MTPT - Leiria, Portugal
September 4, 2006

Cross-Layer Issues in Wireless Networks

Vincent Poor
(poor@princeton.edu)

Wireless in the 21st Century

- High-Level Trends
  - *Dramatic growth rates in capacity demands:*
    - mobile multimedia
    - broadband
    - sensor nets
  - *Increase in shared (multiple-access, interference) channels:*
    - cellular (IS-95, 3G)
    - WiFi/Bluetooth/UWB (unlicensed spectrum)
    - ad hoc networks (large numbers of nodes, flexible transport)
  - *Opportunistic/resource-controlled/cross-layer approaches:*
    - WiMax (IEEE802.15)
    - mobile broadband ('4G')
- Basic Resources
  - *Bandwidth* - tightly constrained ✗
  - *Transmit Power* - tightly constrained ✗
  - *SP Power* - growing exponentially ✓

SP in Wireless Networks

- Advanced node-level processing affords:
  - Mitigation of PHY impairments: dispersion, interference, etc.
  - Exploitation of PHY diversity: spatial, temporal & spectral
  - Compression/collaboration to optimize PHY: batteries & bandwidth
- This affects the overall performance of the network:
  - Spectral Efficiency: bits-per-cycle (users-per-dimension)
  - Energy Efficiency: bits-per-joule
  - Delay: transmission delay and queuing delay
  - Performance in Applications: media transmission, inference, etc.

Today's Talk: Three Topics
A Sampling of Ideas

- Energy Efficiency in Multiple-Access Networks ✓
- Diversity-Multiplexing Tradeoffs in MIMO Systems
- Distributed Inference in Wireless Sensor Networks

ENERGY EFFICIENCY IN MULTIPLE-ACCESS NETWORKS

Outline

- *Multiuser Detection (MUD) - Briefly*
- *The Multiuser Power Control Game*
- *A Unified Power Control Algorithm*
- *Power Control in Multicarrier CDMA*
- *Energy Efficiency and Delay QoS*

Motivation

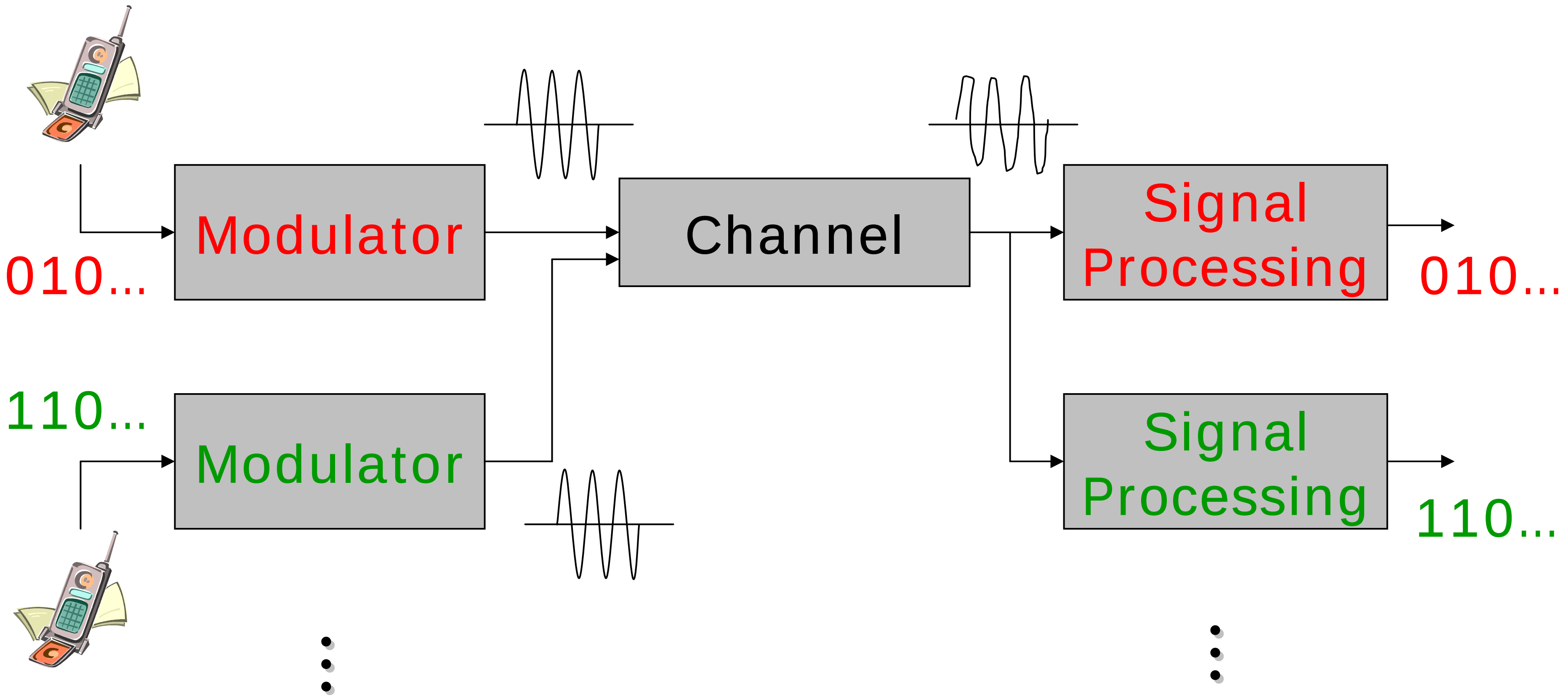
- PHY choices (e.g., modulation, detection scheme, # of antennas, etc.) can affect the energy efficiency of wireless networks.
- This issue can be examined by considering equilibria in a game theoretic framework in which terminals seek to maximize their energy efficiencies while competing for resources.
- First, we digress ...

MULTIUSER DETECTION - BRIEFLY

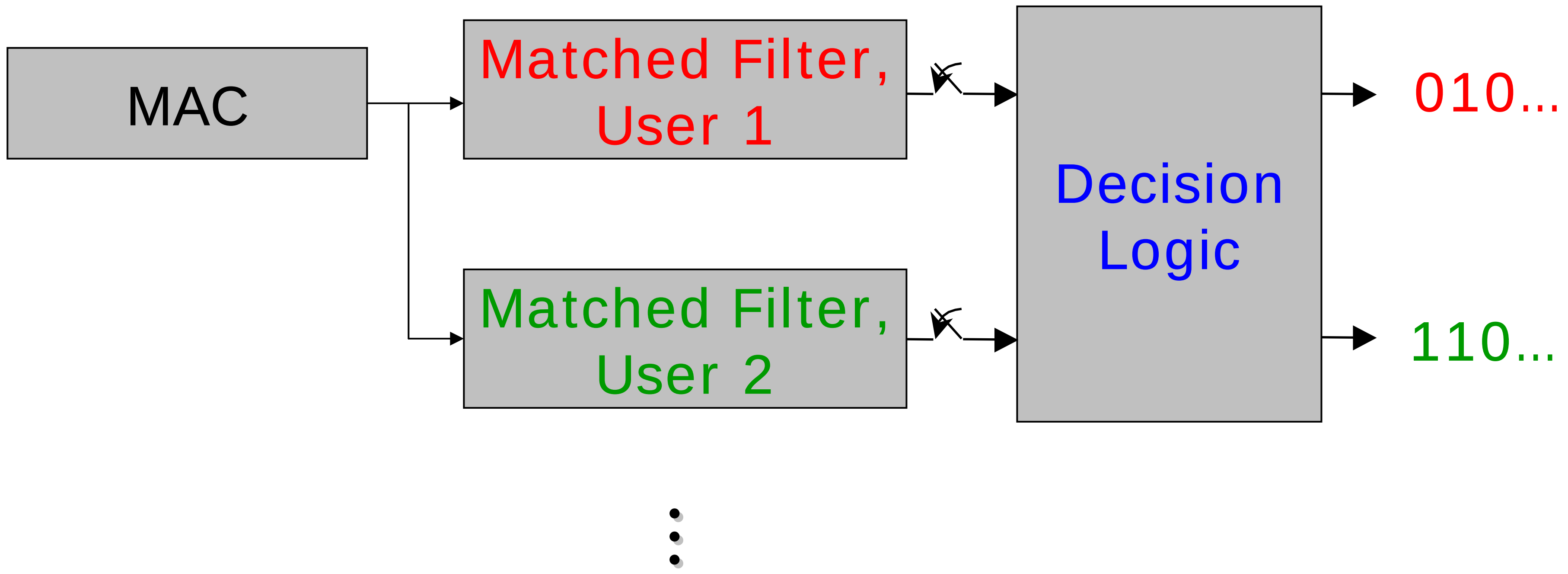
What is MUD Anyway?

- Multiuser detection (MUD) refers to data detection in a non-orthogonal multiplex (e.g., CDMA, TDMA with channel imperfections, etc).
- MUD can potentially increase the capacity (e.g., bits-per-chip) of interference-limited systems significantly.
- MUD comes in various flavors:
  - Optimal
  - Linear
  - Iterative
  - Adaptive

Multiple-Access (MA) Channel



Multiuser Detection (MUD)



- Optimal (maximum likelihood, MAP)
- Linear (zero-forcing, MMSE)
- Iterative (PIC, SIC, linear, nonlinear, EM, turbo)
- Adaptive (LMS, RLS, subspace)

Linear Model

K users, each transmitting a frame of *B* channel symbols, yields a linear model for the decision logic input:

$$\mathbf{y} = \mathbf{H}\,\mathbf{b} + N(0, \sigma^2\mathbf{H})$$

- $\mathbf{y}$ = *KB*-long sufficient statistic vector
- $\mathbf{b}$ = *KB*-long vector of channel symbols
- σ^2 = background noise level
- $\mathbf{H}$ = $KB \times KB$ matrix of cross-correlations
- $\mathbf{b}$ is a function of *RKB* information symbols (*R* = code rate)

Optimal MUD

Maximum Likelihood (ML):

$$\max\{f(\mathbf{y}|\mathbf{b}) \mid \mathbf{b} \in \{-1,+1\}^{KB}\}$$

Maximum *a posteriori* Probability (MAP):

$$\max\{ P(b_k(i)=b|\mathbf{y}) \mid b \in \{-1,+1\} \}$$

$$\mathbf{y} = \mathbf{H}\,\mathbf{b} + N(0, \sigma^2\mathbf{H})$$

$b_k(i)$= i^{th} symbol of user k

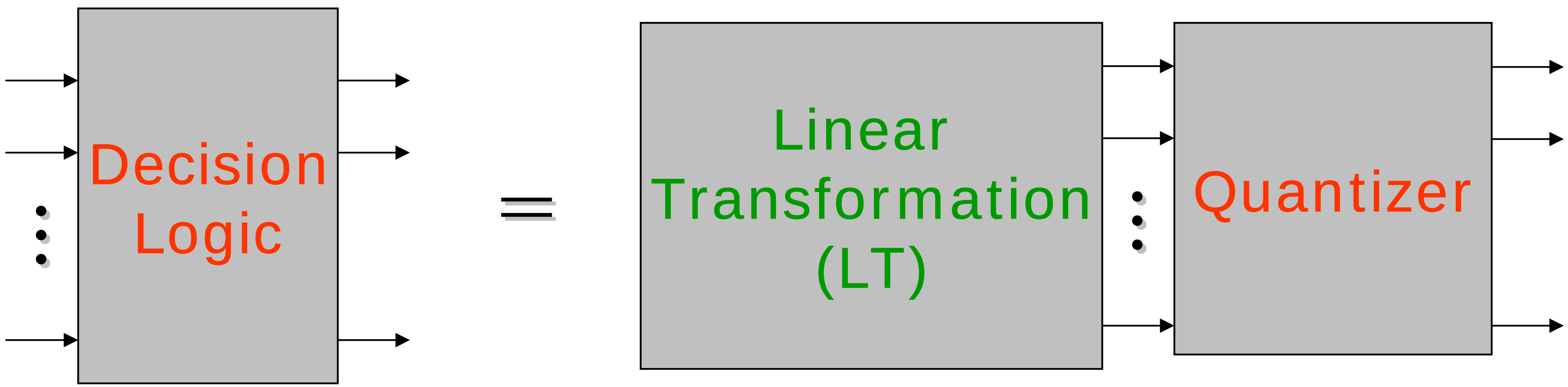
- Optimal MUD can achieve close to single-user performance.
- But, it requires $O(2^{K\Delta})$ complexity, where K is the number of users and Δ is the delay spread of the channel.
- This degree of complexity is prohibitive for most applications

Linear MUD

$$\mathbf{y} = \mathbf{H}\,\mathbf{b} + N(0, \sigma^2\mathbf{H})$$

- Basic Idea: Estimate $\mathbf{b}$ linearly, then quantize.
- Key Examples:
  - Matched filter/RAKE: $\hat{\mathbf{b}} = \operatorname{sgn}\{\mathbf{y}\}$
  - Decorrelator (zero-forcing): $\hat{\mathbf{b}} = \operatorname{sgn}\{\mathbf{H}^{-1}\mathbf{y}\}$
  - MMSE Detector: $\hat{\mathbf{b}} = \operatorname{sgn}\left\{(\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{y}\right\}$
- Complexity: $O((KB)^3)$, or lower.
- Adaptivity: Can be adapted using LMS, RLS & subspace.

Linear MUD: Illustrated



Key Examples:

- Matched Filter/RAKE Receiver: LT = identity
- Decorrelator: LT = channel inverter (i.e., zero-forcing)
- MMSE Detector: LT = MMSE* estimate of the transmitted symbols

*MMSE = Minimum Mean-Square Error

Linear MUD: BER Performance

Question:

- How do these detectors' bit-error rates (BERs) compare with one another?

Partial Answers:

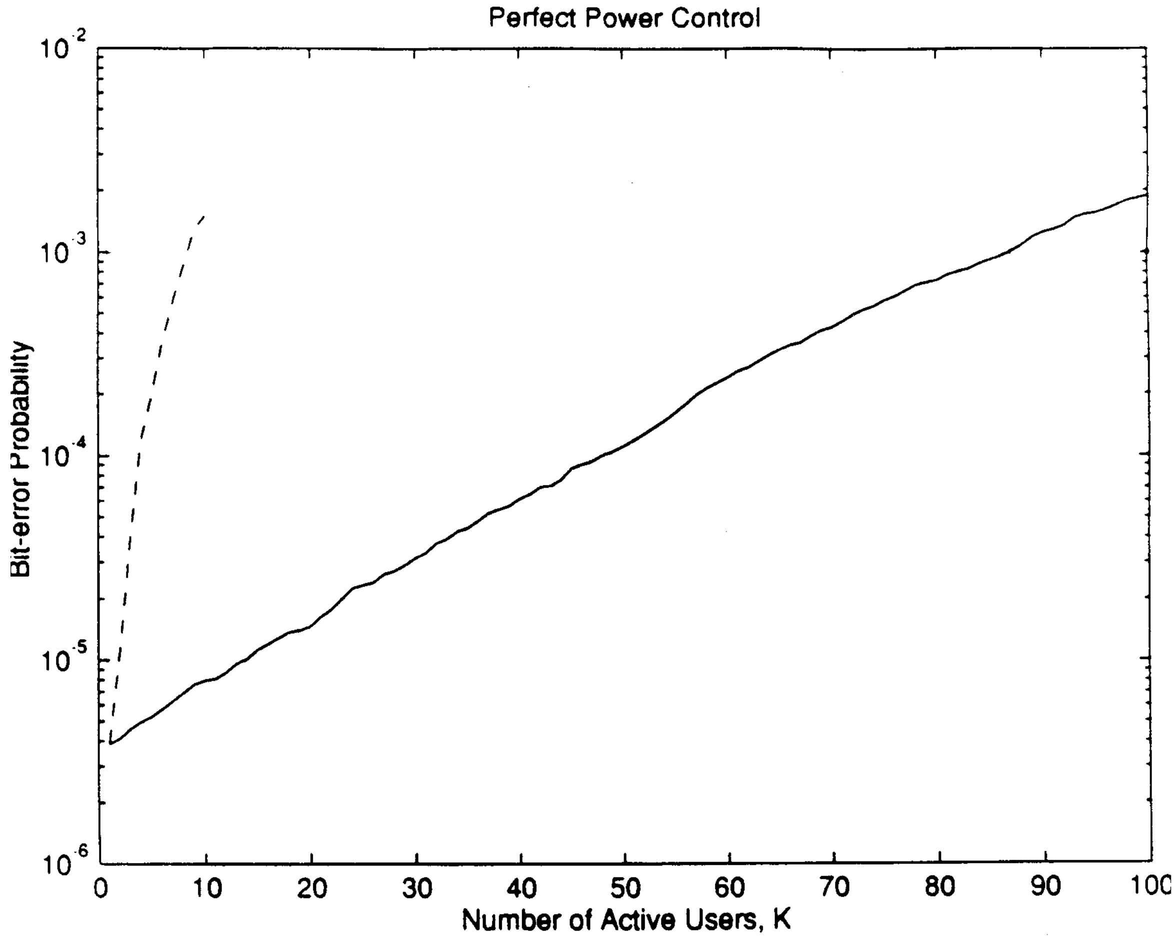
- Under various conditions, the decorrelator and MMSE detectors satisfy

$$P_e \sim Q\left(\sqrt{SINR}\right)$$

where *SINR* = Output Signal-to-Interference-pulse-Noise Ratio. (This is exact for the decorrelator.)

