Capacity and Coding for Multi-Antenna Broadcast Channels

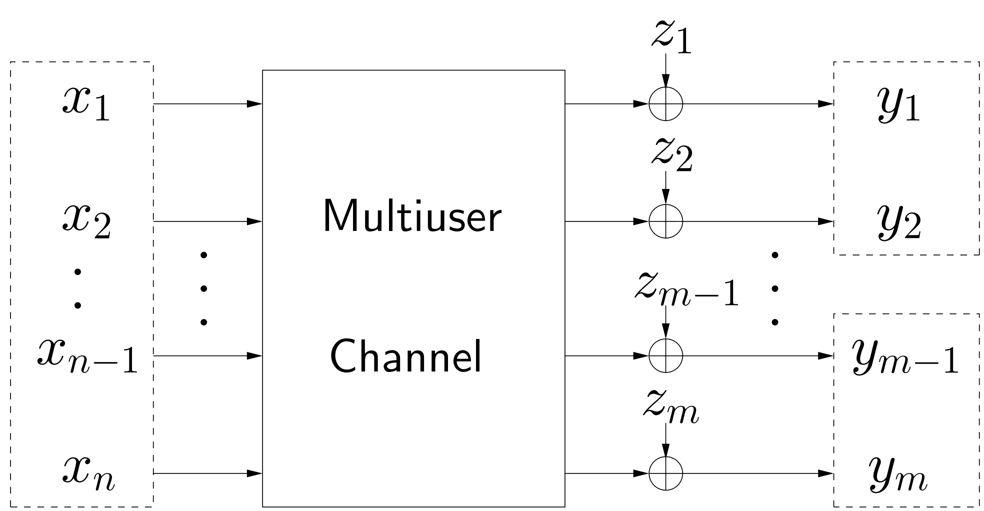
Wei Yu

Electrical Engineering Department
Stanford University

February 20, 2002

Introduction

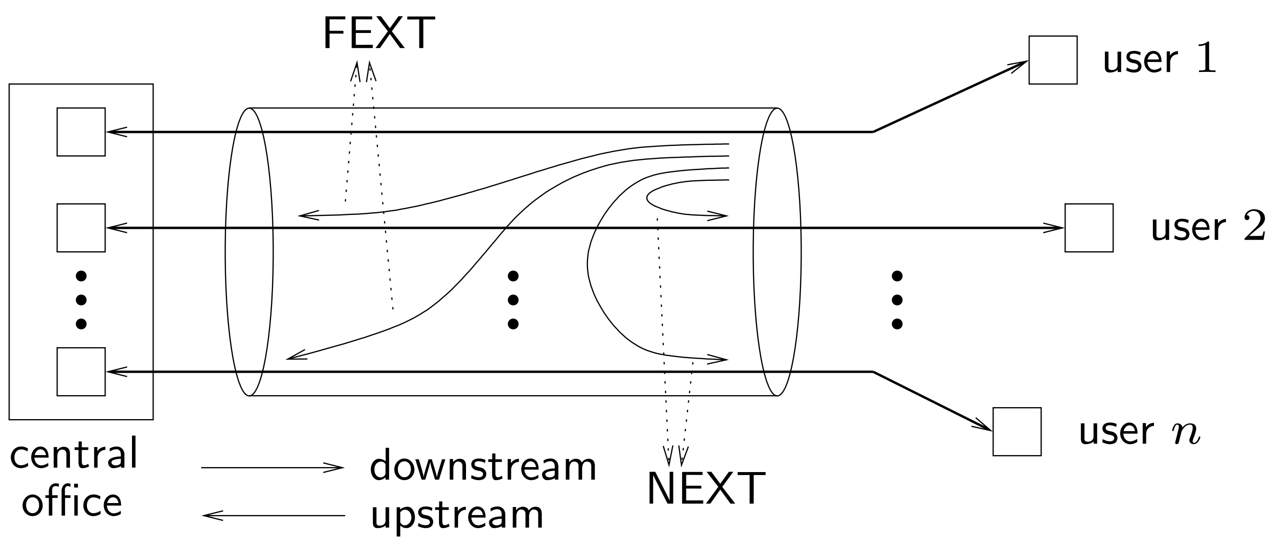
- Consider a communication situation involving mutliple transmitters and receivers:



  – What is the value of cooperation?

Motivation: Multiuser DSL Environment

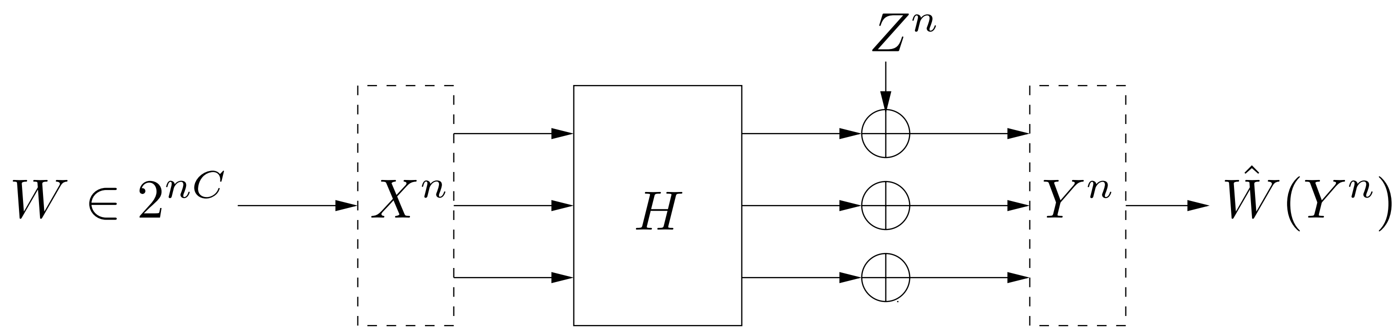
- DSL environment is interference-limited.



  - Explore the benefit of cooperation.

Gaussian Vector Channel

- Capacity: $C = \max I(\mathbf{X};\mathbf{Y})$.

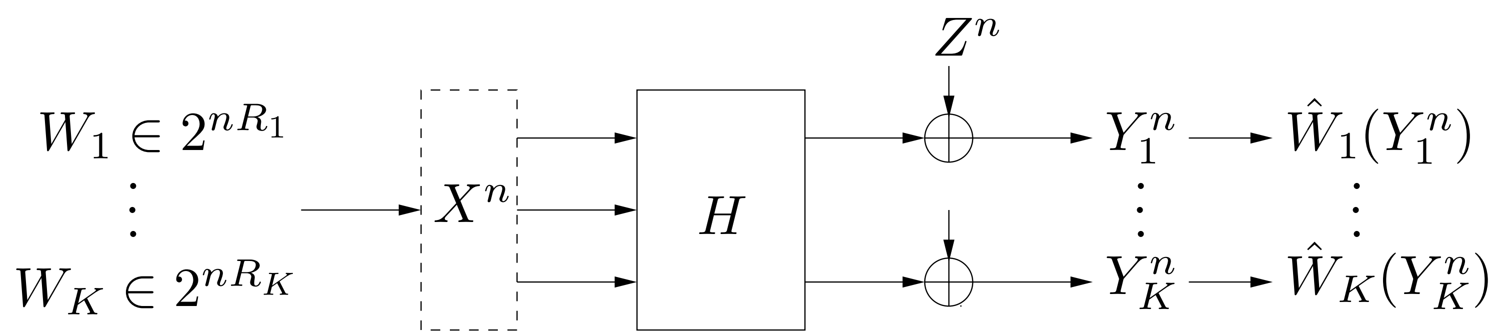


- Optimum Spectrum:

$$\begin{aligned} \text{maximize} \quad & \frac{1}{2}\log\frac{|HK_{xx}H^T + K_{zz}|}{|K_{zz}|} \\ \text{subject to} \quad & \text{tr}(K_{xx}) \le P, \\ & K_{xx} \ge 0. \end{aligned}$$

Gaussian Vector Broadcast Channel

- Capacity Region: $\{(R_1, \cdots, R_K) : \Pr(W_k \neq \hat{W}_k) \to 0,\ k = 1, \cdots K\}$.



