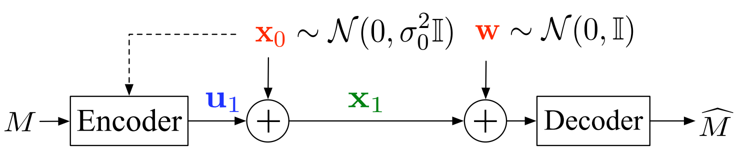
Anant Sahai, Pulkit Grover and Se Yong Park

An information-theoretic view of the Witsenhausen counterexample

Workshop on Information Theory and its Applications (ITA), Feb. 9, 2009.

Slides available: www.eecs.berkeley.edu/~sahai/Presentations/ITA09.pdf
This Handout: .../~sahai/Presentations/ITA09.H.pdf
Further discussion and references can be found in the paper.

1 Towards the "simplest" unsolved IT problem



Dirty paper coding

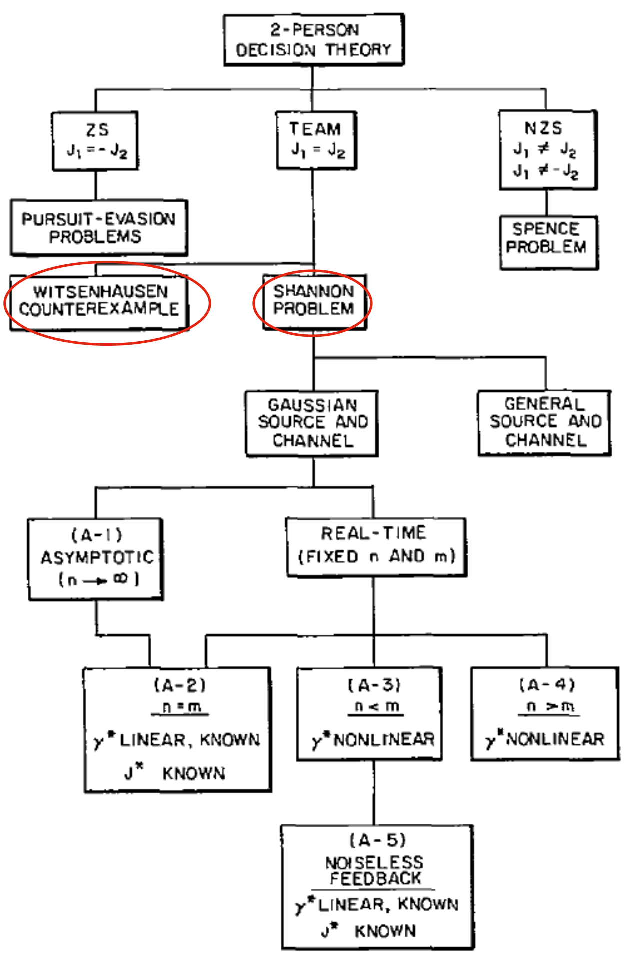
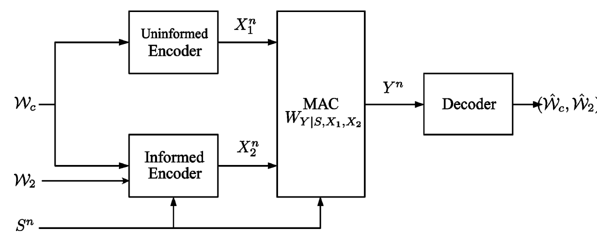
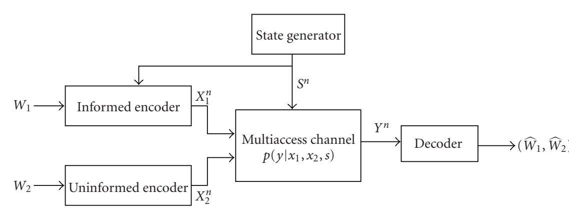


Fig. 1. Teams, signaling, and information theory.

