

Outline:

- 1 The Witsenhausen counterexample
- 2 Towards the “simplest” unsolved IT problem
- 3 Circumstantial evidence for DPC
- 4 A new bound and approximately optimal solution in asymptopia
- 5 Approximate optimality in non-asymptopia

Anant Sahai, Pulkit Grover and Se Yong Park

What does dirty-paper coding have to do with distributed control?

Revisiting the Witsenhausen Counterexample.

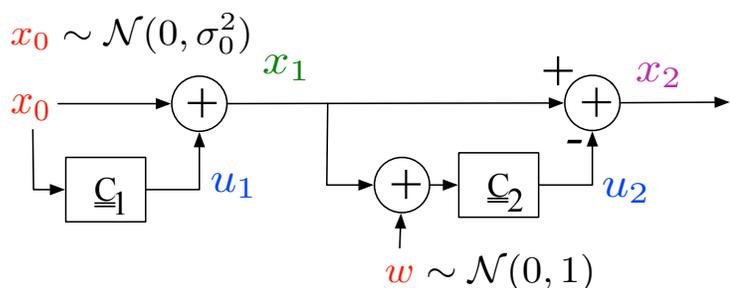
*Systems Seminar at UIUC*, May 8, 2009.

Slides available: [www.eecs.berkeley.edu/~sahai/Presentations/UIUC09.pdf](http://www.eecs.berkeley.edu/~sahai/Presentations/UIUC09.pdf)

This Handout: [.../~sahai/Presentations/UIUC09.H.pdf](http://.../~sahai/Presentations/UIUC09.H.pdf)

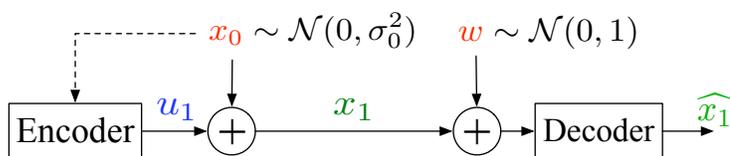
Further discussion and references can be found in the papers.

## 1 The Witsenhausen counterexample



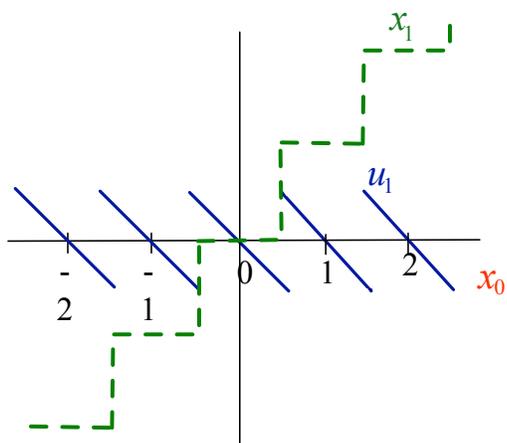
The Witsenhausen counterexample [Witsenhausen '68] is a distributed Linear-Quadratic-Gaussian (LQG) control problem with total cost  $\mathcal{C} = k^2 u_1^2 + x_2^2$ . There are two natural parameters  $\sigma_0^2$  – the initial state variance – and  $k^2$  – the relative importance of the two cost terms. Calculating the optimal linear control law is simple, but Witsenhausen gave a nonlinear control law that performed even better.

It is clear that the second controller should compute  $E[X_1|Y]$  as the MMSE estimate to minimize  $E[X_2^2]$  and thus the problem can be viewed as one of estimation. However, the first controller is from an infinite-dimensional space and the cost turns out to be non-convex over it. The optimization is hard and it is an open problem to find the optimal controller.



The scalar Witsenhausen counterexample

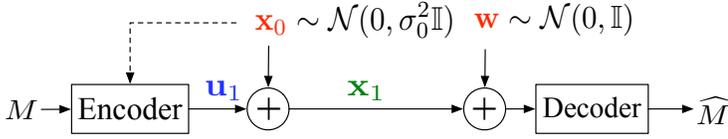
Mathematically, [Bansal and Basar '87] showed that the problem is hard in part because the second stage cost  $x_2^2 = x_0^2 + u_1^2 + u_2^2 + 2x_0u_1 - 2x_0u_2 - 2u_1u_2$  contains a nonzero cross-term  $u_1u_2$ . Without it, linear strategies are optimal. Similarly, [Rotkowitz '06] showed that a deterministic variant of the problem with no stochastic model for  $X_0$  also has linear optimal strategies. [Papadimitriou and Tsitsiklis '86] showed natural discretized versions of the counterexample were NP-complete.