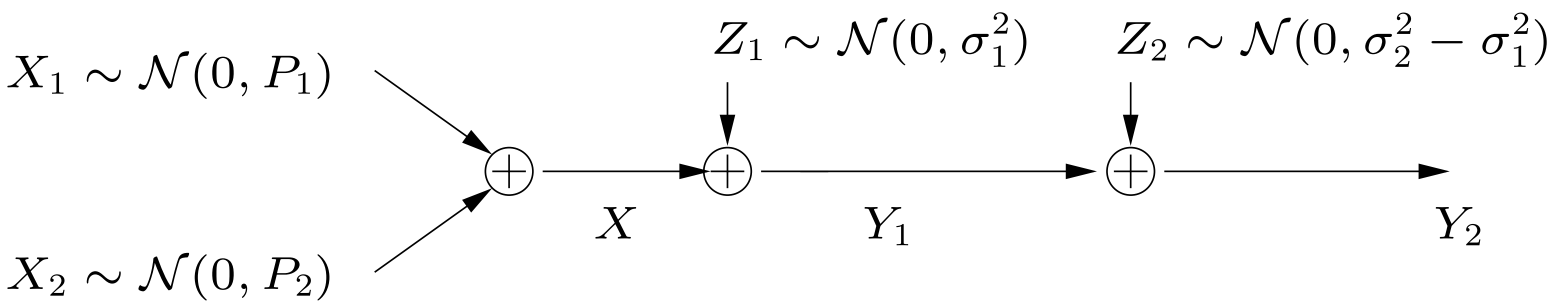
- Capacity is known only in special cases.
  - This talk focuses on sum capacity: $C = \max\{R_1 + \cdots + R_K\}$.

Broadcast Channel: Prior Work

- Introduced by Cover ('72)
  - Superposition coding: Cover ('72).
  - Degraded broadcast channel: Bergman ('74), Gallager ('74)
  - Coding using binning: Marton ('79), El Gamal, van der Meulen ('81)
  - Sum and product channels: El Gamal ('80)
  - Gaussian vector channel, 2×2 case: Caire, Shamai ('00)

- General capacity region remains unknown.

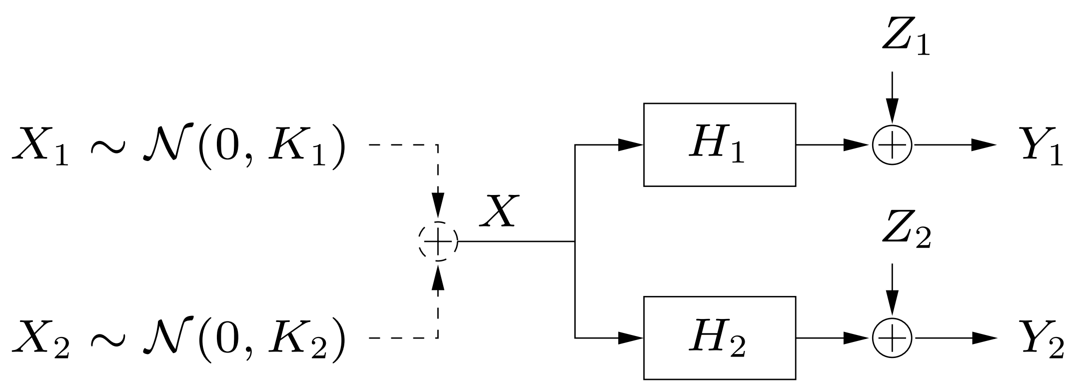
Degraded Broadcast Channel



- Superposition and successive decoding achieve capacity (Cover '72)

$$
\begin{aligned}
R_1 &= I(\mathbf{X_1}; \mathbf{Y_1} | \mathbf{X_2}) = \frac{1}{2} \log \left(1 + \frac{P_1}{\sigma_1^2} \right) \\
R_2 &= I(\mathbf{X_2}; \mathbf{Y_2}) = \frac{1}{2} \log \left(1 + \frac{P_2}{P_1 + \sigma_2^2} \right)
\end{aligned}
$$

Gaussian Vector Broadcast Channel



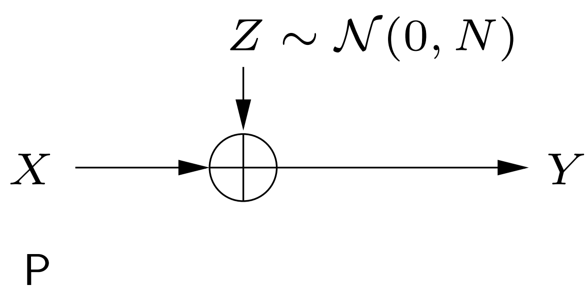
- Superposition coding gives:

$$R_1 = I(\mathbf{X_1}; \mathbf{Y_1}) = \frac{1}{2} \log \frac{|H_1 K_1 H_1^T + H_1 K_2 H_1^T + K_{z_1 z_1}|}{|H_1 K_2 H_1^T + K_{z_1 z_1}|}$$

$$R_2 = I(\mathbf{X_2}; \mathbf{Y_2}) = \frac{1}{2} \log \frac{|H_2 K_2 H_2^T + H_2 K_1 H_2^T + K_{z_2 z_2}|}{|H_2 K_1 H_2^T + K_{z_2 z_2}|}$$

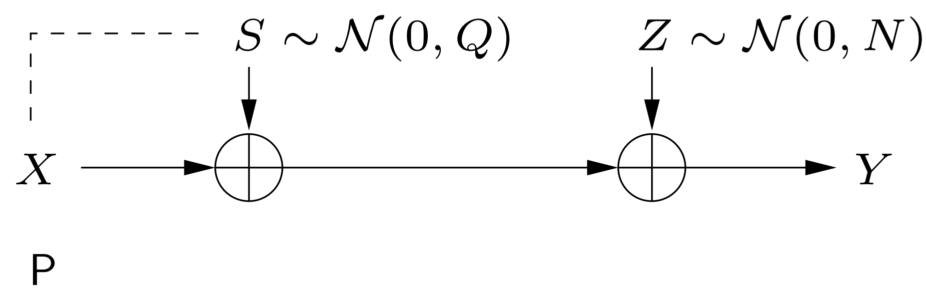
Channel with Transmitter Side Information

