The finite-dimensional Witsenhausen counterexample

Pulkit Grover, Anant Sahai and Se Yong Park
Wireless Foundations, Department of EECS
University of California at Berkeley, CA-94720, USA
{pulkit, sahai, separk}@eecs.berkeley.edu

***Abstract*— Recently, we considered a vector version of Witsenhausen's counterexample and used a new lower bound to show 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 bounds. We also introduce lattice-based quantization strategies for any finite length. Using the new finite-length lower bound, we show that the 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 — which corresponds to the scalar case — lattice-based strategies attain within a factor of 8 of the optimal cost. Based on observations in the scalar case and the infinite-dimensional case, we also conjecture what the optimal strategies could be for any finite vector length.**

I. Introduction

[1]Distributed control problems have long proved challenging for control engineers. In 1968, Witsenhausen [4] 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[2]; the optimal strategy and the optimal costs for the problem are still unknown since the non-convexity of the problem makes the search for an optimal strategy hard [6]–[8]. Discrete approximations of the problem [9] are even NP-complete[3] [10].

In the absence of a solution, research on the counterexample has bifurcated into two different directions. In one direction, the emphasis is to understand what aspect of the problem makes it hard. This is done by modifying the problem in various ways, and classifying the resulting problems into hard and easy ones, e.g. [11]–[14] (see [1] for a survey of other such modifications). In particular, the work of Rotkowitz and Lall in [14] strongly validates the idea that the core challenge in the Witsenhausen counterexample is coming from the fact that distributed controllers have an incentive to "talk to each other" through the plant itself. However, it seems intuitively clear that this feature is likely ubiquitous in any nontrivial distributed control system, and so the counterexample itself deserves to be understood. This is the other direction of research. Since the problem is non-convex, a body of literature (e.g. [7] [8] and the references therein) is dedicated to finding good solutions.

Rather than searching over the space of all possible solutions, a different approach is taken by Witsenhausen [4, Section 6] and Mitter and Sahai [15]. They aim at systematic constructions that perform reasonably well. Witsenhausen's two-point quantization strategy is motivated from the optimal strategy for two-point symmetric distributions of the initial state [4, Section 5] and it outperforms linear strategies for certain parameter choices. Interpreting Witsenhausen's strategy as implicit communication between the two controllers, Mitter and Sahai [15] propose multipoint-quantization strategies. Depending on the problem parameters, these strategies can outperform scalar strategies by an arbitrarily-large factor.

This brings us to a question that has received little attention in the literature — how far are the proposed strategies from the optimal? While the strategies in [8] are believed to be optimal because of the feeling of exhaustiveness in the search procedure, there is no guarantee that they are indeed optimal. Witsenhausen [4, Section 7] derived a lower bound on the costs that is loose in the interesting regimes of small k and large σ_0^2 [1], [3], and hence is insufficient to prove any sort of optimality or goodness for control schemes.

Towards obtaining a guarantee, a strategic simplification of the problem was proposed in [1], [2] 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) [16] model in information theory. There has been a burst of interest in extensions to DPC in large part due to the work of Devroye *et al* [17] on what is now called the "cognitive radio channel." This has inspired many other works in asymmetric cooperation between nodes [18]–[22]. In our work [1], [2], we develop

[1]Because this work conceptually builds upon [1], [2], there is significant overlap in the introduction. For the same reason, there is also overlap in the introduction with [3] although the scope there is narrower and a largely different proof technique is employed.

[2]In words of Yu-Chi Ho [5], "the simplest problem becomes the hardest problem."

[3]More precisely, results in [10] 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 [10] 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).

a new lower bound to the optimal performance of the vector Witsenhausen problem. Using this bound, we show that depending on the problem parameters, either simple linear strategies or the vector analogs of quantization-based strategies achieved within a constant factor of the optimal cost in the limit of infinite vector length. 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 [23]. This constant-factor spirit has also been useful in understanding other stochastic control problems [24], [25] and in asymptotic analysis of problems in multiuser wireless communication [26], [27].

While the lower bound in [1] 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 $k \rightarrow 0$ and $\sigma_0 \rightarrow \infty$. 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 is very loose. This effect is elucidated in [3] by deriving a different lower bound that shows that quantization-based strategies indeed attain within a constant[4] factor of the optimal cost for Witsenhausen's original problem. The bound 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 [1]. Building upon the vector lower bound, we derive a new lower bound that is in the spirit of information-theoretic bounds for finite-length communication problems (e.g. [28]–[31]). In particular, it extends the tools in [31] to a setting with unbounded distortion. The resulting lower bound 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[5] 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 3!

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 that we conjecture are optimal.

[4]The constant is large in [3], but as this paper shows, this is an artifact of the proof rather than reality.

[5]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}.

II. Notation and problem statement

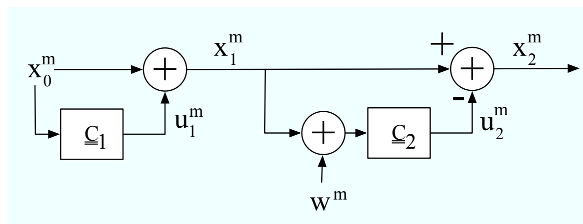

Fig. 1. Block-diagram for vector version of Witsenhausen's counterexample of length m.

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:

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

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

so that they do not necessarily grow with the problem size. 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 $\underline{\underline{\mathrm{C}}}_i$ 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

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

- The *information pattern* represents the information available to each controller at each time it takes an action (it has implicitly been specified above). Following Witsenhausen's notation in [32], the information pattern for the vector problem is

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

  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 above reduces to Witsenhausen's original counterexample [4].

Without loss of generality (as in [4]), we only consider the variance of the observation noise as $\sigma_z^2 = 1$. However, we often need to consider a hypothetical observation noise with variance σ_G^2. The pdf of this test noise is denoted by $f_G(\cdot)$, and the noise of variance 1 has density $f_Z(\cdot)$.

III. Lattice-based quantization strategies

We introduce lattice-based quantization strategies as generalizations of scalar quantization-based strategies [15] to the vector problem. An introduction to lattices can be found in [34], [35]. Relevant definitions are reviewed below. $\mathcal{B}$ denotes the unit ball in $\mathbb{R}^m$.

Definition 1 (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 2 (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 3 (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, $\xi = \frac{r_c}{r_p}$.