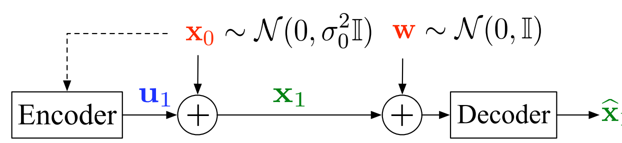
Even in the 70's, IT and Witsenhausen's counterexample were considered related. In the 80's, the new framework of dirty-paper coding (DPC) [Costa '83] emerged as strategically important in IT as a particularly idealized kind of interference. Some recent extensions, state amplification (*communicating* $\mathbf{x}_0$) [Kim, Sutivong, Cover '08] and state masking (*hiding* $\mathbf{x}_0$) [Merhav, Shamai, '06] have been successful.



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.



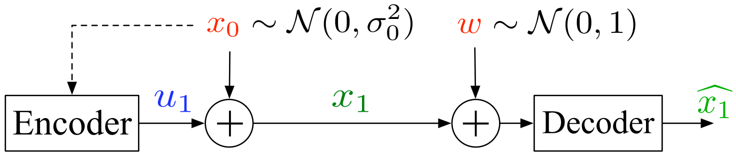
$$\mathbb{E}[\mathcal{C}] = k^2 P + \frac{1}{m} \mathbb{E}\left[\|\mathbf{x}_1 - \widehat{\mathbf{x}}_1\|^2\right]$$

The vector Witsenhausen counterexample

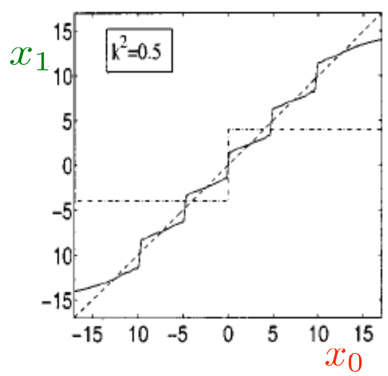
This suggests the above point-to-point natural simplification of the problem in which the encoder attempts to "clean" the "noisy channel" by making the net interference x_1 better estimateable. Amazingly, this is just the Witsenhausen problem.

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

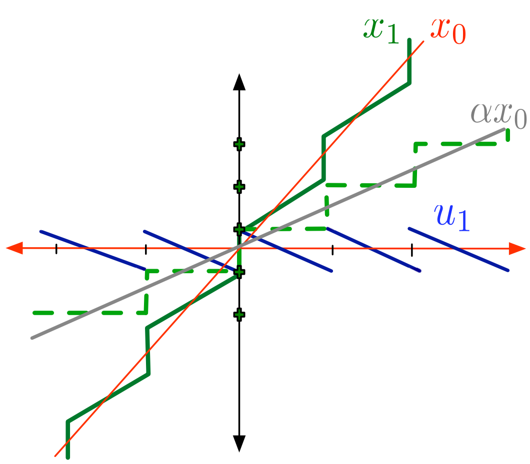
2 Known results for the original (scalar) counterexample