Gaussian Channel



$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right)$$

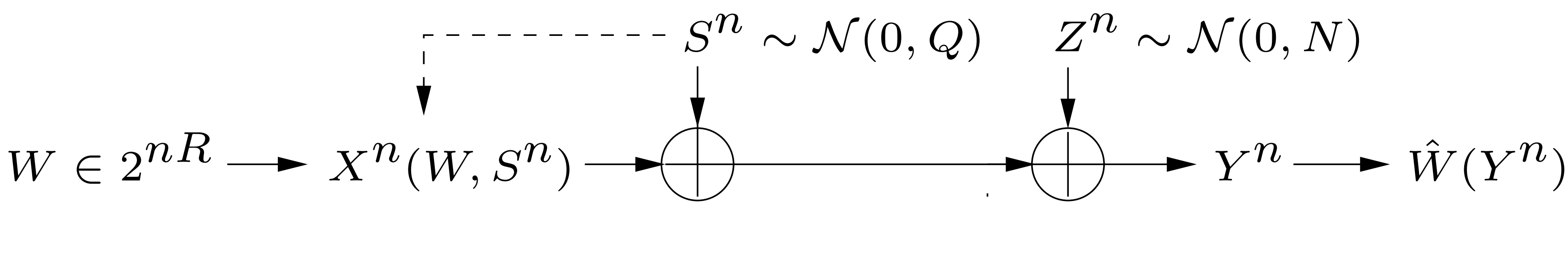
... with Transmitter Side Information



$$C = ?$$

Writing on Dirty Paper

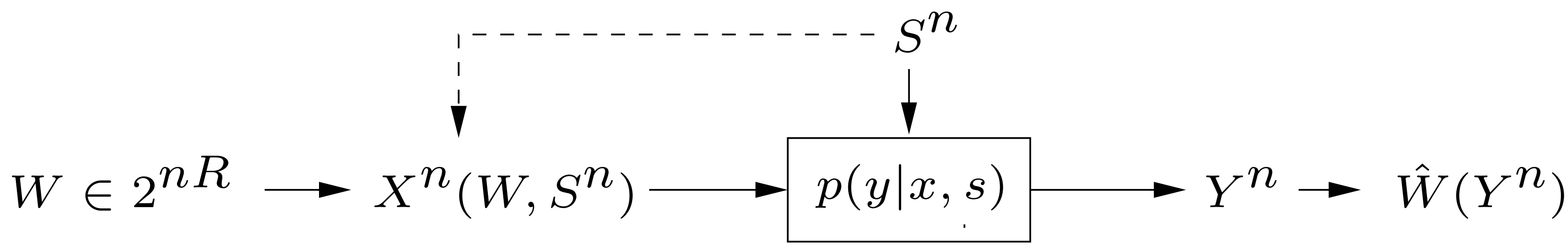
- A surprising result due to Costa ('83):



$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right)$$

- This inspired Caire and Shamai's work on 2x2 broadcast channel ('01).

Channel with Side Information

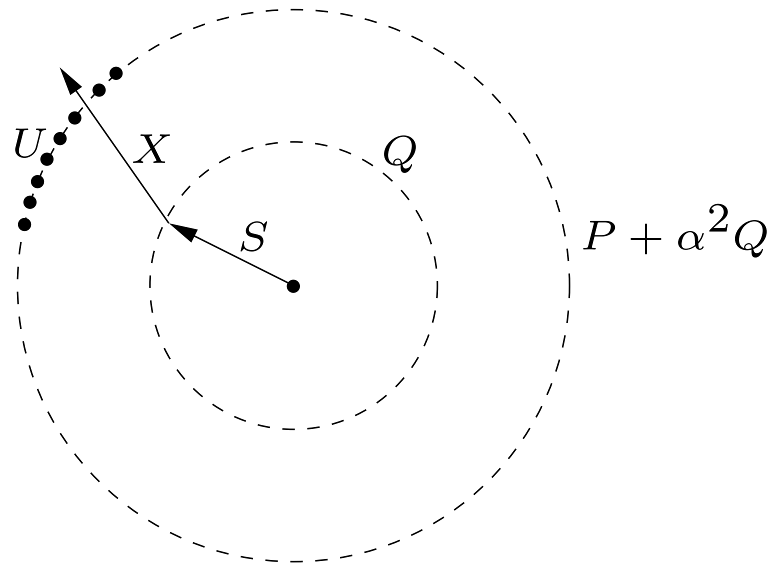


- Gel'fand and Pinsker ('80), Heegard and El Gamal ('83):

$$C = \max_{p(u,x|s)} \{I(U;Y) - I(U;S)\},$$

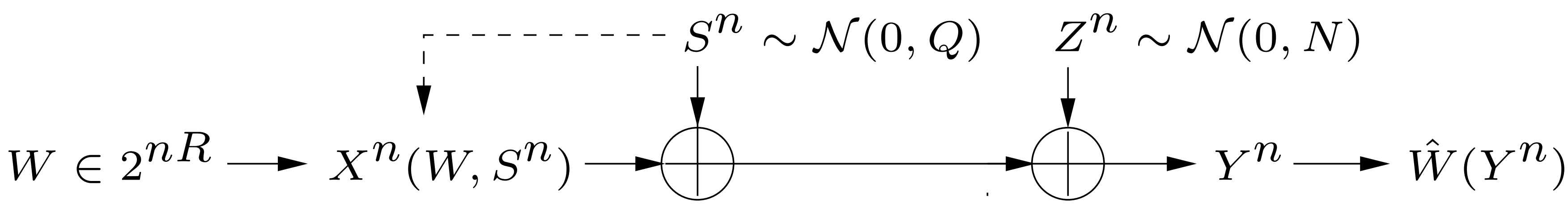
- Key: What is the appropriate auxiliary random variable U?

Random Binning and Joint Typicality



- Randomly choose $u^n(i)$, $i \in 2^{nI(U;Y)}$. Binning using $B : 2^{nI} \longrightarrow 2^{nC}$.
- Encode: Given s^n and message W, find i such that $(u^n(i), s^n)$ is jointly typical, and $B(i) = W$. Send: $x^n = u^n(i) - \alpha s^n$.
- Decode: Find $(y^n, u^n(\hat{i}))$ jointly typical. Recover $\hat{W} = B(\hat{i})$.

Costa's Choice for U

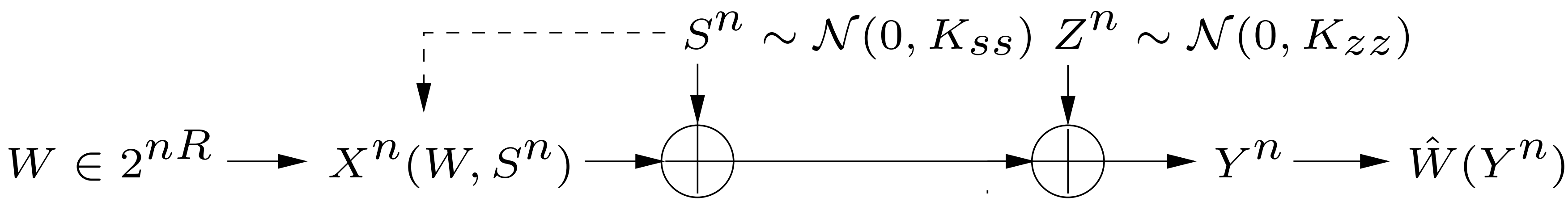


- For i.i.d. S and Z:
  - Let $U = X + \alpha S$, where $\alpha = P/(P+N)$.
  - Let X be independent of S.
  - This gives the optimal joint distribution on (S, X, U, Y, Z).

$$C = I(U;Y) - I(U;S) = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$$

.

Colored Gaussian Channel with Side Information



- For colored S and Z:
  - Let $U = X + FS$, where $F = K_{xx}(K_{xx} + K_{zz})^{-1}$.
  - Let X be independent of S.

$$C = I(U;Y) - I(U;S) = \frac{1}{2}\log\frac{|K_{xx} + K_{zz}|}{|K_{zz}|}$$

Wiener Filtering

- The optimal *non-causal* estimate of X given $X + Z$ is $\hat{X} = F(X + Z)$, where

$$F = K_{xx}(K_{xx} + K_{zz})^{-1}.$$

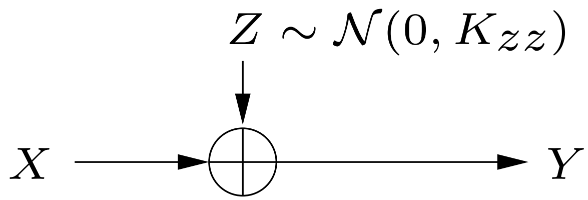
- The optimal auxiliary random variable for channel with *non-causal* transmitter side information is $U = X + FS$, where

$$F = K_{xx}(K_{xx} + K_{zz})^{-1}.$$

- Curiously, the two filters are the same.