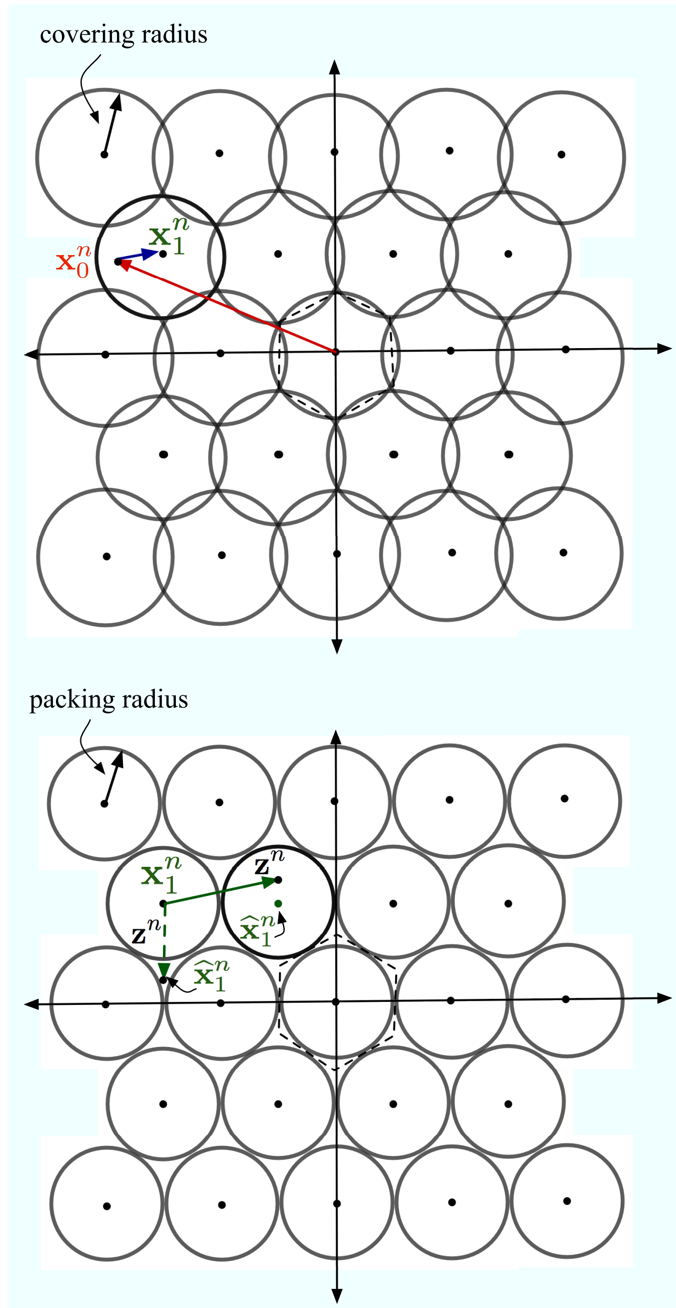

Fig. 2. Covering and packing for the 2-dimensional hexagonal lattice. The packing-covering ratio for this lattice is $\xi = \frac{2}{\sqrt{3}} \approx 1.15$ [33, Appendix C]. The first controller forces the initial state $\mathbf{x}_0^m$ to the lattice point nearest to it. The second controller estimates $\widehat{\mathbf{x}}_1^m$ 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 $\xi = \sqrt{2} \approx 1.41$ [33, Appendix C]) because it has a smaller ξ.

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 dimension m, we use a lattice of covering radius r_c and packing radius r_p. The action $\gamma_1(\cdot)$ of the first controller, $\underline{\underline{\mathrm{C}_1}}$, and $\gamma_2(\cdot)$ of the second controller, $\underline{\underline{\mathrm{C}_2}}$, are then given by

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

In the event that more than one $\widetilde{\mathbf{x}}_1^m$ satisfies $\|\mathbf{y}_2^m - \widetilde{\mathbf{x}}_1^m\|^2 \leq r_p^2$, the decoder chooses the one with the smallest distance from $\mathbf{y}_2^m$. 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

$$\begin{aligned}&\bar{J}^{(\gamma)}(m, k^2, \sigma_0^2) \\ \leq\ &\inf_{P \geq 0} k^2 P + \left(\sqrt{\psi(m+2, r_p)} + \sqrt{\frac{P}{\xi^2}}\sqrt{\psi(m, r_p)}\right)^2,\end{aligned}$$

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

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

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, $\|\mathbf{x}_1^m - \mathbf{x}_0^m\|^2 \leq r_c^2$. 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 $r_c^2 \ll \sigma_0^2$).

We now provide bounds on the second stage cost obtained by using the lattice Λ. Observe that

$$\mathbb{E}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2\right] = \mathbb{E}\left[\mathbb{E}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 | \mathbf{X}_1^m\right]\right]. \qquad (3)$$

We upper bound $\mathbb{E}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 | \mathbf{X}_1^m\right]$ for each lattice point $\mathbf{X}_1^m$. 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 zero second-stage cost. Thus,

$$\mathbb{E}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 | \mathbf{X}_1^m\right] = \mathbb{E}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 \mathbb{1}_{\{\mathcal{E}_m\}} | \mathbf{X}_1^m\right].$$

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 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

$$\mathbb{E}\left[\|\mathbf{X}_1^m - \widehat{\mathbf{X}}_1^m\|^2 | \mathbf{X}_1^m\right] \leq \mathbb{E}\left[\left(\|\mathbf{Z}^m\| + r_p\right)^2 \mathbb{1}_{\{\mathcal{E}_m\}} | \mathbf{X}_1^m\right]. \qquad (4)$$

We choose $r_c^2 = mP$, so that the first stage cost is at most $k^2 P$. Thus, $r_p^2 = \frac{mP}{\xi^2}$. With this choice of r_c, we have the following lemma.

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

$$\begin{aligned}&\frac{1}{m}\mathbb{E}\left[\left(\|\mathbf{Z}^m\| + r_p\right)^2 \mathbb{1}_{\{\mathcal{E}_m\}} | \mathbf{X}_1^m\right] \\ \leq\ &\left(\sqrt{\psi(m+2, r_p)} + \sqrt{\frac{P}{\xi^2}}\sqrt{\psi(m, r_p)}\right)^2.\end{aligned}$$

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

$$\begin{aligned}&\frac{1}{m}\mathbb{E}\left[\left(\|\mathbf{Z}^m\| + r_p\right)^2 \mathbb{1}_{\{\mathcal{E}_m\}} | \mathbf{X}_1^m\right] \\ \leq\ &\left(1 + \sqrt{\frac{P}{\xi^2}}\right)^2 e^{-\frac{mP}{2\xi^2} + \frac{m+2}{2}\left(1 + \ln\left(\frac{P}{\xi^2}\right)\right)}.\end{aligned}$$

Proof: See Appendix I. ■

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

IV. Lower bounds on the cost

Bansal and Basar [6] 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 [1] we derive the following lower bound for Witsenhausen's counterexample itself.

Theorem 2: For $W(m, k^2, \sigma_0^2)$, the following lower bound holds on the total cost for any strategy

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

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

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

Furthermore, $\left(\left(\sqrt{\kappa(P, \sigma_0^2)} - \sqrt{P}\right)^+\right)^2$ is a lower bound on the second-stage cost given that the first stage is constrained to use an input with average power at most P.

Proof: See [1]. ■

The techniques do not yield a tight bound because there is a tension in the Gaussianity of $\mathbf{x}_1^m$. On one hand, aligning

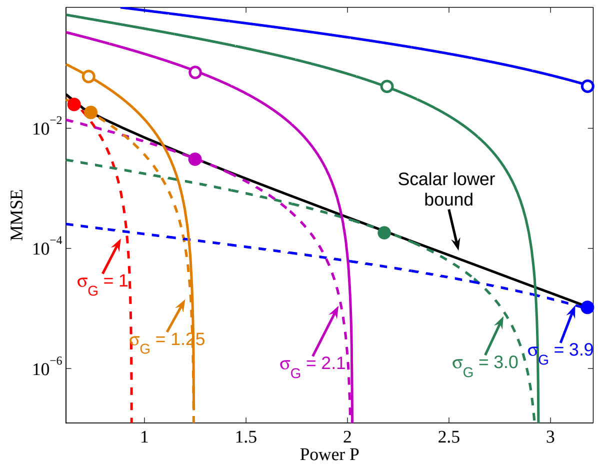

Fig. 3. A pictorial representation of the proof for the lower bound assuming $\sigma_0^2 = 30$. The solid curves show the vector lower bound of [1] for various values of observation noise variances, denoted by σ_G^2. Conceptually, multiplying these curves by the probability of that channel behavior yields 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.

