Implicit Communication in Multiple-Access Settings

Gireeja Ranade and Anant Sahai
Electrical Engineering and Computer Sciences
University of California, Berkeley, CA
Emails: {gireeja, sahai}@eecs.berkeley.edu

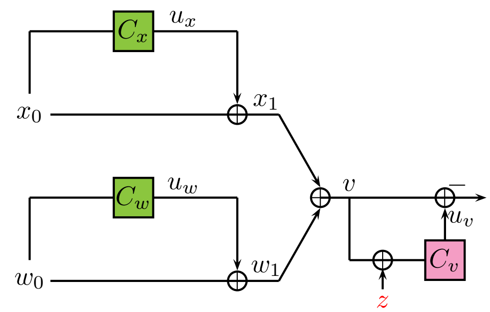
***Abstract*—Optimal control strategies for decentralized control problems may involve internal communication between controllers. We think of such internal communication as *implicit*, since the messages being sent are endogenous to the system and not externally specified. Recently, Grover and Sahai [1] applied information-theoretic techniques to provide an approximately optimal scheme for the Witsenhausen counterexample: one of the simplest models of a decentralized control system. This paper examines a MAC-inspired extension of the Witsenhausen counterexample. Deterministic modeling techniques based on the work by Avestimehr at al. [2] feature centrally in the strategy development. This example illustrates that "Information is in the eye of the beholder", and we find "rate gains" in the context of *implicit* communication. These are not observed in the original Witsenhausen counterexample.**

I. Introduction

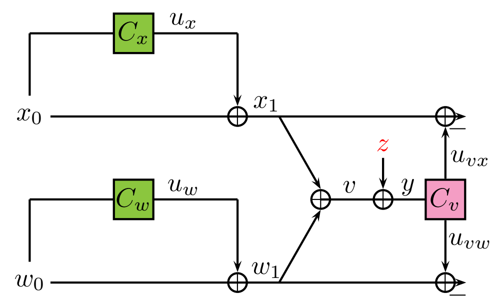
Shannon's groundbreaking 1948 paper built upon an insightful simplification for the communication problem: ignore the content and context of the message while designing communication schemes [3]. Following this, information theoretic problem formulations traditionally focus on *explicit* communication, where the message has been externally specified to the encoder.

Now consider a situation where communication serves as a means to an end. Say the transmitted "messages" are system-generated and possibly even related to the channel over which they are traveling. Can we capitalize on the connections between the system and the message? Recently, Grover and Sahai explored the concept of *implicit* communication in decentralized control systems, where both the message and the channel over which it is to be communicated are endogenous to the system [1]. They provided a provably approximately optimal strategy for an asymptotic vector version of the Witsenhausen counterexample. These strategies were then extended by Grover, Sahai and Park to the scalar counterexample [4]. Another example of *implicit* communication comes from the problem of information embedding with reversible stegotext [5]. The channel state plays double roles: it acts as both a system-generated message and as interference. Finally, Grover, Wagner and Sahai [6] studied the problem of information embedding in control systems, which combined both *implicit* and *explicit* messages in a system.

The communication-centric perspective for decentralized control problems can be illustrated well using deterministic models as in [7]. Deterministic models have given insight into the interference and relay channels [2], [8], but their full potential has not been explored in the context of control problems. The two controller example considered by Grover and Sahai [7] did not involve any interfering states.



(a) The *state-setup* MAW problem, with v as a state to be controlled.



(b) The *observation-setup* MAW problem, with v as an observation. x and w are the states to be controlled.

Fig. 1. The two versions of the MAW problem.

Here, we examine a multiple-access Witsenhausen (MAW) problem (Fig. 1), in order to understand implicit communication in decentralized control systems with interfering states, and to test drive a deterministic approach. The Witsenhausen counterexample can be thought of as a decentralized control parallel to the point-to-point channel. The multiple-access channel (MAC) has traditionally been the first multi-user follow-up to the point-to-point channel, and the MAW problem is a natural exploration. Scenarios where controllers observe linear combinations of state variables could occur frequently in biological or chemical control plants. For instance, molecules A and B may undergo a chemical reaction to produce an observed species C. Based on this observation, the system may need to estimate or control the concentrations of species A, B or C.