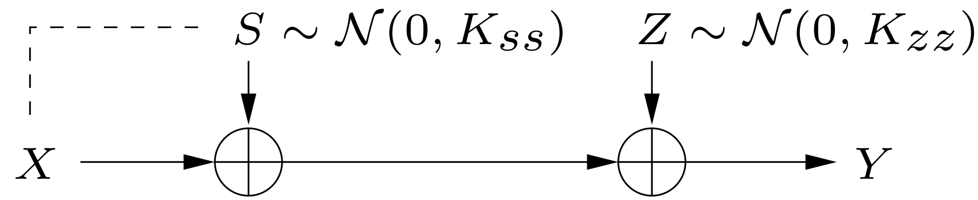
Writing on Colored Paper

Gaussian Channel



$$C = \frac{1}{2} \log \frac{|K_{xx} + K_{zz}|}{|K_{zz}|}$$

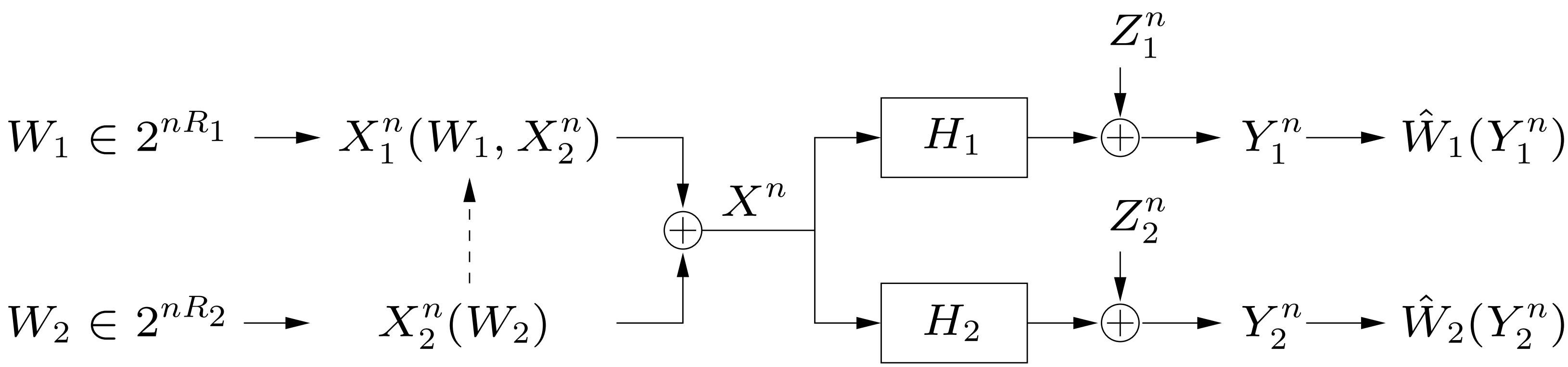
... with Transmitter Side Information



$$C = \frac{1}{2} \log \frac{|K_{xx} + K_{zz}|}{|K_{zz}|}$$

- Capacities are the same if S is known *non-causally* at the transmitter.
  - Several other proofs have been found by Cohen and Lapidoth ('01), and Zamir, Shamai and Erez ('01) under different assumptions

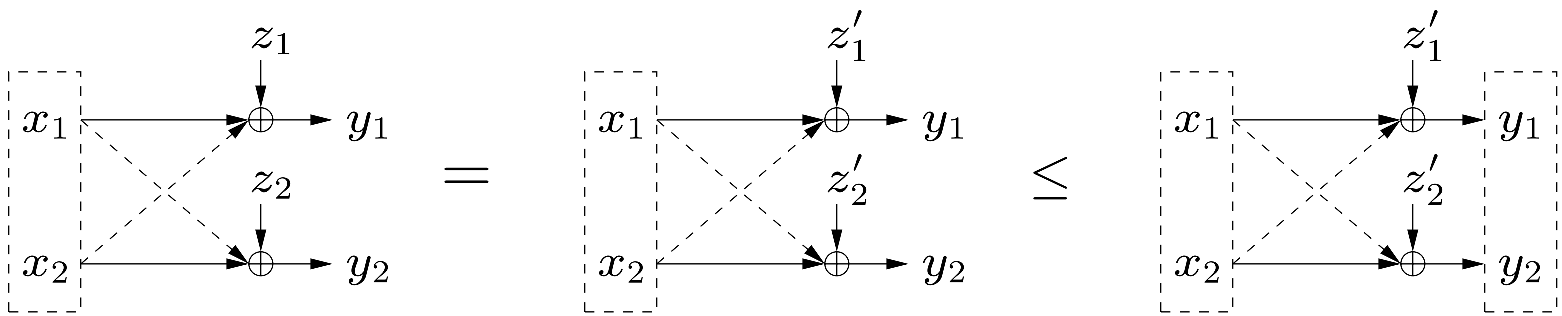
New Achievable Region



$$
\begin{aligned}
R_1 &= I(\mathbf{X_1}; \mathbf{Y_1} | \mathbf{X_2}) = \frac{1}{2} \log \frac{|H_1 K_1 H_1^T + K_{z_1 z_1}|}{|K_{z_1 z_1}|} \\
R_2 &= I(\mathbf{X_2}; \mathbf{Y_2}) = \frac{1}{2} \log \frac{|H_2 K_2 H_2^T + H_2 K_1 H_2^T + K_{z_2 z_2}|}{|H_2 K_1 H_2^T + K_{z_2 z_2}|}
\end{aligned}
$$

Converse

- Broadcast capacity does not depend on noise correlation: Sato ('78).



$$\text{if} \begin{cases} p(z_1) = p(z_1') \\ p(z_2) = p(z_2') \end{cases}, \text{ not necessarily } p(z_1, z_2) = p(z_1', z_2').$$

- Thus, sum-capacity $C \leq \min_{K_{nn}} \max_{K_{xx}} I(\mathbf{X}; \mathbf{Y})$.

Strategy for Proving Achievability

1. Find the worst-case noise correlation $\mathbf{z} \sim \mathcal{N}(0, K_{zz})$.

2. Design an optimal receiver for the vector channel with worst-case noise:

$$\mathbf{y} = H\mathbf{x} + \mathbf{z}$$

3. Precode $\mathbf{x}$ so that receiver coordination is not necessary.

- Tools:
  - Convex optimization
  - Generalized Decision-Feedback Equalization (GDFE)
    Cioffi, Forney ('95), Varanasi, Guess ('97)

Least Favorable Noise

- Fix Gaussian input K_{xx}:

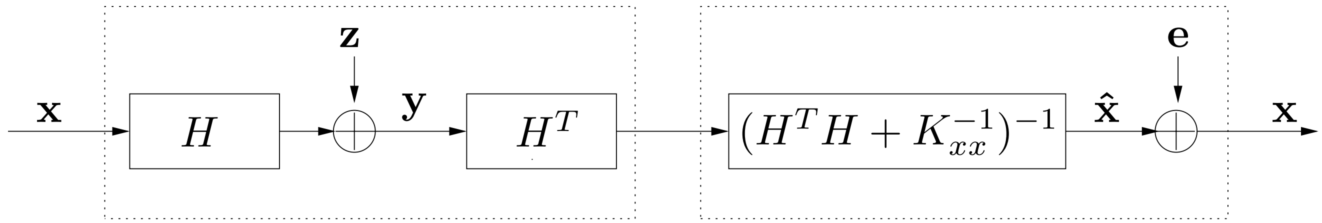
$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \log \frac{|HK_{xx}H^T + K_{zz}|}{|K_{zz}|} \\ \text{subject to} \quad & K_{zz} = \left[\begin{array}{cc} K_{z_1 z_1} & \star \\ \star & K_{z_2 z_2} \end{array} \right] \\ & K_{zz} \geq 0 \end{aligned}$$

- Minimizing a **convex** function over **convex** constraints.

