m\|^2 \le r_c^2$. Thus the first stage cost is bounded above by $\frac{1}{m}k^2 r_c^2$. A tighter bound can be provided for a specific lattice and finite m (for example, for $m = 1$, the first stage cost is approximately $k^2 \frac{r_c^2}{3}$ if $r_c^2 \ll \sigma_0^2$ because the distribution of $\mathbf{x}_0^m$ conditioned on it lying in any of the quantization bins is approximately uniform in the most likely bins).

For the second stage, observe that

$$\mathbb{E}_{\mathbf{X}_1^m, \mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2\right] = \mathbb{E}_{\mathbf{X}_1^m}\left[\mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 | \mathbf{X}_1^m\right]\right]. \tag{3}$$

Denote by $\mathcal{E}_m$ the event $\{\|\mathbf{Z}^m\|^2 \geq r_p^2\}$. Observe that under the event $\mathcal{E}_m^c$, $\widehat{\mathbf{X}}_1^m = \mathbf{X}_1^m$, resulting in a zero second-stage cost. Thus,

$$\begin{aligned}
\mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 | \mathbf{X}_1^m\right] &= \mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 \mathbf{1}_{\{\mathcal{E}_m\}} | \mathbf{X}_1^m\right] + \mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 \mathbf{1}_{\{\mathcal{E}_m^c\}} | \mathbf{X}_1^m\right] \\
&= \mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 \mathbf{1}_{\{\mathcal{E}_m\}} | \mathbf{X}_1^m\right].
\end{aligned}$$

We now bound the squared-error under the error event $\mathcal{E}_m$, when either $\mathbf{x}_1^m$ is decoded erroneously, or there is a decoding failure. If $\mathbf{x}_1^m$ is decoded erroneously to a lattice point $\widetilde{\mathbf{x}}_1^m \neq \mathbf{x}_1^m$, the squared-error can be bounded as follows

$$\begin{aligned}
\|\mathbf{x}_1^m - \widetilde{\mathbf{x}}_1^m\|^2 &= \|\mathbf{x}_1^m - \mathbf{y}_2^m + \mathbf{y}_2^m - \widetilde{\mathbf{x}}_1^m\|^2 \\
&\leq \left(\|\mathbf{x}_1^m - \mathbf{y}_2^m\| + \|\mathbf{y}_2^m - \widetilde{\mathbf{x}}_1^m\|\right)^2 \\
&\leq \left(\|\mathbf{z}^m\| + r_p\right)^2.
\end{aligned}$$

If $\mathbf{x}_1^m$ is decoded as $\mathbf{y}_2^m$, the squared-error is simply $\|\mathbf{z}^m\|^2$, which we also upper bound by $\left(\|\mathbf{z}^m\| + r_p\right)^2$. Thus, under event $\mathcal{E}_m$, the squared error $\|\mathbf{x}_1^m - \widehat{\mathbf{x}}_1^m\|^2$ is bounded above by $\left(\|\mathbf{z}^m\| + r_p\right)^2$, and hence

$$\begin{aligned}
\mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 | \mathbf{X}_1^m\right] &\leq \mathbb{E}_{\mathbf{Z}^m}\left[\left(\|\mathbf{Z}^m\| + r_p\right)^2 \mathbf{1}_{\{\mathcal{E}_m\}} | \mathbf{X}_1^m\right] \\
&\overset{(a)}{=} \mathbb{E}_{\mathbf{Z}^m}\left[\left(\|\mathbf{Z}^m\| + r_p\right)^2 \mathbf{1}_{\{\mathcal{E}_m\}}\right],
\end{aligned} \tag{4}$$

where (a) uses the fact that the pair $(\mathbf{Z}^m, \mathbf{1}_{\{\mathcal{E}_m\}})$ is independent of $\mathbf{X}_1^m$. Now, let $P = \frac{r_c^2}{m}$, so that the first stage cost is at most $k^2 P$. The following lemma helps us derive the upper bound.

Lemma 1: For a given lattice with $r_p^2 = \frac{r_c^2}{\xi^2} = \frac{mP}{\xi^2}$, the following bound holds

$$\frac{1}{m}\mathbb{E}_{\mathbf{Z}^m}\left[\left(\|\mathbf{Z}^m\| + r_p\right)^2 \mathbf{1}_{\{\mathcal{E}_m\}}\right] \leq \left(\sqrt{\psi(m+2, r_p)} + \sqrt{\frac{P}{\xi^2}}\sqrt{\psi(m, r_p)}\right)^2.$$

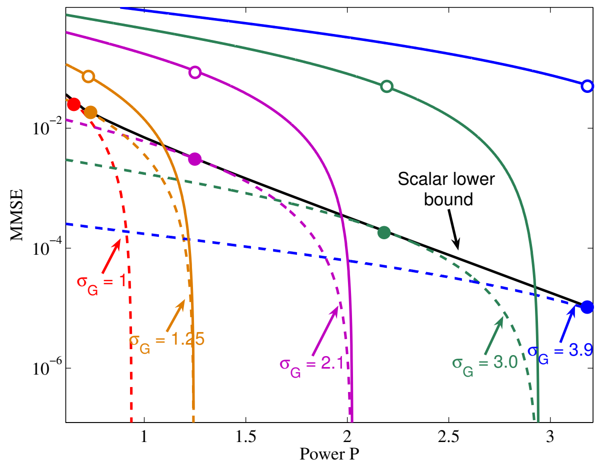

Fig. 3. A pictorial representation of the proof for the lower bound assuming $\sigma_0^2 = 30$. The solid curves show the vector lower bound of [13] for various values of observation noise variances, denoted by σ_G^2. Conceptually, multiplying these curves by the probability of that channel behavior yields the shadow curves for the particular σ_G^2, shown by dashed curves. The scalar lower bound is then obtained by taking the maximum of these shadow curves. The circles at points along the scalar bound curve indicate the optimizing value of σ_G for obtaining that point on the bound.

The following (looser) bound also holds as long as $P > \xi^2$,

$$\frac{1}{m}\mathbb{E}_{\mathbf{Z}^m}\left[\left(\|\mathbf{Z}^m\| + r_p\right)^2 \mathbf{1}_{\{\mathcal{E}_m\}}\right] \leq \left(1 + \sqrt{\frac{P}{\xi^2}}\right)^2 e^{-\frac{mP}{2\xi^2} + \frac{m+2}{2}\left(1 + \ln\left(\frac{P}{\xi^2}\right)\right)}.$$

Proof: See Appendix I. ■

The theorem now follows from (3), (4) and Lemma 1. ■

IV. Lower bounds on the cost

Bansal and Basar [3] use information theoretic techniques related to rate-distortion and channel capacity to show the optimality of linear strategies in a modified version of Witsenhausen's counterexample where the cost function does not contain a product of two decision variables. Following the same spirit, in [13] we derive the following lower bound for Witsenhausen's counterexample itself.

Theorem 2: For $W(m, k^2, \sigma_0^2)$, if for a strategy $\gamma(\cdot)$ the average power $\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{U}_1^m\|^2\right] = P$, the following lower bound holds on the second stage cost

$$\bar{J}_2^{(\gamma)}(m, k^2, \sigma_0^2) \geq \left(\left(\sqrt{\kappa(P, \sigma_0^2)} - \sqrt{P}\right)^+\right)^2,$$

where $(\cdot)^+$ is shorthand for $\max(\cdot, 0)$ and

$$\kappa(P, \sigma_0^2) = \frac{\sigma_0^2}{\sigma_0^2 + P + 2\sigma_0\sqrt{P} + 1}. \tag{5}$$

The following lower bound thus holds on the total cost

$$\bar{J}^{(\gamma)}(m, k^2, \sigma_0^2) \geq \inf_{P \geq 0} k^2 P + \left(\left(\sqrt{\kappa(P, \sigma_0^2)} - \sqrt{P} \right)^+ \right)^2.$$

Proof: We refer the reader to [13] for the full proof. We outline it here because these ideas are used in the derivation of the new lower bound in Theorem 3.

Using a triangle inequality argument, we show

$$\sqrt{\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}^m}\left[\|\mathbf{X}_0^m - \widehat{\mathbf{X}}_1^m\|^2\right]} \leq \sqrt{\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}^m}\left[\|\mathbf{X}_0^m - \mathbf{X}_1^m\|^2\right]} + \sqrt{\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2\right]}. \tag{6}$$

The first term on the RHS is $\sqrt{P}$. It therefore suffices to lower bound the term on the LHS to obtain a lower bound on $\mathbb{E}_{\mathbf{X}_0^m, \mathbf{Z}^m}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2\right]$. To that end, we interpret $\widehat{\mathbf{X}}_1^m$ as an estimate for $\mathbf{X}_0^m$, which is a problem of transmitting a source across a channel. For an iid Gaussian source to be transmitted across a memoryless power constrained additive noise Gaussian channel (with one channel use per source symbol), the optimal strategy that minimizes the mean-square error is merely scaling the source symbol so that the average power constraint is met [40]. The estimation at the second controller is then merely the linear MMSE estimation of $\mathbf{X}_0^m$, and the obtained MMSE is $\kappa(P, \sigma_0^2)$. The lemma now follows from (6). ■

Observe that the lower bound expression is the same for all vector lengths. In the following, sphere-packing style arguments [41], [42] are extended following [33]–[35] to a joint source-channel setting where the distortion measure is unbounded. The obtained bounds are tighter than those in Theorem 2 and depend on the vector length m.