- The MMSE detector maximizes the *SINR* over all detectors, so we would expect that typically it has the lowest error probability of these two. (Not always true.)

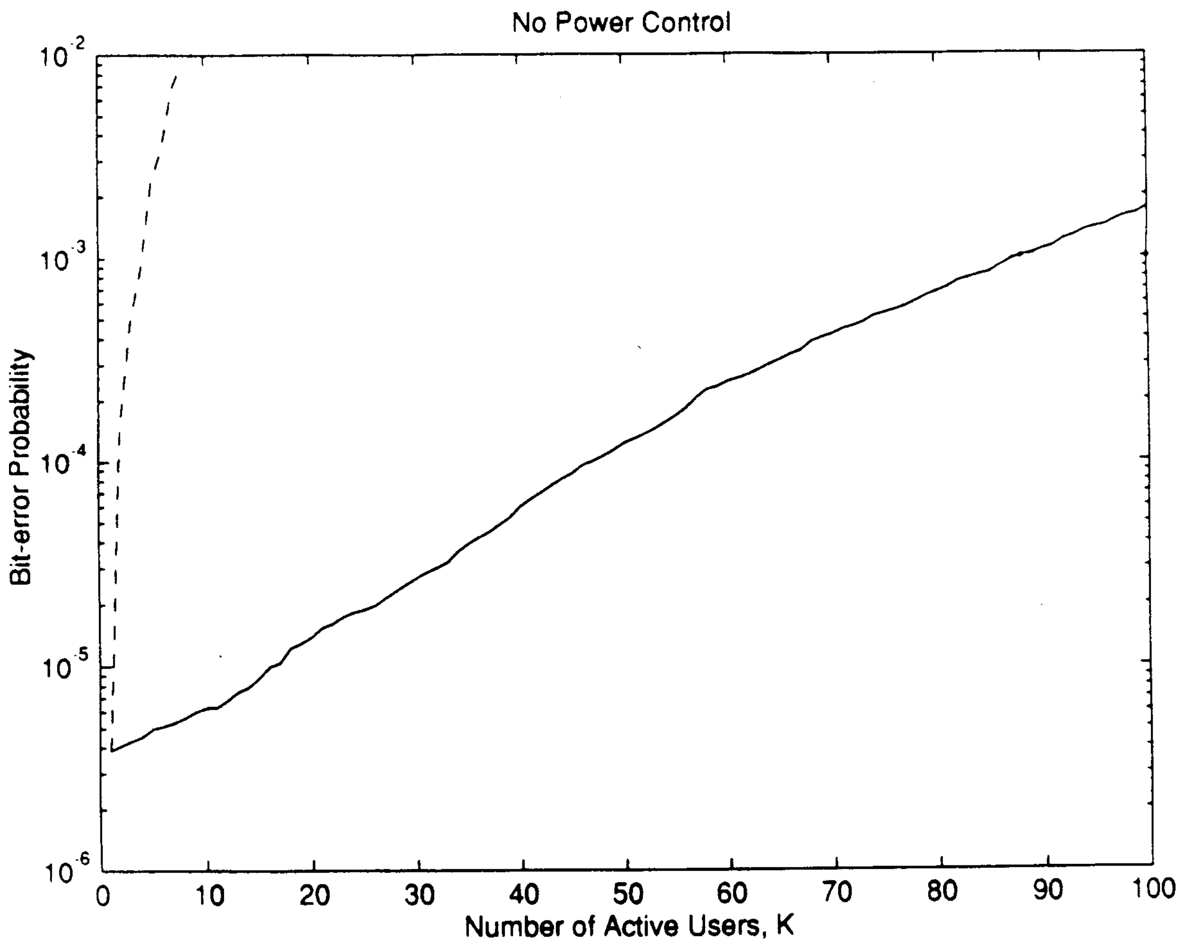
MMSE vs. Matched Filter



MMSE Detector (Solid Line); Matched Filter/RAKE (Dashed Line).

Perfect Power Control: SNR = 10 dB; DS/CDMA: 127-Length Signature Sequences.

MMSE vs. Matched Filter



MMSE Detector (Solid Line); Matched Filter/RAKE (Dashed Line).

No Power Control: SNR = 10 dB; DS/CDMA: 127-Length Signature Sequences.

Iterative MUD

- Basic Idea: Iteratively fit this model.

$$\mathbf{y} = \mathbf{H}\,\mathbf{b} + N(0, \sigma^2\mathbf{H})$$

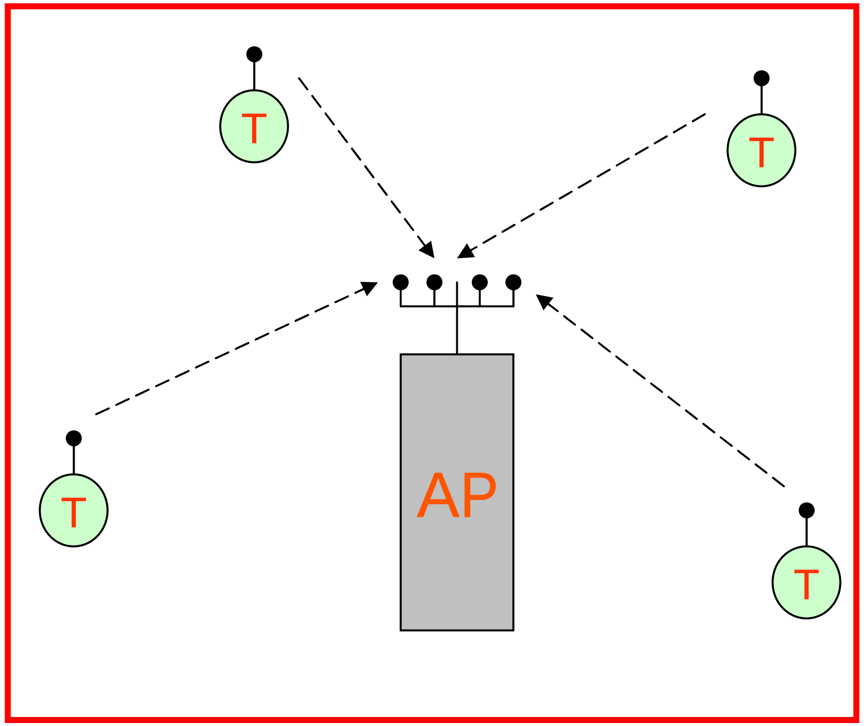
- Key Examples:
  - Linear Interference Cancellers (Gauss-Seidel, Jacobi, etc.)
  - Nonlinear Interference Cancellers (Successive IC; Parallel IC)
  - Expectation-Maximization (EM) Algorithm (Random $\mathbf{b}$)
  - Turbo (Constraints on $\mathbf{b}$ from space-time/error-control coding) ✓
- Complexity: $O(K\Delta n_{iterations})$ for ICs.

THE MULTIUSER POWER CONTROL GAME

[w. Meshkati, et al., *IEEE Trans. Commun.*, Nov. 2005.]

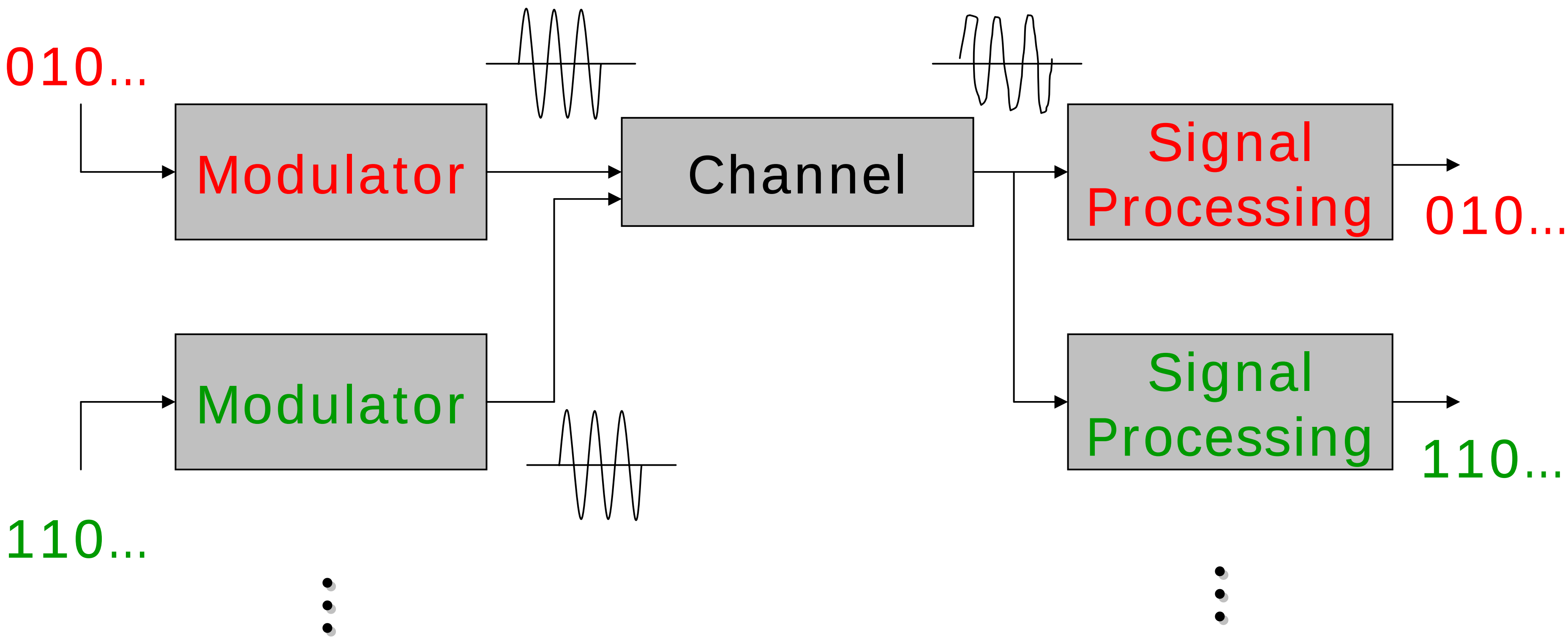
Competition in MA Networks

- Consider a set of terminals transmitting to an access point via a multiple-access channel.
- Terminals are like players in a game, competing for resources to transmit their data to the AP.
- The action of each terminal affects the others.



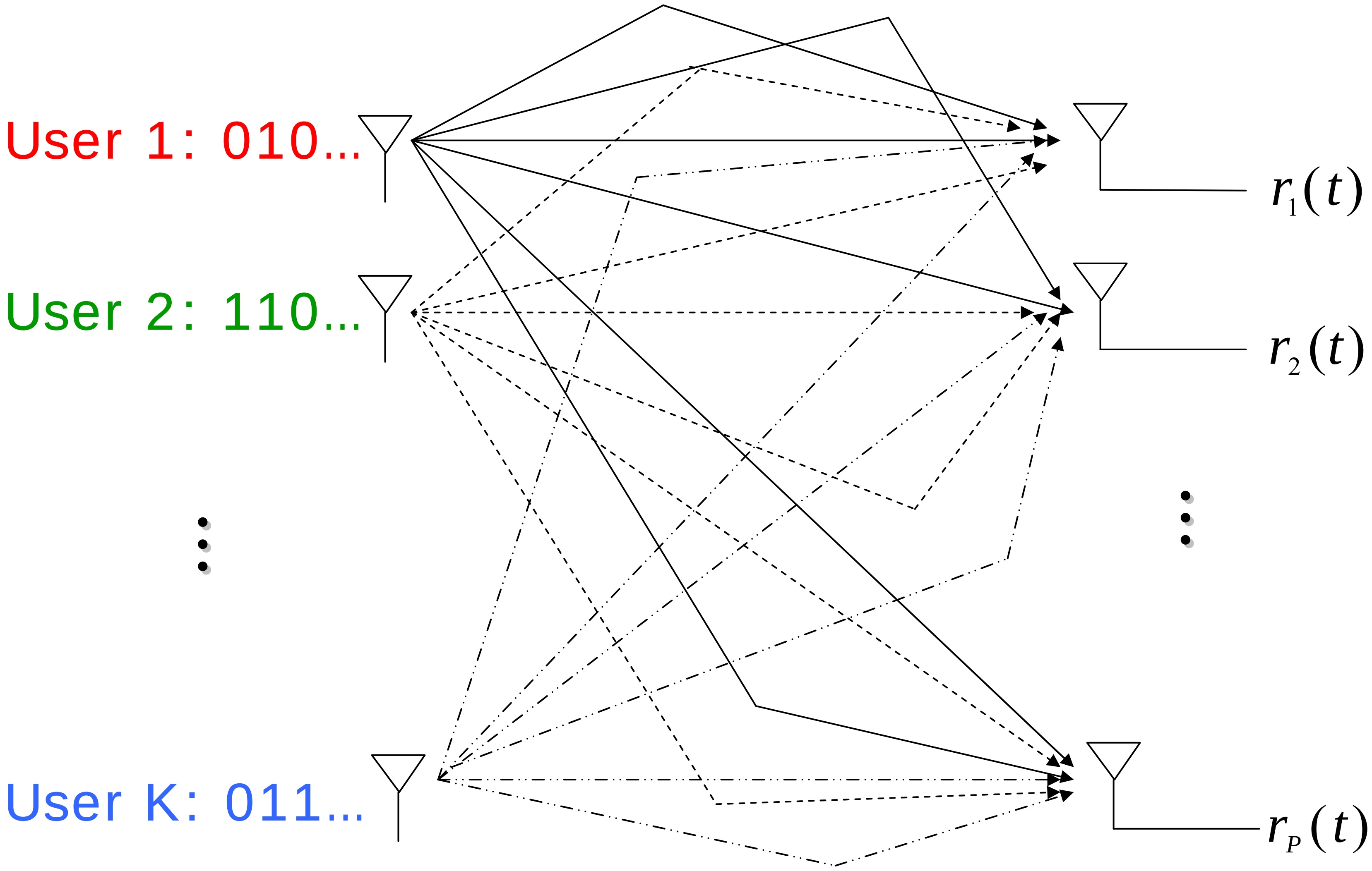
- Can model this as a non-cooperative game, with utility (measured in bits/joule) as a payoff.

Recall, Multiple-Access Channel



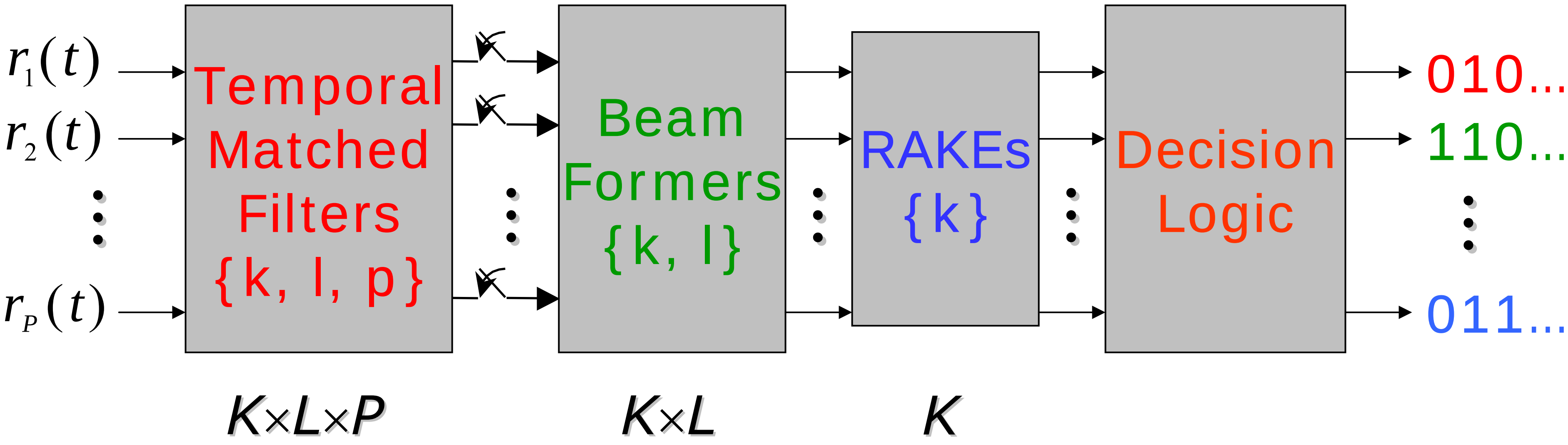
Multiuser Detection: receiver processing for shared-access systems

Multipath, Multi-antenna Case



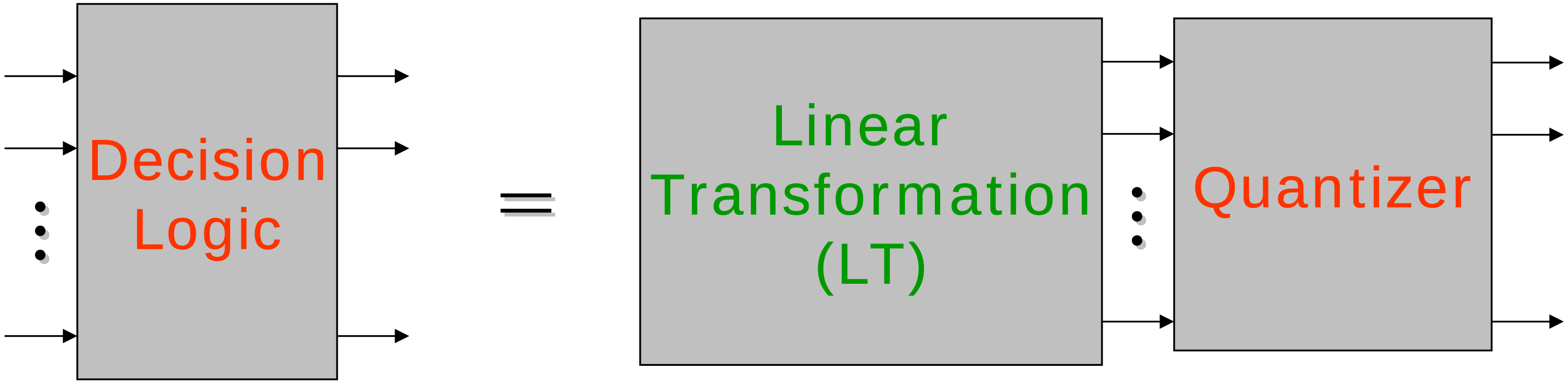
Space-Time MUD Structure

K Users; *P* Receive Antennas; *L* Paths/User/Antenna



- XISO (P=1) requires no beam-formers
- Flat fading (L=1) requires no RAKEs
- Decision logic: Optimal (ML, MAP), linear, iterative, adaptive.