As in [1], we focus on developing approximately optimal control schemes in this paper, and consider two variations of the MAW problem. In the *state-setup*, two states add to form a new state, which is of interest to the second stage controller. Strategies for this first variation turn out to be similar to the original Witsenhausen counterexample. Despite being a multi-user problem it is noise-limited (not interference-limited). Nazer and Gastpar [9] considered a MAC setup where a function of the encoder messages is of interest at the decoder. The structured codes developed for this situation are the crux of the strategy used in the *state-setup* (Sec. IV).

In the *observation-setup* (Sec. V), the second-stage controller must estimate two states from one observation[1]: the sum of states. The two states interfere with each other, just as in a MAC. A slight re-drawing of the problem, as in Fig. 2, highlights that this control problem is closely related to the dirty multiple-access channels considered by Kotagiri and Laneman [10] and by Philosof et al. [11]. The deterministic model proves to be particularly illuminating here, and suggests good lattice and random coding strategies that work well in the Gaussian case.

II. Problem Setup

The MAW problem is a two-time-step control system with two states, x and w. As noted in the introduction, we consider two versions of the problem (Fig. 1). The first stage control is identical for both versions. At time zero, the stage-one controllers, C_x and C_w, observe the initial states $x_0, w_0 \sim \mathcal{N}(0, \sigma^2)$ respectively, and add their own control inputs, u_x and u_w, to get x_1, w_1. x_1 and w_1 add together to give v. Note that u_x and u_w are functions of x_0 and w_0 respectively. The system is described by:

$$x_1 = x_0 + u_x, \;\; w_1 = w_0 + u_w \qquad (1)$$

$$v = x_1 + w_1, \;\; y = v + z \qquad (2)$$

The third controller C_v observes y, which is a noisy version of the sum of the states. The noise z is distributed as $\mathcal{N}(0, 1)$. Depending on the problem setup, v can be thought of as a state of the system or as (the precursor to) an observation. Following the chemistry example in the introduction, if the quantity of interest for control is the molecule C, v is a state. However, if the engineer wants to control the concentration of molecules A and B, then v acts as an observation and is not of inherent interest. This distinction between state and observation can only be seen in multiuser communication and control problems.

This leads to the two problem setups. In the first, v is the system state to drive to zero[2] (the *state-setup*, Fig. 1(a)). In the second, v is just an observation (the *observation-setup*, Fig. 1(b)), and the aim is to drive states x and w to zero. The first stage cost $J_1(||u_x||^2, ||u_w||^2)$ penalizes the power used, and the second stage cost, $J_2(||u_v||^2)$ or $J_2(||u_{vx}||^2, ||u_{vw}||^2)$, penalizes the control error. The total cost to be minimized is a weighted sum of the first and second stage costs, $J(k^2) = k^2 J_1 + J_2$, where k is a parameter. Minimizing mean square error (MMSE) at the second stage error is equivalent to generating the best estimate of the state(s) (v or x_1, w_1). With this perspective, C_v can be seen as a decoder. C_x and C_w then take on the role of encoders (Fig. 2).

In the *state-setup*, controller C_v would like to generate a control input u_v to set state v to zero, and the cost function is $\mathbb{E}\left[k^2||u_x||^2 + k^2||u_w||^2 + ||v - u_v||^2\right]$. In the *observation-setup*, the objective is to minimize $\mathbb{E}\left[k^2||u_x||^2 + k^2||u_w||^2 + ||x_1 - u_{vx}||^2 + ||w_1 - u_{vw}||^2\right]$.