Theorem 3: For $W(m, k^2, \sigma_0^2)$, if for a strategy $\gamma(\cdot)$ the average power $\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{U}_1^m\|^2\right] = P$, 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

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

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

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

$c_m(L) := \frac{1}{\Pr(\|\mathbf{Z}^m\|^2\leq mL^2)} = (1-\psi(m,L\sqrt{m}))^{-1}$, $d_m(L) := \frac{\Pr(\|\mathbf{Z}^{m+2}\|^2\leq mL^2)}{\Pr(\|\mathbf{Z}^m\|^2\leq mL^2)} = \frac{1-\psi(m+2,L\sqrt{m})}{1-\psi(m,L\sqrt{m})}$,

$0 < d_m(L) < 1$, and $\psi(m,r) = \Pr(\|\mathbf{Z}^m\| \geq r)$. Thus the following lower bound holds on the total cost

$$\bar{J}_{\min}(m,k^2,\sigma_0^2) \geq \inf_{P\geq 0} k^2P + \eta(P,\sigma_0^2,\sigma_G^2,L), \tag{7}$$

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 2 for all values of k and σ_0^2.

Proof: From Theorem 2, for a given P, a lower bound on the average second stage cost is $\left(\left(\sqrt{\kappa}-\sqrt{P}\right)^+\right)^2$. We derive another lower bound that is equal to the expression for $\eta(P,\sigma_0^2,\sigma_G^2,L)$. The high-level intuition behind this lower bound is presented in Fig. 3.

Define $\mathcal{S}_L^G := \{\mathbf{z}^m : \|\mathbf{z}^m\|^2 \leq mL^2\sigma_G^2\}$ 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 $\sigma_G^2 \geq 1$.

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

The ratio of the two probability density functions is given by

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

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

$$\frac{f_Z(\mathbf{z}^m)}{f_G(\mathbf{z}^m)} \geq \sigma_G^m e^{-\frac{mL^2\sigma_G^2}{2}\left(1-\frac{1}{\sigma_G^2}\right)} = \sigma_G^m e^{-\frac{mL^2(\sigma_G^2-1)}{2}}. \tag{8}$$

Thus,

$$\begin{aligned}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}^m}\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}^m)\right] &\geq \sigma_G^m e^{-\frac{mL^2(\sigma_G^2-1)}{2}}\int_{\mathbf{z}^m\in\mathcal{S}_L^G}\left(\int_{\mathbf{x}_0^m} J_2^{(\gamma)}(\mathbf{x}_0^m,\mathbf{z}^m)f_0(\mathbf{x}_0^m)d\mathbf{x}_0^m\right)f_G(\mathbf{z}^m)d\mathbf{z}^m\\ &= \sigma_G^m e^{-\frac{mL^2(\sigma_G^2-1)}{2}}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}_G^m)\mathbf{1}_{\{\mathbf{Z}_G^m\in\mathcal{S}_L^G\}}\right]\\ &= \sigma_G^m e^{-\frac{mL^2(\sigma_G^2-1)}{2}}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}_G^m)|\mathbf{Z}_G^m\in\mathcal{S}_L^G\right]\Pr(\mathbf{Z}_G^m\in\mathcal{S}_L^G). \qquad (9)\end{aligned}$$

Analyzing the probability term in (9),

$$\begin{aligned}\Pr(\mathbf{Z}_G^m\in\mathcal{S}_L^G) &= \Pr\left(\|\mathbf{Z}_G^m\|^2\leq mL^2\sigma_G^2\right)\\ &= \Pr\left(\left(\frac{\|\mathbf{Z}_G^m\|}{\sigma_G}\right)^2\leq mL^2\right)\\ &= 1-\Pr\left(\left(\frac{\|\mathbf{Z}_G^m\|}{\sigma_G}\right)^2> mL^2\right)\\ &= 1-\psi(m,L\sqrt{m})=\frac{1}{c_m(L)}, \qquad (10)\end{aligned}$$

because $\frac{\mathbf{Z}_G^m}{\sigma_G}\sim\mathcal{N}(0,\mathbb{I}_m)$. From (9) and (10),

$$\begin{aligned}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}^m}\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}^m)\right] &\geq \sigma_G^m e^{-\frac{mL^2(\sigma_G^2-1)}{2}}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}_G^m)|\mathbf{Z}_G^m\in\mathcal{S}_L^G\right](1-\psi(m,L\sqrt{m}))\\ &= \frac{\sigma_G^m e^{-\frac{mL^2(\sigma_G^2-1)}{2}}}{c_m(L)}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}_G^m)|\mathbf{Z}_G^m\in\mathcal{S}_L^G\right]. \qquad (11)\end{aligned}$$

We now need the following lemma, which connects the new finite-length lower bound to the infinite-length lower bound of [13].

Lemma 2:

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

for any $L>0$.

Proof: See Appendix II. ■

The lower bound on the total average cost now follows from (11) and Lemma 2.

We now verify that $d_m(L)\in(0,1)$. That $d_m(L)>0$ is clear from definition. $d_m(L)<1$ because $\{\mathbf{z}^{m+2}:\|\mathbf{z}^{m+2}\|^2\leq mL^2\sigma_G^2\}\subset\{\mathbf{z}^{m+2}:\|\mathbf{z}^m\|^2\leq mL^2\sigma_G^2\}$, *i.e.*, a sphere sits inside a cylinder.

Finally we verify that this new lower bound is at least as tight as the one in Theorem 2. Choosing $\sigma_G^2 = 1$ in the expression for $\eta(P, \sigma_0^2, \sigma_G^2, L)$,

$$\eta(P, \sigma_0^2, \sigma_G^2, L) \geq \sup_{L>0} \frac{1}{c_m(L)} \left(\left(\sqrt{\kappa_2(P, \sigma_0^2, 1, L)} - \sqrt{P} \right)^+ \right)^2.$$

Now notice that $c_m(L)$ and $d_m(L)$ converge to 1 as $L \to \infty$. Thus $\kappa_2(P, \sigma_0^2, 1, L) \stackrel{L\to\infty}{\longrightarrow} \kappa(P, \sigma_0^2)$ and therefore, $\eta(P, \sigma_0^2, \sigma_G^2, L)$ is lower bounded by $\left(\left(\sqrt{\kappa} - \sqrt{P} \right)^+ \right)^2$, the lower bound in Theorem 2.

■

V. Combination of linear and lattice-based strategies attain within a constant factor of the optimal cost

Theorem 4 (Constant-factor optimality): The costs for $W(m, k^2, \sigma_0^2)$ are bounded as follows

$$\inf_{P\geq 0} \sup_{\sigma_G^2 \geq 1, L>0} k^2 P + \eta(P, \sigma_0^2, \sigma_G^2, L) \leq \bar{J}_{min}(m, k^2, \sigma_0^2) \leq \mu \left(\inf_{P\geq 0} \sup_{\sigma_G^2 \geq 1, L>0} k^2 P + \eta(P, \sigma_0^2, \sigma_G^2, L) \right),$$

where $\mu = 100\xi^2$, ξ is the packing-covering ratio of any lattice in $\mathbb{R}^m$, and $\eta(\cdot)$ is as defined in Theorem 3. For any m, $\mu < 1600$. Further, depending on the (m, k^2, σ_0^2) values, the upper bound can be attained by lattice-based quantization strategies or linear strategies. For $m = 1$, a numerical calculation (MATLAB code available at [43]) shows that $\mu < 8$.

Proof: Let P^* denote the power P in the lower bound in Theorem 3. We show here that for any choice of P^*, the ratio of the upper and the lower bound is bounded.

Consider the two simple linear strategies of zero-forcing ($\mathbf{u}_1^m = -\mathbf{x}_0^m$) and zero-input ($\mathbf{u}_1^m = 0$) followed by LLSE estimation at $\underline{\underline{C}}_2$. It is easy to see [13] that the average cost attained using these two strategies is $k^2\sigma_0^2$ and $\frac{\sigma_0^2}{\sigma_0^2+1} < 1$ respectively. An upper bound is obtained using the best amongst the two linear strategies and the lattice-based quantization strategy.

Case 1: $P^* \geq \frac{\sigma_0^2}{100}$.

The first stage cost is larger than $k^2 \frac{\sigma_0^2}{100}$. Consider the upper bound of $k^2\sigma_0^2$ obtained by zero-forcing. The ratio of the upper bound and the lower bound is no larger than 100.

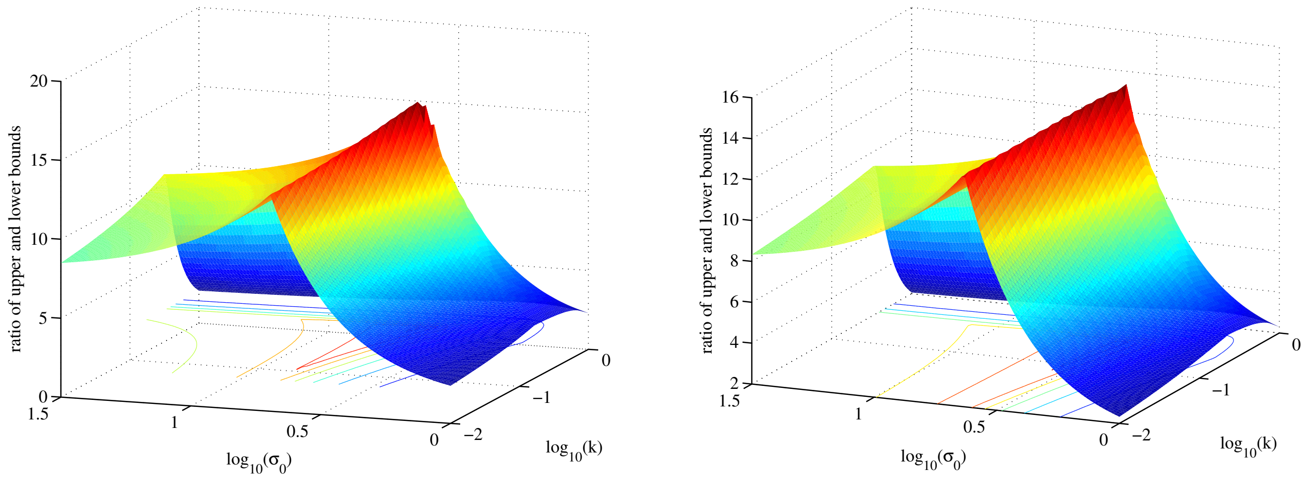

Fig. 4. The ratio of the upper and the lower bounds for the scalar Witsenhausen problem (top), and the 2-D Witsenhausen problem (bottom, using hexagonal lattice of $\xi = \frac{2}{\sqrt{3}}$) for a range of values of k and σ_0. The ratio is bounded above by 17 for the scalar problem, and by 14.75 for the 2-D problem.