Space-Time Linear MUD



<u>Key Examples</u>:

- <u>Matched Filter/RAKE Receiver</u>: LT = identity
- <u>Decorrelator</u>: LT = channel inverter (i.e., zero-forcing)
- <u>MMSE Detector</u>: LT = MMSE estimate of the transmitted symbols

Game Theoretic Framework

Game: $G = [\{1,\ldots, K\}, \{A_k\}, \{u_k\}]$

K: total number of users

A_k: set of strategies for user k

u_k: utility function for user k

$$u_k = utility = \frac{throughput}{transmit\ power} = \frac{T_k}{p_k} \left[\frac{bits}{Joule}\right]$$

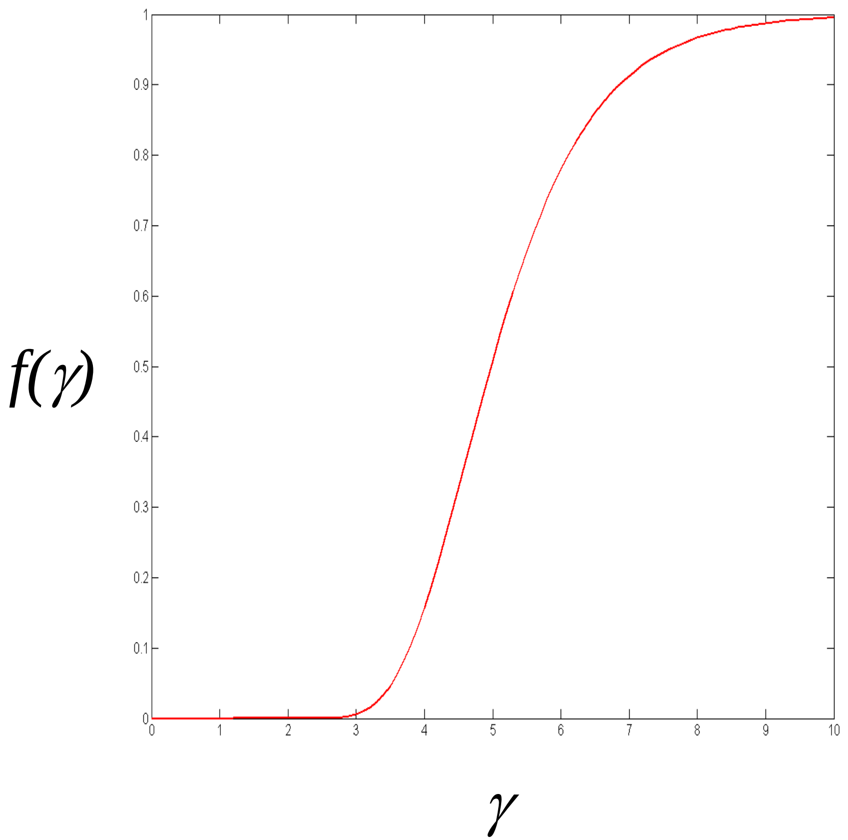
$T_k = R_k f(\gamma_k)$, where $f(\gamma_k)$ is the frame success rate, and γ_k is the received SIR of user k.

Efficiency Function

- $f(\cdot)$ is the efficiency function.
- It is assumed to be increasing and "S-shaped"
  - A useful choice is

$$f(\gamma) = (1 - e^{-\gamma})^M$$

(M = packet length)



An Uplink SIMO Game

- Game: Each user selects its transmit power and uplink linear MUD to maximize its own utility.
- Nash equilibrium (i.e., no user can unilaterally improve its utility) is reached when each user:
  - chooses the MMSE detector as its receiver, and
  - chooses a transmit power that achieves γ^*, the solution to:

$$\boxed{f(\gamma) = \gamma f'(\gamma)}$$

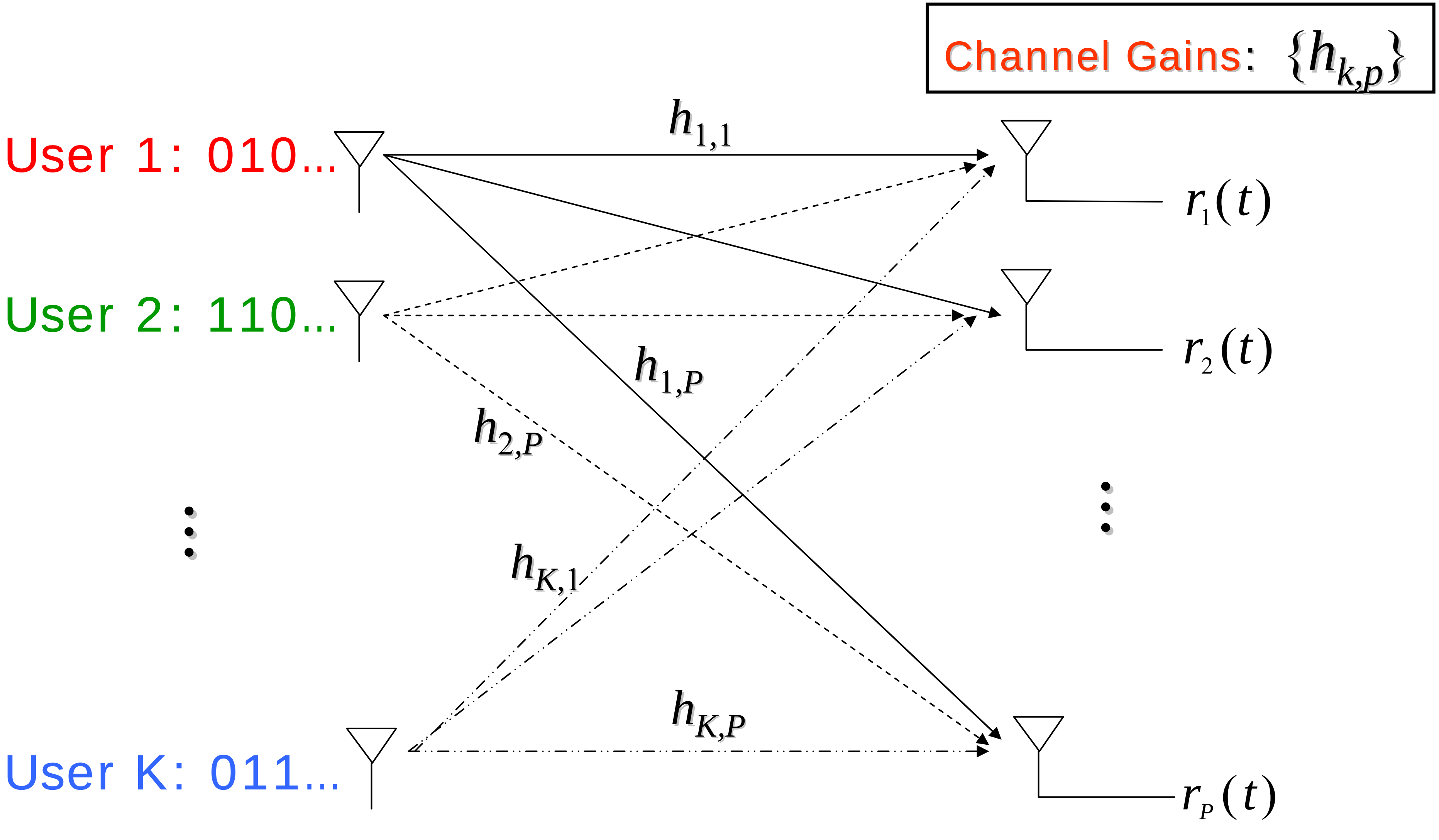
Outline of Proof

- Regularity conditions on f imply first-order stationary point of utility is a Nash equilibrium.
- This leads to SIR balancing as a necessary and sufficient condition for a Nash equilibrium for any linear MUD. (This relies on the linearity of γ_k in p_k.)
- Since MMSE maximizes SIR, this does it.

Remarks

- Nash equilibrium (NE) requires SIR balancing.
- The NE is unique, and can be reached iteratively as the unique fixed point of a nonlinear map.
- Effects on Energy Efficiency of Detector Choice:
  - If we were to fix the uplink detectors to be linear detectors other than MMSE detectors, the corresponding NE still requires SIR balancing, with the same target SIR.
  - Of interest are the classical matched filter and the (zero-forcing) decorrelator.

Flat SIMO Model



Large-System Analysis

- Consider R-CDMA with spreading gain N.
- As $K,N \to \infty$ with $K/N = \alpha$, NE utilities are:

$$u_k = \frac{R_k f(\gamma^*)}{\gamma^* \sigma^2} \bar{h}_k \Gamma \qquad \text{where}$$

$$\Gamma^{MF} = 1 - \bar{\alpha}\gamma^* \qquad \text{for } \bar{\alpha} < \frac{1}{\gamma^*}$$

$$\Gamma^{DE} = 1 - \alpha \qquad \text{for } \alpha < 1$$

$$\Gamma^{MMSE} = 1 - \bar{\alpha}\frac{\gamma^*}{1+\gamma^*} \qquad \text{for } \bar{\alpha} < 1 + \frac{1}{\gamma^*}$$

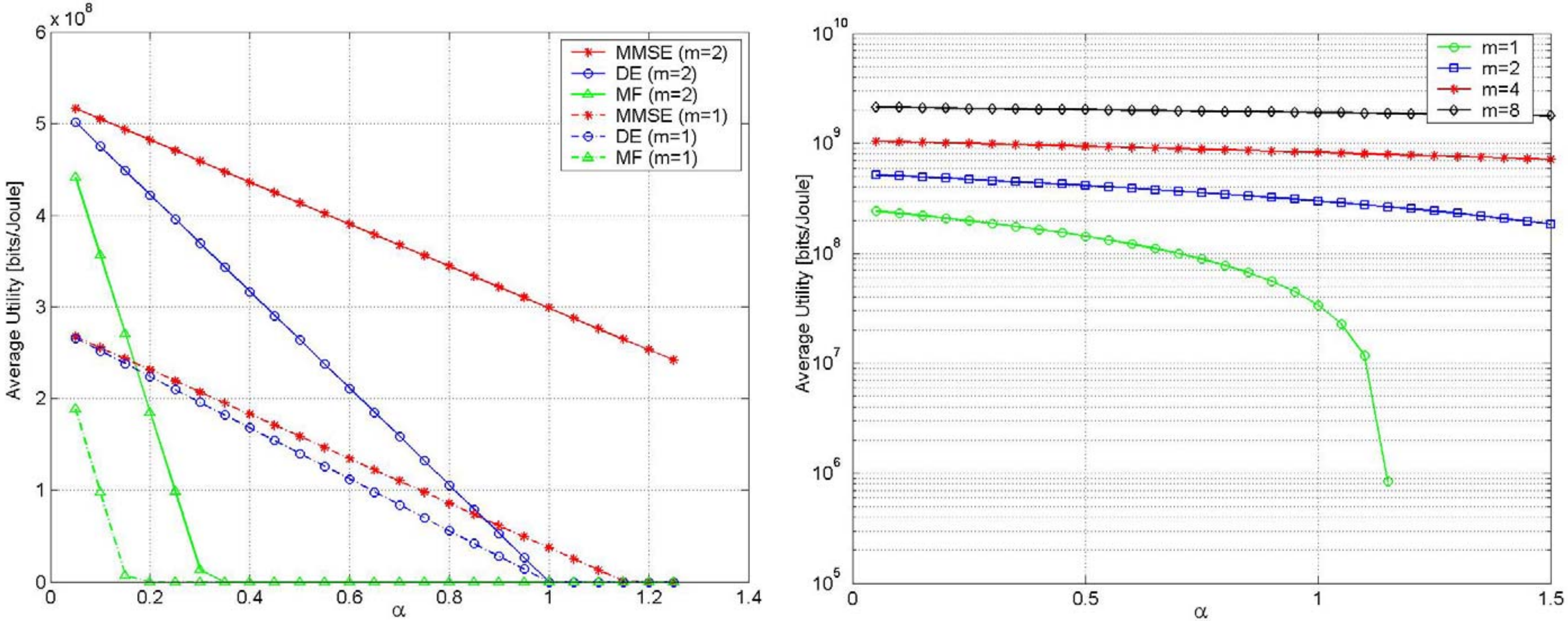
$$\text{with} \qquad \bar{h}_k = \sum_{p=1}^{P} h_{kp}^2 \qquad \text{and} \qquad \bar{\alpha} = \frac{\alpha}{P}$$

Two mechanisms:
- power pooling
- interference reduction

Example: Parameters

- Packet length: $M = 100$.
- Rate: 100 kbps
- Thermal noise level: 5×10^{-14} W
- Equilibrium SIR: $\gamma^* = 8.1$ dB
- Channel gains: Rayleigh

Example: Utility vs. Load

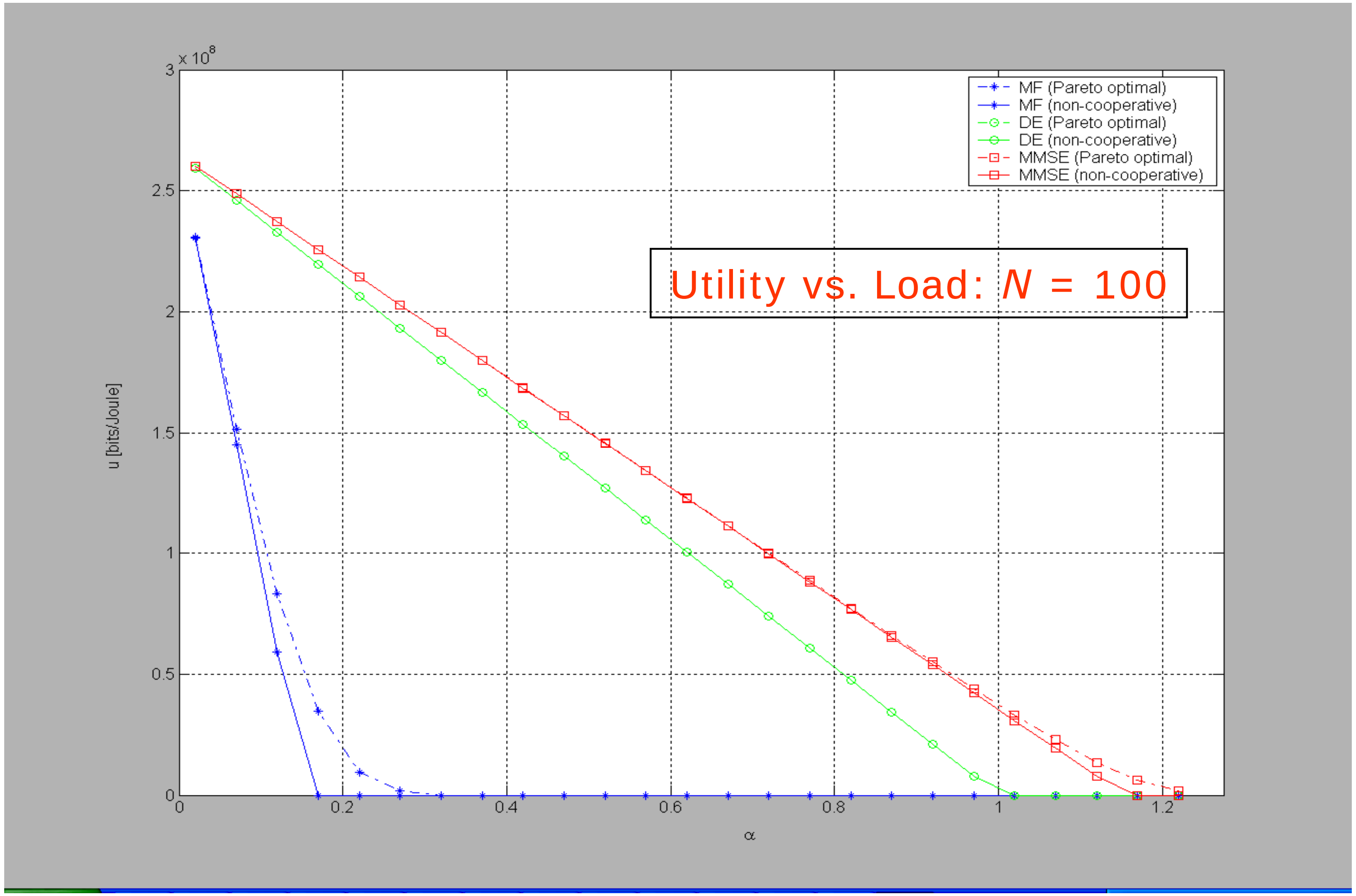


- Multiuser detectors achieve higher utility and can accommodate more users compared to the matched filter.
- Significant performance improvements are achieved when multiple antennas are used compared to single antenna case.