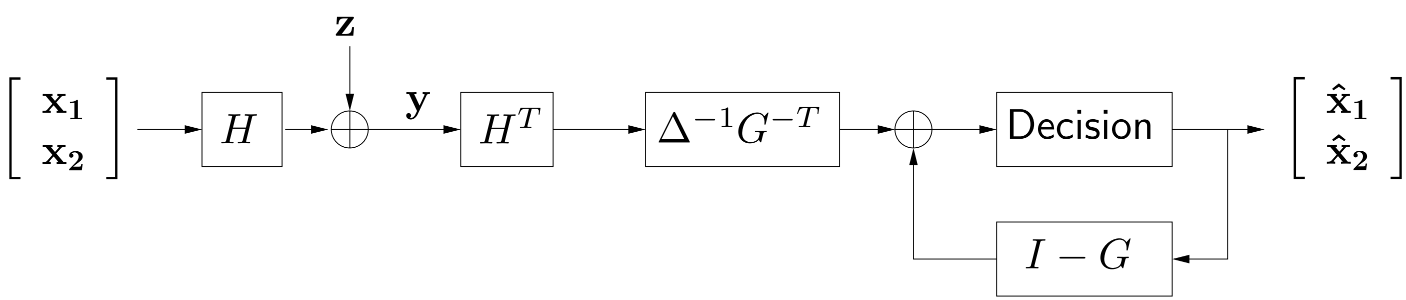
- Optimality condition: $K_{zz}^{-1} - (HK_{xx}H^T + K_{zz})^{-1} = \left[\begin{array}{cc} \Psi_1 & 0 \\ 0 & \Psi_2 \end{array} \right]$.

Generalized Decision Feedback Equalizer

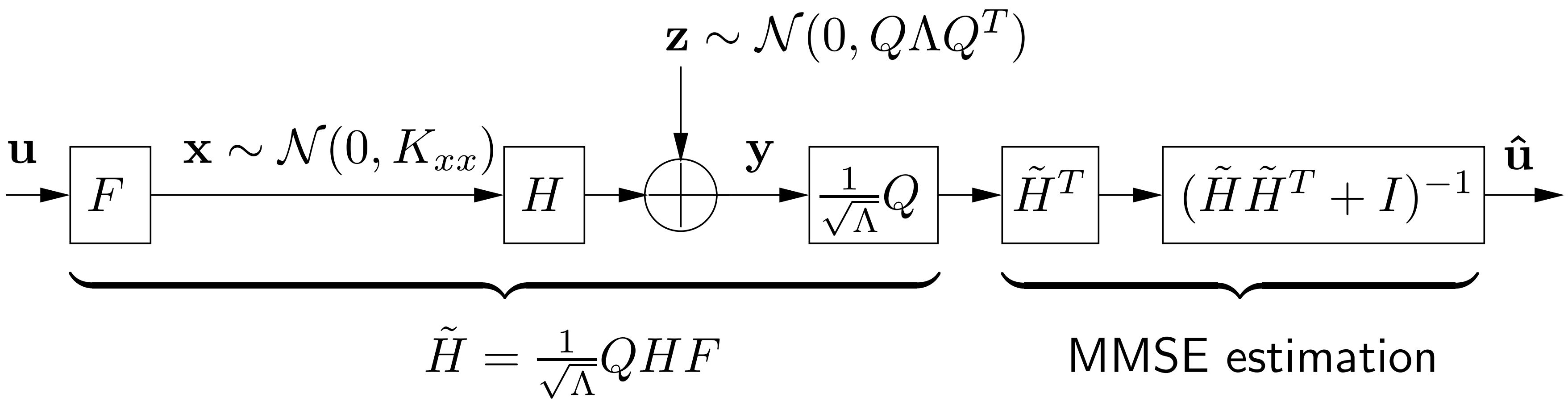
- Key idea: MMSE estimation is capacity-lossless



- Channel can be triangularized: $(H^T H + K_{xx}^{-1})^{-1} = G^{-1}\Delta^{-1}G^{-T}$.

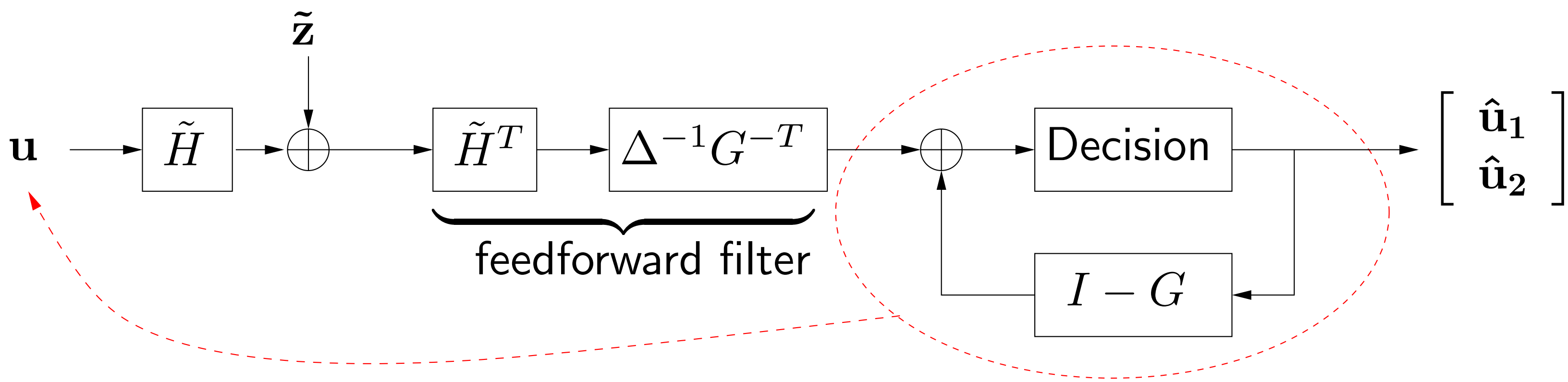


GDFE with Transmit Filter



- Set $\mathbf{z} \sim \mathcal{N}(0, K_{zz})$ to be the least favorable noise.
- Fix $\mathbf{x} \sim \mathcal{N}(0, K_{xx})$, and $\mathbf{u} \sim \mathcal{N}(0, I)$. Choose a transmit filter F.