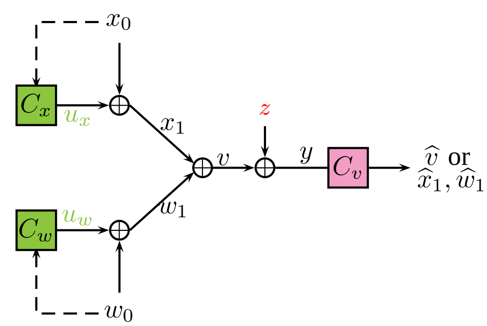


Fig. 2. x_0 and w_0 can be thought of as known interference in the communication from C_x and C_w to C_v, similar to the setup of a doubly dirty MAC [11].

Just as in the original Witsenhausen counterexample, the goal is to minimize $\mathbb{E}J(k^2)$. In both setups, the first stage cost is a symmetric function of the power P used by C_x and C_w. Thus, it is logical for both controllers C_x and C_w to use the same power P. Indeed, say C_x used power $P + \epsilon$. C_w could also use the same power without changing the order of magnitude of J_1, and will not make a significant difference in an approximately optimal scheme. For now, we focus our attention on this symmetric problem. We also note that minimizing the total cost J is equivalent to minimizing J_2 such that $\mathbb{E}||u_x||^2, \; \mathbb{E}||u_w||^2 \leq P$ (for all P) and we interchangably deal with the two formulations.

As traditional in information theory, in sections IV-B and V-B we consider a vector version of the problem, where the variables $\mathbf{x}_0, \mathbf{w}_0$ etc. are vectors of length l, and let $l \to \infty$. From a control-theoretic perspective it is important to also consider the scalar version of the problem, as in [4], but we leave this for future work. To start off, we examine deterministic versions of the problem.

III. A deterministic interpretation

Grover and Sahai [7] used a deterministic model based on the ideas from [2] to analyze the Witsenhausen counterexample with an added explicit message. Building on this, we consider a deterministic MAW problem, as detailed below (Fig. 3).

- This simplified version replaces each system variable with a binary string. For instance, x_0 is replaced by an infinite binary string $b_n b_{n-1} ... b_1 b_0 . b_{-1} b_{-2} ...$. The representation beyond b_0 (deemed the least significant bit (LSB)) turns

[1] In linear control theory, such a system would be unobservable, since there is no way to uniquely determine two summands from their sum. However, if we restrict the sets in which the initial states might lie, we could resolve them from the sum. This paper builds on a similar idea.

[2] or other value of interest.

out to not be relevant in our case. b_n is the most significant bit (MSB).

- The highest bit level (in the representation of a random variable) that can be 1 is determined by the power of the random variable. For example, $x_0 \sim \mathcal{N}(0, \sigma^2)$ can have highest representation at the $\log_2 \sigma^2$ level. Similarly, to control the value of a bit at level l, a controller needs power $P = 2^l$.
- The bitwise XOR operation replaces real addition.
- The noise is represented by an additive Bernoulli($\frac{1}{2}$).
- The noise z only affects bits b_0 and below since $z \sim \mathcal{N}(0, 1)$. Bits below the noise floor are not available to the decoder, but are part of the model.
- We replace the 2-norm with the power function. $pow(x) = 2^m$, where m is the highest non-zero bit level that x can take. The cost function takes the form $k^2 pow(u_x) + k^2 pow(u_w) + pow(error)$ depending on the case we are looking at.

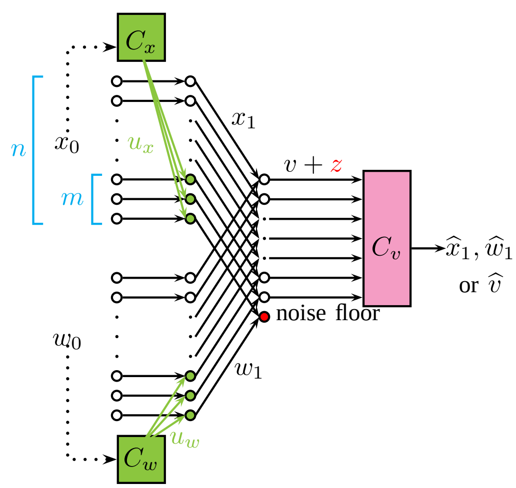


Fig. 3. A deterministic model for the MAW problem. Each circle represents a bit level. Green (filled) circles at the first stage show bit levels that C_x and C_w can control. Bits below the noise floor are not received by C_v. x_0 and w_0 have n bits above the noise floor, of which C_x and C_w can control m bits.