$\mathbf{u}_1^m$ with $\mathbf{x}_0^m$ yields a large power Gaussian distribution on $\mathbf{x}_1^m$ that maximizes the capacity of the implicit channel, potentially reducing the estimation error. On the other, large power Gaussian sources are also the hardest to estimate across a Gaussian channel, and thus non-Gaussian (probably discrete) distributions on $\mathbf{x}_1^m$ might perform better. Our bounding technique ignores this tension.

Observe that the lower bound expression is the same for all vector lengths. In the following, the sphere-packing style arguments [36] are extended based on [29]–[31] 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 $W(m, k^2, \sigma_0^2)$, the following lower bound holds on the total cost

$$\bar{J}_{\min}(m, k^2, \sigma_0^2) \geq \inf_{P \geq 0} k^2 P + \eta(P, \sigma_0^2), \tag{6}$$

where

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

where

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

where $c_m(L) := \left(1 - \psi(m, L\sqrt{m})\right)^{-1}$ and $d_m(L) := c_m(L)\left(1 - \psi(m+2, L\sqrt{m})\right) = \frac{1 - \psi(m+2, L\sqrt{m})}{1 - \psi(m, L\sqrt{m})} > 0$. Recall that $\psi(m, r) = \Pr(\|\mathbf{Z}^m\| \geq r)$. Further, this bound is at least as tight as that of Theorem 2 for all values of k and σ_0^2.

Proof: Define $P := \frac{1}{m}\mathbb{E}\left[\|\mathbf{U}_1^m\|^2\right]$ as the average power of the input at the first stage. For given P, a lower bound on the average second stage cost is given by $\left(\left(\sqrt{\kappa} - \sqrt{P}\right)^+\right)^2$ (see Theorem 2). We derive another lower bound that is equal to the expression for $\eta(P, \sigma_0^2)$. The 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 σ_G^2. In the following, expectation over $\mathbf{x}_0^m$ is not denoted explicitly in the notation for clearer exposition.

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

The ratio of the two probability density functions is given by

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

Thus for $\mathbf{z}^m \in \mathcal{S}_L^G$,

$$\frac{f_Z(\mathbf{z}^m)}{f_G(\mathbf{z}^m)} \geq \sigma_G^m e^{-\frac{mL^2\sigma_G^2}{2}\left(1 - \frac{1}{\sigma_G^2}\right)} \geq \sigma_G^m e^{-\frac{mL^2(\sigma_G^2 - 1)}{2}}. \tag{7}$$

Thus,

$$\begin{aligned}
&\mathbb{E}_Z\left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}^m)\right] \\
\geq\ &\sigma_G^m e^{-\frac{mL^2(\sigma_G^2 - 1)}{2}} \\
&\int_{\mathbf{z}^m \in \mathcal{S}_L^G}\left(\int_{\mathbf{x}_0^m} J_2^{(\gamma)}(\mathbf{x}_0^m, \mathbf{z}^m) f_0(\mathbf{x}_0^m) d\mathbf{x}_0^m\right) f_G(\mathbf{z}^m) d\mathbf{z}^m \\
=\ &\sigma_G^m e^{-\frac{mL^2(\sigma_G^2 - 1)}{2}} \mathbb{E}_G\left[J_2^{(\gamma)}(\mathbf{X}_0^m, \mathbf{Z}^m) \mathbb{1}_{\{\mathbf{Z}^m \in \mathcal{S}_L^G\}}\right]
\end{aligned} \tag{8}$$

But

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

because $\frac{\mathbf{Z}^m}{\sigma_G} \sim \mathcal{N}(0, \mathbb{I}_m)$. Thus,

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

We now need the following lemma, which connects the new finite-length lower bound to the infinite-length lower bound of [1].

Lemma 2:

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

for any $L > 0$.

Proof: See Appendix II. ∎

The lower bound on the total average cost now follows from (9) and Lemma 2. We now 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 $\eta(P, \sigma_0^2)$,

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

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, $\eta(P, \sigma_0^2)$ 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

It has been observed that for the infinite-length vector case [1] and for the scalar case [7] that in some regimes, linear strategies perform better than quantization based strategies. Thus for an upper bound on the optimal cost, we consider the minimum of that achieved by lattice-based strategies and the optimal linear strategy. In this section we show that the cost attained by this upper bound is within a constant factor of the lower bound of Section IV uniformly over all m, k^2, σ_0^2.

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

$$\begin{aligned} \inf_{P\geq 0} k^2P + \eta(P, \sigma_0^2) &\leq \bar{J}_{min}(m, k^2, \sigma_0^2) \\ &\leq \mu \left(\inf_{P\geq 0} k^2P + \eta(P, \sigma_0^2) \right), \end{aligned}$$

where $\mu = 300\xi^2$, ξ is the packing-covering ratio of any lattice in $\mathbb{R}^m$, and $\eta(\cdot, \cdot)$ is as defined in Theorem 3. 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 computer calculation shows that $\mu < 8$.

Proof: Let P^* denote the optimum P that attains the

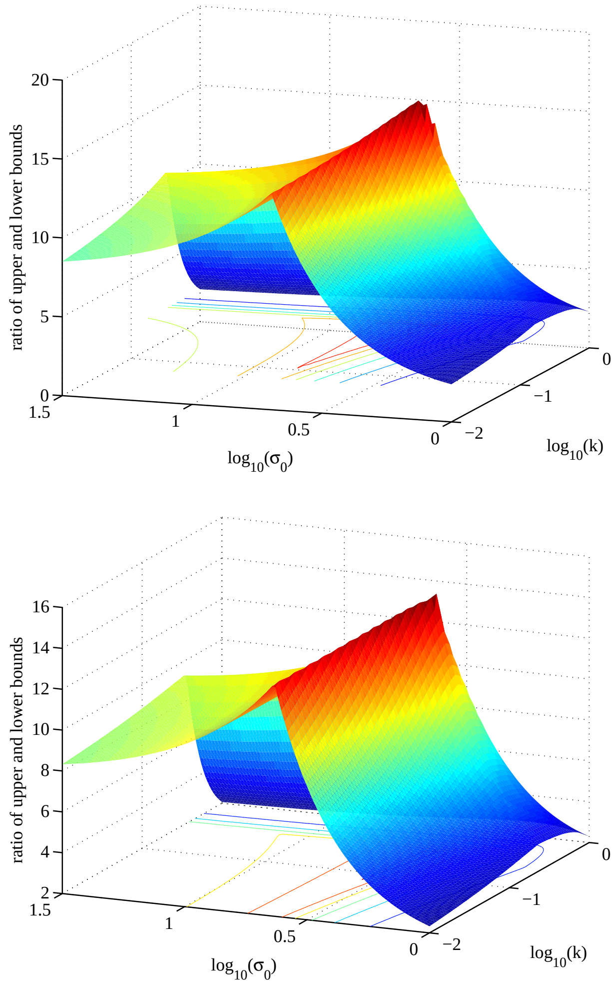

Fig. 4. The ratio of the upper and the lower bounds for the scalar Witsenhausen problem (top), and the 2-D Witsenhausen problem (bottom, using hexagonal lattice of $\xi = \frac{2}{\sqrt{3}}$) for a range of values of k and σ_0. The ratio is bounded above by 17 for the scalar problem, and by 14.75 for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