Intuitively, it is the possibility of “signaling” or implicit communication – using control signals to talk to other controllers through changing the system state – that makes the Witsenhausen counterexample hard; but this is also a ubiquitous feature in distributed control. [Rotkowitz and Lall '06] made this intuition precise for larger distributed systems by introducing the concept of quadratic invariance and using it to show that if controllers can talk to each other before their outputs effect each other’s inputs, then the problem is convex. [Mitter and Sahai '99] instead used this intuition to design an explicit communication-inspired control strategy where the first controller quantizes the state  $x_0$  using a regular lattice. This established that nonlinear control can result in a cost that can be an arbitrarily-high factor better than the optimal linear strategy — so the possibility of implicit communication cannot be safely ignored.

## 2 Towards the “simplest” unsolved IT problem

Even in the 70’s, Witsenhausen’s counterexample and information theory were considered related. In the 80’s, the new framework of **dirty-paper coding (DPC)** [Costa ’83] emerged as strategically important in IT as a particularly idealized kind of interference. [Devroye, Mitran, Tarokh ’06] crucially used it in the analysis of the “cognitive radio channel.” Inspired by this, some recent extensions to DPC, **state amplification** (*communicating  $x_0$* ) [Kim, Sutivong, Cover ’08] and **state masking** (*hiding  $x_0$* ) [Merhav, Shamai ’06] have been successful.



Dirty paper coding

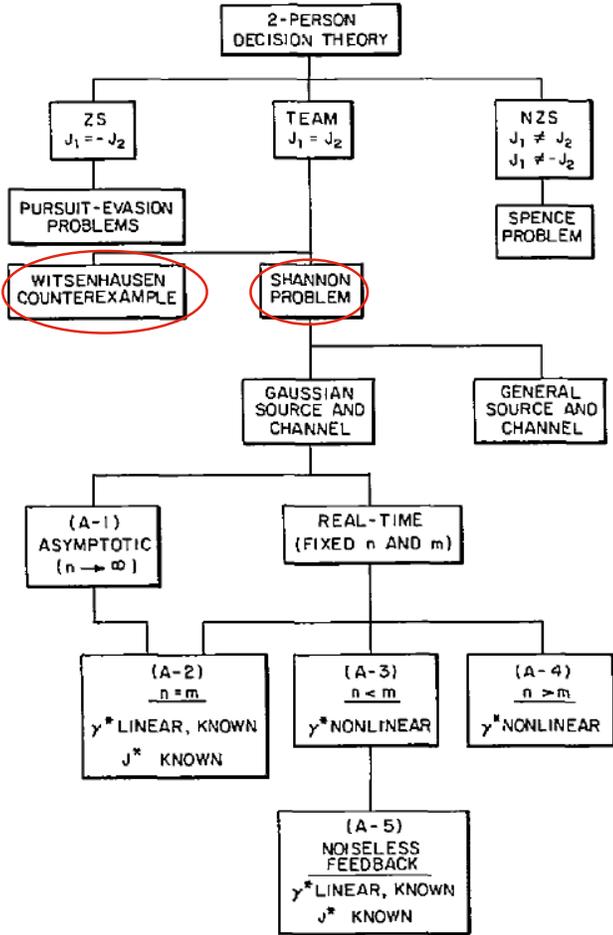
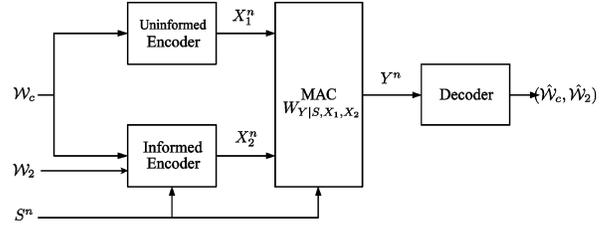
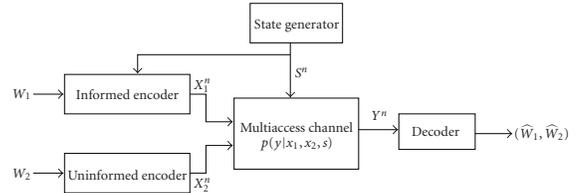


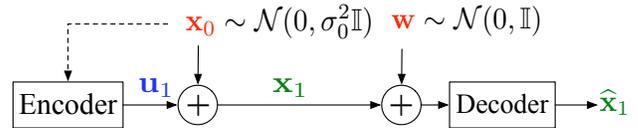
Fig. 1. Teams, signaling, and information theory.