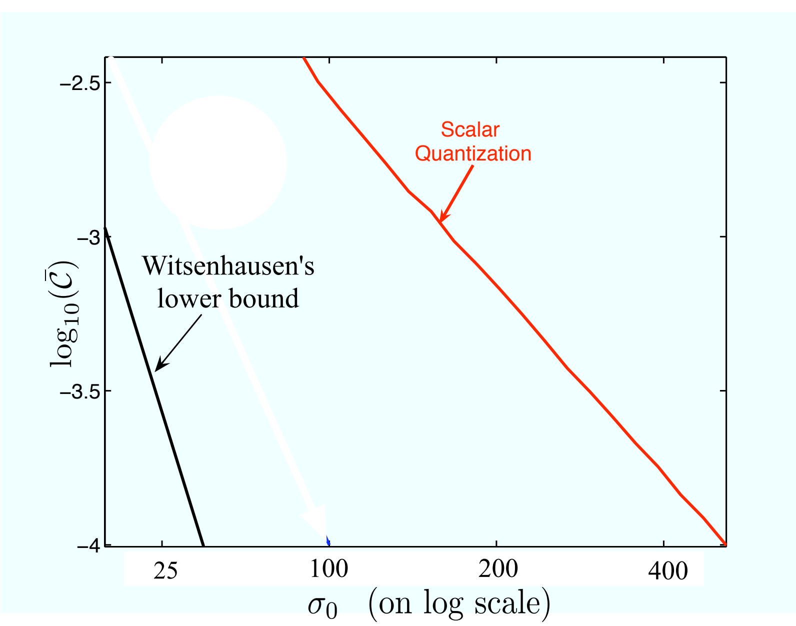
The scalar Witsenhausen counterexample



Witsenhausen's counterexample [Witsenhausen '68] is a distributed Linear-Quadratic-Gaussian (LQG) control problem with total cost $\mathcal{C} = k^2 {u_1}^2 + (x_1 - \widehat{x}_1)^2$. Contrary to non-distributed LQG systems, the optimal control law for the counterexample is nonlinear, and is still unknown. Quantization-based signaling strategies can outperform all linear strategies by an arbitrarily large factor [Mitter, Sahai 99].



Further, numerical search results in [Baglietto, Parisini, Zoppoli][Lee, Lau, Ho] suggest that in an interesting regime of small k and large σ_0^2, *soft-quantization based strategies* might be optimal. 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!
The middle figure is taken from [Baglietto, Parisini, Zoppoli].

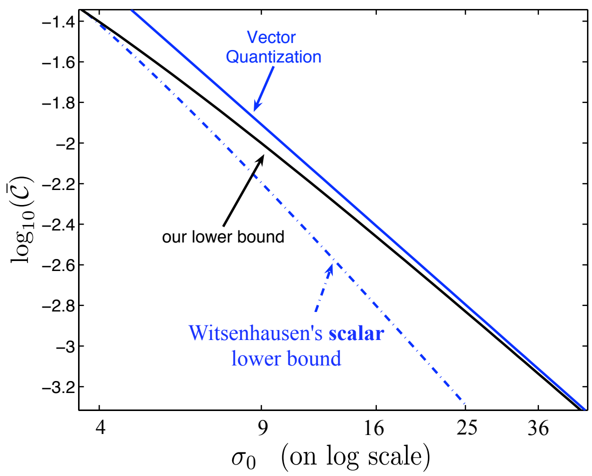


Witsenhausen derived the following lower bound to the total cost for the scalar problem.

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

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 $h(a) := \sqrt{2\pi}a^2\phi(a)\int_{-\infty}^{+\infty}\frac{\phi(y)}{\cosh(ay)}dy$.
However, the derived bound holds only for the scalar case. Further, the technique is loose, and the bound is at a substantial gap from the performance of quantization-based strategies.

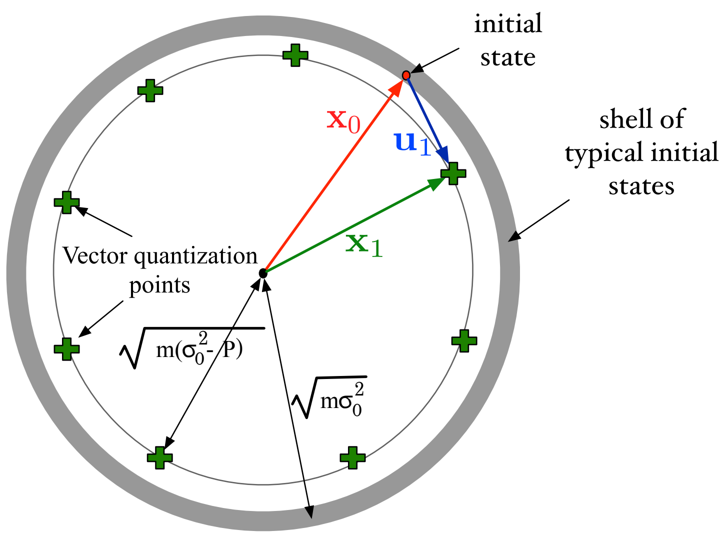
3 Approximately (within a constant factor) optimal solution