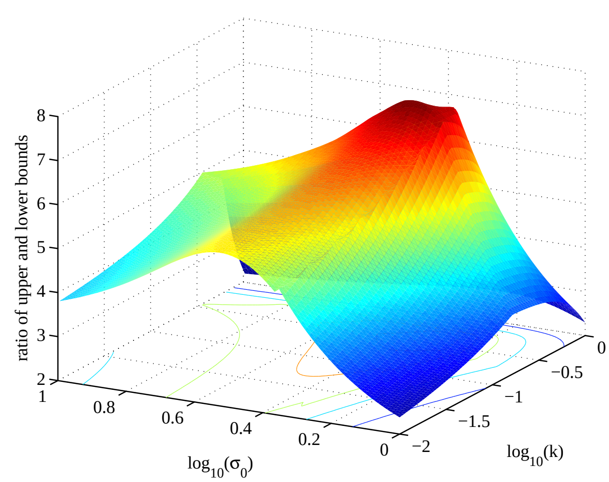

Fig. 5. An exact calculation of the first and second stage costs yields an improved maximum ratio smaller than 8 for the scalar Witsenhausen problem.

lower bound in Theorem 3. 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 MMSE estimation at $\underline{\underline{\mathrm{C}_2}}$. It is easy to see [1] that the 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. We show that the ratio of this upper bound and the lower bound in Theorem 3 is bounded.

Consider $P^* \geq \frac{\sigma_0^2}{300}$. Then the first stage cost is larger than $k^2\frac{\sigma_0^2}{300}$. Consider the upper bound of $k^2\sigma_0^2$ obtained by zero-forcing the input. The ratio of the upper bound and the lower bound is no larger than 300. Thus, for the rest of the analysis, we assume that $P^* < \frac{\sigma_0^2}{300}$.

Now consider $\sigma_0^2 < 50$, and $P^* \leq \frac{\sigma_0^2}{300}$. Then, using the bound from Theorem 2 (which is a special case of the bound in Theorem 3),

$$\kappa = \frac{\sigma_0^2}{(\sigma_0+\sqrt{P^*})^2+1} \overset{(P^*<\frac{\sigma_0^2}{300},\sigma_0^2<50)}{\geq} \frac{\sigma_0^2}{50\left(1+\frac{1}{\sqrt{300}}\right)^2+1} \approx \frac{\sigma_0^2}{56.94} \geq \frac{\sigma_0^2}{57}.$$

Thus, for $\sigma_0^2 < 50$ and $P^* \leq \frac{\sigma_0^2}{300}$,

$$\bar{J}_{min} \geq \left((\sqrt{\kappa}-\sqrt{P^*})^+\right)^2 \geq \sigma_0^2\left(\frac{1}{\sqrt{57}}-\frac{1}{\sqrt{300}}\right)^2 \geq \frac{\sigma_0^2}{180}.$$

Using the zero-forcing upper bound of $\frac{\sigma_0^2}{\sigma_0^2+1}$, the ratio of the upper and lower bounds is at most $\frac{180}{\sigma_0^2+1} \leq 180$.

From now on, we assume that $P^* \leq \frac{\sigma_0^2}{300}$ and $\sigma_0^2 > 50$. We divide the rest of the analysis into two cases.

Case 1: $P^* \leq \frac{1}{4}$.

In this case,

$$\kappa = \frac{\sigma_0^2}{(\sigma_0+\sqrt{P^*})^2+1} \overset{P^*\leq\frac{1}{4}}{\geq} \frac{\sigma_0^2}{(\sigma_0+0.5)^2+1} \overset{(a)}{\geq} \frac{50}{(\sqrt{50}+0.5)^2+1} \geq 0.75,$$

where (a) follows from $\sigma_0^2 > 50$ 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,

$$\left((\sqrt{\kappa}-\sqrt{P})^+\right)^2 \geq (\sqrt{0.75}-0.5)^2 \geq 0.13.$$

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 10.

We observe that the proof until here does not use the new lower bound, and hence works for any vector length m.

Case 2: $\sigma_0^2 > 50$, $\frac{1}{4} < P^* \leq \frac{\sigma_0^2}{300}$

Since $d_m(L) > 0$, we further lower bound the lower bound in Theorem 3 by replacing $d_m(L)$ in the expression by zero. Now, using $L = 2$,

$$c_m(L) = \frac{1}{\Pr_G(\|\mathbf{z}^m\|^2 \leq mL^2\sigma_G^2)} \overset{(a)}{\leq} \frac{1}{1-\frac{m\sigma_G^2}{mL^2\sigma_G^2}} \overset{L=2}{\leq} \frac{4}{3},$$

where (a) is obtained using Markov's inequality. In the bound, we are free to use any $\sigma_G^2 \geq 1$. Using $\sigma_G^2 = 4c_m^{\frac{2}{m}}(2)P^* > 1$ and $c_m(2) \leq \frac{4}{3}$ yields

$$\begin{aligned}
&\kappa_2 \\
&= \frac{4c_m^{\frac{2}{m}}(2)P^*\sigma_0^2}{\left((\sigma_0+\sqrt{P^*})^2+4c_m^{1+\frac{2}{m}}(2)P^*\right)c_m^{\frac{2}{m}}(2)e^{1-d_m}} \\
&\overset{\left(P^*<\frac{\sigma_0^2}{300}\right)}{\geq} \frac{4P^*}{\left(\left(1+\frac{1}{\sqrt{300}}\right)^2+4\left(\frac{4}{3}\right)^{1+\frac{2}{m}}\frac{1}{300}\right)e} \\
&\overset{(m\geq 1)}{\geq} 1.294P^*.
\end{aligned}$$

Thus for this case,

$$\left((\sqrt{\kappa_2}-\sqrt{P^*})^+\right)^2 \geq P^*(\sqrt{1.294}-1)^2 \geq \frac{P^*}{60}. \tag{10}$$