IV. State-setup: sum estimation

A. The deterministic case

Here, the cost function we want to minimize is $k^2 pow(u_x) + k^2 pow(u_w) + pow(v - \widehat{v})$. We focus on minimizing $pow(v - \widehat{v})$ with $pow(u_x), pow(u_w) \leq P$.

Theorem 4.1: Let $\log_2 \sigma^2 = n$, $\log_2 P = m$ and $J_{2,min}$ denote the minimum stage-two cost. Let $pow(u_x),\ pow(u_w) \leq 2^m$. Then,

1) if $m \geq n,\ J_{2,min} = 0$.
2) if $n > m > 0, J_{2,min} = 0$.
3) if $m < 0, J_{2,min} = 0.5$.

Proof: Let b_0 to b_n, c_0 to c_n and d_0 to d_n represent the bit levels of x_1, w_1 and v respectively. The noise power is at the level of the lowest bits, b_0, c_0, d_0. d_0 (the red bit at the noise floor in Fig. 3) is not received by the decoder, while d_1 to d_n are received noiselessly.

Case 1 (*zero-forcing*): If C_x, C_w can control all bit levels, i.e. $m \geq n$ then the obvious strategy is to set all bits for both x_1 and w_1 to zero at stage-one. Clearly, $J_{2,min} = 0$, since C_v sets $\widehat{v} = 0$.

Case 2 (*quantization*): In the deterministic model, all bits of v above the noise level (see Fig. 2) are received noiselessly by C_v. The only challenge is to correctly decode the bits below the noise floor. Hence if $m > \log_2 1 = 0$, then C_x and C_w can set all bits of x_1, w_1 below the noise floor to zero (see Fig. 4), effectively setting d_0 to 0. C_v reads d_1 to d_n noiselessly, and knows $d_0 = 0$ in advance. Thus, $J_{2,min} = 0$.

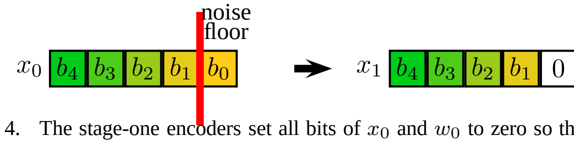


Fig. 4. The stage-one encoders set all bits of x_0 and w_0 to zero so that they are noiselessly recovered at stage-two.

Case 3 (*do-nothing*): Unlike Case 2, here the stage-one controllers cannot set the value of d_0. C_v will correctly guess the value with probablility 0.5, and make an error of power 1 otherwise. Thus, $J_{2,min} = 0.5$. ■

Remark 4.1: Just as in the original Witsenhausen counterexample, the bit bottleneck in the *state-setup* comes from noise. Therefore, the problem is *noise-limited*. Hence, from the perspective of the second stage decoder, the second stage estimation problem is identical to that in the point-to-point counterexample [1].

B. The vector Gaussian case

In the Gaussian setup, we look at all variables as being vectors of length l, and we change notation appropriately.

Theorem 4.2: An achievable cost for the vector Gaussian case state-setup MAW problem is $\min \left\{ 2k^2\sigma^2, \frac{2\sigma^2}{1+2\sigma^2}, 2k^2 \right\}$ as $l \to \infty$.