Social Optimality

- The Pareto (or socially) optimal solution, chooses the transmit power so that no user's utility can be improved without decreasing that of another.
- The Pareto solution is generally hard to find.
- The Nash equilibrium solution not generally Pareto optimal.
- But, it's close.

Example: Nash & Pareto Optima



A MIMO Game

- Game: Each user selects its transmit power, uplink linear detector, and distribution of power among transmit antennas to maximize its own utility.
- Conjecture: Nash equilibrium is reached when each user:
  - chooses the MMSE detector as its receiver,
  - transmits to achieve SIR γ^*, and
  - uses *spatial waterfilling* (i.e., transmits in the direction of the principal eigenvector of an effective channel matrix.)

UNIFIED POWER CONTROL

[w D. Guo, et al., *IEEE Trans. Wireless Commun.,* to appear]

Nonlinear MUD

- That SIR-balancing leads to a Nash equilibrium for a given detector follows from the following property:

$$\boxed{\partial \gamma_k / \partial p_k = \gamma_k / p_k}$$

- For linear MUD, this property always holds.
- What about nonlinear MUD (e.g., ML MUD)?

Large-System Analysis of Nonlinear MUD

- Consider (SISO) R-CDMA in the large-system limit.
- Asymptotically, many MUDs (linear detectors, ML MUD, MAP MUD, PIC, etc.) have the property:

$$\gamma_k = \eta_k SNR_{received} = \eta_k h_k p_k / \sigma^2$$

where η_k is the multiuser efficiency of MUD. ($h_k = h_{k,1}$)

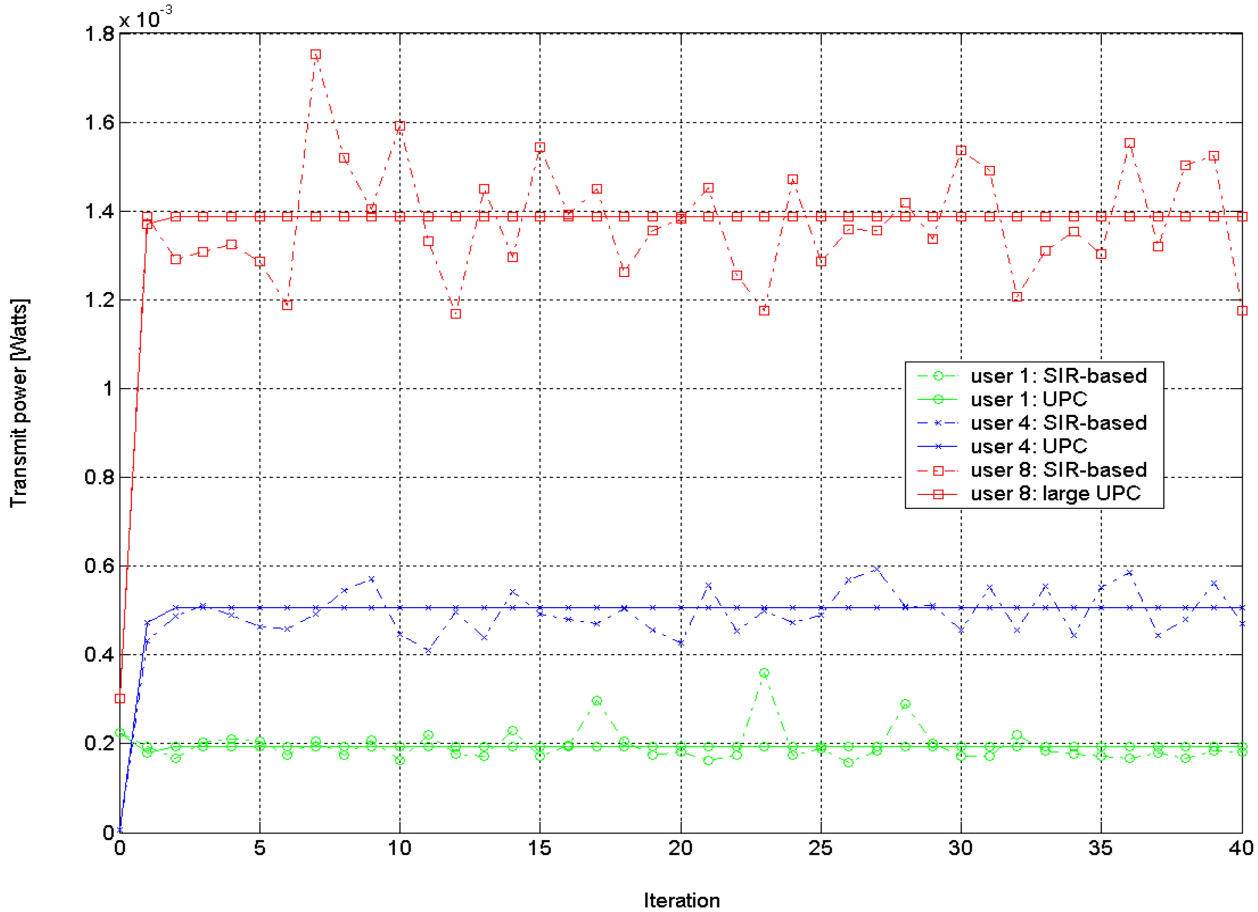
- So, $\partial\gamma_k / \partial p_k = \gamma_k / p_k$, holds asymp. for all such MUDs.

UPC Algorithm

- Conclude: In the large-system limit, SIR balancing leads to a Nash equilibrium for all such detectors, linear and nonlinear.
- For a fixed detector, the NE can be reached iteratively via the unified power control (UPC) algorithm:

$$\boxed{p_k(n+1) = \frac{\gamma^* \sigma^2}{h_k \eta_k(n)}}$$

UPC Iteration: MMSE

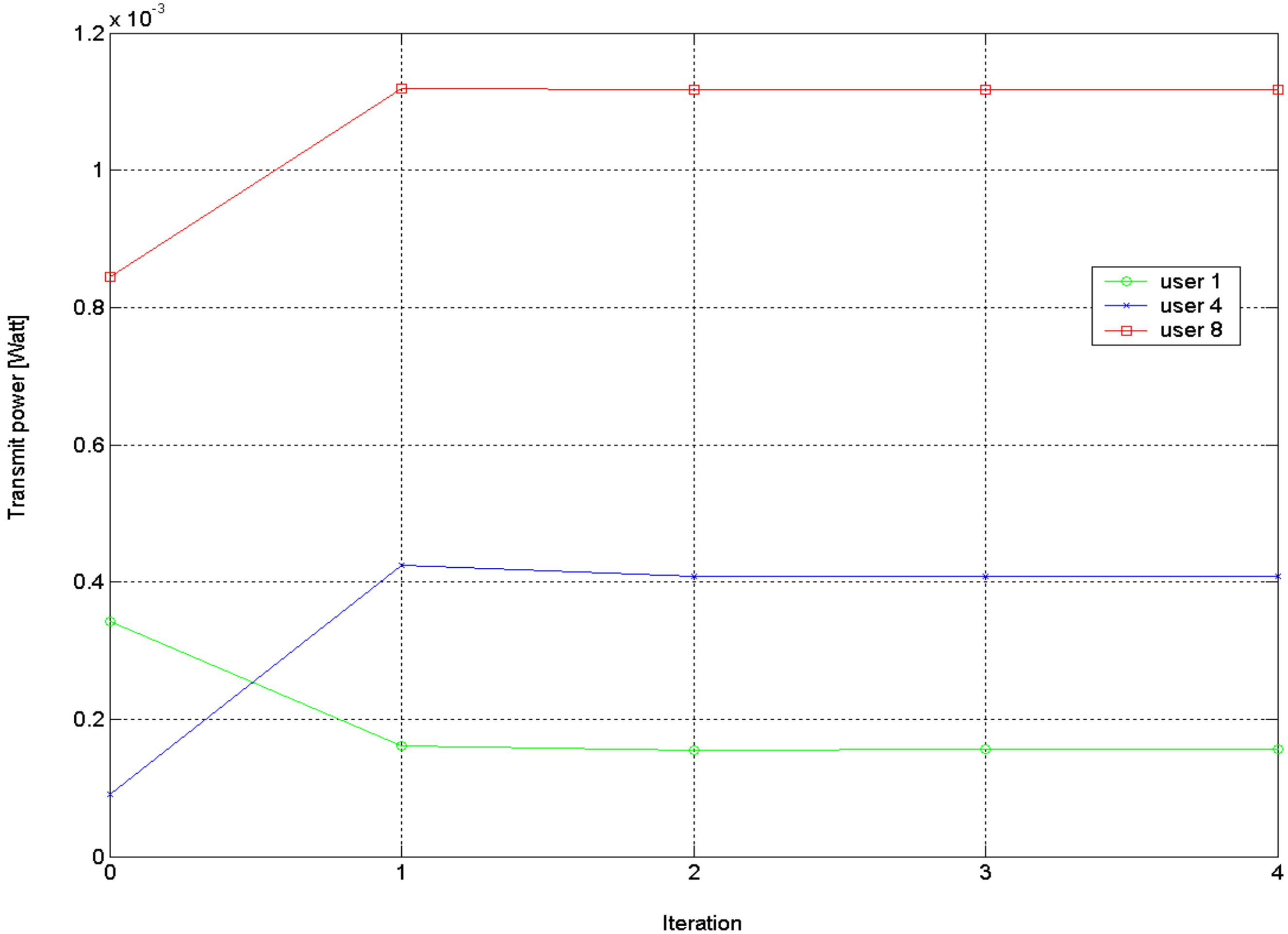


- N = 32
- K = 8
- P = 1
- Inverse-quartic path loss
- Rayleigh fading
- Distances differ

- UPC: $p_k(n+1) = \gamma^* \sigma^2 / h_k \eta_k(n)$
- SIR [Foschini, et al.]: $p_k(n+1) = \gamma^* p_k(n) / \gamma_k(n)$

Unified Power Control

UPC Iteration: ML



- Recall for MMSE: $p_1 = 0.2 \times 10^{-3}$; $p_4 = 0.5 \times 10^{-3}$; $p_8 = 1.4 \times 10^{-3}$

POWER CONTROL IN MULTICARRIER CDMA (BRIEFLY)

[w. M. Chiang, et al., *IEEE JSAC*, June 2006]

Multicarrier CDMA

- Now we have K users, D carriers, and processing gain N for each carrier.
- User k now chooses D powers, $p_k^{(1)}$, $p_k^{(2)}$, … , $p_k^{(D)}$, resulting in D throughputs, $T_k^{(1)}$, $T_k^{(2)}$, … , $T_k^{(D)}$.
- The resulting utility is

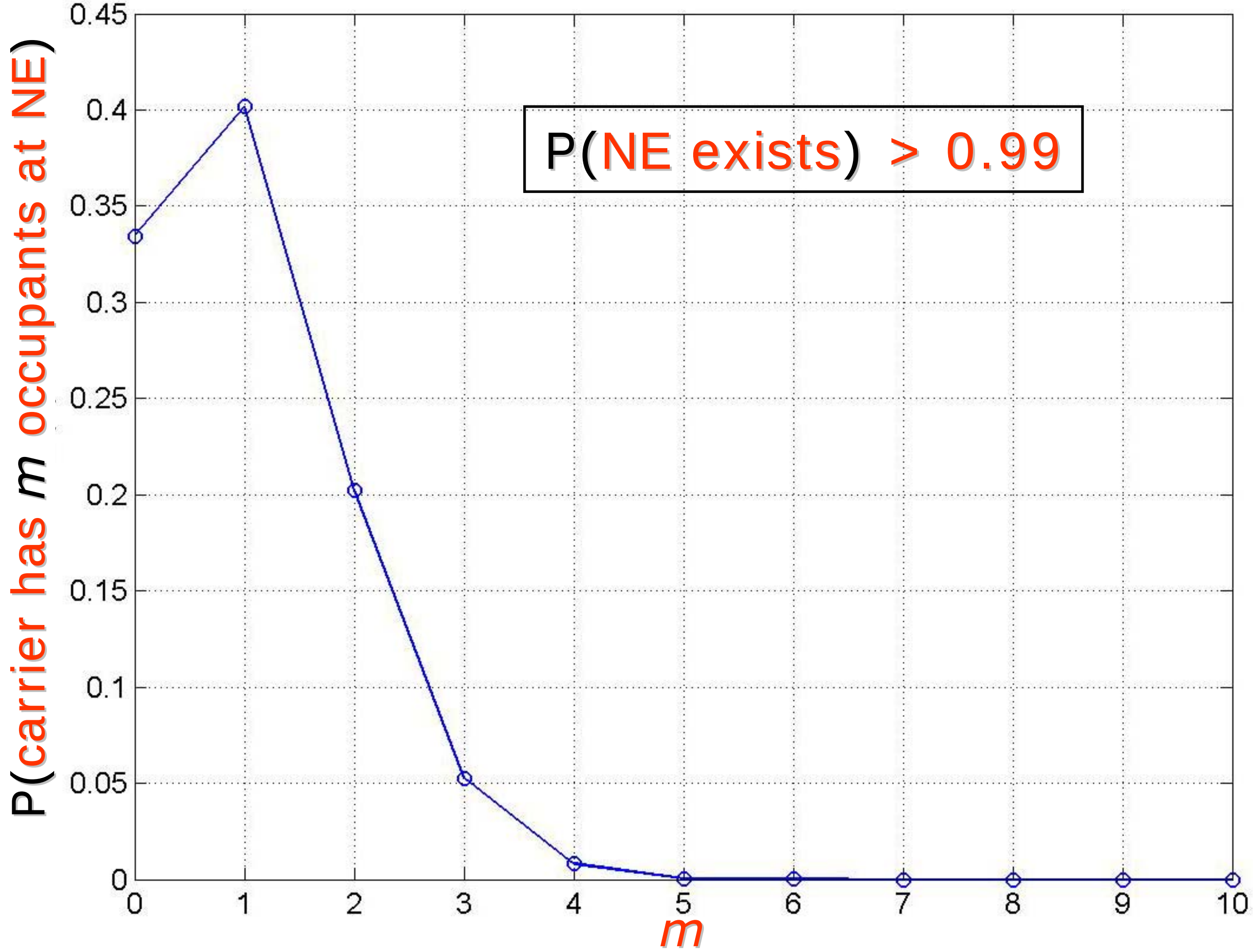
$$u_k = \frac{T_k^{(1)} + \ldots + T_k^{(D)}}{p_k^{(1)} + \ldots + p_k^{(D)}}$$

where $T_k^{(d)} = R_k^{(d)} f(\gamma_k^{(d)})$.

Nash Equilibrium

- For simplicity, we assume all users use MFs.
- u_k is maximized when all k's power is transmitted on its "best" carrier, and so as to achieve SIR γ^*.
- NE $\Leftrightarrow$ all users achieve this state simultaneously.
- NE may not exist, and may not be unique.
- Depends on the channel gains (a set of nec. & suff. inequalities can be derived).