Now, using the full form of the lower bound in Theorem 3, and substituting $L = 2$,

$$\begin{aligned}
&\bar{J}_{min}(m, k^2, \sigma_0^2) \\
&\geq k^2P^* + \frac{\sigma_G^m}{c_m(2)}\exp\left(-\frac{mL^2(\sigma_G^2-1)}{2}\right) \\
&\quad \left(\left(\sqrt{\kappa_2}-\sqrt{P^*}\right)^+\right)^2 \\
&\overset{(\sigma_G^2=4c_m^{\frac{2}{m}}(2)P^*)}{\geq} k^2P^* + 4^{\frac{m}{2}}P^{*\frac{m}{2}}e^{2n}e^{-8nc_m^{\frac{2}{m}}P^*}\frac{P^*}{60} \\
&\overset{(c_m(2)\leq\frac{4}{3})}{\geq} k^2P^* + (4P^*)^{\frac{m}{2}}e^{2n}e^{-8n(\frac{4}{3})^{\frac{2}{m}}P^*}\frac{P^*}{60} \\
&\overset{(m\geq 1, 4P^*>1)}{\geq} k^2P^* + e^2e^{-m\frac{128}{9}P^*}\frac{P^*}{60} \\
&\overset{(P^*\geq\frac{1}{4},\frac{128}{9}<15)}{>} k^2P^* + \frac{e^2}{60\times 4}e^{-15mP^*} \\
&> k^2P^* + \frac{1}{33}e^{-15mP^*}.
\end{aligned} \tag{11}$$

We now concentrate on the lattice-based upper bound from Theorem 1. Here, P is a part of the optimization:

$$\begin{aligned}
& \bar{J}_{min}(m,k^2,\sigma_0^2) \\
\leq \quad & \inf_{P>\xi^2} k^2P + \left(1+\sqrt{\frac{P}{\xi^2}}\right)^2 e^{-\frac{mP}{2\xi^2}+\frac{m+2}{2}\left(1+\ln\left(\frac{P}{\xi^2}\right)\right)} \\
= \quad & \inf_{P>\xi^2} k^2P + \frac{1}{33}e^{-\frac{mP}{20\xi^2}} \\
& \times e^{-m\left(\frac{9P}{20\xi^2}-\frac{1+\frac{2}{m}}{2}(1+\ln\left(\frac{P}{\xi^2}\right))-\frac{2}{m}\ln\left(1+\sqrt{\frac{P}{\xi^2}}\right)-\frac{\ln(33)}{m}\right)} \\
\overset{(m\geq 1)}{\leq} \quad & \inf_{P>\xi^2} k^2P + \frac{1}{33}e^{-\frac{mP}{20\xi^2}} \\
& \times e^{-m\left(\frac{9P}{20\xi^2}-\frac{3}{2}\left(1+\ln\left(\frac{P}{\xi^2}\right)\right)-2\ln\left(1+\sqrt{\frac{P}{\xi^2}}\right)-\ln(33)\right)} \\
\leq \quad & \inf_{P\geq 31\xi^2} k^2P + \frac{1}{33}e^{-\frac{mP}{20\xi^2}},
\end{aligned}$$

where the last inequality follows from the fact that $\frac{9P}{20\xi^2} > \frac{3}{2}\left(1+\ln\left(\frac{P}{\xi^2}\right)\right) + 2\ln\left(1+\sqrt{\frac{P}{\xi^2}}\right) + \ln(33)$ for $\frac{P}{\xi^2} > 31$. This can be checked easily by plotting it.[6]

Using $P = 300\xi^2P^* \geq 75\xi^2 > 31\xi^2$ (since $P^* \geq \frac{1}{4}$) in (12),

$$\begin{aligned}
\bar{J}_{min}(m,k^2,\sigma_0^2) \quad & \leq \quad k^2 300\xi^2P^* + \frac{1}{33}e^{-m\frac{300\xi^2P^*}{20\xi^2}} \\
& = \quad k^2 300\xi^2P^* + \frac{1}{33}e^{-15mP^*}. \qquad (12)
\end{aligned}$$

Using (11) and (12), the ratio of the upper and the lower bounds is bounded for all m since

$$\mu \leq \frac{k^2 300\xi^2P^* + \frac{1}{33}e^{-15mP^*}}{k^2P^* + \frac{1}{33}e^{-15mP^*}} \leq \frac{k^2 300\xi^2P^*}{k^2P^*} = 300\xi^2. \tag{13}$$

For $m=1$, $\xi=1$, and thus in the proof the ratio $\mu \leq 300$. For m large, $\xi \approx 2$ [35], and $\mu \lesssim 1200$. For arbitrary m, using the recursive construction in [37, Theorem 8.18], $\xi \leq 4$, and thus $\mu \leq 4800$ regardless of m. We also observe here that the simple grid-lattice has ξ that scales as $\Theta(\sqrt{m})$, and thus the lattice strategy that is good for scalars is not good for the vector problem, an observation consistent with [2].

The entire proof above is admittedly an ugly and coarse calculation without much intuitive appeal. However, it does the job and since the underlying performance bounds themselves can probably be tightened a bit more, it is not worth optimizing the proof for increased elegance at this time. Therefore, even though this ratio of $300\xi^2$ seems large, it is clear that it is loose. Computer calculation (see Fig. 4) shows that for $m=1$ (the original Witsenhausen problem), $\mu < 17$ even for the current bounds. Interestingly, for $m=2$ using the hexagonal lattice, $\mu < 14.75$, even though the packing-covering ratio for hexagonal lattice is larger than that for the uniform lattice for $n=1$. This occurs because the large-deviation bounds in the sphere-packing argument tighten as n gets large.

[6] It can also be verified symbolically by examining the expression $g(b) = \frac{9}{20}b^2 - \frac{3}{2}(1+\ln b^2) - 2\ln(1+b) - \ln 33$, taking its derivative $g'(b) = \frac{18}{20}b - \frac{3}{b} - \frac{2}{1+b}$, and second derivative $g''(b) = \frac{18}{20} + \frac{3}{b^2} + \frac{2}{(1+b)^2} > 0$. The g function is convex-$\cup$ and the first derivative is clearly positive whenever $b > 3$. Evaluating $g(\sqrt{31}) \approx 0.03$ and so $g(b) > 0$ whenever $b \geq \sqrt{31}$.

For the lattice strategy for $m=1$, the first stage cost can be evaluated explicitly because the lattice corresponds to uniform quantization. Similarly, the second stage cost can also be evaluated explicitly by weighted summation of probabilities of $\mathbf{y}_2^m$ falling in a bin other than the transmitted bin[7], where the weights are $\|\widehat{\mathbf{x}}_1^m - \mathbf{x}_1^m\|^2$. The ratio of the upper bound thus obtained and the lower bound is bounded above by 8, as shown in Fig. 5. ■

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 [8] and the infinite-length case [1]. It is shown in [1] that the strategy of Lee, Lau and Ho [8] 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, where they achieve costs within a factor of 2 of the optimal. This suggests that DPC-based strategies might be very good for finite lengths as well.

A DPC-based strategy would work as follows. Given the initial state $\mathbf{x}_0^m$, scale it by a factor $\alpha < 1$ and quantize it using the lattice to a quantization point $\mathbf{x}_q^m$. Now use $\mathbf{u}_1^m = \mathbf{x}_q^m - \alpha\mathbf{x}_0^m$ as the first stage input, producing $\mathbf{x}_1^m = \mathbf{x}_q^m + (1-\alpha)\mathbf{x}_0^m$. The controller $\underline{\underline{C_2}}$ can now perform an MMSE estimation. Alternatively, as a low-complexity estimation algorithm, $\underline{\underline{C_2}}$ can estimate the lattice point $\mathbf{x}_q^m$ followed by linear estimation for the Gaussian $(1-\alpha)\mathbf{x}_0^m$.