Proof: (Sketch) Following the deterministic model, the Gaussian sum estimation problem employs three control strategies: *zero-forcing*, *lattice-point quantization*, and *do-nothing*. In each of the strategies, we examine average costs over the vector length (i.e. $\frac{1}{l}\mathbb{E}[J(k^2)]$ as $l \to \infty$ etc.), but drop the l to simplify the explantion). The *zero-forcing* strategy pushes both states to zero at the first stage, with $\mathbb{E}[J_1] = 2\sigma^2$ at the first stage and $\mathbb{E}[J_2] = 0$, so $\mathbb{E}[J] = 2k^2\sigma^2$. In the *do-nothing* strategy, $\mathbb{E}[J_1] = 0$ and C_v performs MMSE estimation to get $\hat{\mathbf{v}}$. Hence, $\mathbb{E}[J_2] \leq \frac{2\sigma^2}{1+2\sigma^2}$. The *lattice-point quantization* strategy is inspired by the work of Nazer and Gastpar [9]. Lattice codebooks of second moment 1 (i.e. the cell radius is matched to the noise power, 1) are used to quantize $\mathbf{x_0}$ and $\mathbf{w_0}$ to $\mathbf{x_1}$ and $\mathbf{w_1}$ respectively. Then, $\mathbb{E}[||\mathbf{u_x}||^2] \leq 1$. The sum of two lattice-points is also a lattice-point, and since the second moment of the lattice and the noise variance are both 1, the noise will not push the received codeword outside the decoding region of $\mathbf{v}$. Hence, $\mathbb{E}[J_2] = 0$, and we achieve $\mathbb{E}[J(k^2)] \leq 2k^2$ as $l \to \infty$. ■

Theorem 4.3: $\min\limits_{P \geq 0} 2k^2 P + \left(\left[\sqrt{\frac{2\sigma^2}{1+4(\sigma+\sqrt{P})^2}} - \sqrt{4P} \right]_+ \right)^2$ is a lower bound on the cost in the vector Gaussian case state-setup MAW problem as the vector length $l \to \infty$. The scheme presented in Thm. 4.1 achieves this bound to within a constant factor.

We leave the proof of this to an extended version of the paper [12].

V. Observation-setup: separate estimation

A. The deterministic case

In this setup, we are working to minimize $J(k^2) = k^2 pow(u_x) + k^2 pow(u_w) + pow(x_1 - \widehat{x}_1) + pow(w_1 - \widehat{w}_1)$. We'll use the same notation as in the *state-setup*. Looking at Fig. 3, the deterministic model highlights that the separate estimation problem is *interference-limited*. Notice that the bits of v are a communication resource shared by x_1 and w_1. As in the MAC, there is a tension between the variables.

Theorem 5.1: Let $\log_2 \sigma^2 = n$, $\log_2 P = m$ and $J_{2,min}$ denote the minimum achievable stage-two cost. Let $pow(u_x),\ pow(u_w) \leq 2^m$. Then,

1) if $m < 0,\ J_{2,min} = 2^n$.
2) if $n/2 > m > 0,\ J_{2,min} = 2^{n-m}$.
3) if $n > m \geq n/2 > 0,\ J_{2,min} = 0$.
4) if $m \geq n,\ J_{2,min} = 0$.

Proof: Case 1: (*do-nothing*) Recall that the stage-two cost focuses on the highest erroneous bit. No bits of x_1 or w_1 (above the noise level if $n > 0$, and below the noise level if $n < 0$) can be reproduced at the decoder, since they are all killed by interfering with each other. We will have errors on x_1 and w_1 at level n with probability 0.5 each. Hence, $J_{2,min} = 0.5 \cdot 2^n + 0.5 \cdot 2^n = 2^n$.

Case 2: (*quantization*) To be able to communicate any information across the interference-limited channel, one of the controllers, say C_w, needs to "make room" for C_x to communicate. We have the ability to control m bits of x and w (upto level m, and above the noise level), but cannot modify anything beyond level m. To understand the strategy, consider $n = 6, m = 2$, as in Fig. 5. the optimal strategy is for C_x to set b_1, b_2 to equal b_5, b_6 and use these two bits to communicate the two MSBs of x_0. On the other hand, C_w sets c_1, c_2 to zero. C_v receives b_5, b_6, as well as $b_5 + c_5$ and $b_6 + c_6$ noiselessly, and decodes the four MSBs perfectly. The highest bits not recovered in the example are b_4 and c_4, and $J_{2,min} = 2 \cdot \frac{1}{2} \cdot 2^4 = 2^4$. Using an identical scheme for any $m < n/2$ we can achieve $J_{2,min} = 2^{n-m}$.