Case 2: $P^* < \frac{\sigma_0^2}{100}$ and $\sigma_0^2 < 16$.

Using the bound from Theorem 2 (which is a special case of the bound in Theorem 3),

$$\begin{aligned}
\kappa &= \frac{\sigma_0^2}{(\sigma_0+\sqrt{P^*})^2+1} \\
&\overset{\left(P^*<\frac{\sigma_0^2}{100}\right)}{\geq} \frac{\sigma_0^2}{\sigma_0^2\left(1+\frac{1}{\sqrt{100}}\right)^2+1} \\
&\overset{(\sigma_0^2<16)}{\geq} \frac{\sigma_0^2}{16\left(1+\frac{1}{\sqrt{100}}\right)^2+1} = \frac{\sigma_0^2}{20.36} \geq \frac{\sigma_0^2}{21}.
\end{aligned}$$

Thus, for $\sigma_0^2 < 16$ and $P^* \leq \frac{\sigma_0^2}{100}$,

$$\begin{aligned}
\bar{J}_{min} &\geq \left((\sqrt{\kappa}-\sqrt{P^*})^+\right)^2 \geq \sigma_0^2\left(\frac{1}{\sqrt{21}}-\frac{1}{\sqrt{100}}\right)^2 \\
&\approx 0.014\sigma_0^2 \geq \frac{\sigma_0^2}{72}.
\end{aligned}$$

Using the zero-input upper bound of $\frac{\sigma_0^2}{\sigma_0^2+1}$, the ratio of the upper and lower bounds is at most $\frac{72}{\sigma_0^2+1} \leq 72$.

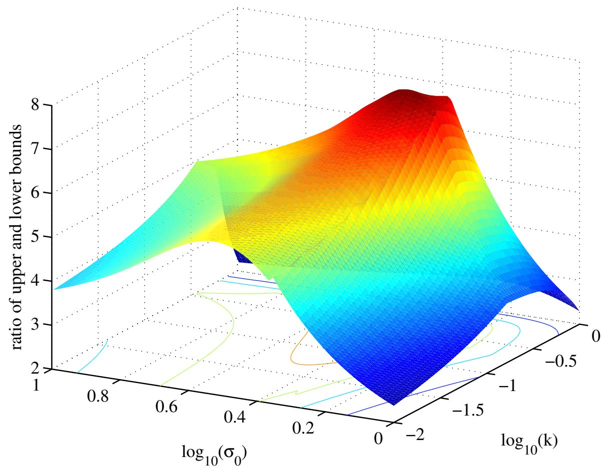

Fig. 5. An exact calculation of the first and second stage costs yields an improved maximum ratio smaller than 8 for the scalar Witsenhausen problem.

Case 3: $P^* \leq \frac{\sigma_0^2}{100}, \sigma_0^2 \geq 16, P^* \leq \frac{1}{2}$.

In this case,

$$\begin{aligned} \kappa &= \frac{\sigma_0^2}{(\sigma_0+\sqrt{P^*})^2+1} \\ &\overset{(P^*\leq\frac{1}{2})}{\geq} \frac{\sigma_0^2}{(\sigma_0+\sqrt{0.5})^2+1} \\ &\overset{(a)}{\geq} \frac{16}{(\sqrt{16}+\sqrt{0.5})^2+1} \approx 0.6909 \geq 0.69, \end{aligned}$$

where (a) uses $\sigma_0^2 \geq 16$ and the observation that $\frac{x^2}{(x+b)^2+1} = \frac{1}{\left(1+\frac{b}{x}\right)^2+\frac{1}{x^2}}$ is an increasing function of x for $x, b > 0$. Thus,

$$\left((\sqrt{\kappa}-\sqrt{P})^+\right)^2 \geq ((\sqrt{0.69}-\sqrt{0.5})^+)^2 \approx 0.0153 \geq 0.015.$$

Using the upper bound of $\frac{\sigma_0^2}{\sigma_0^2+1} < 1$, the ratio of the upper and the lower bounds is smaller than $\frac{1}{0.015} < 67$.

Case 4: $\sigma_0^2 > 16$, $\frac{1}{2} < P^* \leq \frac{\sigma_0^2}{100}$

Using $L = 2$ in the lower bound,

$$\begin{aligned} c_m(L) &= \frac{1}{\Pr(\|\mathbf{Z}^m\|^2 \leq mL^2)} = \frac{1}{1-\Pr(\|\mathbf{Z}^m\|^2 > mL^2)} \\ &\overset{\text{(Markov's ineq.)}}{\leq} \frac{1}{1-\frac{m}{mL^2}} \overset{(L=2)}{=} \frac{4}{3}, \end{aligned}$$

Similarly,

$$\begin{aligned} d_m(2) &= \frac{\Pr(\|\mathbf{Z}^{m+2}\|^2 \le mL^2)}{\Pr(\|\mathbf{Z}^{m}\|^2 \le mL^2)} \\ &\ge \Pr(\|\mathbf{Z}^{m+2}\|^2 \le mL^2) \\ &= 1 - \Pr(\|\mathbf{Z}^{m+2}\|^2 > mL^2) \\ &\overset{\text{(Markov's ineq.)}}{\ge} 1 - \frac{m+2}{mL^2} \\ &= 1 - \frac{1+\frac{2}{m}}{4} \overset{(m\ge1)}{\ge} 1 - \frac{3}{4} = \frac{1}{4}. \end{aligned}$$

In the bound, we are free to use any $\sigma_G^2 \ge 1$. Using $\sigma_G^2 = 6P^* > 1$,

$$\begin{aligned} \kappa_2 &= \frac{\sigma_G^2\sigma_0^2}{\left((\sigma_0+\sqrt{P^*})^2 + d_m(2)\sigma_G^2\right) c_m^{\frac{2}{m}}(2) e^{1-d_m(2)}} \\ &\overset{(a)}{\ge} \frac{6P^*\sigma_0^2}{\left((\sigma_0+\frac{\sigma_0}{10})^2 + \frac{6\sigma_0^2}{100}\right)\left(\frac{4}{3}\right)^{\frac{2}{m}} e^{\frac{3}{4}}} \overset{(m\ge1)}{\ge} 1.255P^*. \end{aligned}$$

where (a) uses $\sigma_G^2 = 6P^*$, $P^* < \frac{\sigma_0^2}{100}$, $c_m(2) \le \frac{4}{3}$ and $1 > d_m(2) \ge \frac{1}{4}$. Thus,

$$\left((\sqrt{\kappa_2} - \sqrt{P^*})^+\right)^2 \ge P^*(\sqrt{1.255}-1)^2 \ge \frac{P^*}{70}. \tag{12}$$

Now, using the lower bound on the total cost from Theorem 3, and substituting $L = 2$,

$$\begin{aligned} \bar{J}_{min}(m, k^2, \sigma_0^2) &\ge k^2P^* + \frac{\sigma_G^m}{c_m(2)} \exp\left(-\frac{mL^2(\sigma_G^2-1)}{2}\right)\left(\left(\sqrt{\kappa_2}-\sqrt{P^*}\right)^+\right)^2 \\ &\overset{(\sigma_G^2=6P^*)}{\ge} k^2P^* + \frac{(6P^*)^m}{c_m(2)} \exp\left(-\frac{4m(6P^*-1)}{2}\right) \frac{P^*}{70} \\ &\overset{(a)}{\ge} k^2P^* + \frac{3^m}{\frac{4}{3}} e^{2m} e^{-12P^*m} \frac{1}{70\times 2} \\ &\overset{(m\ge1)}{\ge} k^2P^* + \frac{3\times3\times e^2}{4\times70\times2} e^{-12mP^*} \\ &> k^2P^* + \frac{1}{9} e^{-12mP^*}, \end{aligned} \tag{13}$$

where (a) uses $c_m(2) \leq \frac{4}{3}$ and $P^* \geq \frac{1}{2}$. We loosen the lattice-based upper bound from Theorem 1 and bring it in a form similar to (13). Here, P is a part of the optimization:

$$\begin{aligned}
&\bar{J}_{min}(m, k^2, \sigma_0^2) \\
\leq \quad & \inf_{P>\xi^2} k^2 P + \left(1 + \sqrt{\frac{P}{\xi^2}}\right)^2 e^{-\frac{mP}{2\xi^2} + \frac{m+2}{2}\left(1+\ln\left(\frac{P}{\xi^2}\right)\right)} \\
\leq \quad & \inf_{P>\xi^2} k^2 P + \frac{1}{9} e^{-\frac{0.5mP}{\xi^2} + \frac{m+2}{2}\left(1+\ln\left(\frac{P}{\xi^2}\right)\right) + 2\ln\left(1+\sqrt{\frac{P}{\xi^2}}\right) + \ln(9)} \\
\leq \quad & \inf_{P>\xi^2} k^2 P + \frac{1}{9} e^{-m\left(\frac{0.5P}{\xi^2} - \frac{m+2}{2m}\left(1+\ln\left(\frac{P}{\xi^2}\right)\right) - \frac{2}{m}\ln\left(1+\sqrt{\frac{P}{\xi^2}}\right) - \frac{\ln(9)}{m}\right)} \\
= \quad & \inf_{P>\xi^2} k^2 P + \frac{1}{9} e^{-\frac{0.12mP}{\xi^2}} \times e^{-m\left(\frac{0.38P}{\xi^2} - \frac{1+\frac{2}{m}}{2}\left(1+\ln\left(\frac{P}{\xi^2}\right)\right) - \frac{2}{m}\ln\left(1+\sqrt{\frac{P}{\xi^2}}\right) - \frac{\ln(9)}{m}\right)} \\
\overset{(m\geq 1)}{\leq} \quad & \inf_{P>\xi^2} k^2 P + \frac{1}{9} e^{-\frac{0.12mP}{\xi^2}} e^{-m\left(\frac{0.38P}{\xi^2} - \frac{3}{2}\left(1+\ln\left(\frac{P}{\xi^2}\right)\right) - 2\ln\left(1+\sqrt{\frac{P}{\xi^2}}\right) - \ln(9)\right)} \\
\leq \quad & \inf_{P\geq 34\xi^2} k^2 P + \frac{1}{9} e^{-\frac{0.12mP}{\xi^2}}, \qquad (14)
\end{aligned}$$

where the last inequality follows from the fact that $\frac{0.38P}{\xi^2} > \frac{3}{2}\left(1 + \ln\left(\frac{P}{\xi^2}\right)\right) + 2\ln\left(1 + \sqrt{\frac{P}{\xi^2}}\right) + \ln(9)$ for $\frac{P}{\xi^2} > 34$. This can be checked easily by plotting it.[8]

Using $P = 100\xi^2 P^* \geq 50\xi^2 > 34\xi^2$ (since $P^* \geq \frac{1}{2}$) in (14),

$$\begin{aligned}
\bar{J}_{min}(m, k^2, \sigma_0^2) \quad &\leq \quad k^2 100\xi^2 P^* + \frac{1}{9} e^{-m\frac{0.12\times 100\xi^2 P^*}{\xi^2}} \\
&= \quad k^2 100\xi^2 P^* + \frac{1}{9} e^{-12mP^*}. \qquad (15)
\end{aligned}$$

Using (13) and (15), the ratio of the upper and the lower bounds is bounded for all m since

$$\mu \leq \frac{k^2 100\xi^2 P^* + \frac{1}{9}e^{-12mP^*}}{k^2 P^* + \frac{1}{9}e^{-12mP^*}} \leq \frac{k^2 100\xi^2 P^*}{k^2 P^*} = 100\xi^2. \qquad (16)$$

For $m = 1$, $\xi = 1$, and thus in the proof the ratio $\mu \leq 100$. For m large, $\xi \approx 2$ [39], and $\mu \lesssim 400$. For arbitrary m, using the recursive construction in [44, Theorem 8.18], $\xi \leq 4$, and thus $\mu \leq 1600$ regardless of m.

■

[8] It can also be verified symbolically by examining the expression $g(b) = 0.38b^2 - \frac{3}{2}(1 + \ln b^2) - 2\ln(1 + b) - \ln(9)$, taking its derivative $g'(b) = 0.76b - \frac{3}{b} - \frac{2}{1+b}$, and second derivative $g''(b) = 0.76 + \frac{3}{b^2} + \frac{2}{(1+b)^2} > 0$. Thus $g(\cdot)$ is convex-$\cup$. Further, $g'(\sqrt{34}) \approx 3.62 > 0$, and $g(\sqrt{34}) \approx 0.09$ and so $g(b) > 0$ whenever $b \geq \sqrt{34}$.

Though the proof above succeeds in showing that the ratio is uniformly bounded by a constant, it is not very insightful and the constant is large. However, since the underlying vector bound can be tightened (as shown in [26]), it is not worth improving the proof for increased elegance at this time. The important thing is that such a constant exists.

A numerical evaluation of the upper and lower bounds (of Theorem 1 and 3 respectively) shows that the ratio is smaller than 17 for $m = 1$ (see Fig. 4). A precise calculation of the cost of quantization strategy improves the upper bound to yield a maximum ratio smaller than 8 (see Fig. 5).

Simple grid lattice has a packing-covering ratio $\xi = \sqrt{m}$. Therefore, while the grid lattice has the best possible packing-covering ratio of 1 in the scalar case, it has a rather large packing covering ratio of $\sqrt{2}$ (≈ 1.41) for $m = 2$. On the other hand, a hexagonal lattice (for $m = 2$) has an improved packing-covering ratio of $\frac{2}{\sqrt{3}} \approx 1.15$. In contrast with $m = 1$, where the ratio of upper and lower bounds of Theorem 1 and 3 is approximately 17, a hexagonal lattice yields a ratio smaller than 14.75, despite having a larger packing-covering ratio. This is a consequence of tightening of the sphere-packing lower bound (Theorem 3) as m gets large[9].

VI. Discussions and Conclusions

Though lattice-based quantization strategies allow us to get within a constant factor of the optimal cost for the vector Witsenhausen problem, they are not optimal. This is known for the scalar [5] and the infinite-length case [13]. It is shown in [13] that the "slopey-quantization" strategy of Lee, Lau and Ho [5] that is believed to be very close to optimal in the scalar case can be viewed as an instance of a linear scaling followed by a dirty-paper coding (DPC) strategy. Such DPC-based strategies are also the best known strategies in the asymptotic infinite-dimensional case, requiring optimal power P to attain 0 asymptotic mean-square error in the estimation of $\mathbf{x}_1^m$, and attaining costs within a factor[10] of 1.3 of the

[9] Indeed, in the limit $m \to \infty$, the ratio of the asymptotic average costs attained by a vector-quantization strategy and the vector lower bound of Theorem 2 is bounded by 4.45 [13].

[10] Because of the looseness in the lower bound of [26], the ratio of the costs attained by DPC to the optimal cost is even smaller.

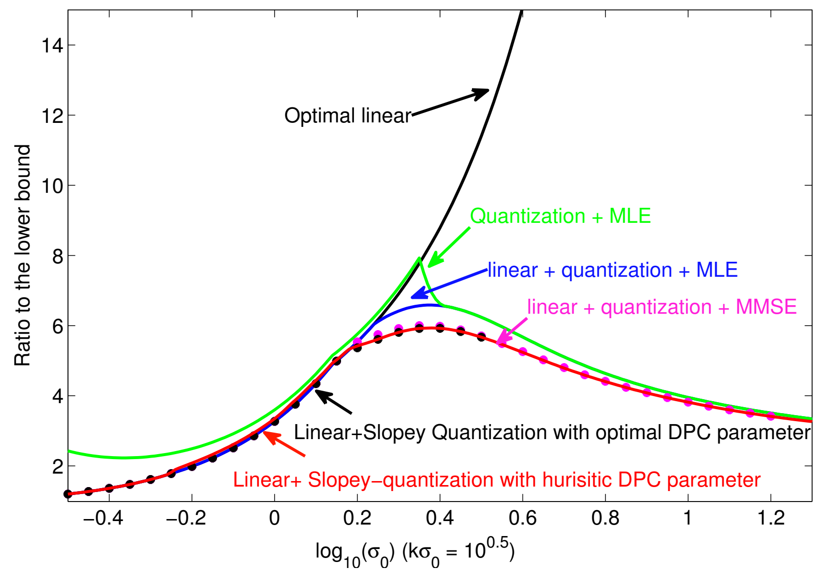

Fig. 6. Ratio of the achievable costs to the scalar lower bound along $k\sigma_0 = 10^{-0.5}$ for various strategies. Quantization with MMSE-estimation at the second controller performs visibly better than quantization with MLE, or even scaled MLE. For slopey-quantization with heuristic DPC-parameter, the parameter α in DPC-based scheme is borrowed from the infinite-length analysis. The figure suggests that along this path ($k\sigma_0 = \sqrt{10}$), the difference between optimal-DPC and heuristic DPC is not substantial. However, Fig. 7 shows that this is not true in general.

optimal [26] for all (k, σ_0^2). This leads us to conjecture that a DPC-based strategy would be optimal for finite-vector lengths as well.

It is natural to ask how much is there to gain using a DPC-based strategy over a simple quantization strategy? Notice that the DPC-strategy gains not only from the slopey quantization, but also from the MMSE-estimation at the second controller. In Fig. 6, we eliminate the latter advantage by considering first a quantization-based strategy with an appropriate scaling of the MLE so that it approximates the MMSE-estimation performance, and then the actual MMSE-estimation strategy. Along the curve $k\sigma_0 = \sqrt{10}$, there is significant gain in using this approximate-MMSE estimation over MLE, and further gain in using MMSE-estimation itself, bringing out a tradeoff between the complexity of the second controller and the performance.