Ex.: K=D=10; Rayleigh



ENERGY EFFICIENCY AND DELAY QOS

[w F. Meshkati, et al., *IEEE 2005 ISIT, Adelaide*]

Delay Model (Infinite Backlog)

- X is a r.v. representing the number of transmissions needed for a packet to be received error-free. Then

$$P(X=m) = f(\gamma)\,[1 - f(\gamma)]^{m-1}\ ,\ m = 0, 1, \ldots$$

- *i.e.*, X is a geometric r.v. with parameter $f(\gamma)$
- Specify the delay requirements by a pair (D, β):

$$P(X \le D) \ge \beta$$

- The delay requirements translate to a lower bound on SIR:

$$P(X \le D_k) \ge \beta_k \Leftrightarrow \gamma_k \ge \gamma_k'$$

Nash Equilibrium

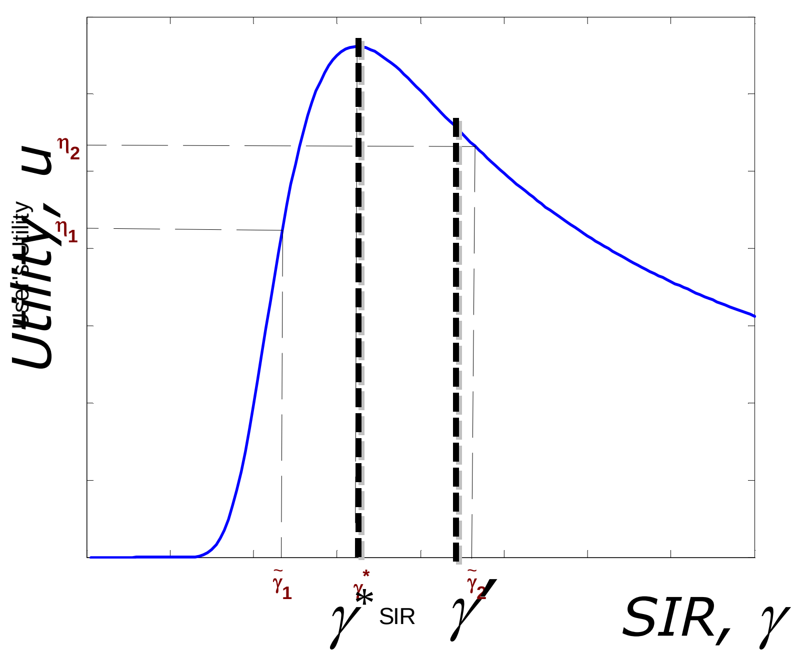
- Proposed delay-constrained power control game: $\max_{p_k \geq 0,\ \gamma_k \geq \gamma_k'} u_k$

- Th'm: For all linear multiuser receivers, the proposed game has a unique Nash equilibrium. At NE, each user transmits at a power level that achieves an output SIR equal to:

$$\max\{\gamma^*, \gamma_k'\}$$

where γ^* is the solution to

$$\boxed{f(\gamma) = \gamma f'(\gamma)}$$



Multi-class Networks

- Consider a network with C classes of users and K_c users in class c.
- All users in class c have the same delay requirements (D_c, β_c).
- At NE, all users in class c have the same output SIR $\gamma_c{}^*$ where

$$\boxed{\gamma_c{}^* = \max\{\gamma^*, \gamma_c{}'\}}$$

- Note: $\gamma_c{}^*$ depends on the delay constraints through $\gamma_c{}'$ and on physical layer parameters such as modulation, coding and packet size through γ^*.

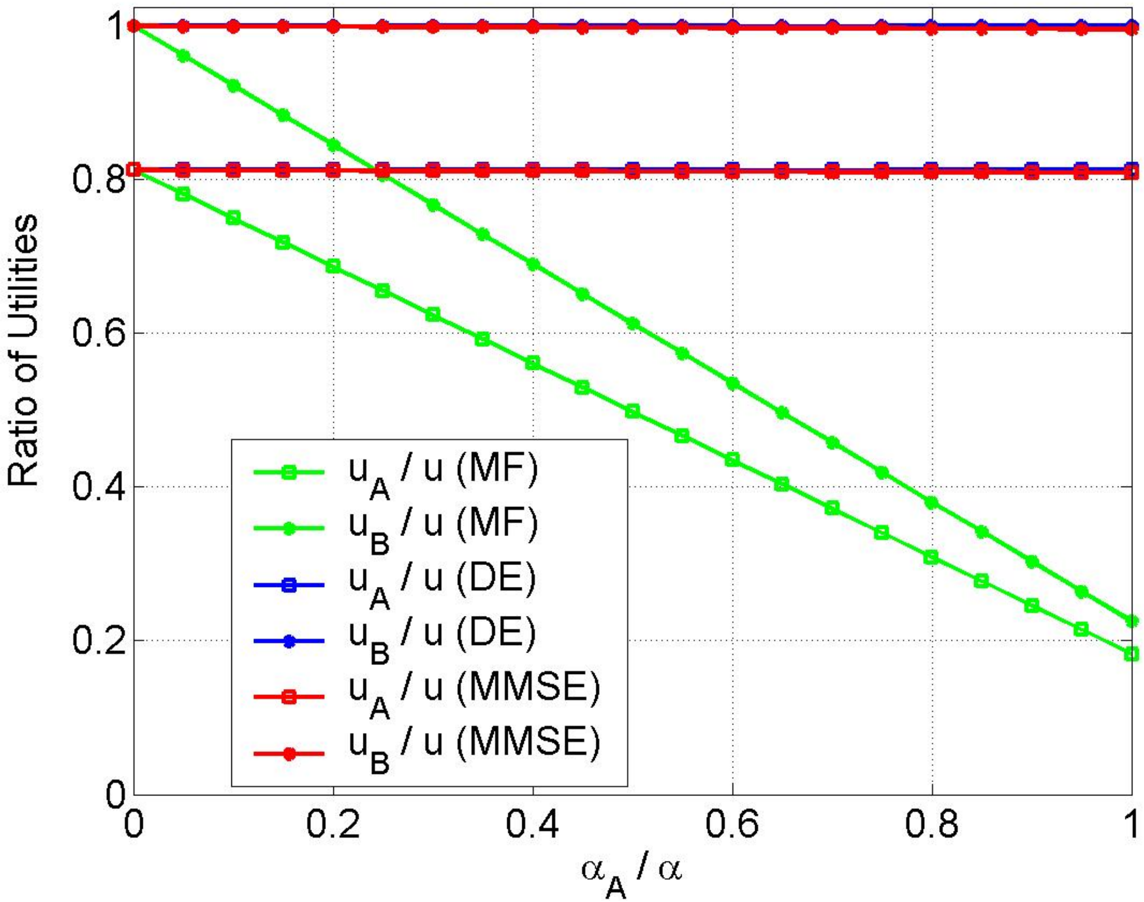
Large-System Analysis

- Large-system assumptions:

K_c, $N \to \infty$, with $\alpha_c = K_c/N$ fixed, $c = 1, 2, \ldots C$; $\alpha_1 + \alpha_2 + \ldots + \alpha_C = \alpha$

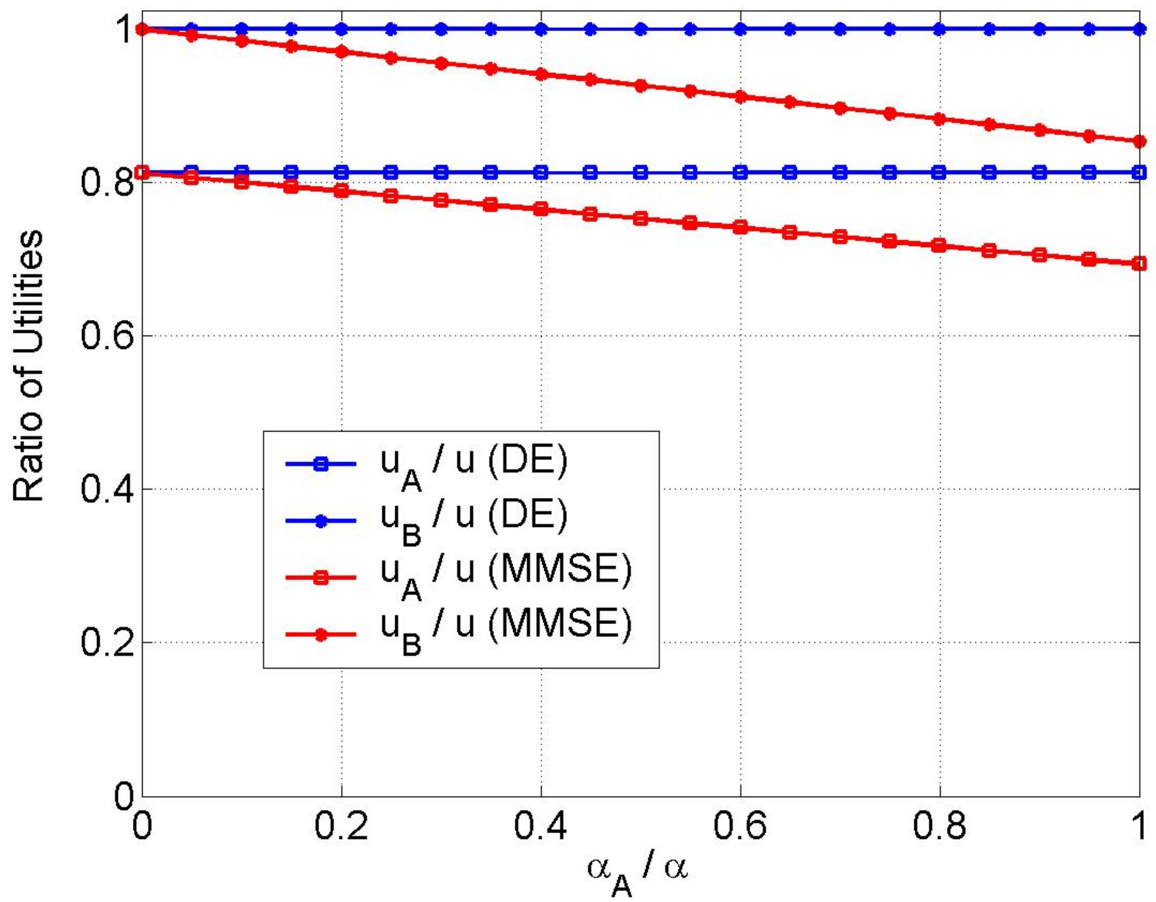
- Equilibrium utilities for matched filter, decorrelator and MMSE detector can be written in closed form (see paper).
- Observations:
  - Presence of users with stringent delay requirements affects the energy efficiency of all users in the network
  - Two factors contribute to the reduction in utility
    - Increase of target SIR (only for delay-sensitive users)
    - Increase in multiple-access interference (for all users)
  - The energy efficiency and network capacity are larger for MMSE detector as compared to decorrelator and matched filter

Numerical Results



Effect of delay on utility with low network load (α=0.1). Users in Class A are delay-sensitive ($D_A=1$, $\beta_A=0.99$) and users in Class B are delay-tolerant ($D_B=3$, $\beta_B=0.90$)

Numerical Results (Cont'd)



Effect of delay on utility with high network load (α =0.9). Users in Class A are delay-sensitive (D_A=1, β_A=0.99) and users in Class B are delay-tolerant (D_B=3, β_B=0.90)

Summary

- The Multiuser Power Control Game
- Unified Power Control
- Power Control in Multi-carrier CDMA
- Energy Efficiency and Delay QoS
- Other Interesting Problems:
  - Delay with Finite-Backlog (w. R. Balan et al.)
  - Adaptive modulation (w. A. Goldsmith, et al.)
  - Formalism for ad hoc networks (w. S. Betz)

Today's Talk: Three Topics
A Sampling of Ideas

- Energy Efficiency in Multiple-Access Networks
- Diversity-Multiplexing Tradeoffs in MIMO Systems ✓
- Distributed Inference in Wireless Sensor Networks

DIVERSITY-MULTIPLEXING TRADEOFFS IN MIMO SYSTEMS

[w T. Holliday & A. Goldsmith, *Proc. 2006 IEEE ICC*, Istanbul]

Introduction

- Multiple antennas provide a multiplexing versus diversity tradeoff in wireless channels.
- The "sweet spot" on this tradeoff curve is driven by higher-layer protocol performance metrics.
  - Cross-layer design with multiple antennas
- We investigate this sweet spot in the context of joint source and channel coding.
- The framework can be extended to include queueing delay and ARQ.

Diversity and Freedom

- Two fundamental resources in a MIMO wireless channel:
  - Diversity – improve error probability
  - Degrees of freedom – increase rate
- Most traditional coding formulations attempt to maximize only one of these.
- Recent results allow us to characterize optimal combinations of the two.

Diversity and Freedom

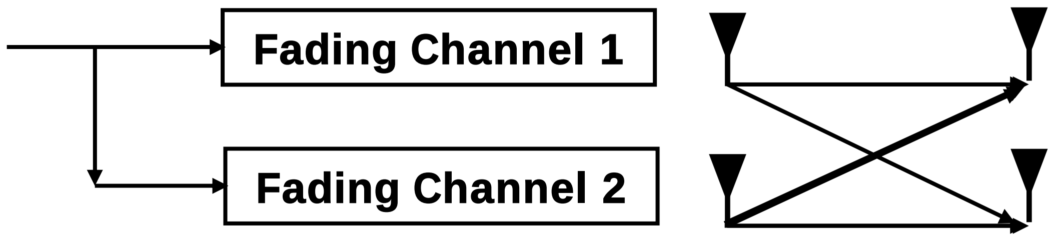
- Two fundamental resources in a MIMO wireless channel:
  - Diversity – improve error probability
  - Degrees of freedom – increase rate

- Most traditional coding formulations attempt to maximize only one of these.

- Recent results allow us to characterize optimal combinations of the two.

Diversity

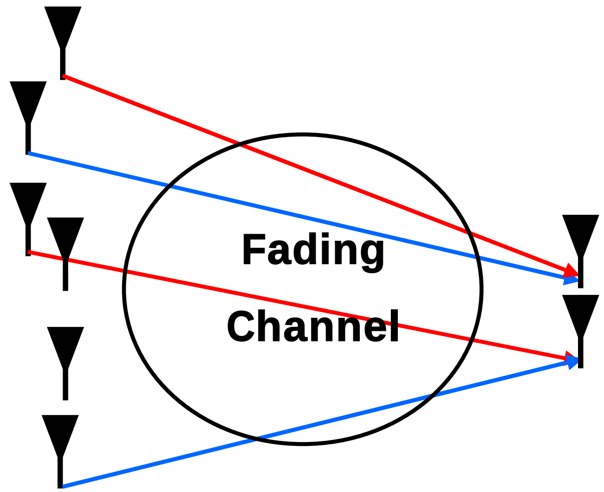
- Two independent fading channels increase diversity.



- Spatial Diversity:
  - Transmit, receive diversity, or both
- For a channel with *M* receive antennas and *N* transmit antennas the total diversity is *MN.*

Degrees of Freedom

- Arrivals from different directions provide extra degrees of freedom.



- We can achieve the same results with a scattering environment.
- In a M-by-N channel with scattering there are $min\{M,N\}$ degrees of freedom.

DEI SUB NUMINE VIGET

Diversity-Multiplexing Tradeoff

- A space-time code achieves a diversity-multiplexing tradeoff with rate *r* if

$$\boxed{R(SNR) \sim r \log SNR}$$

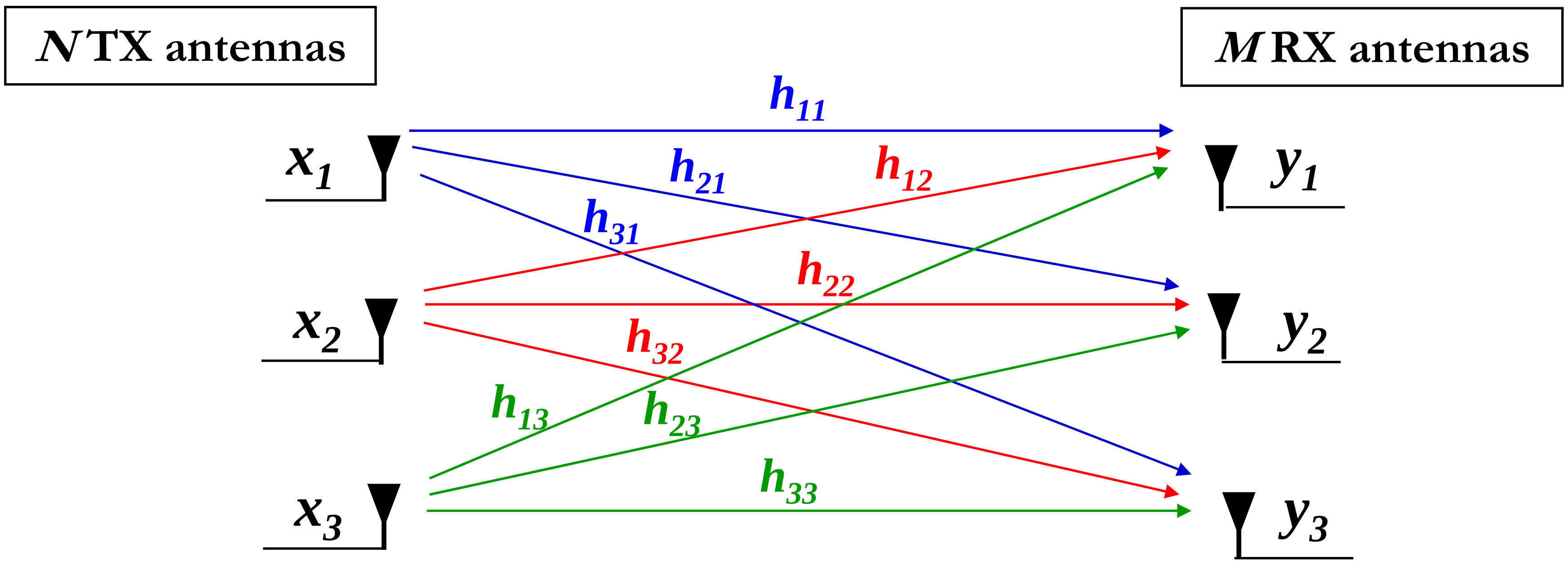
and the diversity gain *d(r)* satisfies

$$\boxed{P_e(SNR) \sim SNR^{-d(r)}}$$

- The largest rate is *min{M,N}.*

- The largest diversity gain is *MN.*

MIMO Channel Models



$$y = \sqrt{\frac{SNR}{N}} Hx + n$$

$$h_{ij} \sim CN(0,1)$$

$$n \sim N(0, I)$$

Tradeoff at High SNR*

- Define a family of block codes $\{C(SNR)\}$ of length T with rate $R(SNR) \sim r \log SNR$.