GDFE Precoder



- Decision-feedback may be moved to the transmitter by precoding.
- Least Favorable Noise $\Longleftrightarrow$ Feedforward/whitening filter is diagonal!

$$C = \min_{K_{nn}} I(\mathbf{X}; \mathbf{Y})$$ *(i.e. with least favorable noise) is achievable.*

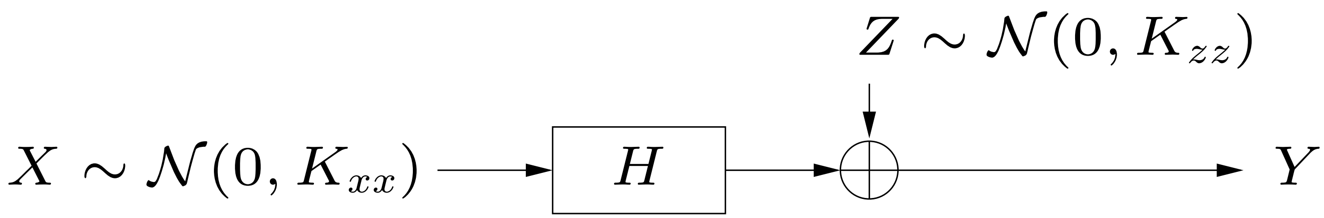
Gaussian Broadcast Channel Sum Capacity

- Achievability: $C \geq \max_{K_{xx}} \min_{K_{zz}} I(\mathbf{X}; \mathbf{Y})$.
- Converse (Sato): $C \leq \min_{K_{zz}} \max_{K_{xx}} I(\mathbf{X}; \mathbf{Y})$.
- (Diggavi, Cover '98): $\min_{K_{zz}} \max_{K_{xx}} I(\mathbf{X}; \mathbf{Y}) = \max_{K_{xx}} \min_{K_{zz}} I(\mathbf{X}; \mathbf{Y})$.

Theorem 1. *Gaussian vector broadcast channel sum capacity is:*

$$C = \max_{K_{xx}} \min_{K_{zz}} \frac{1}{2} \log \frac{|HK_{xx}H^T + K_{zz}|}{|K_{zz}|}$$

Gaussian Mutual Information Game

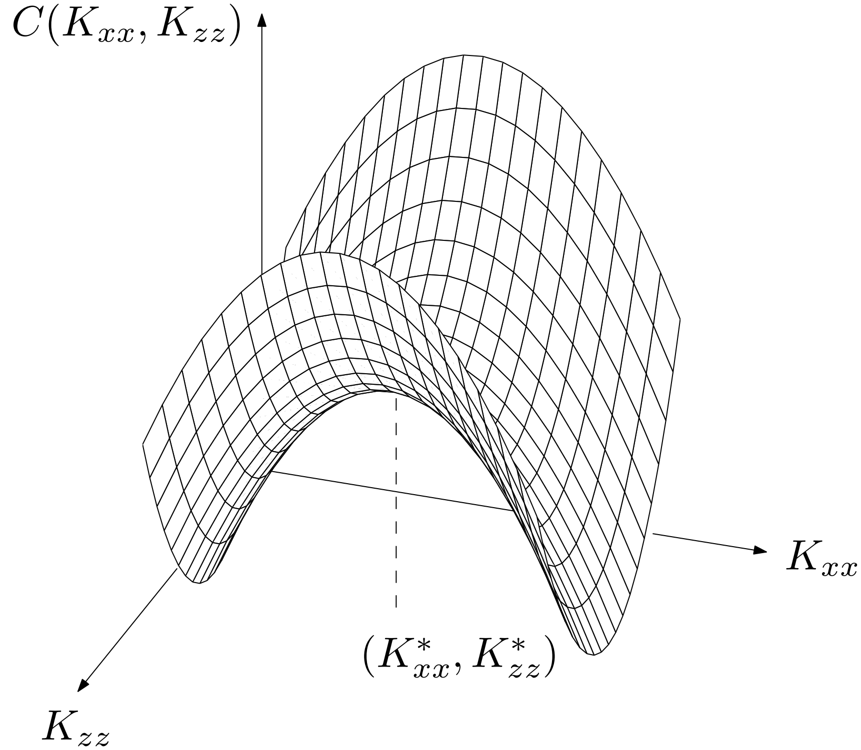


	Strategy	Objective
Signal Player	$\{K_{xx} : \text{trace}(K_{xx}) \leq P\}$	$\max I(\mathbf{X};\mathbf{Y})$
Fictitious Noise Player	$\left\{K_{zz} : K_{zz} = \begin{bmatrix} K_{z_1 z_1} & \star \\ \star & K_{z_2 z_2} \end{bmatrix} \geq 0 \right\}$	$\min I(\mathbf{X};\mathbf{Y})$

Nash equilibrium exists.

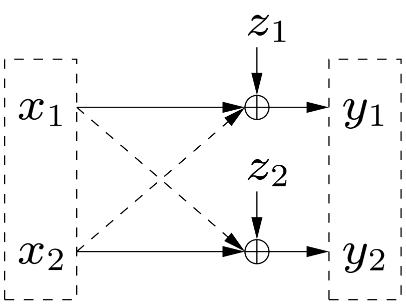
Saddle-Point is the Broadcast Capacity

- The optimum K_{xx}^* is a water-filling covariance against K_{zz}^*.
- The optimum K_{zz}^* is a least-favorable noise for K_{xx}^*.



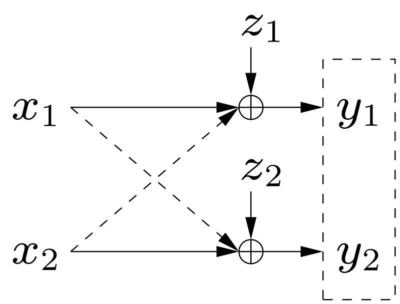
Broadcast Channel Sum Capacity = Nash Equilibrium

The Value of Cooperation



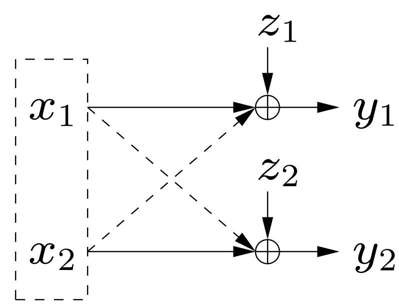
$$\max_{K_{xx}} I(\mathbf{X};\mathbf{X}+\mathbf{Z})$$

$$\text{s.t. } \operatorname{trace}(K_{xx}) \le P$$



$$\max_{K_{xx}} I(\mathbf{X};\mathbf{X}+\mathbf{Z})$$

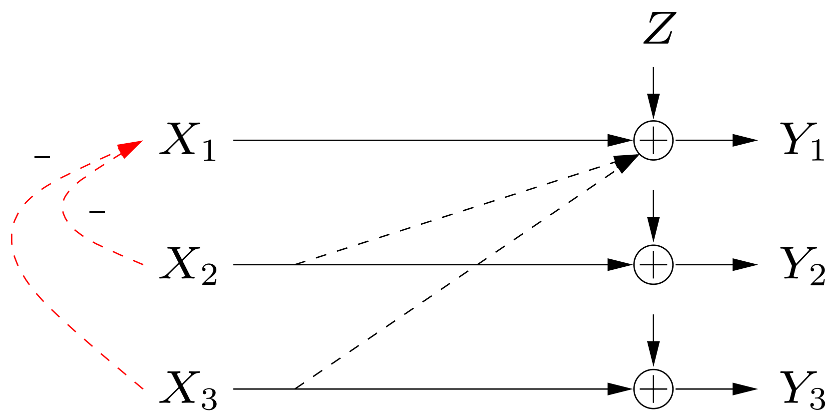
$$\text{s.t. } \begin{array}{c} K_{xx} = \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix} \\ \operatorname{trace}(K_i) \le P_i, \end{array}$$