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 and obtaining corresponding finite length results using the sphere-packing tools developed here. In [1] we showed the correspondence between the vector Witsenhausen problem and the communication problem of Assisted Interference Suppression (AIS). Further work on the closely related “distributed dirty-paper coding” problem of Kotagiri and Laneman [22] has appeared in [21] where Philosof *et al* derive new outer bounds. These techniques could help tighten the lower bounds on the vector Witsenhausen problem as well.

Tightening the upper bound can be performed by using the DPC-based technique over lattices, as outlined above. Further, an exact analysis of the required first-stage power when using a lattice would yield an improvement (as pointed out earlier, for $m=1$, $\frac{1}{m}k^2r_c^2$ 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, with noisy state evolution and noisy observations at both controllers. The hope is that

[7] We note that nearest neighbor decoding is not the MMSE strategy at the second controller.

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 [38], [39].

Acknowledgments

We gratefully acknowledge the support of the National Science Foundation (CNS-403427 and CCF-729122) and Sumitomo Electric and Samsung.

Appendix I
Proof of Lemma 1

$$\begin{aligned}
&\mathbb{E}\left[(\|\mathbf{Z}^m\|+r_p)^2\mathbb{1}_{\{\mathcal{E}_m\}}|\mathbf{X}_1^m\right]\\
\overset{(a)}{=}\ &\mathbb{E}\left[(\|\mathbf{Z}^m\|+r_p)^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]\\
=\ &\mathbb{E}\left[\|\mathbf{Z}^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]+r_p^2\Pr(\mathcal{E}_m)\\
&+2r_p\mathbb{E}\left[\left(\mathbb{1}_{\{\mathcal{E}_m\}}\right)\left(\|\mathbf{Z}^m\|\mathbb{1}_{\{\mathcal{E}_m\}}\right)\right]\\
\overset{(b)}{\leq}\ &\mathbb{E}\left[\|\mathbf{Z}^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]+r_p^2\Pr(\mathcal{E}_m)\\
&+2r_p\sqrt{\mathbb{E}\left[\mathbb{1}_{\{\mathcal{E}_m\}}\right]}\sqrt{\mathbb{E}\left[\|\mathbf{Z}^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]}\\
=\ &\left(\sqrt{\mathbb{E}\left[\|\mathbf{Z}^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]}+r_p\sqrt{\Pr(\mathcal{E}_m)}\right)^2, \qquad (14)
\end{aligned}$$

where (a) follows from the independence of $(\mathbf{Z}^m,\mathcal{E}_m)$ and $\mathbf{X}_1^m$, and (b) from the Cauchy-Schwartz inequality [40, Pg. 13].

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

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 $\mathbb{R}^m$ [41, 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,

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{Z}^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right] &= \int_{\|\mathbf{z}^m\|\geq r_p}\|\mathbf{z}^m\|^2\frac{e^{-\frac{\|\mathbf{z}^m\|^2}{2}}}{(\sqrt{2\pi})^m}d\mathbf{z}^m\\
&= \int_{r\geq r_p} r^2\frac{e^{-\frac{r^2}{2}}}{(\sqrt{2\pi})^m}\mathcal{A}_m(r)dr\\
&= \int_{r\geq r_p} r^2\frac{e^{-\frac{r^2}{2}}}{(\sqrt{2\pi})^m}\frac{2\pi^{\frac{m}{2}}r^{m-1}}{\Gamma\left(\frac{m}{2}\right)}dr\\
&= \int_{r\geq r_p}\frac{e^{-\frac{r^2}{2}}2\pi}{(\sqrt{2\pi})^{m+2}}\frac{2\pi^{\frac{m+2}{2}}r^{m+1}}{\pi\Gamma\left(\frac{m+2}{2}\right)\frac{2}{m}}dr\\
&= m\psi(m+2,r_p). \qquad (15)
\end{aligned}$$

Using (14), (15), and $r_p=\frac{mP}{\xi^2}$

$$\begin{aligned}
&\mathbb{E}\left[(\|\mathbf{Z}^m\|+r_p)^2\mathbb{1}_{\{\mathcal{E}_m\}}|\mathbf{X}_1^m\right]\\
&\leq m\left(\sqrt{\psi(m+2,r_p)}+\sqrt{\frac{P}{\xi^2}}\sqrt{\psi(m,r_p)}\right)^2,
\end{aligned}$$

which yields the first part of Lemma 1 on using $r_p=\sqrt{\frac{mP}{\xi^2}}$. To obtain a closed-form upper bound we consider $P>\xi^2$. It now suffices to bound $\psi(\cdot,\cdot)$.

$$\begin{aligned}
\psi(m,r_p) &= \Pr(\|\mathbf{Z}^m\|^2\geq r_p^2)\\
&= \Pr(\exp(\rho\sum_{i=1}^m Z_i^2)\geq\exp(\rho r_p^2))\\
&\overset{(a)}{\leq} \mathbb{E}\left[\exp(\rho\sum_{i=1}^m Z_i^2)\right]e^{-\rho r_p^2}\\
&= \mathbb{E}\left[\exp(\rho z_1^2)\right]^m e^{-\rho r_p^2}\\
&\overset{(\text{for } 0<\rho<0.5)}{=} \frac{1}{(1-2\rho)^{\frac{m}{2}}}e^{-\rho r_p^2},
\end{aligned}$$

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 $\rho\in(0,0.5)$ [42, 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,

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

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

$$\begin{aligned}
\Pr(\mathcal{E}_m) &= \psi(m,r_p)=\psi\left(m,\sqrt{\frac{mP}{\xi^2}}\right)\\
&\leq e^{-\frac{mP}{2\xi^2}+\frac{m}{2}+\frac{m}{2}\ln\left(\frac{P}{\xi^2}\right)}. \qquad (16)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left[\|\mathbf{Z}^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]\leq m\psi\left(m+2,\sqrt{\frac{mP}{\xi^2}}\right)\\
&\leq me^{-\frac{mP}{2\xi^2}+\frac{m+2}{2}+\frac{m+2}{2}\ln\left(\frac{mP}{(m+2)\xi^2}\right)}. \qquad (17)
\end{aligned}$$

From (14), (16) and (17),

$$\begin{aligned}
&\mathbb{E}\left[(\|\mathbf{Z}^m\|+r_p)^2\mathbb{1}_{\{\mathcal{E}_m\}}|\mathbf{x}_1^m\right]\\
\leq\ &\left(\sqrt{m}e^{-\frac{mP}{4\xi^2}+\frac{m+2}{4}+\frac{m+2}{4}\ln\left(\frac{mP}{(m+2)\xi^2}\right)}\right.\\
&\left.+\sqrt{\frac{mP}{\xi^2}}e^{-\frac{mP}{4\xi^2}+\frac{m}{4}+\frac{m}{4}\ln\left(\frac{P}{\xi^2}\right)}\right)^2\\
\overset{(\text{since } P>\xi^2)}{<}\ &\left(\sqrt{m}\left(1+\sqrt{\frac{P}{\xi^2}}\right)e^{-\frac{mP}{4\xi^2}+\frac{m+2}{4}+\frac{m+2}{4}\ln\left(\frac{P}{\xi^2}\right)}\right)^2\\
=\ &m\left(1+\sqrt{\frac{P}{\xi^2}}\right)^2e^{-\frac{mP}{2\xi^2}+\frac{m+2}{2}+\frac{m+2}{2}\ln\left(\frac{P}{\xi^2}\right)}.
\end{aligned}$$

Appendix II
Proof of Lemma 2

The following lemma is taken from [1].