Remark 5.1: Even though we focused all available resources on communicating two bits of x_1, we noiselessly decoded four bits at the decoder. Since w_1 acts as both message and state, we are able to effectively "double" our communication rate. We'll revisit this idea in the Gaussian case.

Case 3: (*coding*): This case is identical to the previous case, except that since $m \geq \frac{n}{2}$, we can decode all $\frac{n}{2}$ MSBs of $x_1,\ w_1$. Thanks to the "room" created by C_w, c_1 to $c_{\frac{n}{2}}$ are known to be zero, and b_1 to $b_{\frac{n}{2}}$ are noiselessly received. All bits above the noise level can be recovered at the decoder and $J_{2,min} = 0$. We call this a coding strategy (as opposed

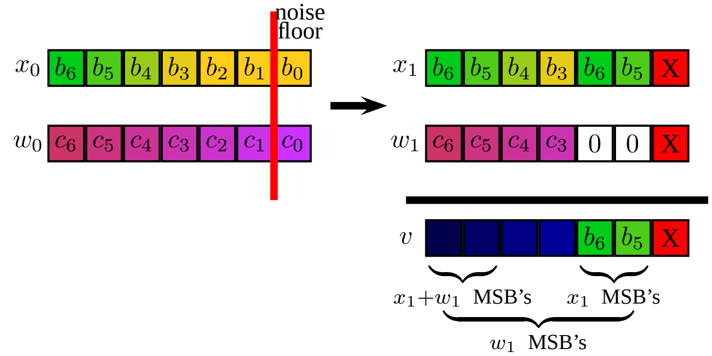


Fig. 5. The deterministic quantization strategy used when $\frac{n}{2} > m > 0$. $b_4,\ c_4$ are the highest bits which are not decodable.

to quantization) since here C_v perfectly recovers both x_1 and w_1.

Case 4: (*zero-forcing*) If $m > n$, then we can set all bits of $x_1,\ w_1$ to zero at stage-one, regardless of whether $n > 0$ or $n \leq 0$. Clearly, $J_{2,min} = 0$. ■

B. The vector Gaussian case

Theorem 5.2: Let $\zeta(P) = 2k^2 P + 2\left(\sqrt{\frac{\sigma^2}{.5+P}} + \sqrt{P}\right)^2 + 1$. An achievable cost for the vector Gaussian observation-setup MAW problem as $l \to \infty$ is $\min\left\{\sigma^2, 2k^2\sigma^2, 2k^2\sigma, \min_{P\geq 1} \zeta(P)\right\}$.

Proof: (Sketch) Once again, all costs are averaged over vector length l as $l \to \infty$, but we drop the explicit notation for simplicity.

Case 1: (*do-nothing*, cost σ^2) This strategy involves no action on the part of the stage-one controllers, and is useful when σ^2 is comparable to the noise power. The second stage controller takes the cost hit, by estimating $\widehat{\mathbf{x}}_1 = \widehat{\mathbf{w}}_1 = 0$, and we have $\mathbb{E}[J_1] = 0$ and $\mathbb{E}[J_2] = \sigma^2$.

Case 2: (*lattice quantization*, cost $\min_{P\geq 1} \zeta(P)$) This strategy is useful when the available power, P, is greater than the noise power 1. Consider two nested lattices, Λ_P (the coarse lattice) and Λ_1 (the fine lattice), of second moments (i.e. Voronoi cell radii) P and 1 respectively. Intuitively, following the deterministic model, the "helper" C_w can clear out up to $\log_2 P$ bits by pushing $\mathbf{w_0}$ to one of the points of Λ_P, so that $\mathbf{w_1}$ disappears modulo Λ_P.