$$\min_{K_{zz}} \max_{K_{xx}} I(\mathbf{X};\mathbf{X}+\mathbf{Z})$$

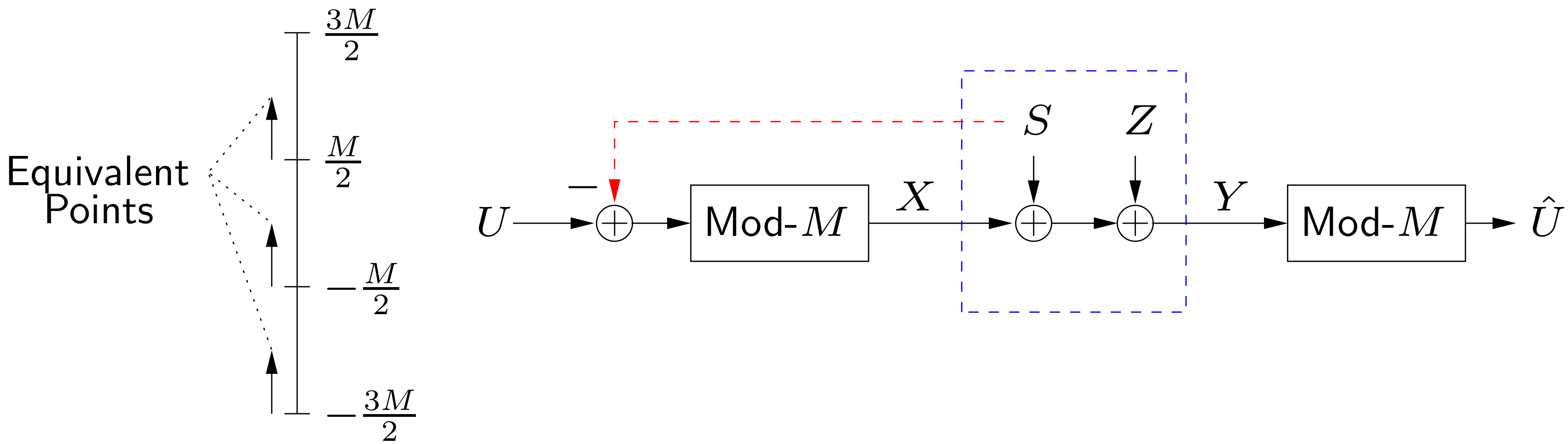
$$\text{s.t. } \begin{array}{c} K_{zz} = \begin{bmatrix} K_{z_1 z_1} & \star \\ \star & K_{z_2 z_2} \end{bmatrix} \\ \operatorname{trace}(K_{xx}) \le P \end{array}$$

Application: Vector Transmission in DSL



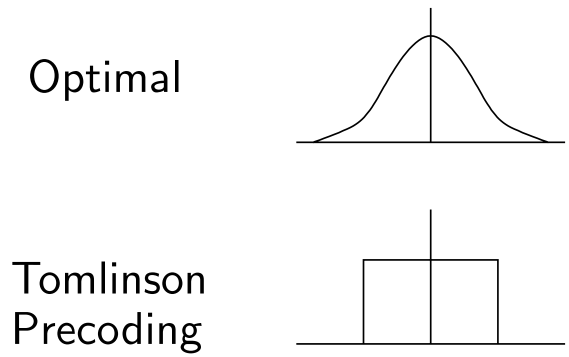
- If interference is known in advance, it can be pre-subtracted:
  - Send $X_1' = X_1 - X_2 - X_3$.
- Problem: energy enhancement $||X_1'||^2 = ||X_1||^2 + ||X_2||^2 + ||X_3||^2$.

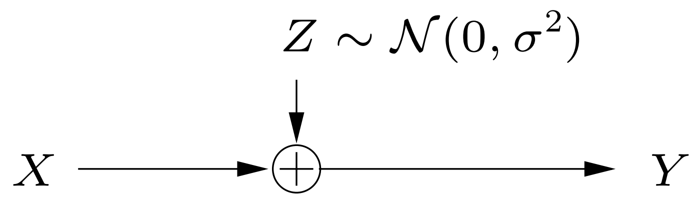
Reducing Energy Enhancement: Tomlinson Precoder



- Key idea: Use modulo operation to reduce energy enhancement
  - X is uniformly distributed in $[-\frac{M}{2}, \frac{M}{2}]$.
- Capacity loss due to shaping: 1.53dB. (Erez, Shamai, Zamir '00)

Shaping Loss

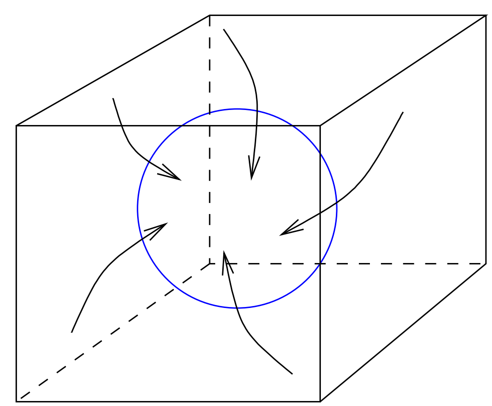




- Gaussian input distribution is optimum in a Gaussian channel.
  - But, Tomlinson-Harashima precoding produces uniform distribution.
- Need to use shaping techniques to recover shaping loss.

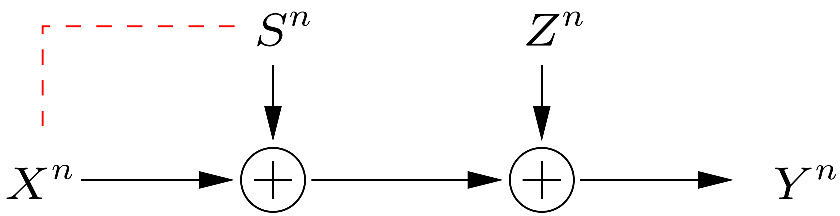
Shaping: Modulo a Sphere

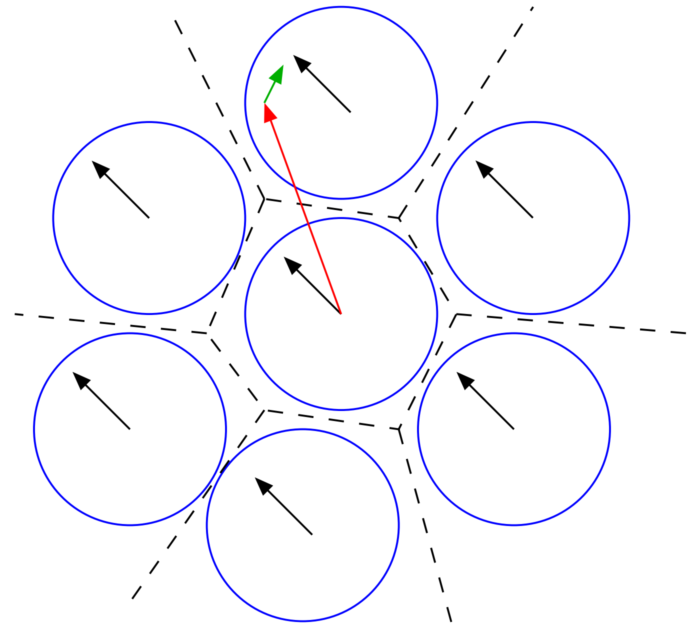
- High dimensional Gaussian = Uniform distribution in a sphere.
  - Uniform distribution can be produced by modulo operation



- Shaping can be done by expanding the constellation modulo a sphere.