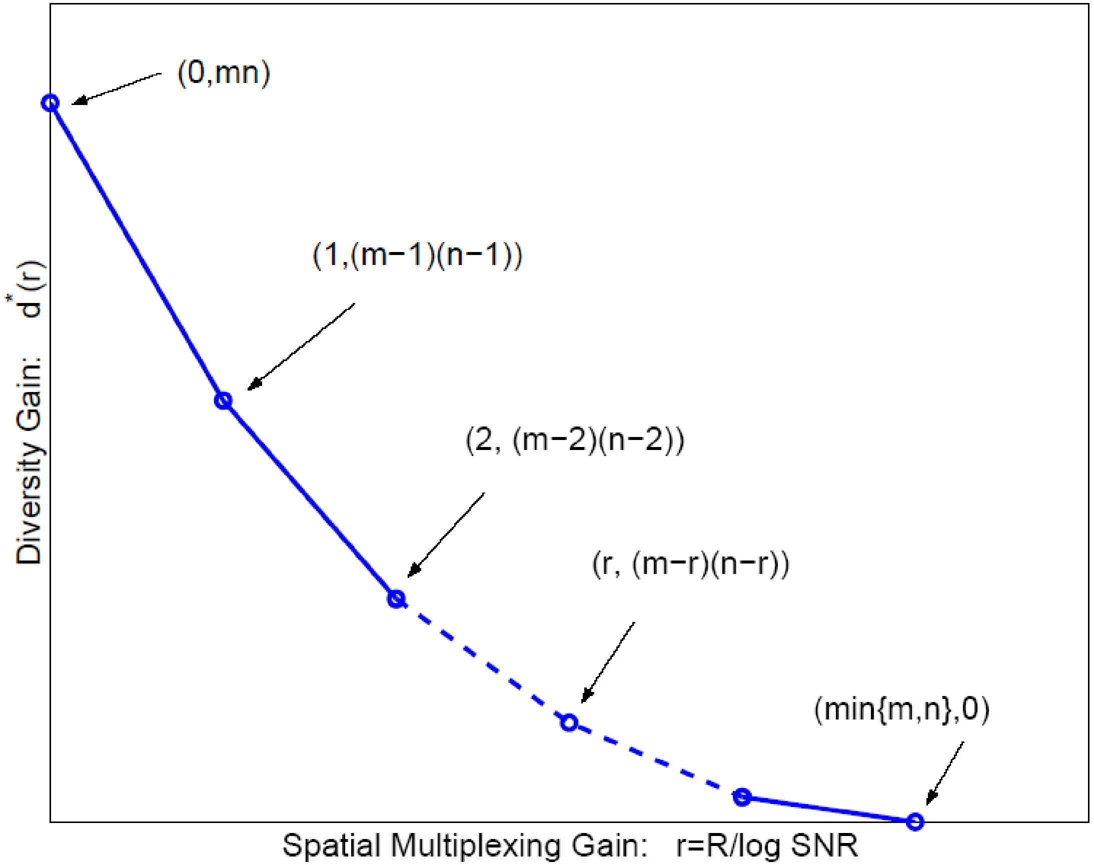
- Given $\{C(SNR)\}$, define diversity and multiplexing gains asymptotically.

$$\boxed{\lim_{SNR\to\infty} \frac{\log P_e(SNR)}{\log SNR} = -d} \qquad \boxed{\lim_{SNR\to\infty} \frac{R(SNR)}{\log SNR} = r}$$

$$\boxed{d^*(r) = (M-r)(N-r)}$$

*Zheng/Tse 2002

Diversity/Multiplexing Tradeoff



For an integer *r* it's as though *r* transmit and receive antennas are used for multiplexing, and the rest for diversity.

Applications

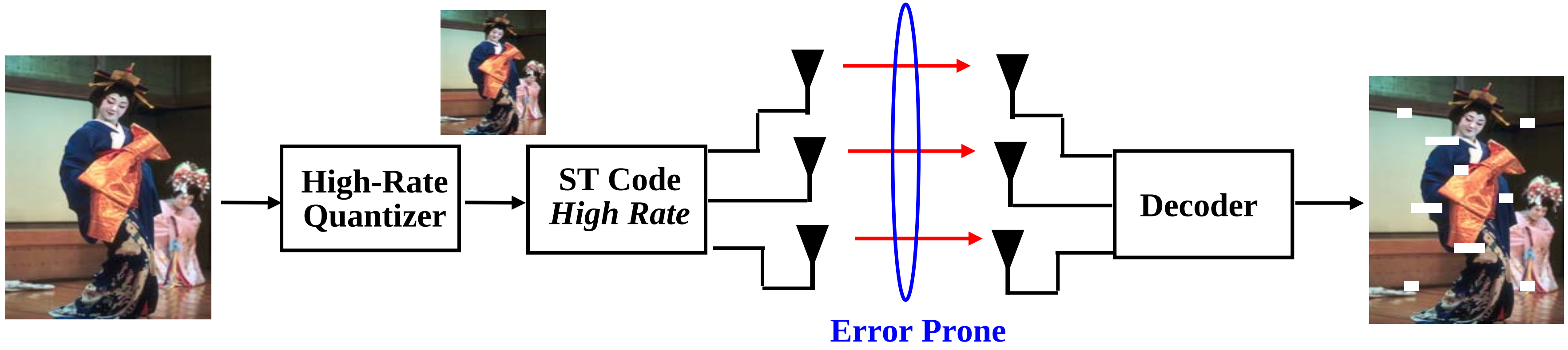
- Codes that achieve the tradeoff
  - El Gamal, Caire, *IEEE Trans. Inform. Theory,* 2004
  - Tavildar, Viswanath, *IEEE Trans. Inform. Theory,* 2004

- Incremental Redundancy and ARQ
  - Discuss later in the talk

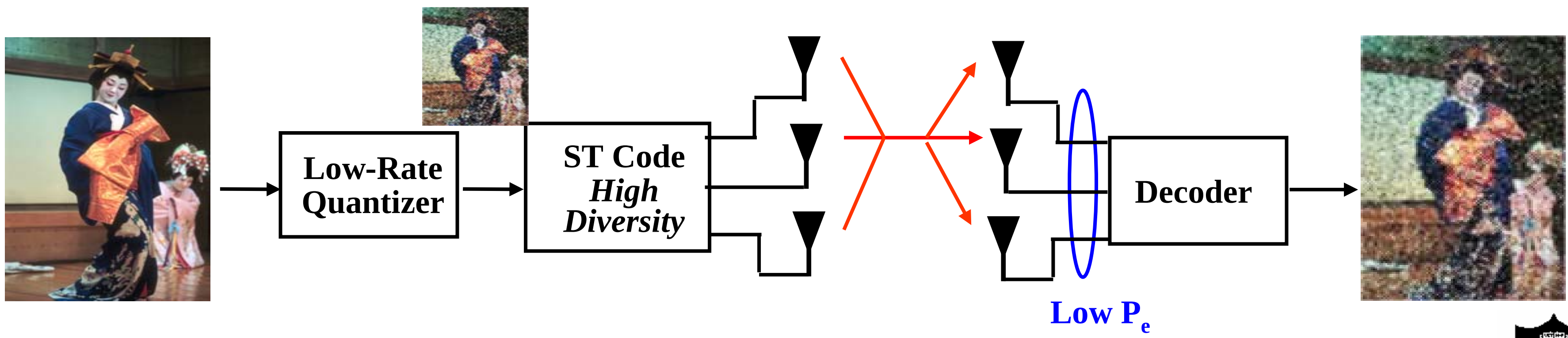
- Cross-layer design problems
  - Joint-source channel coding
  - Rate, compression, and ARQ adaptation for delay sensitive traffic

Joint S/C Coding with MIMO

- Use antennas for multiplexing:



- Use antennas for diversity:



Traditional Formulation

- Code over many (asymptotically infinite) source and channel coding blocks.
- Channel error goes to zero with blocklength via error exponent.
- Asymptotically optimal to encode source at channel capacity (full multiplexing).
  - No channel distortion
- Leads to optimal separation of source and channel code designs.

What about finite blocklengths?

Finite Block Lengths

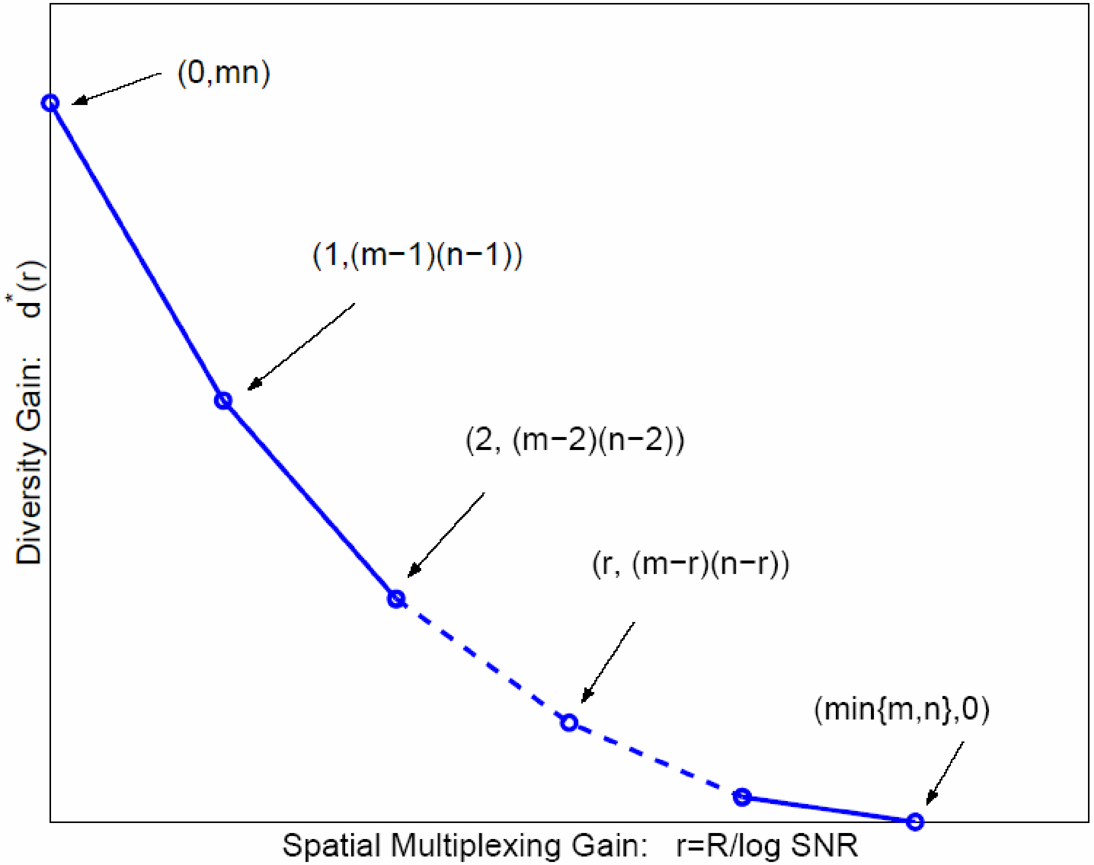
- Required under delay constraints.
- Can't drive channel error to zero:
  - Induces diversity/multiplexing tradeoff
- To optimize this tradeoff analytically, we require a high SNR regime:
  - Make use of Zheng/Tse results

Distortion Exponent

- Suppose we have a source with some notion of a distortion measure.
- Define the average distortion as $\overline{D}(r)$
- Then the optimal distortion exponent is

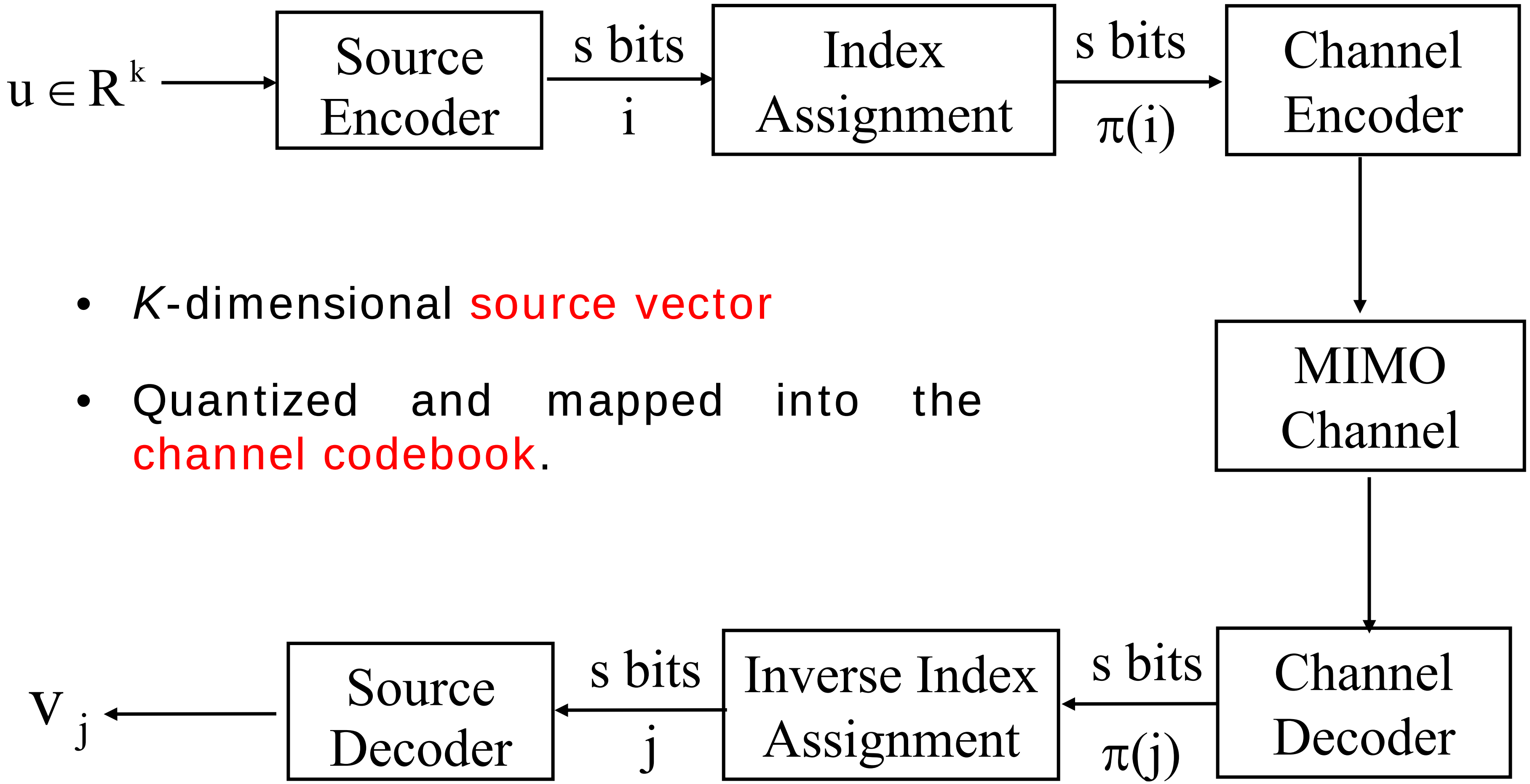
$$\boxed{\min_{r}\left[\lim_{SNR\to\infty}\frac{\log\overline{D}(r)}{\log SNR}\right]}$$