Distributed DPC is more interesting. If the informed encoder knows all the messages, [Somekh-Baruch, Shamai, Verdú ’08] solve the problem.



But if the informed encoder does not know the other message [Kotagiri, Laneman ’08], the problem is unsolved even if the informed encoder has no message of its own. This suggests a further simplification:

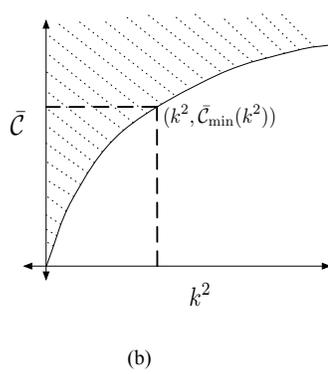
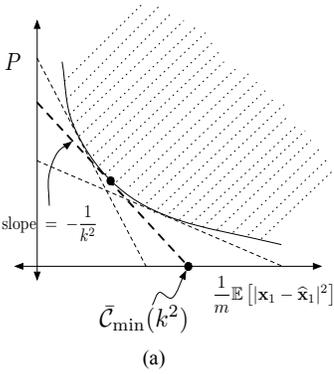
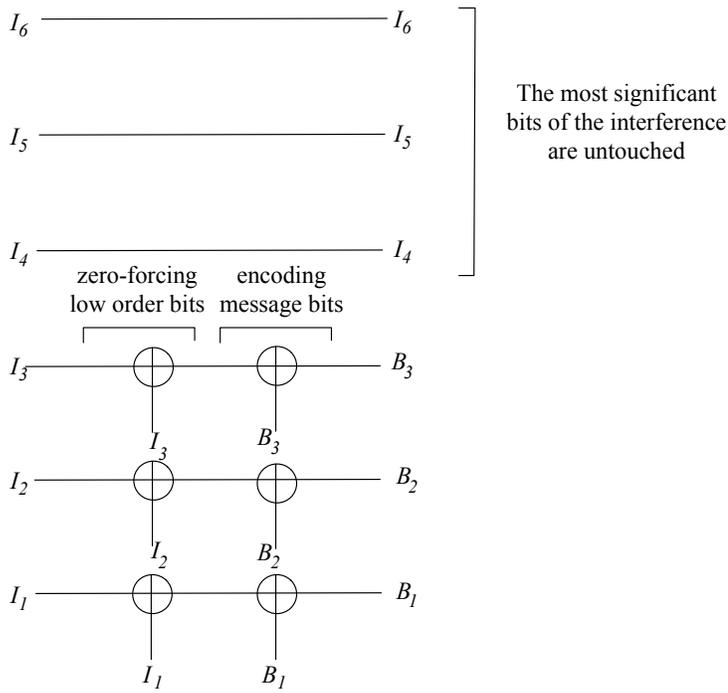


$$\mathbb{E}[C] = k^2 P + \frac{1}{m} \mathbb{E} [\|\mathbf{x}_1 - \hat{\mathbf{x}}_1\|^2]$$

The vector Witsenhausen counterexample

All messages are now eliminated and we have a point-to-point problem where the encoder just attempts to “clean” the “noisy channel” so the net interference  $x_1$  better estimateable. Amazingly, this is the Witsenhausen problem viewed as “assisted interference suppression.”

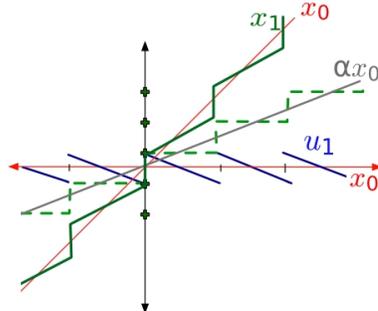
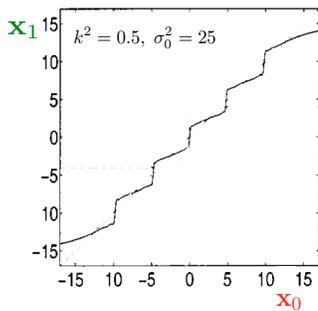
The left figure is taken from [Ho, Kastner, Wong ’78].