C_x then communicates bits of $\mathbf{x_0}$ in this cleared out space using the points of Λ_1. A modulo-lattice decoder can decode (possibly with errors) the fine and coarse lattice points separately. However, unless P is large enough, the resolution provided by the fine lattice may not be sufficient to decode $\mathbf{x}_0$ noiselessly. The focus on $\mathbf{x_0}$ instead of $\mathbf{x_1}$ leads to an addition to the second stage cost, which we quantify below.

We omit the details of this strategy here (see [12]). This requires the use of inflated versions of Λ_P and Λ_1, as in Philosof et al. [11]. If the decoder would like to recover explicit messages (say $\mathcal{M}_1, \mathcal{M}_2$) from two encoders (with power P), in a doubly dirty MAC, at rates R_1, R_2, then all rate pairs $R_1 + R_2 \leq \left[\frac{1}{2}\log_2(\frac{1}{2} + P)\right]_+$ are achievable [11].

$\mathbf{x}_0$ can be thought of as message $\mathcal{M}_1$. Based on the above achievability result, we know that $R_{x_0} = R_1 = \frac{1}{2}\log_2(\frac{1}{2}+P)$ is achievable, if $R_2 = 0$ and $\mathbb{E}||\mathbf{u}_x||^2, \mathbb{E}||\mathbf{u}_w||^2 \leq P$. The rate distortion theorem ([13], Thm. 10.3.2) gives $\mathbb{E}||\mathbf{x_0} - \widehat{\mathbf{x}}_\mathbf{0}||^2 = \frac{\sigma^2}{.5+P}$. Since, $\mathbb{E}||\mathbf{x_0} - \mathbf{x_1}||^2 \leq P$, we use the triangle inequality as in [1], and hence, $\mathbb{E}||\mathbf{x_1} - \widehat{\mathbf{x}}_\mathbf{1}||^2 \leq \left(\sqrt{\frac{\sigma^2}{.5+P}} + \sqrt{P}\right)^2$. Using this, $\widehat{\mathbf{x}}_\mathbf{1}$, the MMSE estimation of $\mathbf{w_1}$, as $\widehat{\mathbf{w}}_\mathbf{1} = \mathbf{y} - \widehat{\mathbf{x}}_\mathbf{1}$ gives, $\mathbb{E}||\mathbf{w_1} - \widehat{\mathbf{w}}_\mathbf{1}||^2 \leq \left(\sqrt{\frac{\sigma^2}{.5+P}} + \sqrt{P}\right)^2 + 1$.

Thus, an achievable cost for every P is $2k^2P + 2\left(\sqrt{\frac{\sigma^2}{.5+P}} + \sqrt{P}\right)^2 + 1$. Minimizing over P gives the optimal tradeoff between first and second stage costs.

Remark 5.2: We manage to recover $\mathbf{x_1}$ (acting as the message) and $\mathbf{w}_1$ (acting as the state) to almost the same fidelity. The *implicit* nature of the communication and the message-state double role played by $\mathbf{w}_1$ help us get two for the price of one.

Case 3: (*coding*, cost $2k^2\sigma$) The coding strategy is really a special case of the lattice strategy in Case 2, where P can "make enough room" for $\mathbf{x}_1$ to be communicated noiselessly. We can check by comparing the rate region of the MAC to the rate-distortion function that if $P \geq \sigma$, when $\sigma > 1$, the stage one controllers can compress $\mathbf{x}_0, \mathbf{w}_0$ to $\mathbf{x}_1, \mathbf{w}_1$, which have entropies small enough that they can be communicated noiselessly over the MAC (details omitted, see [12]). In this case, $\mathbb{E}[J_1] = 2k^2\sigma$ and $\mathbb{E}[J_2] = 0$.

Case 4: (*zero-forcing*, cost $2k^2\sigma^2$) When $P \geq \sigma^2$ we can push the state to zero at the first stage. Again, the second stage controller estimates both $\widehat{\mathbf{x}}_1 = \widehat{\mathbf{w}}_1 = 0$, but this time at no cost. The costs achieved are $\mathbb{E}[J_1] = 2k^2\sigma^2$ and $\mathbb{E}[J_2] = 0$. ■

Theorem 5.3: $\min_{P\geq 0}\left\{2k^2P + \left(\left[\sqrt{\frac{\sigma^2}{1+4P}} - \sqrt{P}\right]_+\right)^2\right\}$ is a lower bound on the Gaussian case observation-setup MAW problem. The scheme presented in Thm. 5.2 achieves this bound to within a constant factor, $\rho \leq 100$.