Diversity/Multiplexing Tradeoff



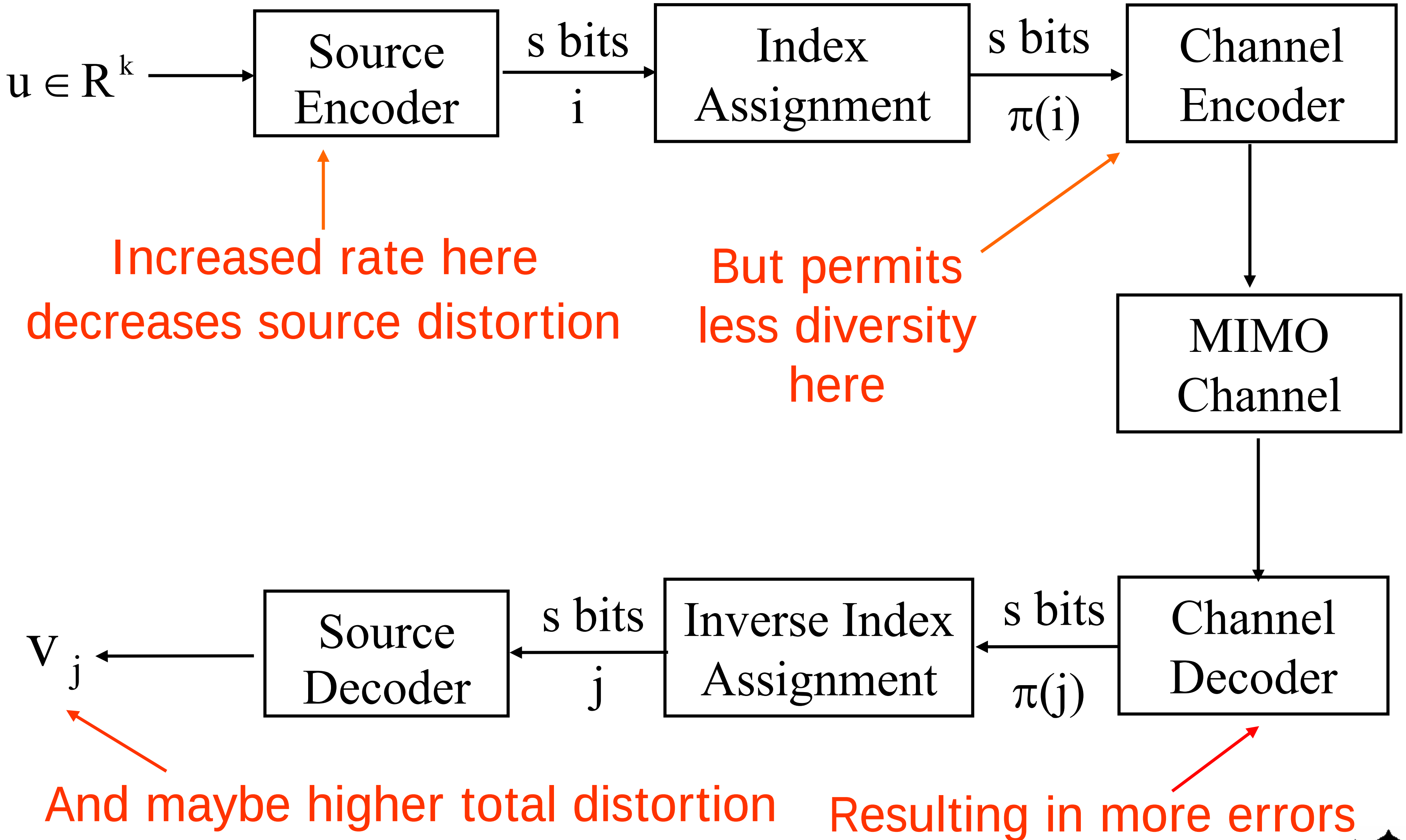
How do we map the distortion exponent to the diversity-multiplexing tradeoff curve?

Source and Channel Coding



- *K*-dimensional source vector
- Quantized and mapped into the channel codebook.

End-to-End Tradeoffs

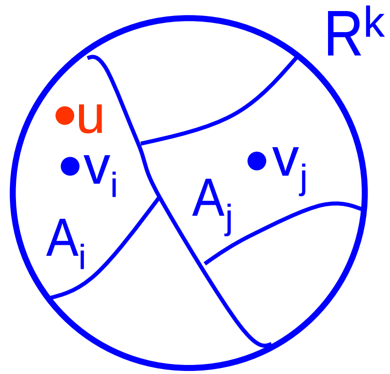


Source Code Construction

- Vector source $u \in R^k$ encoded into s bits by a quantizer Q with distortion $D_s(Q)$

$$Q(u) = \sum_{i=1}^{2^s} v_i 1_{A_i(u)}$$

$$D_s(Q) = \sum_{i=1}^{2^s} \int_{A_i} \|u - v_i\|^p f(u)du$$



- Distortion at asymptotically high source resolution satisfies‡

$$\boxed{D_s(Q) = 2^{-ps/k+O(1)} \text{ as } s \to \infty}$$

‡Gersho 1979

End-to-End Distortion

- Define $q(\pi(j)|\pi(i))$ as the probability that the channel decoder selects codeword j when codeword i was sent.

$$D_T(Q,SNR,\pi) = \sum_{i,j=1}^{2^s} (q(\pi(j)\,|\,\pi(i)) \int_{A_i} \|u - v_j\|^p f(x)dx$$

- From [Hochwald, 1998], this can be split into two pieces:

$$D_T(Q,SNR,\pi) \le \underbrace{D_s(Q)}_{\text{Correct decoding}} + \underbrace{O(1)}_{\text{Decoding failure}} \max_{1\le i\le 2^s} P_{e|\pi(i)}$$

Bound on Total Distortion

- We use the high SNR bound (finite blocks):

$$D_T(Q, SNR, \pi) \le D_s(Q) + O(1) \max_{1 \le i \le 2^s} P_{e|\pi(i)}$$

$$= 2^{-ps/k + O(1)} + 2^{-d^*(r)\log SNR + o(\log SNR)}$$

- Equating source and channel rates and assuming high SNR (order terms negligible):

$$D_T(Q, SNR, \pi) \le 2^{-\frac{p}{k}Tr\log SNR} + 2^{-d^*(r)\log SNR}$$

$$= SNR^{-\frac{p}{k}Tr} + SNR^{-d^*(r)}$$

- As SNR→∞, equal exponents minimizes distortion.

Asymptotic Upper Bound

- With matching exponents,

$$D_T(Q, SNR, \pi) \le SNR^{-\frac{p}{k}Tr} + SNR^{-d^*(r)} = 2SNR^{-d^*(r^*)}$$

- The distortion is a simple function of SNR and the optimal diversity/multiplexing point.

- Asymptotically, leads to a familiar form:

$$\boxed{\min_r \left[\lim_{SNR \to \infty} \frac{\log D_T(Q, SNR, \pi)}{\log SNR} \right] \le -d^*(r^*)}$$

Tight Results

- A lower bound can also be constructed for the average distortion (w.r.t π)

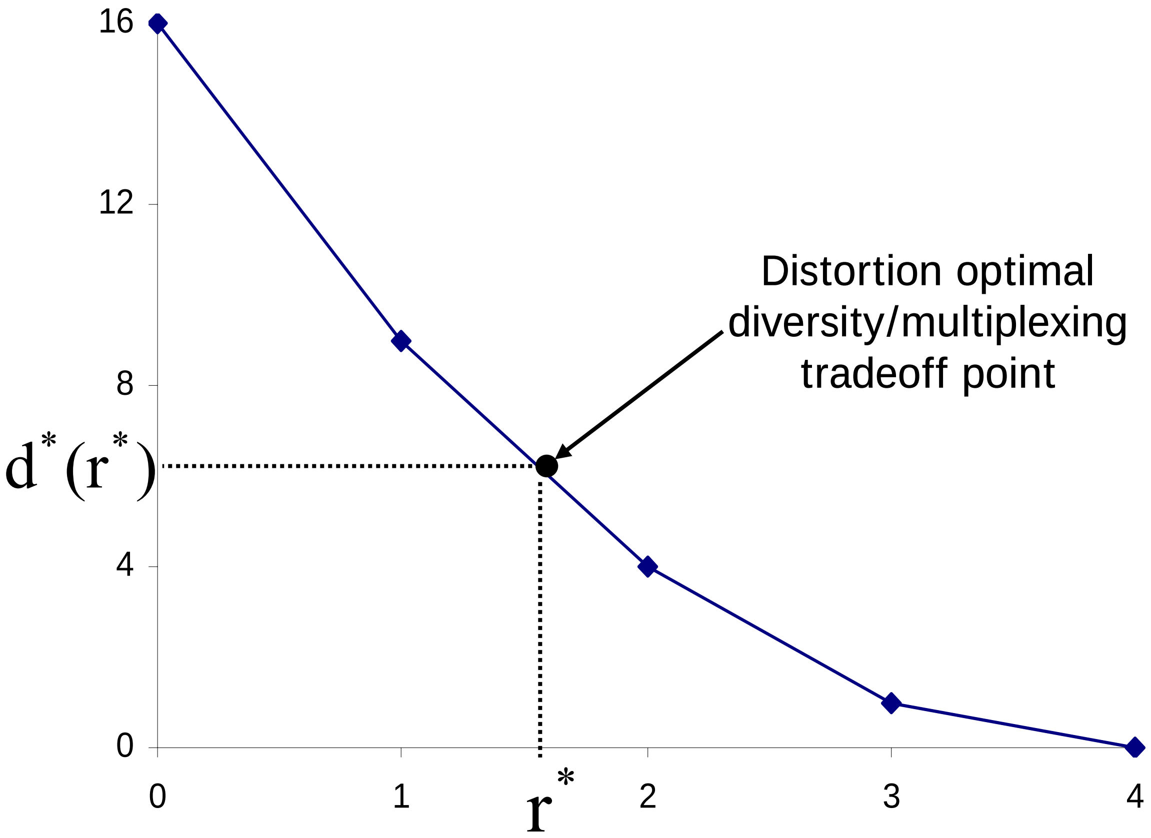
- The optimal distortion exponent is:

$$-d^*(r^*) = \min_r \left[\lim_{SNR \to \infty} \frac{\log \overline{D}_T(Q, SNR)}{\log SNR} \right]$$

$$d^*(r^*) = \frac{pTr^*}{k}$$

Solve this to find r*.

Optimal Tradeoff Point



$$d^*(r^*) = \frac{pTr^*}{k} = 4r^*$$

$$r^* = \frac{14}{9}$$

4x4 MIMO, with T=16, p=2, and k=8

Minimizing Distortion

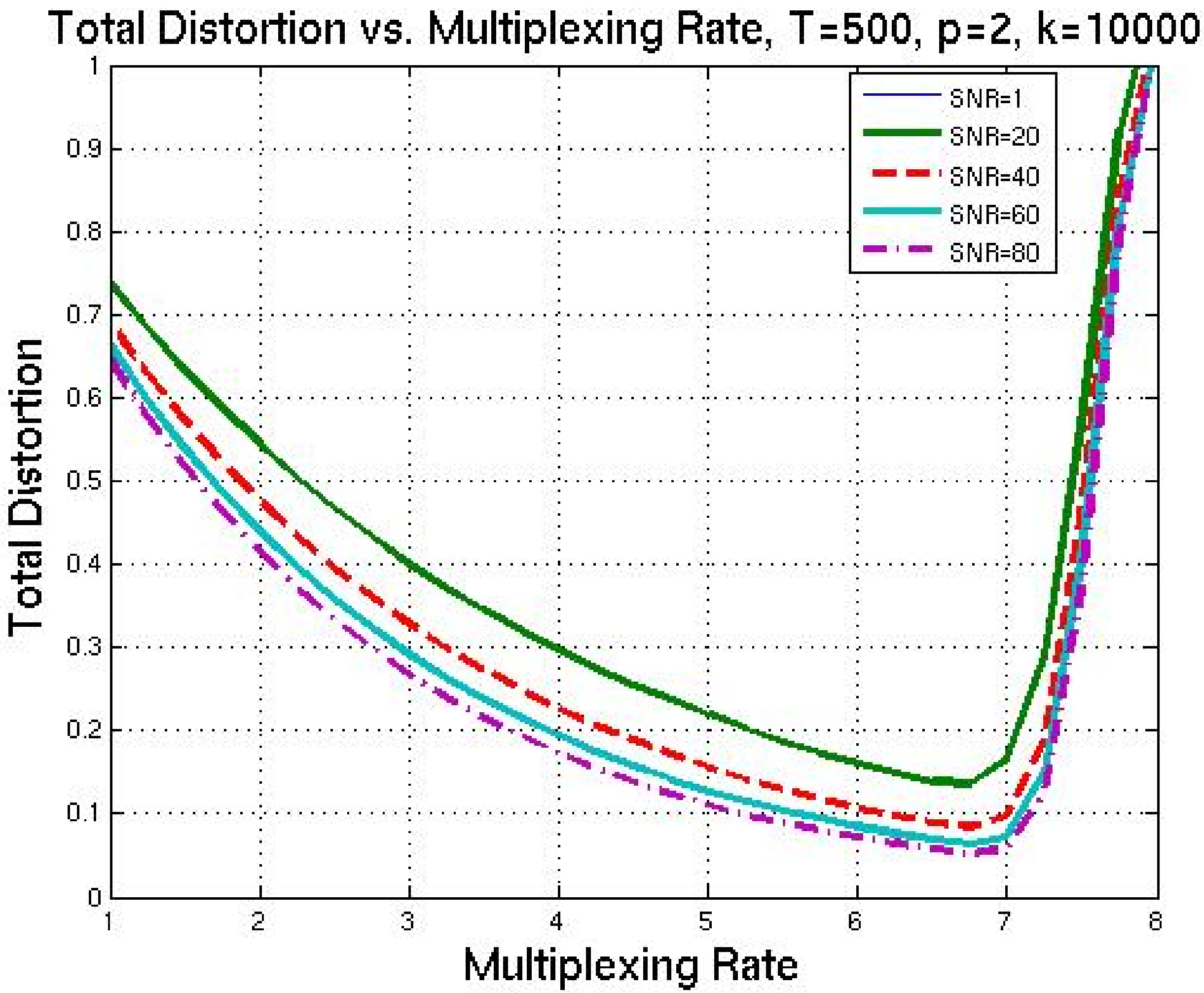
$$\min_{r} \; \mathrm{SNR}^{-\frac{p}{k}Tr} + \mathrm{SNR}^{-d^*(r)}$$

$$s.t. \quad d^*(r) = (N - r)(M - r), \; \textit{piecewise linear}$$

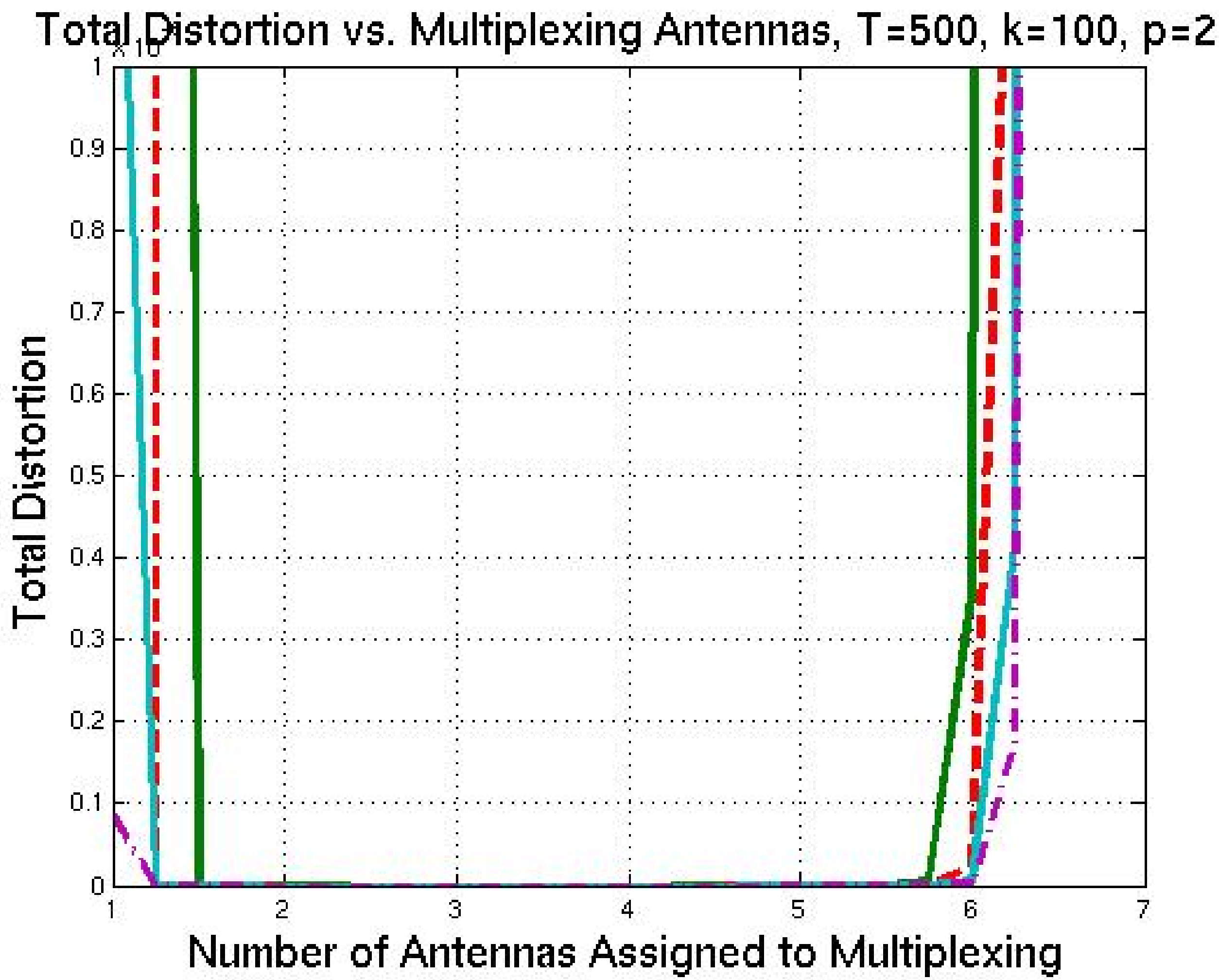
$$0 \le r \le \min(M, N)$$

- Convex problem:
  - Source dimension *k*
  - Channel code block size *T*
  - Distortion order *p*
- Solution shows how to best use antennas.