In addition to the progress inspired by DPC and the “cognitive radio channel,” the community’s approaches to understanding communication problems have also changed. [Gupta and Kumar '00] introduced the idea of “scaling laws” instead of exact capacities to the communication community, and since then we’ve also thought about other sorts of scaling like what happens in high-SNR limits. Recently, [Etkin, Tse, Wang '07] in the context of the interference channel introduced the idea of understanding a problem by approximating capacity to within a constant number of bits regardless of the problem parameters. [Avestimehr, Diggavi, Tse '07] further introduced the idea of using simple deterministic models to guide our intuitions about interference: noise is suppressed and real addition is replaced by binary XOR at the bit-wise level.

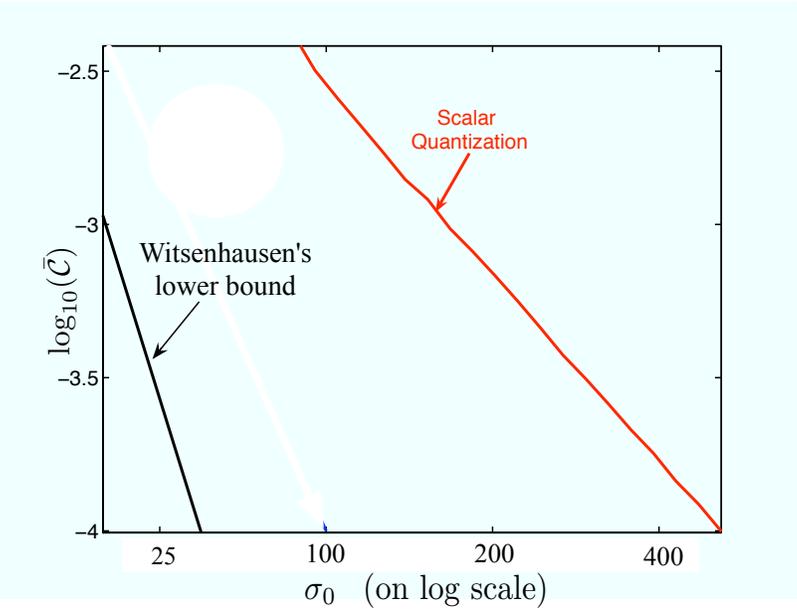
The deterministic model of distributed dirty-paper coding suggests that it should be possible to distribute the tasks of “cleaning the dirt” and communicating the message. This suggests that an approximate optimality result should be possible for the Witsenhausen counterexample in its assisted interference suppression manifestation. A simple convexity argument shows that the Witsenhausen perspective (b) of  $k$  vs total cost is equivalent to understanding the (a) tradeoff between first-stage input power and estimation error.

### 3 Circumstantial evidence for dirty-paper-coding



Numerical search results in [Baglietto, Parisini, Zoppoli '97][Lee, Lau, Ho '01] strongly suggest that in an interesting regime of small  $k$  and large  $\sigma_0^2$ , *soft-quantization based strategies* might be optimal for the original Witsenhausen counterexample. The strategy can be interpreted as quantizing a scaled down  $x_0$ , and adding the resulting input  $u_1$ . This is precisely the DPC-technique applied to scalars!

Left figure from [Baglietto, Parisini, Zoppoli '97].



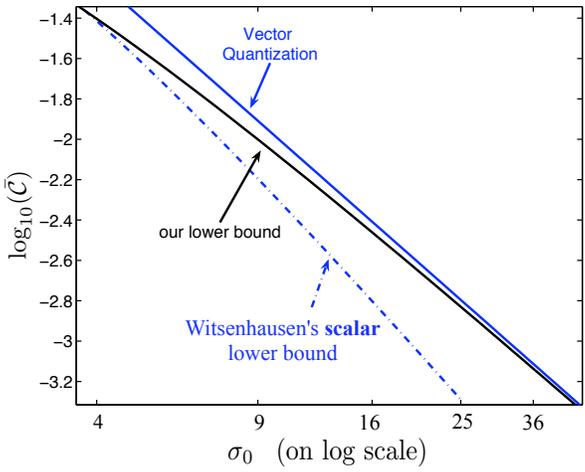
Witsenhausen derived the following lower bound to the total cost for the scalar problem.

$$\bar{C}_{\min}^{\text{scalar}} \geq \frac{1}{\sigma_0} \int_{-\infty}^{+\infty} \phi\left(\frac{\xi}{\sigma_0}\right) V_k(\xi) d\xi,$$

where  $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})$ ,  $V_k(\xi) := \min_a [k^2(a - \xi)^2 + h(a)]$ , and  $h(a) := \sqrt{2\pi} a^2 \phi(a) \int_{-\infty}^{+\infty} \frac{\phi(y)}{\cosh(ay)} dy$ .