From Fig. 6, DPC strategy performs only negligibly better than a quantization-based strategy with

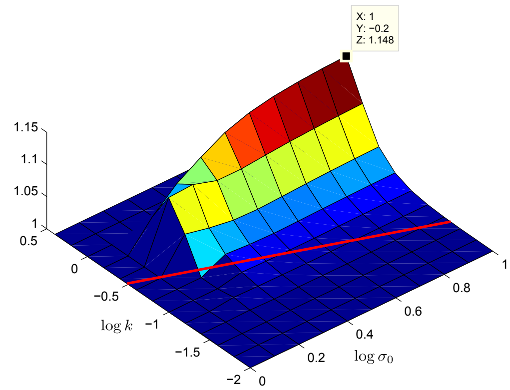

Fig. 7. Ratio of cost attained by linear+quantization (with MMSE decoding) to DPC with parameter α obtained by brute-force optimization. DPC can do up to 15% better than the optimal quantization strategy. Also the maximum is attained along $k \approx 0.6$ which is different from $k = 0.2$ of the benchmark problem [5].

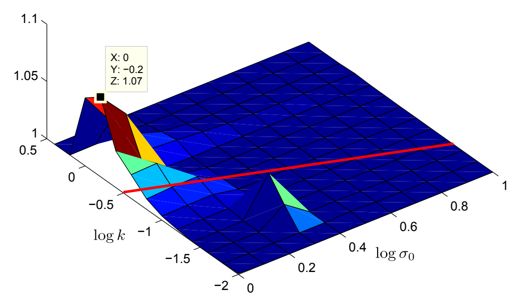

Fig. 8. Ratio of cost attained by linear+quantization to DPC with α borrowed from infinite-length optimization. Heuristic DPC does not outperform linear+quantization substantially.

MMSE estimation along $k\sigma_0 = \sqrt{10}$. Fig. 7 shows that this is not true in general. A DPC-based strategy can perform up to 15% better than a simple quantization-based scheme depending on the problem parameters. Interestingly, the advantage of using DPC at the benchmark case of $k = 0.2, \sigma_0 = 5$ [5], [8] is quite small! The maximum gain of about 15% is obtained at $k \approx 10^{-0.2} \approx 0.63$, and $\sigma_0 > 1$.

Given that there may be substantial advantage in using the DPC strategy, an interesting question is whether the DPC parameter α that optimizes the DPC-strategy's performance at infinite-lengths gives good performance for the scalar case as well. The answer to this question turns out to be negative.

Finally, it is questionable whether our strategies (quantization or DPC) that use uniform bin-size are almost as good as using nonuniform bins. Table I compares the cost obtained for uniform-bin strategies (plain quantization and DPC) with the cost attained in [5], which allows for nonuniform quantization bins. Clearly, the advantage in having nonuniform bins is not substantial, at least for this benchmark case. This observation is consistent with that in [8].

TABLE I

COSTS ATTAINED FOR THE BENCHMARK CASE OF $k = 0.2$, $\sigma_0 = 5$.

	linear+quantization	Slopey-quantization
Lee, Lau and Ho [5]	0.171394644442	0.167313205368
This paper	0.171533547912493	0.167365453179507

There are plenty of open problems that arise naturally. Both the lower and the upper bounds have room for improvement. The lower bound can be improved by tightening the lower bound on the infinite-length problem (one such tightening is performed in [26]) and obtaining corresponding finite-length results using the sphere-packing tools developed here.

Tightening the upper bound can be performed by using the DPC-based technique over lattices. Further, an exact analysis of the required first-stage power when using a lattice would yield an improvement (as pointed out earlier, for $m = 1$, $\frac{1}{m}k^2 r_c^2$ overestimates the required first-stage cost), especially for small m. Improved lattice designs with better packing-covering ratios would also improve the upper bound.

Perhaps a more significant set of open problems are the next steps in understanding more realistic versions of Witsenhausen's problem, specifically those that include costs on all the inputs and all the states [12], with noisy state evolution and noisy observations at both controllers. The hope is that solutions to these problems can then be used as the basis for provably-good nonlinear controller synthesis in larger distributed systems. Further, tools developed for solving these problems could help address multiuser problems in information theory, in the spirit of [45], [46].

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the National Science Foundation (CNS-403427, CNS-093240, CCF-0917212 and CCF-729122), Sumitomo Electric and Samsung. We thank Amin Gohari and Bobak Nazer for helpful discussions, and Gireeja Ranade for suggesting improvements in the paper.

APPENDIX I

PROOF OF LEMMA 1

$$\begin{aligned}
\mathbb{E}_{\mathbf{Z}^m}\left[(\|\mathbf{Z}^m\|+r_p)^2\mathbf{1}_{\{\mathcal{E}_m\}}\right] &= \mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{Z}^m\|^2\mathbf{1}_{\{\mathcal{E}_m\}}\right]+r_p^2\Pr(\mathcal{E}_m)+2r_p\mathbb{E}_{\mathbf{Z}^m}\left[\left(\mathbf{1}_{\{\mathcal{E}_m\}}\right)\left(\|\mathbf{Z}^m\|\mathbf{1}_{\{\mathcal{E}_m\}}\right)\right]\\
&\overset{(a)}{\leq} \mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{Z}^m\|^2\mathbf{1}_{\{\mathcal{E}_m\}}\right]+r_p^2\Pr(\mathcal{E}_m)+2r_p\sqrt{\mathbb{E}_{\mathbf{Z}^m}\left[\mathbf{1}_{\{\mathcal{E}_m\}}\right]}\sqrt{\mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{Z}^m\|^2\mathbf{1}_{\{\mathcal{E}_m\}}\right]}\\
&= \left(\sqrt{\mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{Z}^m\|^2\mathbf{1}_{\{\mathcal{E}_m\}}\right]}+r_p\sqrt{\Pr(\mathcal{E}_m)}\right)^2, \qquad (17)
\end{aligned}$$

where (a) uses the Cauchy-Schwartz inequality [47, Pg. 13].

We wish to express $\mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{Z}^m\|^2\mathbf{1}_{\{\mathcal{E}_m\}}\right]$ in terms of $\psi(m,r_p):=\Pr(\|\mathbf{Z}^m\|\geq r_p)=\int_{\|\mathbf{z}^m\|\geq r_p}\frac{e^{-\frac{\|\mathbf{z}^m\|^2}{2}}}{\left(\sqrt{2\pi}\right)^m}d\mathbf{z}^m$.

Denote by $\mathcal{A}_m(r):=\frac{2\pi^{\frac{m}{2}}r^{m-1}}{\Gamma\left(\frac{m}{2}\right)}$ the surface area of a sphere of radius r in $\mathbb{R}^m$ [48, Pg. 458], where $\Gamma(\cdot)$ is the Gamma-function satisfying $\Gamma(m)=(m-1)\Gamma(m-1)$, $\Gamma(1)=1$, and $\Gamma(\frac{1}{2})=\sqrt{\pi}$. Dividing the space $\mathbb{R}^m$ into shells of thickness dr and radii r,

$$\begin{aligned}
\mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{Z}^m\|^2\mathbf{1}_{\{\mathcal{E}_m\}}\right] &= \int_{\|\mathbf{z}^m\|\geq r_p}\|\mathbf{z}^m\|^2\frac{e^{-\frac{\|\mathbf{z}^m\|^2}{2}}}{\left(\sqrt{2\pi}\right)^m}d\mathbf{z}^m\\
&= \int_{r\geq r_p}r^2\frac{e^{-\frac{r^2}{2}}}{\left(\sqrt{2\pi}\right)^m}\mathcal{A}_m(r)dr\\
&= \int_{r\geq r_p}r^2\frac{e^{-\frac{r^2}{2}}}{\left(\sqrt{2\pi}\right)^m}\frac{2\pi^{\frac{m}{2}}r^{m-1}}{\Gamma\left(\frac{m}{2}\right)}dr\\
&= \int_{r\geq r_p}\frac{e^{-\frac{r^2}{2}}2\pi}{\left(\sqrt{2\pi}\right)^{m+2}}\frac{2\pi^{\frac{m+2}{2}}r^{m+1}}{\pi\frac{2}{m}\Gamma\left(\frac{m+2}{2}\right)}dr=m\psi(m+2,r_p). \qquad (18)
\end{aligned}$$

Using (17), (18), and $r_p=\sqrt{\frac{mP}{\xi^2}}$

$$\mathbb{E}_{\mathbf{Z}^m}\left[(\|\mathbf{Z}^m\|+r_p)^2\mathbf{1}_{\{\mathcal{E}_m\}}\right]\leq m\left(\sqrt{\psi(m+2,r_p)}+\sqrt{\frac{P}{\xi^2}}\sqrt{\psi(m,r_p)}\right)^2,$$

which yields the first part of Lemma 1. To obtain a closed-form upper bound we consider $P > \xi^2$. It suffices to bound $\psi(\cdot,\cdot)$.

$$\begin{aligned}\psi(m, r_p) &= \Pr(\|\mathbf{Z}^m\|^2 \geq r_p^2) = \Pr(\exp(\rho\sum_{i=1}^{m} Z_i^2) \geq \exp(\rho r_p^2)) \\ &\stackrel{(a)}{\leq} \mathbb{E}_{\mathbf{Z}^m}\left[\exp(\rho\sum_{i=1}^{m} Z_i^2)\right] e^{-\rho r_p^2} = \mathbb{E}_{Z_1}\left[\exp(\rho Z_1^2)\right]^m e^{-\rho r_p^2} \stackrel{(\text{for } 0<\rho<0.5)}{=} \frac{1}{(1-2\rho)^{\frac{m}{2}}} e^{-\rho r_p^2},\end{aligned}$$

where (a) follows from the Markov inequality, and the last inequality follows from the fact that the moment generating function of a standard χ_2^2 random variable is $\frac{1}{(1-2\rho)^{\frac{1}{2}}}$ for $\rho \in (0, 0.5)$ [49, Pg. 375]. Since this bound holds for any $\rho \in (0, 0.5)$, we choose the minimizing $\rho^* = \frac{1}{2}\left(1 - \frac{m}{r_p^2}\right)$. Since $r_p^2 = \frac{mP}{\xi^2}$, ρ^* is indeed in $(0, 0.5)$ as long as $P > \xi^2$. Thus,

$$\begin{aligned}\psi(m, r_p) &\leq \frac{1}{(1-2\rho^*)^{\frac{m}{2}}} e^{-\rho^* r_p^2} \\ &= \left(\frac{r_p^2}{m}\right)^{\frac{m}{2}} e^{-\frac{1}{2}\left(1-\frac{m}{r_p^2}\right) r_p^2} = e^{-\frac{r_p^2}{2} + \frac{m}{2} + \frac{m}{2}\ln\left(\frac{r_p^2}{m}\right)}.\end{aligned}$$