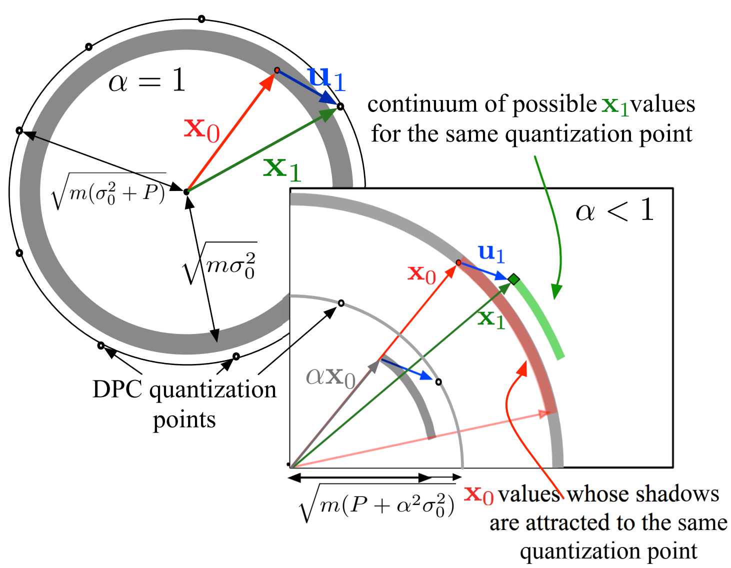
For maximum capacity across the channel, $\mathbf{x}_1$ should be Gaussian and large. For easy reconstruction, $\mathbf{x}_1$ should be non-Gaussian (probably discrete) and small. Our lower bound ignores this tension but is valid for all vector lengths $m \geq 1$,

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

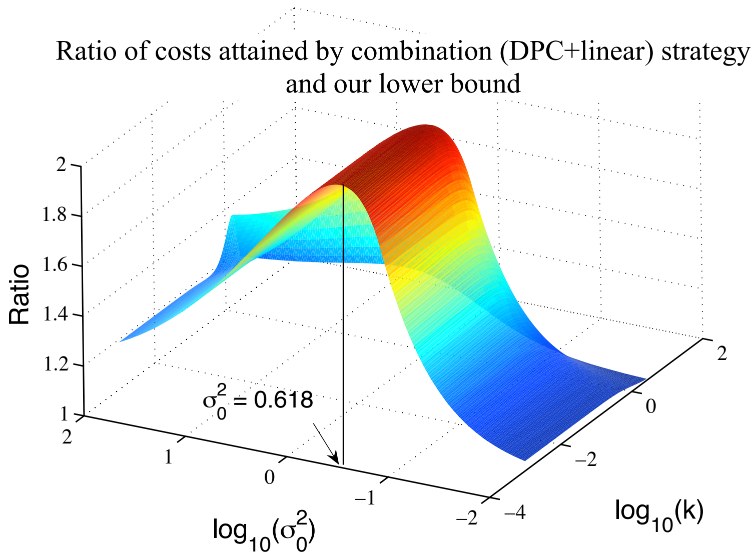
where $\kappa(P) = \frac{\sigma_0^2}{\sigma_0^2 + 2\sigma_0\sqrt{P} + P + 1}$.



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 σ_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.



A DPC-based strategy where *the shadow state $\alpha\mathbf{x}_0$ is driven to the nearest quantization point* is a natural generalization of strategies in [Baglietto et al][Lee, Lau, Ho]. 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 figure shows the ratio of the asymptotic cost attained by the combination strategy and our lower bound. This ratio is uniformly bounded by 2 for all values of k and σ_0^2.

4 Summary

This talk intends to bring out the following ideas:

- 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 since it is just point-to-point.
- Returning to the control problem, more involved arguments tell us that a similar constant-factor result is true even for the scalar problem.