His bound holds only for the scalar case and furthermore, there is a substantial and growing gap from the performance of quantization-based strategies. (plot keeps  $\sigma_0 k = 1$ )

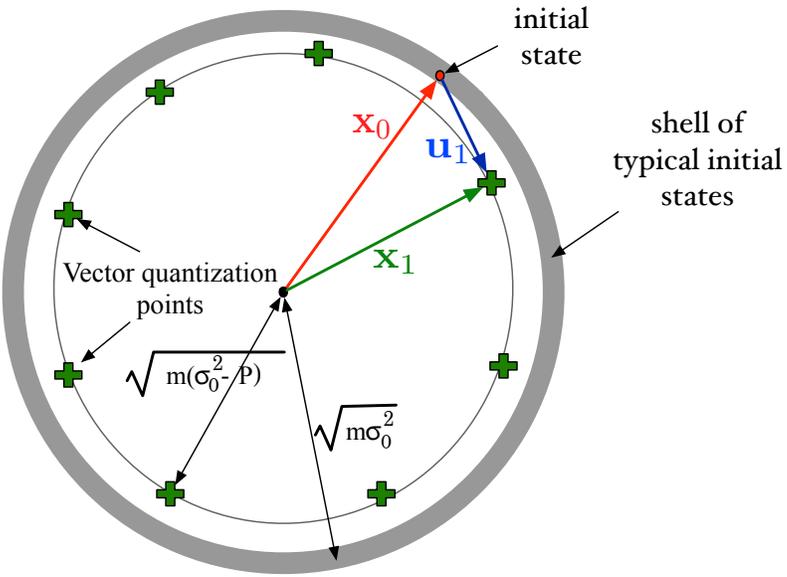
### 4 A new bound and approximately optimal solution in asymptopia



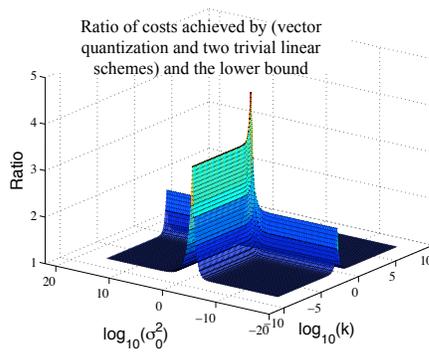
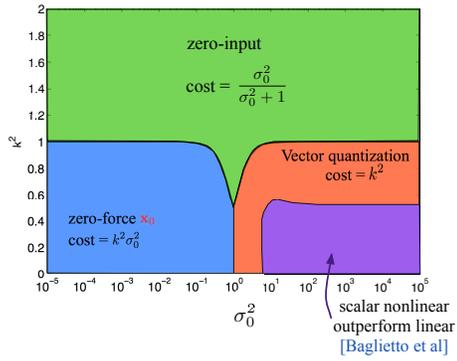
An information-theoretic bound can be derived by using rate-distortion theory and treating the second-stage input noise as a Gaussian channel. For maximum capacity across the channel,  $\mathbf{x}_1$  should be Gaussian and large. For easy reconstruction,  $\mathbf{x}_1$  should be non-Gaussian and small. Our lower bound ignores this tension but is valid for all vector lengths  $m \geq 1$ ,

$$\bar{C}_{\min} \geq \inf_{P \geq 0} k^2 P + \left( (\sqrt{\kappa(P)} - \sqrt{P})^+ \right)^2,$$

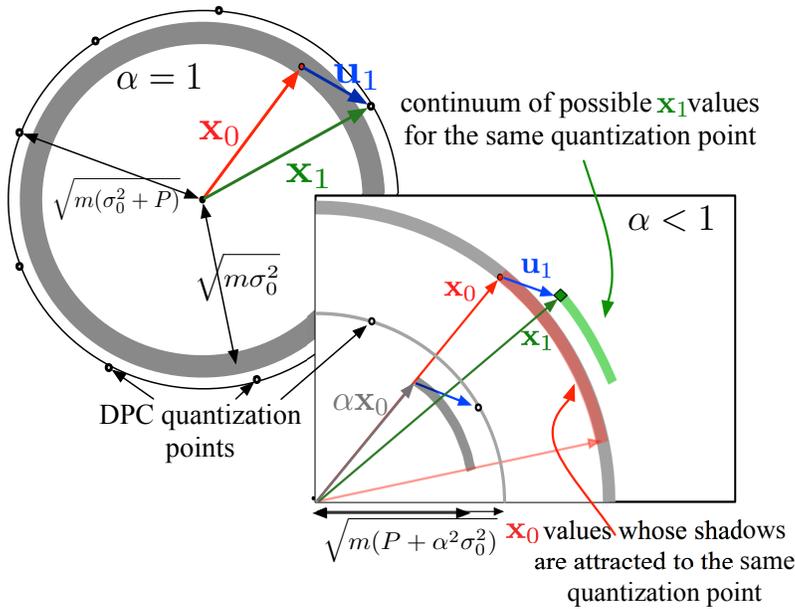
where  $\kappa(P) = \frac{\sigma_0^2}{\sigma_0^2 + 2\sigma_0 \sqrt{P} + P}$ .