Using the substitutions $r_c^2 = mP$, $\xi = \frac{r_c}{r_p}$ and $r_p^2 = \frac{mP}{\xi^2}$,

$$\Pr(\mathcal{E}_m) = \psi(m, r_p) = \psi\left(m, \sqrt{\frac{mP}{\xi^2}}\right) \leq e^{-\frac{mP}{2\xi^2} + \frac{m}{2} + \frac{m}{2}\ln\left(\frac{P}{\xi^2}\right)}, \text{ and} \tag{19}$$

$$\mathbb{E}_{\mathbf{Z}^m}\left[\|\mathbf{Z}^m\|^2 \mathbf{1}_{\{\mathcal{E}_m\}}\right] \leq m\psi\left(m+2, \sqrt{\frac{mP}{\xi^2}}\right) \leq m e^{-\frac{mP}{2\xi^2} + \frac{m+2}{2} + \frac{m+2}{2}\ln\left(\frac{mP}{(m+2)\xi^2}\right)}. \tag{20}$$

From (17), (19) and (20),

$$\begin{aligned}\mathbb{E}_{\mathbf{Z}^m}\left[(\|\mathbf{Z}^m\| + r_p)^2 \mathbf{1}_{\{\mathcal{E}_m\}}\right] &\leq \left(\sqrt{m} e^{-\frac{mP}{4\xi^2} + \frac{m+2}{4} + \frac{m+2}{4}\ln\left(\frac{mP}{(m+2)\xi^2}\right)} \sqrt{\frac{mP}{\xi^2}} e^{-\frac{mP}{4\xi^2} + \frac{m}{4} + \frac{m}{4}\ln\left(\frac{P}{\xi^2}\right)}\right)^2 \\ &\stackrel{(\text{since } P>\xi^2)}{<} \left(\sqrt{m}\left(1 + \sqrt{\frac{P}{\xi^2}}\right) e^{-\frac{mP}{4\xi^2} + \frac{m+2}{4} + \frac{m+2}{4}\ln\left(\frac{P}{\xi^2}\right)}\right)^2 \\ &= m\left(1 + \sqrt{\frac{P}{\xi^2}}\right)^2 e^{-\frac{mP}{2\xi^2} + \frac{m+2}{2} + \frac{m+2}{2}\ln\left(\frac{P}{\xi^2}\right)}.\end{aligned}$$

APPENDIX II

PROOF OF LEMMA 2

The following lemma is taken from [13].

Lemma 3: For any three random variables A, B and C,

$$\mathbb{E}\left[\|B-C\|^2\right] \geq \left(\left(\sqrt{\mathbb{E}\left[\|A-C\|^2\right]}-\sqrt{\mathbb{E}\left[\|A-B\|^2\right]}\right)^+\right)^2.$$

Proof: See [13, Appendix II]. ■

Choosing $A = \mathbf{X}_0^m$, $B = \mathbf{X}_1^m$ and $C = \widehat{\mathbf{X}}_1^m$,

$$\begin{aligned}
& \mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}_G^m)|\mathbf{Z}_G^m \in \mathcal{S}_L^G\right] \\
= \ & \frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[\|\mathbf{X}_1^m-\widehat{\mathbf{X}}_1^m\|^2|\mathbf{Z}_G^m \in \mathcal{S}_L^G\right] \\
\geq \ & \left(\left(\sqrt{\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[\|\mathbf{X}_0^m-\widehat{\mathbf{X}}_1^m\|^2|\mathbf{Z}_G^m \in \mathcal{S}_L^G\right]}-\sqrt{\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[\|\mathbf{X}_0^m-\mathbf{X}_1^m\|^2|\mathbf{Z}_G^m \in \mathcal{S}_L^G\right]}\right)^+\right)^2 \\
= \ & \left(\left(\sqrt{\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[\|\mathbf{X}_0^m-\widehat{\mathbf{X}}_1^m\|^2|\mathbf{Z}_L^m \in \mathcal{S}_L^G\right]}-\sqrt{P}\right)^+\right)^2,
\end{aligned} \tag{21}$$

since $\mathbf{X}_0^m - \mathbf{X}_1^m = \mathbf{U}_1^m$ is independent of $\mathbf{Z}_G^m$ and $\mathbb{E}\left[\|\mathbf{U}_1^m\|^2\right] = mP$. Define $\mathbf{Y}_L^m := \mathbf{X}_1^m + \mathbf{Z}_L^m$ to be the output when the observation noise $\mathbf{Z}_L^m$ is distributed as a truncated Gaussian distribution:

$$f_{Z_L}(\mathbf{z}_L^m) = \begin{cases} c_m(L)\dfrac{e^{-\frac{\|\mathbf{z}_L^m\|^2}{2\sigma_G^2}}}{\left(\sqrt{2\pi\sigma_G^2}\right)^m} & \mathbf{z}_L^m \in \mathcal{S}_L^G \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Let the estimate at the second controller on observing $\mathbf{y}_L^m$ be denoted by $\widehat{\mathbf{X}}_L^m$. Then, by the definition of conditional expectations,

$$\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[\|\mathbf{X}_0^m-\widehat{\mathbf{X}}_1^m\|^2|\mathbf{Z}^m \in \mathcal{S}_L^G\right] = \mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[\|\mathbf{X}_0^m-\widehat{\mathbf{X}}_L^m\|^2\right]. \tag{23}$$

To get a lower bound, we now allow the controllers to optimize themselves with the additional knowledge that the observation noise $\mathbf{z}^m$ must fall in $\mathcal{S}_L^G$. In order to prevent the first controller from "cheating" and allocating different powers to the two events (*i.e.* $\mathbf{z}^m$ falling or not falling in $\mathcal{S}_L^G$), we enforce the constraint that the power P must not change with this additional knowledge. Since the controller's observation $\mathbf{X}_0^m$ is independent of $\mathbf{Z}^m$, this constraint is satisfied by the original controller (without the additional knowledge) as well, and hence the cost for the system with the additional knowledge is still a valid lower bound to that of the original system.

The rest of the proof uses ideas from channel coding and rate-distortion theorem [50, Ch. 13] from information theory. We view the problem as a problem of implicit communication from the first controller to the second. Notice that for a given $\gamma(\cdot)$, $\mathbf{X}_1^m$ is a function of $\mathbf{X}_0^m$, $\mathbf{Y}_L^m = \mathbf{X}_1^m + \mathbf{Z}_L^m$ is conditionally independent of $\mathbf{X}_0^m$ given $\mathbf{X}_1^m$ (since the noise $\mathbf{Z}_L^m$ is additive and independent of $\mathbf{X}_1^m$ and $\mathbf{X}_0^m$). Further, $\widehat{\mathbf{X}}_L^m$ is a function of $\mathbf{Y}_L^m$. Thus $\mathbf{X}_0^m - \mathbf{X}_1^m - \mathbf{Y}_L^m - \widehat{\mathbf{X}}_L^m$ form a Markov chain. Using the data-processing inequality [50, Pg. 33],

$$I(\mathbf{X}_0^m; \widehat{\mathbf{X}}_L^m) \leq I(\mathbf{X}_1^m; \mathbf{Y}_L^m), \tag{24}$$

where $I(A, B)$ is the expression for mutual information expression between two random variables A and B (see, for example, [50, Pg. 18, Pg. 231]). To estimate the distortion to which $\mathbf{X}_0^m$ can be communicated across this truncated Gaussian channel (which, in turn, helps us lower bound the MMSE in estimating $\mathbf{X}_1^m$), we need to upper bound the term on the RHS.

Lemma 4:

$$\frac{1}{m} I(\mathbf{X}_1^m; \mathbf{Y}_L^m) \leq \frac{1}{2}\log_2\left(\frac{e^{1-d_m(L)}(\bar{P} + d_m(L)\sigma_G^2)c_m^{\frac{2}{m}}(L)}{\sigma_G^2}\right).$$

Proof: We first obtain an upper bound the power of $\mathbf{X}_1^m$ (this bound is the same as that used in [13]):

$$\begin{aligned}
\mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{X}_1^m\|^2\right] &= \mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{X}_0^m + \mathbf{U}_1^m\|^2\right] \\
&= \mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{X}_0^m\|^2\right] + \mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{U}_1^m\|^2\right] + 2\mathbb{E}_{\mathbf{X}_0^m}\left[\mathbf{X}_0^{mT}\mathbf{U}_1^m\right] \\
&\overset{(a)}{\leq} \mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{X}_0^m\|^2\right] + \mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{U}_1^m\|^2\right] + 2\sqrt{\mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{X}_0^m\|^2\right]}\sqrt{\mathbb{E}_{\mathbf{X}_0^m}\left[\|\mathbf{U}_1^m\|^2\right]} \\
&\leq m(\sigma_0 + \sqrt{P})^2,
\end{aligned}$$

where (a) follows from the Cauchy-Schwartz inequality. We use the following definition of *differential entropy* $h(A)$ of a continuous random variable A [50, Pg. 224]:

$$h(A) = -\int_S f_A(a)\log_2\left(f_A(a)\right) da, \tag{25}$$

where $f_A(a)$ is the pdf of A, and S is the support set of A. Conditional differential entropy is defined similarly [50, Pg. 229].