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

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

Proof: See [1, Appendix II]. ∎

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

$$\begin{aligned}
&\mathbb{E}_G\left[J_2^{(\gamma)}(\mathbf{X}_0^m,\mathbf{Z}^m)|\mathbf{Z}^m\in\mathcal{S}_L^G\right]\\
=\;& \frac{1}{m}\mathbb{E}_G\left[\|\mathbf{X}_1^m-\widehat{\mathbf{X}}_1^m\|^2|\mathbf{Z}^m\in\mathcal{S}_L^G\right]\\
\geq\;& \left(\left(\sqrt{\frac{1}{m}\mathbb{E}_G\left[\|\mathbf{X}_0^m-\widehat{\mathbf{X}}_1^m\|^2|\mathbf{Z}^m\in\mathcal{S}_L^G\right]}\right.\right.\\
&\left.\left.-\sqrt{\frac{1}{m}\mathbb{E}_G\left[\|\mathbf{X}_0^m-\mathbf{X}_1^m\|^2|\mathbf{Z}^m\in\mathcal{S}_L^G\right]}\right)^+\right)^2\\
=\;& \left(\left(\sqrt{\frac{1}{m}\mathbb{E}_G\left[\|\mathbf{X}_0^m-\widehat{\mathbf{X}}_1^m\|^2|\mathbf{Z}^m\in\mathcal{S}_L^G\right]}-\sqrt{P}\right)^+\right)^2, \qquad (18)
\end{aligned}$$

since $\mathbf{X}_0^m-\mathbf{X}_1^m=\mathbf{U}_1^m$ is independent of $\mathbf{Z}^m$ 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 follows

$$f_{Z_L}(\mathbf{z}_L^m)=\begin{cases} c_m(L)\dfrac{e^{-\frac{\|\mathbf{z}_L^m\|^2}{2\sigma_G^2}}}{\left(\sqrt{2\pi\sigma_G^2}\right)^m} & \mathbf{z}_L^m\in\mathcal{S}_L^G\\ 0 & \text{otherwise.}\end{cases} \qquad (19)$$

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,

$$\mathbb{E}_G\left[\|\mathbf{X}_0^m-\widehat{\mathbf{X}}_1^m\|^2|\mathbf{Z}^m\in\mathcal{S}_L^G\right]=\mathbb{E}_G\left[\|\mathbf{X}_0^m-\widehat{\mathbf{X}}_L^m\|^2\right]. \qquad (20)$$

To get a lower bound, we now allow the controllers to optimize themselves with the additional knowledge that the observation noise must fall in $\mathcal{S}_L^G$. We follow the rate-distortion spirit and consider the mutual information between the initial state $\mathbf{X}_0^m$ and the estimate $\widehat{\mathbf{X}}_L^m$. Notice that there is a Markov chain $\mathbf{X}_0^m-\mathbf{X}_1^m-\mathbf{Y}_L^m-\widehat{\mathbf{X}}_L^m$. Using the data-processing inequality [43],

$$I(\mathbf{X}_0^m;\widehat{\mathbf{X}}_L^m)\leq I(\mathbf{X}_1^m;\mathbf{Y}_L^m). \qquad (21)$$

To upper bound the term on the RHS, we first upper bound the power of $\mathbf{X}_1^m$.

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right] &= \mathbb{E}\left[\|\mathbf{X}_0^m+\mathbf{U}_1^m\|^2\right]\\
&= \mathbb{E}\left[\|\mathbf{X}_0^m\|^2\right]+\mathbb{E}\left[\|\mathbf{U}_1^m\|^2\right]+2\mathbb{E}\left[{\mathbf{X}_0^m}^T\mathbf{U}_1^m\right]\\
&\leq \mathbb{E}\left[\|\mathbf{X}_0^m\|^2\right]+\mathbb{E}\left[\|\mathbf{U}_1^m\|^2\right]\\
&\quad +2\sqrt{\mathbb{E}\left[\|\mathbf{X}_0^m\|^2\right]}\sqrt{\mathbb{E}\left[\|\mathbf{U}_1^m\|^2\right]}\\
&\leq m(\sigma_0+\sqrt{P})^2.
\end{aligned}$$

Let $\bar{P}:=(\sigma_0+\sqrt{P})^2$. In the following, we derive an expression for $C_G^{(m)}$, an upper bound on $\frac{1}{m}I(\mathbf{X}_1^m;\mathbf{Y}_L^m)$.

$$\begin{aligned}
C_G^{(m)} &\leq \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right]\leq m\bar{P}} \frac{1}{m}I(\mathbf{X}_1^m;\mathbf{Y}_L^m)\\
&= \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right]\leq m\bar{P}} \frac{1}{m}h(\mathbf{Y}_L^m)-\frac{1}{m}h(\mathbf{Y}_L^m|\mathbf{X}_1^m)\\
&= \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right]\leq m\bar{P}} \frac{1}{m}h(\mathbf{Y}_L^m)-\frac{1}{m}h(\mathbf{Z}_L^m)\\
&\leq \sup_{p(\mathbf{X}_1^m):\mathbb{E}\left[\|\mathbf{X}_1^m\|^2\right]\leq m\bar{P}} \frac{1}{m}\sum_{i=1}^m h(Y_{L,i})-\frac{1}{m}h(\mathbf{Z}_L^m)\\
&\overset{(a)}{\leq} \frac{1}{m}\sum_{i=1}^m\frac{1}{2}\log_2\left(2\pi e(\bar{P}_i+\sigma_{G,i}^2)\right)-\frac{1}{m}h(\mathbf{Z}_L^m)\\
&\overset{(b)}{\leq} \frac{1}{2}\log_2\left(2\pi e(\bar{P}+c_m(L)\sigma_G^2)\right)-\frac{1}{m}h(\mathbf{Z}_L^m). \qquad (22)
\end{aligned}$$

In (a), we used the fact that Gaussian random variables maximize differential entropy and used the notation $\bar{P}_i=\mathbb{E}\left[X_{1,i}^2\right]$ and $\sigma_{G,i}^2=\mathbb{E}\left[Z_{L,i}^2\right]$ (by symmetry, $Z_{L,i}$ are zero mean random variables). The inequality (b) follows from the concavity of the $\log(\cdot)$ function. We also use the fact that $\frac{1}{m}\sum_{i=1}^m\sigma_{G,i}^2\leq c_m(L)\sigma_G^2$, which can be proven as follows.