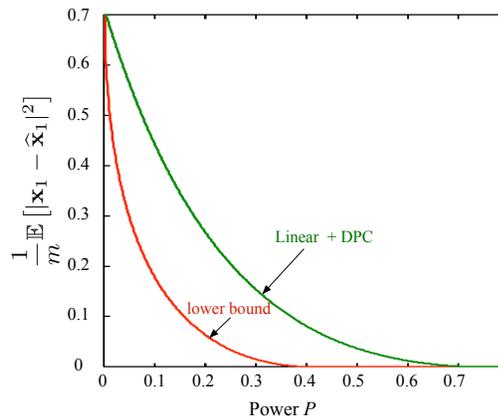
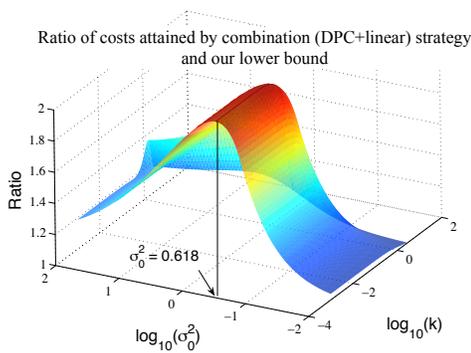
A **vector quantization** based strategy has *the encoder drive the state  $\mathbf{x}_0$  to the nearest quantization point*. These quantization points have power smaller than  $\sigma_0^2$ . Provided the number of quantization points is sufficiently small, they can be decoded correctly at the second controller. The asymptotic cost is  $k^2 \sigma_w^2$  and 0 for the first and the second stage respectively.



The quantization strategy can be combined with simple linear strategies depending on which performs best. This results in a cost that is provably within a constant factor of the new lower bound, regardless of the problem parameters. Numerically, the constant is something less than 5 for this strategy.

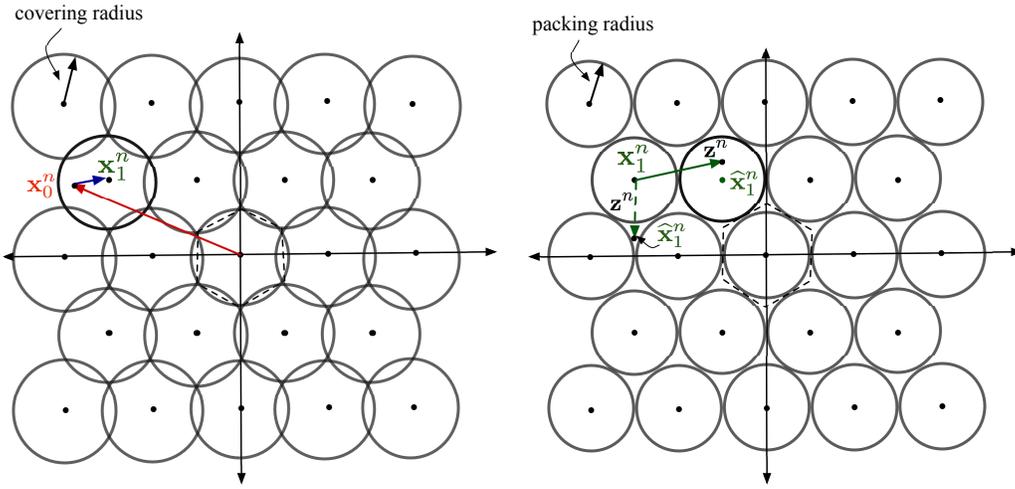


A DPC-based strategy where *the shadow state  $\alpha x_0$  is driven to the nearest quantization point* is a natural generalization of strategies in [Baglietto et al '97][Lee, Lau, Ho '01]. The first stage cost can be lowered at the expense of nonzero second stage costs. A combination strategy can *divide its power between a linear strategy and the DPC strategy*. It performs at least as well, and in some cases strictly better than the DPC strategy alone.