Let $\bar{P} := (\sigma_0 + \sqrt{P})^2$. Now, $\mathbb{E}\left[Y_{L,i}^2\right] = \mathbb{E}\left[X_{1,i}^2\right] + \mathbb{E}\left[Z_{L,i}^2\right]$ (since $X_{1,i}$ is independent of $Z_{L,i}$ and by symmetry, $Z_{L,i}$ are zero mean random variables). Denote $\bar{P}_i = \mathbb{E}\left[X_{1,i}^2\right]$ and $\sigma_{G,i}^2 = \mathbb{E}\left[Z_{L,i}^2\right]$. In the following, we derive an upper bound $C_G^{(m)}$ on $\frac{1}{m}I(\mathbf{X}_1^m; \mathbf{Y}_L^m)$.

$$
\begin{aligned}
C_G^{(m)} &:= \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right] \leq m\bar{P}} \frac{1}{m} I(\mathbf{X}_1^m; \mathbf{Y}_L^m) \\
&\overset{(a)}{=} \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right] \leq m\bar{P}} \frac{1}{m} h(\mathbf{Y}_L^m) - \frac{1}{m} h(\mathbf{Y}_L^m | \mathbf{X}_1^m) \\
&= \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right] \leq m\bar{P}} \frac{1}{m} h(\mathbf{Y}_L^m) - \frac{1}{m} h(\mathbf{X}_1^m + \mathbf{Z}_L^m | \mathbf{X}_1^m) \\
&\overset{(b)}{=} \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right] \leq m\bar{P}} \frac{1}{m} h(\mathbf{Y}_L^m) - \frac{1}{m} h(\mathbf{Z}_L^m | \mathbf{X}_1^m) \\
&\overset{(c)}{=} \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right] \leq m\bar{P}} \frac{1}{m} h(\mathbf{Y}_L^m) - \frac{1}{m} h(\mathbf{Z}_L^m) \\
&\overset{(d)}{\leq} \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right] \leq m\bar{P}} \frac{1}{m} \sum_{i=1}^{m} h(Y_{L,i}) - \frac{1}{m} h(\mathbf{Z}_L^m) \\
&\overset{(e)}{\leq} \sup_{\bar{P}_i:\sum_{i=1}^m \bar{P}_i \leq m\bar{P}} \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \log_2\left(2\pi e(\bar{P}_i + \sigma_{G,i}^2)\right) - \frac{1}{m} h(\mathbf{Z}_L^m) \\
&\overset{(f)}{\leq} \frac{1}{2} \log_2\left(2\pi e(\bar{P} + d_m(L)\sigma_G^2)\right) - \frac{1}{m} h(\mathbf{Z}_L^m).
\end{aligned} \tag{26}
$$

Here, (a) follows from the definition of mutual information [50, Pg. 231], (b) follows from the fact that translation does not change the differential entropy [50, Pg. 233], (c) uses independence of $\mathbf{Z}_L^m$ and $\mathbf{X}_1^m$, and (d) uses the chain rule for differential entropy [50, Pg. 232], and the fact that conditioning reduces entropy [50, Pg. 232]. In (e), we used the fact that Gaussian random variables maximize differential entropy. The inequality (f) follows from the concavity of the $\log(\cdot)$ function and an application of Jensen's inequality [50, Pg. 25]. We also use the fact that $\frac{1}{m}\sum_{i=1}^m \sigma_{G,i}^2 = d_m(L)\sigma_G^2$, which can be proven as follows

$$\begin{aligned}
\frac{1}{m}\mathbb{E}\left[\sum_{i=1}^{m} Z_{L,i}^2\right] &\overset{\text{(using (22))}}{=} \frac{\sigma_G^2}{m}\int_{\mathbf{z}^m\in\mathcal{S}_L^G}\frac{\|\mathbf{z}^m\|^2}{\sigma_G^2}c_m(L)\frac{\exp\left(-\frac{\|\mathbf{z}_G^m\|^2}{2\sigma_G^2}\right)}{\left(\sqrt{2\pi\sigma_G^2}\right)^m}d\mathbf{z}_G^m \\
&= \frac{c_m(L)\sigma_G^2}{m}\mathbb{E}\left[\|\mathbf{Z}_G^m\|^2\mathbf{1}_{\left\{\|\mathbf{Z}_G^m\|\le\sqrt{mL^2\sigma_G^2}\right\}}\right] \\
&\overset{(\widetilde{\mathbf{Z}}^m:=\frac{\mathbf{Z}_G^m}{\sigma_G})}{=} \frac{c_m(L)\sigma_G^2}{m}\mathbb{E}\left[\|\widetilde{\mathbf{Z}}^m\|^2\mathbf{1}_{\left\{\|\widetilde{\mathbf{Z}}^m\|\le\sqrt{mL^2}\right\}}\right] \\
&= \frac{c_m(L)\sigma_G^2}{m}\left(\mathbb{E}\left[\|\widetilde{\mathbf{Z}}^m\|^2\right]-\mathbb{E}\left[\|\widetilde{\mathbf{Z}}^m\|^2\mathbf{1}_{\left\{\|\widetilde{\mathbf{Z}}^m\|>\sqrt{mL^2}\right\}}\right]\right) \\
&\overset{\text{(using (18))}}{=} \frac{c_m(L)\sigma_G^2}{m}\left(m-m\psi(m+2,\sqrt{mL^2})\right) \\
&= c_m(L)\left(1-\psi(m+2,L\sqrt{m})\right)\sigma_G^2 = d_m(L)\sigma_G^2.
\end{aligned}$$

We now compute $h(\mathbf{Z}_L^m)$

$$\begin{aligned}
h(\mathbf{Z}_L^m) &= \int_{\mathbf{z}^m\in\mathcal{S}_L^G} f_{Z_L}(\mathbf{z}^m)\log_2\left(\frac{1}{f_{Z_L}(\mathbf{z}^m)}\right)d\mathbf{z}^m \\
&= \int_{\mathbf{z}^m\in\mathcal{S}_L^G} f_{Z_L}(\mathbf{z}^m)\log_2\left(\frac{\left(\sqrt{2\pi\sigma_G^2}\right)^m}{c_m(L)e^{-\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}}}\right)d\mathbf{z}^m \\
&= -\log_2(c_m(L))+\frac{m}{2}\log_2\left(2\pi\sigma_G^2\right)+\int_{\mathbf{z}^m\in\mathcal{S}_L^G} c_m(L)f_G(\mathbf{z}^m)\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}\log_2(e)\,d\mathbf{z}^m. \qquad (27)
\end{aligned}$$

Analyzing the last term of (27),

$$\begin{aligned}
\int_{\mathbf{z}^m\in\mathcal{S}_L^G} c_m(L)f_G(\mathbf{z}^m)\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}\log_2(e)\,d\mathbf{z}^m &= \frac{\log_2(e)}{2\sigma_G^2}\int_{\mathbf{z}^m\in\mathcal{S}_L^G} c_m(L)\frac{e^{-\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}}}{\left(\sqrt{2\pi\sigma_G^2}\right)^m}\|\mathbf{z}^m\|^2 d\mathbf{z}^m \\
&= \frac{\log_2(e)}{2\sigma_G^2}\int_{\mathbf{z}^m} f_{Z_L}(\mathbf{z}^m)\|\mathbf{z}^m\|^2 d\mathbf{z}^m \\
&\overset{\text{(using (22))}}{=} \frac{\log_2(e)}{2\sigma_G^2}\mathbb{E}_G\left[\|\mathbf{Z}_L^m\|^2\right] = \frac{\log_2(e)}{2\sigma_G^2}\mathbb{E}_G\left[\sum_{i=1}^m Z_{L,i}^2\right] \\
&\overset{\text{(using (27))}}{=} \frac{\log_2(e)}{2\sigma_G^2}md_m(L)\sigma_G^2 = \frac{m\log_2\left(e^{d_m(L)}\right)}{2}. \qquad (28)
\end{aligned}$$

The expression $C_G^{(m)}$ can now be upper bounded using (26), (27) and (28) as follows.

$$\begin{aligned}
C_G^{(m)} &\le \frac{1}{2}\log_2\left(2\pi e(\bar{P}+d_m(L)\sigma_G^2)\right)+\frac{1}{m}\log_2(c_m(L))-\frac{1}{2}\log_2\left(2\pi\sigma_G^2\right)-\frac{1}{2}\log_2\left(e^{d_m(L)}\right) \\
&= \frac{1}{2}\log_2\left(2\pi e(\bar{P}+d_m(L)\sigma_G^2)\right)+\frac{1}{2}\log_2\left(c_m^{\frac{2}{m}}(L)\right)-\frac{1}{2}\log_2\left(2\pi\sigma_G^2\right)-\frac{1}{2}\log_2\left(e^{d_m(L)}\right) \\
&= \frac{1}{2}\log_2\left(\frac{2\pi e(\bar{P}+d_m(L)\sigma_G^2)c_m^{\frac{2}{m}}(L)}{2\pi\sigma_G^2 e^{d_m(L)}}\right) = \frac{1}{2}\log_2\left(\frac{e^{1-d_m(L)}(\bar{P}+d_m(L)\sigma_G^2)c_m^{\frac{2}{m}}(L)}{\sigma_G^2}\right). \qquad (29)
\end{aligned}$$