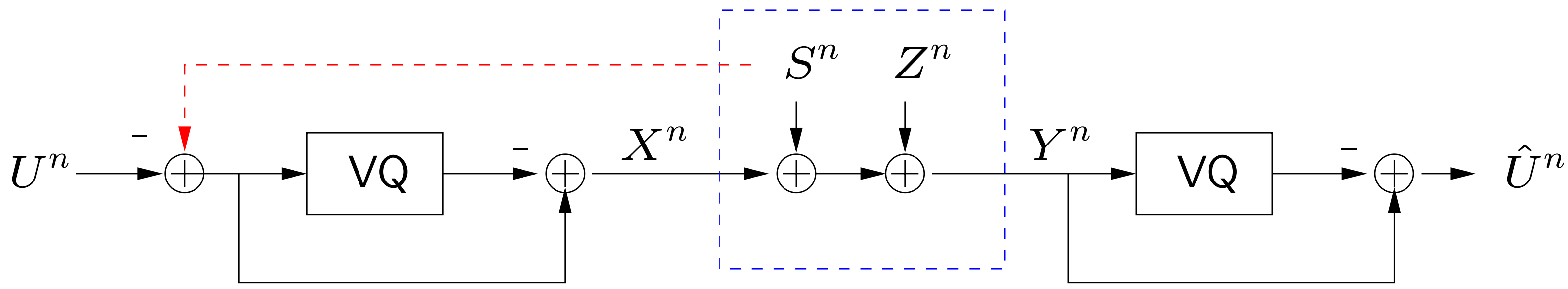
Precoding with Spherical Shape





- Precoding the entire S^n sequence.
  - X^n is uniformly distributed in the sphere $=$ Gaussian distribution.

Precoding via Vector Quantization

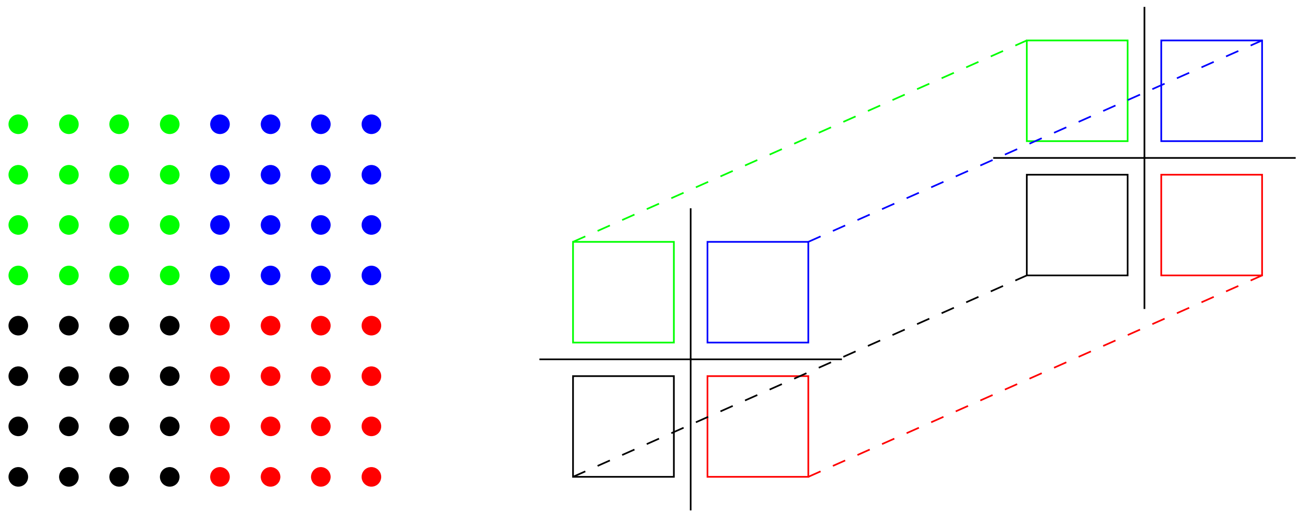


- Use the Voronoi region of a vector quantizer as the sphere.
  - Quantization is a generalization of Modulo-M operation.
  - Special case of lattice precoding by Zamir, Shamai, Erez ('01).

- At high SNR, shaping gain is completely recovered.

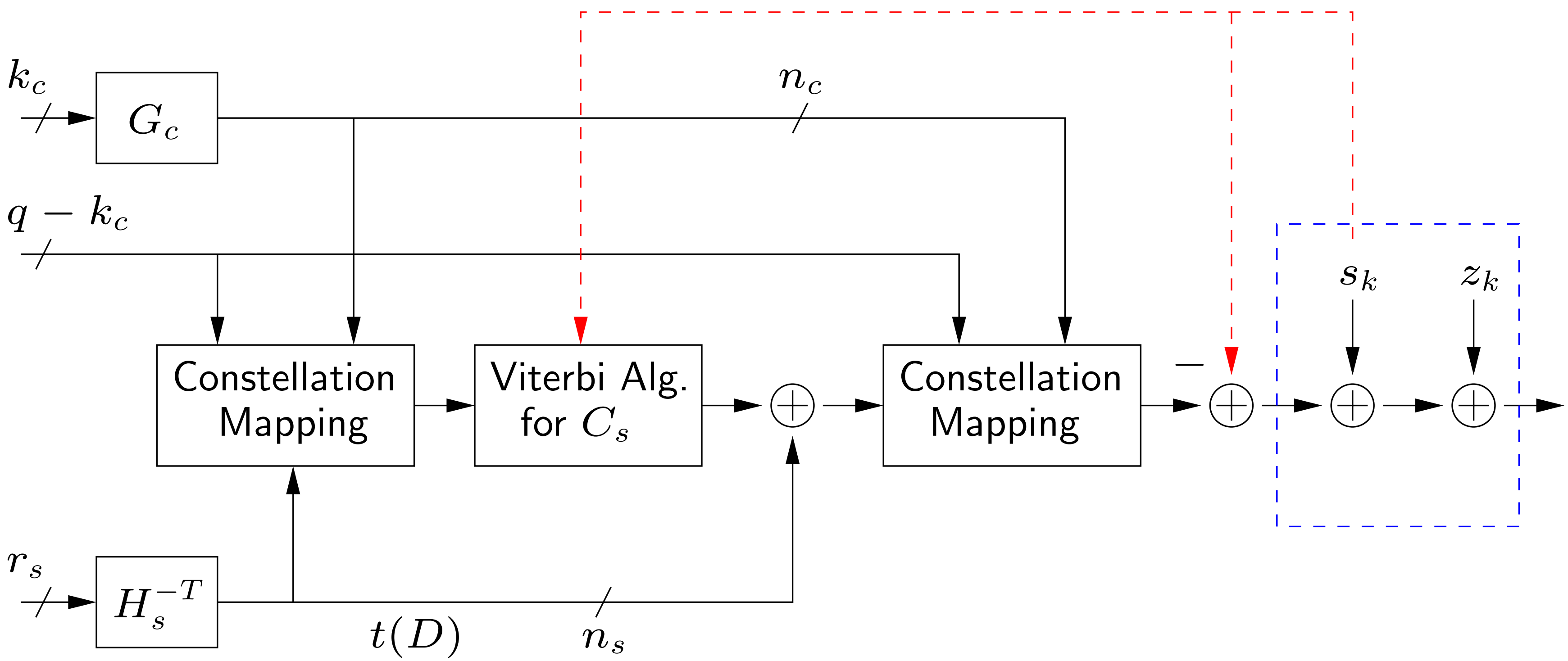
Voronoi Shaping using Nested Trellis Codes

- Inner trellis error correcting code + Outer trellis shaping code.



- Use the Voronoi region of shaping code to approximate the sphere.

Trellis Precoding



Trellis shaping (Forney, Eyuboglu '92): 1dB shaping gain with 4-state code.

Summary

- Sum capacity of a Gaussian vector broadcast channel is:

$$C = \max_{K_{xx}} \min_{K_{zz}} \frac{1}{2} \log \frac{|HK_{xx}H^T + K_{zz}|}{|K_{zz}|}$$

  - "Dirty-paper" coding is applicable to non-degraded channels.
  - Generalized decision-feedback equalizer is an optimal receiver.

- Practical precoding methods are proposed:

  - Tomlinson precoder gets within 1.53dB of capacity.
  - Trellis shaping codes can be used to approach capacity.