The ratio of the cost attained by the combination strategy and our lower bound is uniformly bounded by 2 for all values of  $k$  and  $\sigma_0^2$ . However, a look in the tradeoff between first-stage power and estimation error shows that there is still significant room for improvement in these bounds.

## 5 Approximate optimality in non-asymptopia

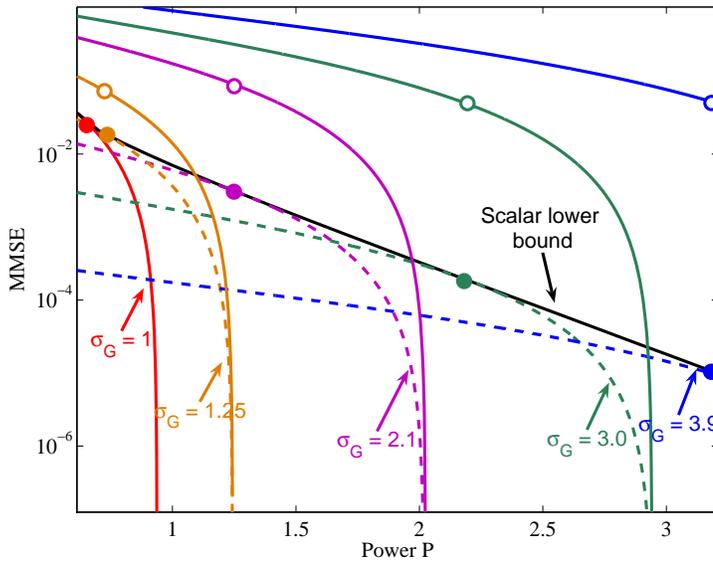


There are natural finite-dimensional generalizations of the scalar quantization strategy. The first-stage cost is bounded by the square of the covering-radius and the estimation-error cost will be zero as long as the noise stays within the packing radius. The key is to use a lattice with a good covering-to-packing ratio. Constructions are known so that this ratio is no more than 4 for all dimensions. However, the average estimation-error cost will not truly be zero unless the first-stage completely subtracts the state. There is always the probability of a large observation noise.

In infinite-dimensional asymptopia, the probability of unusually large observation noise is zero. However, this will happen in finite-dimensions with some small probability. The lower bound needs to account for this.

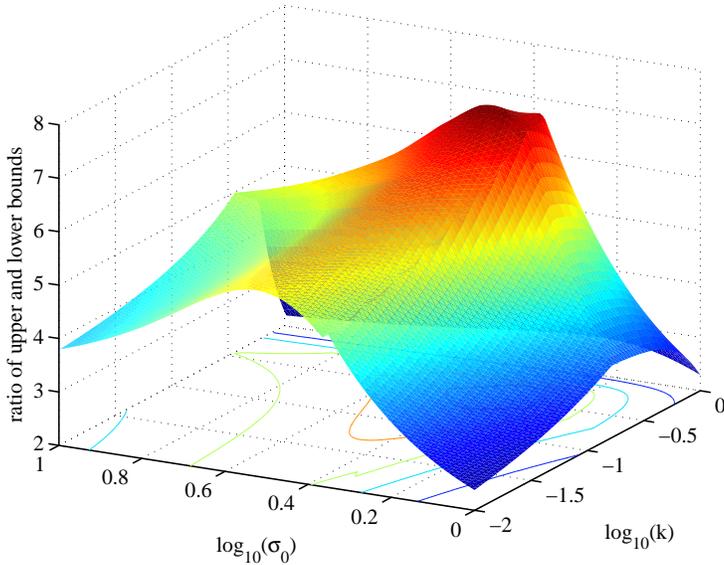
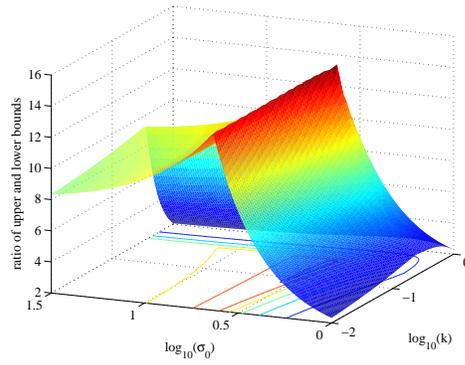
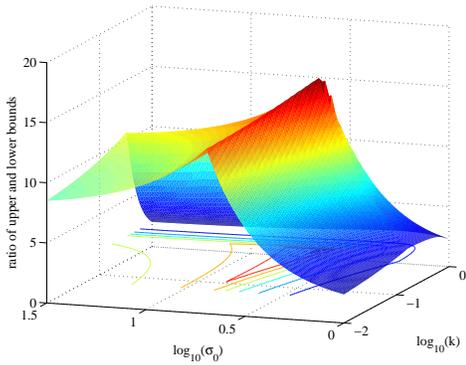
$$\begin{aligned}
 E_b(R) &= \min_{Q: H(Q) \geq R} D(Q||P) \\
 &= \sup_{\rho \geq 0} \rho R - E_0(\rho) \\
 E_0(\rho) &= \ln \left[ \sum_x P(x)^{\frac{1}{1+\rho}} \right]^{1+\rho}
 \end{aligned}$$