■

Now, recall that the rate-distortion function $D_m(R)$ for squared error distortion for source $\mathbf{X}_0^m$ and reconstruction $\widehat{\mathbf{X}}_L^m$ is,

$$D_m(R) := \inf_{\substack{p(\widehat{\mathbf{X}}_L^m|\mathbf{X}_0^m) \\ \frac{1}{m}I(\mathbf{X}_0^m;\widehat{\mathbf{X}}_L^m)\leq R}} \frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[\|\mathbf{X}_0^m - \widehat{\mathbf{X}}_L^m\|^2\right], \tag{30}$$

which is the dual of the rate-distortion function [50, Pg. 341]. Since $I(\mathbf{X}_0^m;\widehat{\mathbf{X}}_L^m) \leq mC_G^{(m)}$, using the converse to the rate distortion theorem [50, Pg. 349] and the upper bound on the mutual information represented by $C_G^{(m)}$,

$$\frac{1}{m}\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[\|\mathbf{X}_0^m - \widehat{\mathbf{X}}_L^m\|^2\right] \geq D_m(C_G^{(m)}). \tag{31}$$

Since the Gaussian source is iid, $D_m(R) = D(R)$, where $D(R) = \sigma_0^2 2^{-2R}$ is the distortion-rate function for a Gaussian source of variance σ_0^2 [50, Pg. 346]. Thus, using (21), (23) and (31),

$$\mathbb{E}_{\mathbf{X}_0^m,\mathbf{Z}_G^m}\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}^m)|\mathbf{Z}^m \in \mathcal{S}_L^G\right] \geq \left(\left(\sqrt{D(C_G^{(m)})} - \sqrt{P}\right)^+\right)^2.$$

Substituting the bound on $C_G^{(m)}$ from (29),

$$D(C_G^{(m)}) = \sigma_0^2 2^{-2C_G^{(m)}} = \frac{\sigma_0^2\sigma_G^2}{c_m^{\frac{2}{m}}(L)e^{1-d_m(L)}(\bar{P} + d_m(L)\sigma_G^2)}$$

Using (21), this completes the proof of the lemma. Notice that $c_m(L) \to 1$ and $d_m(L) \to 1$ for fixed m as $L \to \infty$, as well as for fixed $L > 1$ as $m \to \infty$. So the lower bound on $D(C_G^{(m)})$ approaches κ of Theorem 2 in both of these two limits.

REFERENCES

[1] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, Jan. 1968.

[2] Y.-C. Ho, "Review of the Witsenhausen problem," *Proceedings of the 47th IEEE Conference on Decision and Control (CDC)*, pp. 1611–1613, 2008.

[3] R. Bansal and T. Basar, "Stochastic teams with nonclassical information revisited: When is an affine control optimal?" *IEEE Trans. Automat. Contr.*, vol. 32, pp. 554–559, Jun. 1987.

[4] M. Baglietto, T. Parisini, and R. Zoppoli, “Nonlinear approximations for the solution of team optimal control problems,” *Proceedings of the IEEE Conference on Decision and Control (CDC)*, pp. 4592–4594, 1997.

[5] J. T. Lee, E. Lau, and Y.-C. L. Ho, “The Witsenhausen counterexample: A hierarchical search approach for nonconvex optimization problems,” *IEEE Trans. Automat. Contr.*, vol. 46, no. 3, pp. 382–397, 2001.

[6] 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.

[7] C. H. Papadimitriou and J. N. Tsitsiklis, “Intractable problems in control theory,” *SIAM Journal on Control and Optimization*, vol. 24, no. 4, pp. 639–654, 1986.

[8] N. Li, J. R. Marden, and J. S. Shamma, “Learning approaches to the Witsenhausen counterexample from a view of potential games,” *Proceedings of the 48th IEEE Conference on Decision and Control (CDC)*, 2009.

[9] T. Basar, “Variations on the theme of the Witsenhausen counterexample,” *Proceedings of the 47th IEEE Conference on Decision and Control (CDC)*, pp. 1614–1619, 2008.

[10] M. Rotkowitz, “On information structures, convexity, and linear optimality,” *Proceedings of the 47th IEEE Conference on Decision and Control (CDC)*, pp. 1642–1647, 2008.

[11] 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.

[12] 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.

[13] P. Grover and A. Sahai, “Vector Witsenhausen counterexample as assisted interference suppression,” *To appear in the special issue on Information Processing and Decision Making in Distributed Control Systems of the International Journal on Systems, Control and Communications (IJSCC)*, Sep. 2009. [Online]. Available: http://www.eecs.berkeley.edu/∼sahai/

[14] 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.

[15] S. K. Mitter and A. Sahai, “Information and control: Witsenhausen revisited,” in *Learning, Control and Hybrid Systems: Lecture Notes in Control and Information Sciences 241*, Y. Yamamoto and S. Hara, Eds. New York, NY: Springer, 1999, pp. 281–293.

[16] S. Y. Park, P. Grover, and A. Sahai, “A constant-factor approximately optimal solution to the witsenhausen counterexample,” *Submitted to the 48th IEEE Conference on Decision and Control (CDC)*, 2009.

[17] P. Grover and A. Sahai, “A vector version of Witsenhausen’s counterexample: Towards convergence of control, communication and computation,” *Proceedings of the 47th IEEE Conference on Decision and Control (CDC)*, Oct. 2008.

[18] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[19] H. Weingarten, Y. Steinberg, and S. Shamai, “The capacity region of the gaussian multiple-input multiple-output broadcast channel,” *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[20] N. Devroye, P. Mitran, and V. Tarokh, “Achievable rates in cognitive radio channels,” *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp.

1813–1827, May 2006.

[21] A. Jovicic and P. Viswanath, “Cognitive radio: An information-theoretic perspective,” in *Proceedings of the 2006 International Symposium on Information Theory, Seattle, WA*, Seattle, WA, Jul. 2006, pp. 2413–2417.

[22] Y.-H. Kim, A. Sutivong, and T. M. Cover, “State amplification,” *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1850–1859, May 2008.

[23] N. Merhav and S. Shamai, “Information rates subject to state masking,” *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 2254–2261, Jun. 2007.

[24] T. Philosof, A. Khisti, U. Erez, and R. Zamir, “Lattice strategies for the dirty multiple access channel,” in *Proceedings of the IEEE Symposium on Information Theory*, Nice, France, Jul. 2007, pp. 386–390.

[25] S. Kotagiri and J. Laneman, “Multiaccess channels with state known to some encoders and independent messages,” *EURASIP Journal on Wireless Communications and Networking*, no. 450680, 2008.

[26] P. Grover, A. B. Wagner, and A. Sahai, “Information embedding meets distributed control,” *In preparation for submission to IEEE Transactions on Information Theory*, 2009.

[27] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi, *Complexity and Approximation: Combinatorial optimization problems and their approximability properties*. Springer Verlag, 1999.

[28] R. Cogill and S. Lall, “Suboptimality bounds in stochastic control: A queueing example,” in *American Control Conference, 2006*, Jun. 2006, pp. 1642–1647.

[29] R. Cogill, S. Lall, and J. P. Hespanha, “A constant factor approximation algorithm for event-based sampling,” in *American Control Conference, 2007. ACC ’07*, Jul. 2007, pp. 305–311.

[30] 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.

[31] A. Avestimehr, S. Diggavi, and D. Tse, “A deterministic approach to wireless relay networks,” in *Proc. of the Allerton Conference on Communications, Control and Computing*, October 2007.

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

[33] M. S. Pinsker, “Bounds on the probability and of the number of correctable errors for nonblock codes,” *Problemy Peredachi Informatsii*, vol. 3, no. 4, pp. 44–55, Oct./Dec. 1967.

[34] A. Sahai, “Why block-length and delay behave differently if feedback is present,” *IEEE Trans. Inform. Theory*, no. 5, pp. 1860–1886, May 2008.

[35] A. Sahai and P. Grover, “The price of certainty : “waterslide curves” and the gap to capacity,” Dec. 2007. [Online]. Available: http://arXiv.org/abs/0801.0352v1

[36] H. S. Witsenhausen, “Separation of estimation and control for discrete time systems,” *Proceedings of the IEEE*, vol. 59, no. 11, pp. 1557–1566, Nov. 1971.

[37] R. F. H. Fisher, *Precoding and Signal Shaping for Digital Transmission*. New York, NY: John Wiley, 2002.

[38] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*. New York: Springer-Verlag, 1988.

[39] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[40] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inform. Theory*, vol. 11, no. 4, Oct. 1965.

[41] R. Blahut, "A hypothesis testing approach to information theory," Ph.D. dissertation, Cornell University, Ithaca, NY, 1972.

[42] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[43] "Code for performance of lattice-based strategies for Witsenhausen's counterexample." [Online]. Available: http://www.eecs.berkeley.edu/∼pulkit/FiniteWitsenhausenCode.htm

[44] D. Micciancio and S. Goldwasser, *Complexity of Lattice Problems: A Cryptographic Perspective*. Springer, 2002.

[45] W. Wu, S. Vishwanath, and A. Arapostathis, "Gaussian interference networks with feedback: Duality, sum capacity and dynamic team problems," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2005.

[46] N. Elia, "When Bode meets Shannon: control-oriented feedback communication schemes," *IEEE Trans. Automat. Contr.*, vol. 49, no. 9, pp. 1477–1488, Sep. 2004.

[47] R. Durrett, *Probability: Theory and Examples*, 1st ed. Belmont, CA: Brooks/Cole, 2005.

[48] R. Courant, F. John, A. A. Blank, and A. Solomon, *Introduction to Calculus and Analysis*. Springer, 2000.

[49] S. M. Ross, *A first course in probability*, 6th ed. Prentice Hall, 2001.

[50] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley, 1991.