$$\begin{aligned}
&\frac{1}{m}\mathbb{E}_G\left[\sum_{i=1}^m Z_{L,i}^2\right]\\
=\;& \frac{1}{m}\int_{\mathbf{z}^m\in\mathcal{S}_L^G}\|\mathbf{z}^m\|^2 c_m(L)\frac{\exp\left(-\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}\right)}{\left(\sqrt{2\pi\sigma_G^2}\right)^m}d\mathbf{z}^m\\
\leq\;& \frac{c_m(L)}{m}\int_{\mathbf{z}^m\in\mathbb{R}^m}\|\mathbf{z}^m\|^2\frac{\exp\left(-\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}\right)}{\left(\sqrt{2\pi\sigma_G^2}\right)^m}d\mathbf{z}^m\\
\leq\;& c_m(L)\sigma_G^2,
\end{aligned}$$

since $\mathbb{E}_G\left[\|\mathbf{Z}^m\|^2\right]=m\sigma_G^2$. We now lower bound $h(\mathbf{Z}_L^m)$

$$\begin{aligned}
&h(\mathbf{Z}_L^m)\\
=\;& \int_{\mathbf{z}^m\in\mathcal{S}_L^G} f_{Z_L}(\mathbf{z}^m)\log_2\left(\frac{1}{f_{Z_L}(\mathbf{z}^m)}\right)d\mathbf{z}^m\\
=\;& \int_{\mathbf{z}^m\in\mathcal{S}_L^G} f_{Z_L}(\mathbf{z}^m)\log_2\left(\frac{\left(\sqrt{2\pi\sigma_G^2}\right)^m}{c_m(L)e^{-\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}}}\right)d\mathbf{z}^m\\
=\;& -\log_2\left(c_m(L)\right)+\frac{m}{2}\log_2\left(2\pi\sigma_G^2\right)\\
&+\int_{\mathbf{z}^m\in\mathcal{S}_L^G} c_m(L)f_G(\mathbf{z}^m)\frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2}\log_2(e)\,d\mathbf{z}^m. \qquad (23)
\end{aligned}$$

Analyzing the last term,

$$\begin{aligned}
&\int_{\mathbf{z}^m \in \mathcal{S}_L^G} c_m(L) f_G(\mathbf{z}^m) \frac{\|\mathbf{z}^m\|^2}{2\sigma_G^2} \log_2(e)\, d\mathbf{z}^m \\
=\;& \frac{c_m(L)\log_2(e)}{2} \mathbb{E}_G\left[\left(\frac{\|\mathbf{Z}^m\|}{\sigma_G}\right)^2 \mathbb{1}_{\left\{\frac{\|\mathbf{Z}^m\|}{\sigma_G} \le \sqrt{mL^2}\right\}}\right] \\
\overset{\widetilde{\mathbf{Z}}^m := \frac{\mathbf{Z}^m}{\sigma_G}}{=}\;& \frac{c_m(L)\log_2(e)}{2} \mathbb{E}\left[\|\widetilde{\mathbf{Z}}^m\|^2 \mathbb{1}_{\{\|\widetilde{\mathbf{Z}}^m\| \le \sqrt{mL^2}\}}\right] \\
=\;& \frac{c_m(L)\log_2(e)}{2} \left(\mathbb{E}\left[\|\widetilde{\mathbf{Z}}^m\|^2\right] \right. \\
& \left. -\mathbb{E}\left[\|\widetilde{\mathbf{Z}}^m\|^2 \mathbb{1}_{\{\|\widetilde{\mathbf{Z}}^m\| > \sqrt{mL^2}\}}\right]\right) \\
\overset{(\text{using } (15))}{=}\;& \frac{c_m(L)\log_2(e)}{2}\left(m - m\psi(m+2, \sqrt{mL^2})\right) \\
=\;& \frac{m\log_2(e)}{2} c_m(L)\left(1 - \psi(m+2, L\sqrt{m})\right). \qquad (24)
\end{aligned}$$

Define $d_m(L) := c_m(L)\left(1-\psi(m+2, L\sqrt{m})\right)$. Then, the expression $C_G^{(m)}$ can be upper bounded using (22), (23) and (24) as follows.

$$\begin{aligned}
C_G^{(m)} \le\;& \frac{1}{2}\log_2\left(2\pi e(\bar{P} + c_m(L)\sigma_G^2)\right) + \frac{1}{m}\log_2(c_m(L)) \\
& -\frac{1}{2}\log_2\left(2\pi\sigma_G^2\right) - \frac{1}{2}\log_2\left(e^{d_m(L)}\right) \\
=\;& \frac{1}{2}\log_2\left(2\pi e(\bar{P} + c_m(L)\sigma_G^2)\right) + \frac{1}{2}\log_2\left(c_m^{\frac{2}{m}}(L)\right) \\
& -\frac{1}{2}\log_2\left(2\pi\sigma_G^2\right) - \frac{1}{2}\log_2\left(e^{d_m(L)}\right) \\
=\;& \frac{1}{2}\log_2\left(\frac{2\pi e(\bar{P} + c_m(L)\sigma_G^2)c_m^{\frac{2}{m}}(L)}{2\pi\sigma_G^2 e^{d_m(L)}}\right) \\
=\;& \frac{1}{2}\log_2\left(\frac{e^{1-d_m(L)}(\bar{P} + c_m(L)\sigma_G^2)c_m^{\frac{2}{m}}(L)}{\sigma_G^2}\right). \qquad (25)
\end{aligned}$$

Now, recall that the Gaussian rate-distortion function $D_m(R)$ is defined as follows

$$D_m(R) := \inf_{\substack{p(\widehat{\mathbf{X}}_L^m | \mathbf{X}_0^m) \\ \frac{1}{m} I(\mathbf{X}_0^m; \widehat{\mathbf{X}}_L^m) \le R}} \frac{1}{m}\mathbb{E}\left[\|\mathbf{X}_0^m - \widehat{\mathbf{X}}_L^m\|^2\right] \qquad (26)$$

Since $I(\mathbf{X}_0^m; \widehat{\mathbf{X}}_L^m) \le mC_G^{(m)}$, using the converse to the rate distortion theorem [43, Pg. 349] and the upper bound on the mutual information,

$$\frac{1}{m}\mathbb{E}\left[\|\mathbf{X}_0^m - \widehat{\mathbf{X}}_L^m\|^2\right] \ge D_m(C_G^{(m)}). \qquad (27)$$

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 [43, Pg. 346]. Thus, using (18), (20) and (27),

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

Substituting the bound on $C_G^{(m)}$ from (25),

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

Using (18), this completes the proof of the lemma. Notice that $c_m(L) \to 1$, $c_m(L) > 1$, and $d_m(L) \to 1$ both for fixed m as $L \to \infty$ and for fixed $L > 1$ as $m \to \infty$. So $D(C_G^{(m)})$ approaches κ of Theorem 2 in both of these two limits.

REFERENCES

[1] 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/

[2] ——, "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.

[3] 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.

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

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] G. Lipsa and N. Martins, "Finite horizon optimal memoryless control of a delay in Gaussian noise: A simple counterexample," *Proceedings of the 47th IEEE Conference on Decision and Control (CDC)*, pp. 1628–1635, 2008.

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

[14] 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.

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

[17] 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.

[18] 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.

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

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

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

[26] 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.

[27] 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.

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

[29] 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.

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

[31] 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

[32] 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.

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

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

[35] 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.

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

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

[38] 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.

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

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

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

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

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