Here, the Stein's Lemma approach of [Blahut '72] to error-exponents needs to be adapted. The idea is that the sources and channels can “behave like” other sources and channels with some probability with an exponent governed by the KL divergence. We had built upon this approach earlier in our asymptotic study of delay and then non-asymptotically as we explored fundamental bounds for iterative decoding to understand decoder power consumption.

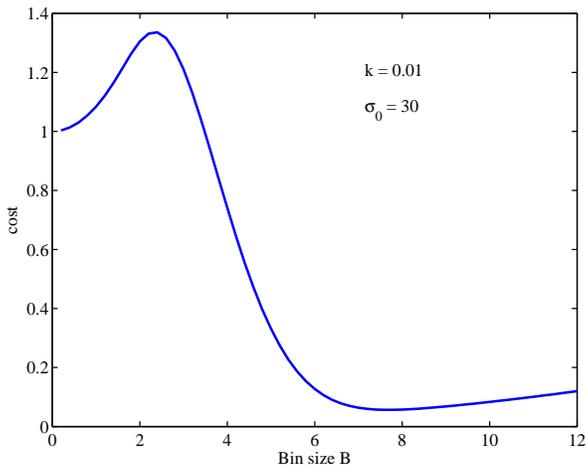


This “Platonic” approach treats the finite-dimensional world as shadows of the infinite-dimensional world. The earlier bound can be re-done for different values of the observation noise variance  $\sigma_G^2$  and then pulled back to the scalar case roughly by multiplying the lower bound by the probability of that atypical channel behavior. The upper-envelope of all such bounds then provides a bound for the scalar case.

The resulting bound turns out to be good enough to say that linear-strategies plus finite-dimensional lattice-based control strategies get within a constant-factor of the true optimal cost. This works even with a coarse covering/packing-ratio based upper-bound, and numerically seems to get better as we increase from 1 to 2 dimensions and beyond.



A more careful evaluation of the quantization-based strategy in the scalar case shows that the approximation ratio is something less than 8. We were surprised that this large-deviations based bound was so tight even when  $m = 1$ .



This result suggests a new way forward for distributed control. We can restrict attention to tractable nonlinear control strategies and still get provable results. However, at least for quantization, the problem of choosing the optimal bin-size appears nonconvex with a local minimum. There is probably a principled way around this.

## 6 Summary

This talk intends to convey the following ideas:

- Witsenhausen’s counterexample brings out the core issue of implicit communication or signaling in distributed control systems.
- Inspired by the “cognitive radio channel,” Witsenhausen’s counterexample can be viewed as an oversimplification that might contain the essence of why distributed DPC is hard. It focuses on the problem of channel cleaning or active interference suppression.
- Standard information-theoretic tools are able to provide an approximately optimal solution (within a factor of two in cost) to the problem in the asymptotic limit of long block lengths. But the problem remains open and is arguably the simplest open problem in information theory since it is just point-to-point.
- Returning to the control problem, a careful “Platonic perspective” on sphere-packing bounds gives a new bound that reveals that a similar constant-factor result is true even for the scalar problem, giving us the first *provably* positive result on the original Witsenhausen counterexample in 40+ years.
- This is important because it shows that we can get a handle on the most significant effects of implicit communication in distributed control systems. As they say, “science is the art of the solvable” and these issues just got moved onto the solvable side of things from where they firmly sat earlier on the intractable side.
- Many incremental improvements are likely possible. This does not look like an impossible problem for information theory.