We defer the proof of this to the extended version of the paper [12]. Since the bounding techniques used in the proof are loose we only prove $\rho \leq 100$, however numerical calculations show that the ratio is in fact ≤ 4.

VI. Conclusions and Future Work

With insight gained from the deterministic model, we were able to provide strategies for the multiple-access Witsenhausen problem. The deterministic model suggested key ideas (e.g. quantization and "making room"), which hinted at the use of structured codes in the Gaussian problem. Lattice coding was essential in both versions of the problem, reemphasizing its importance as a technique in multi-user communication.

Communication problems focus on sending a message from an encoder to a decoder. Control problems, on the other hand, focus on state control and estimation. In the MAW problem, we find that system states can play double roles: they act as both messages and interference. This double role was not observed in the original Witsenhausen counterexample. Gelfand-Pinsker [14] channels assume the interference, S, is independent of other variables in the problem and uninteresting to the decoder. Communication rate is measured in terms of the bits of the encoder message that are communicated to the decoder. However, The MAW problem highlights that "information is in the eye of the beholder". Since the MAW decoder cares about the internally generated interference/state, it learns more relevant information and we achieve a "higher rate" of *implicit* communication.

Versions of the MAW problem with more than two users, as well as problems with asymmetries between the stage-one power costs and initial state variances remain open for future work. Systems with explicit messages and channels as in [6], [7] can also be explored. Based on preliminary investigations, we do not expect significant qualitative differences in the result in any of these cases.

VII. Acknowledgments

We would like to thank Pulkit Grover for suggesting the direction of research and helpful discussions, as well as Pulkit Grover, Kristen Woyach and Hari Palaiyanur for comments on the manuscript. Special thanks to the anonymous reviewers whose detailed comments were very useful. We also thank the NSF for a GRF and grant CNS-0932410, which supported this research.

References

[1] P. Grover and A. Sahai, "The vector Witsenhausen problem as assisted interference cancelation," *International Journal on Systems, Control and Communications (IJSCC)*, 2008.

[2] S. Avestimehr, S. Diggavi, and D. Tse, "Wireless Network Information Flow: A Deterministic Approach," *Arxiv preprint arXiv:0906.5394*, 2009.

[3] C. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[4] P. Grover, A. Sahai, and S. Park, "The finite-dimensional Witsenhausen counterexample," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on*. IEEE, 2009, pp. 1–10.

[5] O. Sumszyk and Y. Steinberg, "Information embedding with reversible stegotext," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*. IEEE, 2009, pp. 2728–2732.

[6] P. Grover, A. Wagner, and A. Sahai, "Information embedding meets distributed control," in *Information Theory Workshop (ITW), 2010 IEEE*. IEEE, 2010, pp. 1–5.

[7] P. Grover and A. Sahai, "Implicit and explicit communication in decentralized control," in *48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois.*, 2010.

[8] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *Information Theory, IEEE Transactions on*, vol. 54, no. 12, pp. 5534–5562, 2008.

[9] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498–3516, 2007.

[10] S. Kotagiri and J. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, pp. 1–14, 2008.

[11] T. Philosof, R. Zamir, U. Erez, and A. Khisti, "Lattice Strategies for the Dirty Multiple Access Channel," *Arxiv preprint arXiv:0904.1892v1*, 2009.

[12] G. Ranade and A. Sahai, "Implicit Communication in Multiple-Access Settings," *To be submitted.*

[13] T. Cover and J. Thomas, *Elements of information theory*, 2nd ed. Wiley, 2006.

[14] S. Gelfand and M. Pinsker, "Coding for a channel with random parameters," *Probl. Control and Inf. Theory*, vol. 9, pp. 19–31, 1980.