Distortion vs. Multiplexing N=M=8 k>>T

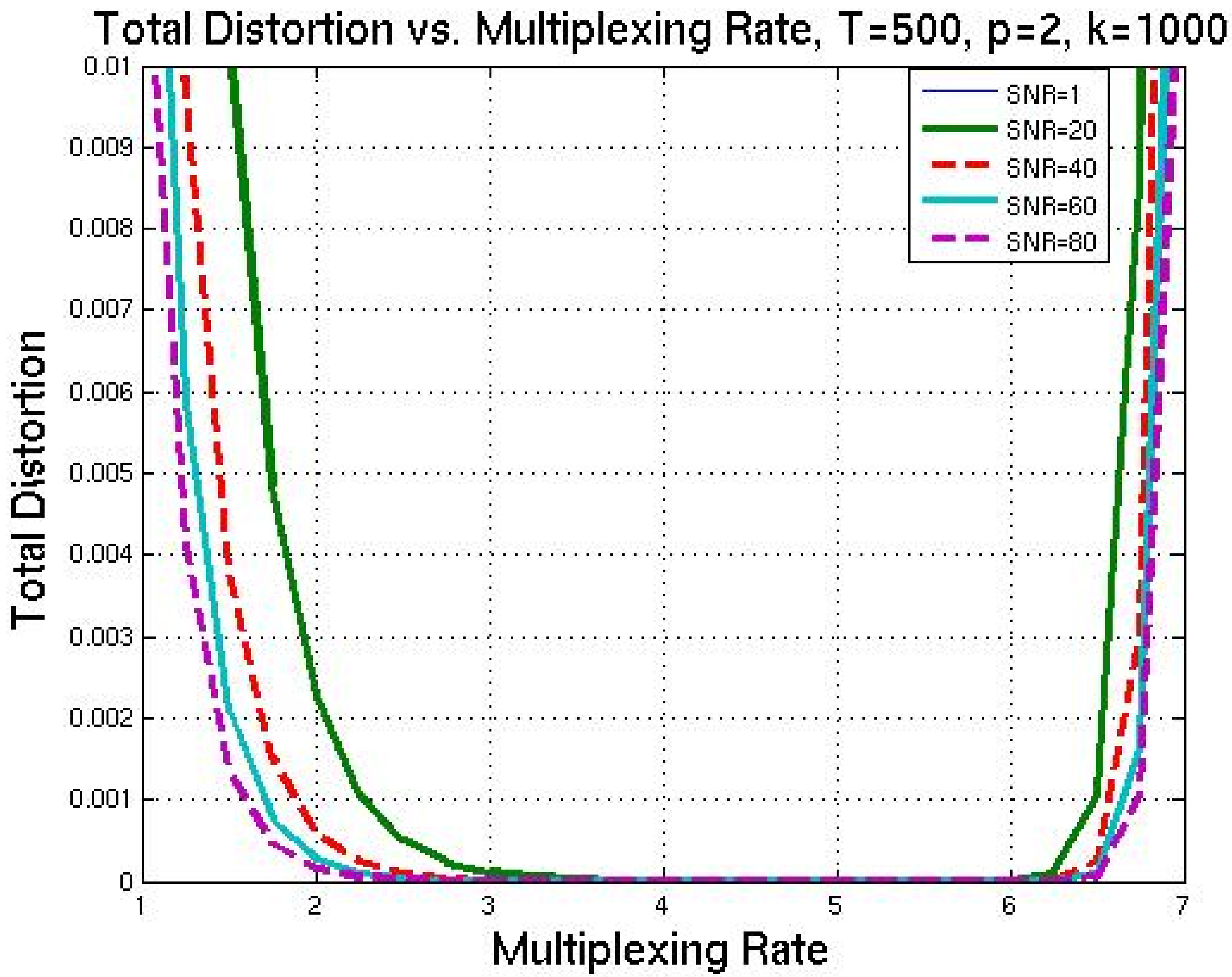


Distortion vs. Multiplexing N=M=8 k<<T



Diversity-Multiplexing Tradeoffs in MIMO Systems

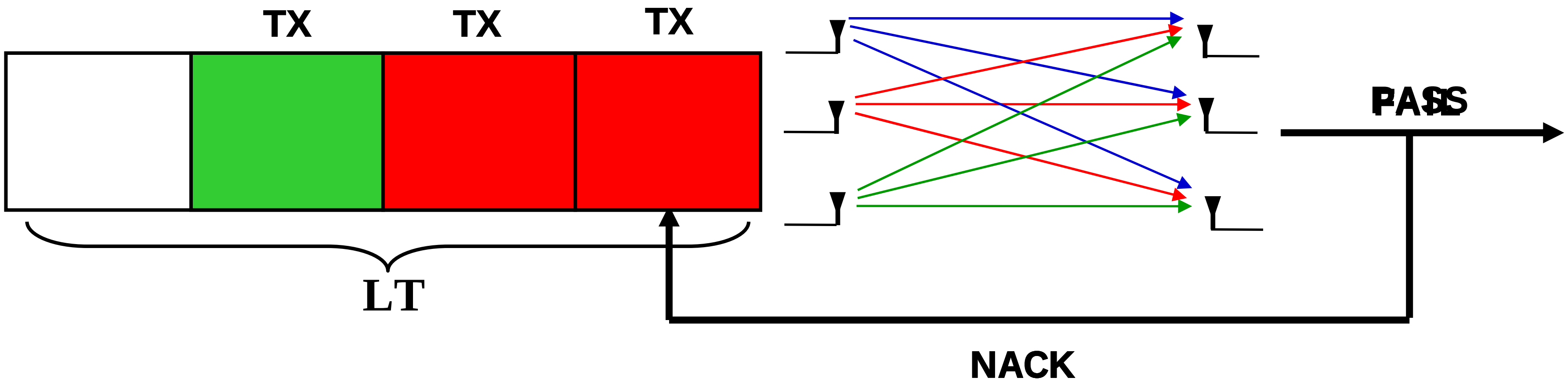
Distortion vs. Multiplexing N=M=8
k = O(T)



Separation

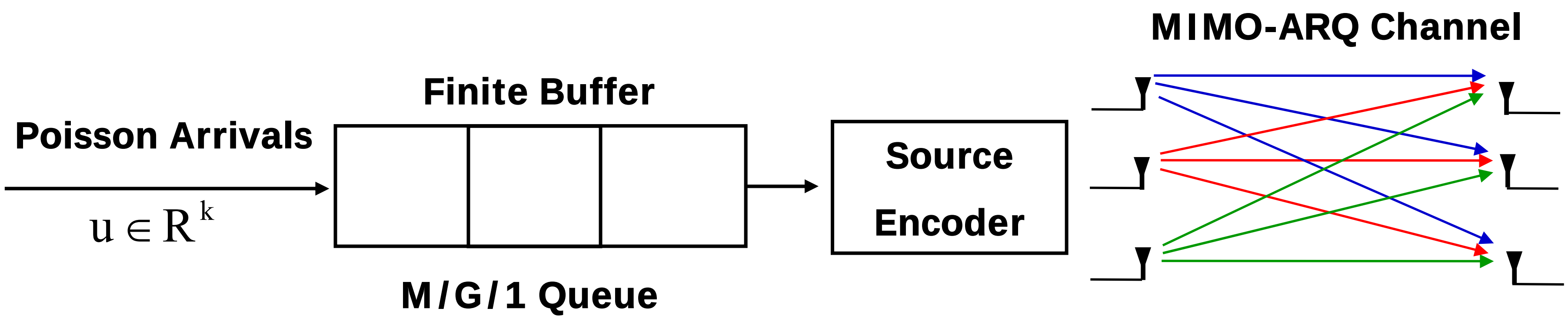
- Assume the source and channel encoder know M, N, T, p, and k.
- Then we can encode the source and channel at r^* without coordination.
- Provides a version of a separation theorem.
- What happens when we permit ARQ?

MIMO-ARQ Channel



- Suppose we correct errors through incremental redundancy.
- Define a family of block codes $\{C(\mathrm{SNR})\}$ of length LT:
  - Block fading channel with block size T
  - ARQ window size L
- Retransmissions reduce the average rate.

Delay System Model



- Arriving data have deadlines.
- Errors result from ARQ failure or deadline expiration.
- What is the optimal tradeoff?

Tradeoff Options

- New average distortion measure:

$$\overline{D}_T(Q,\mathrm{SNR}) \le \overline{D}_S(Q) + P_e(\mathrm{SNR}) + P[Delay > t]$$

- Find a fixed allocation of rate, diversity, and ARQ that minimizes average distortion:

- Or we could try to adapt to the random queue:
  – Large queue -> high compression, high diversity
  – Empty queue -> low compression, high multiplexing

- Fixed scheme is separable, adaptive scheme is not.

Adaptive Solution

- The M/G/1 queue is a stochastic process.
- Control the size of a job by changing the source compression
- Control the service time with the multiplexing rate r and ARQ window L
- Formulate and solve a standard dynamic program to find the optimal solution.‡

‡Holliday, Goldsmith, Poor ICC 2006

Numerical Example

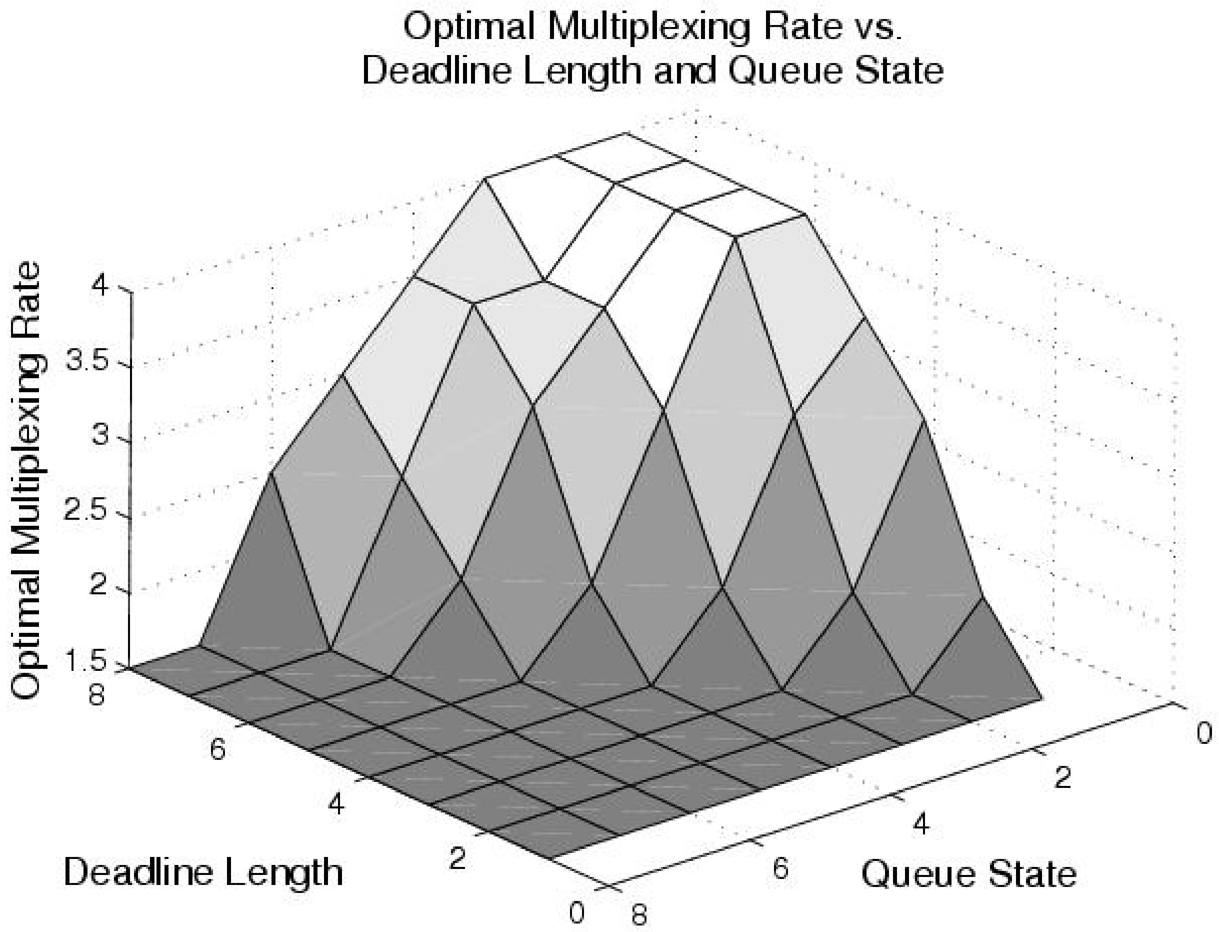
- 4x4 MIMO–ARQ channel with block fading

- $\mathrm{L} \in \{1, 2, 3, 4\}$

- Arrival rate λ =0.9

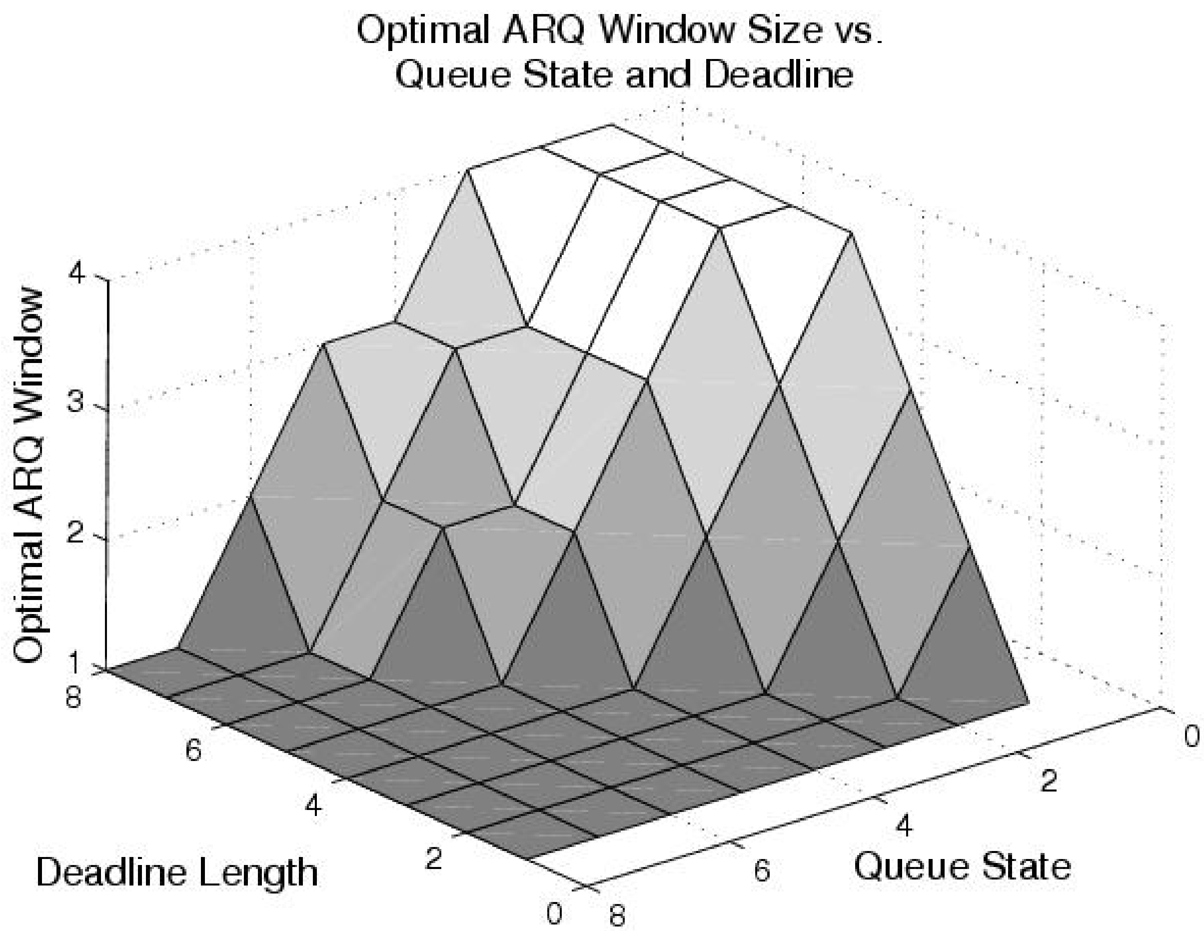
- Examine the impact of deadlines ranging from short (d=2) to long (d=16)

Optimal Multiplexing Rate



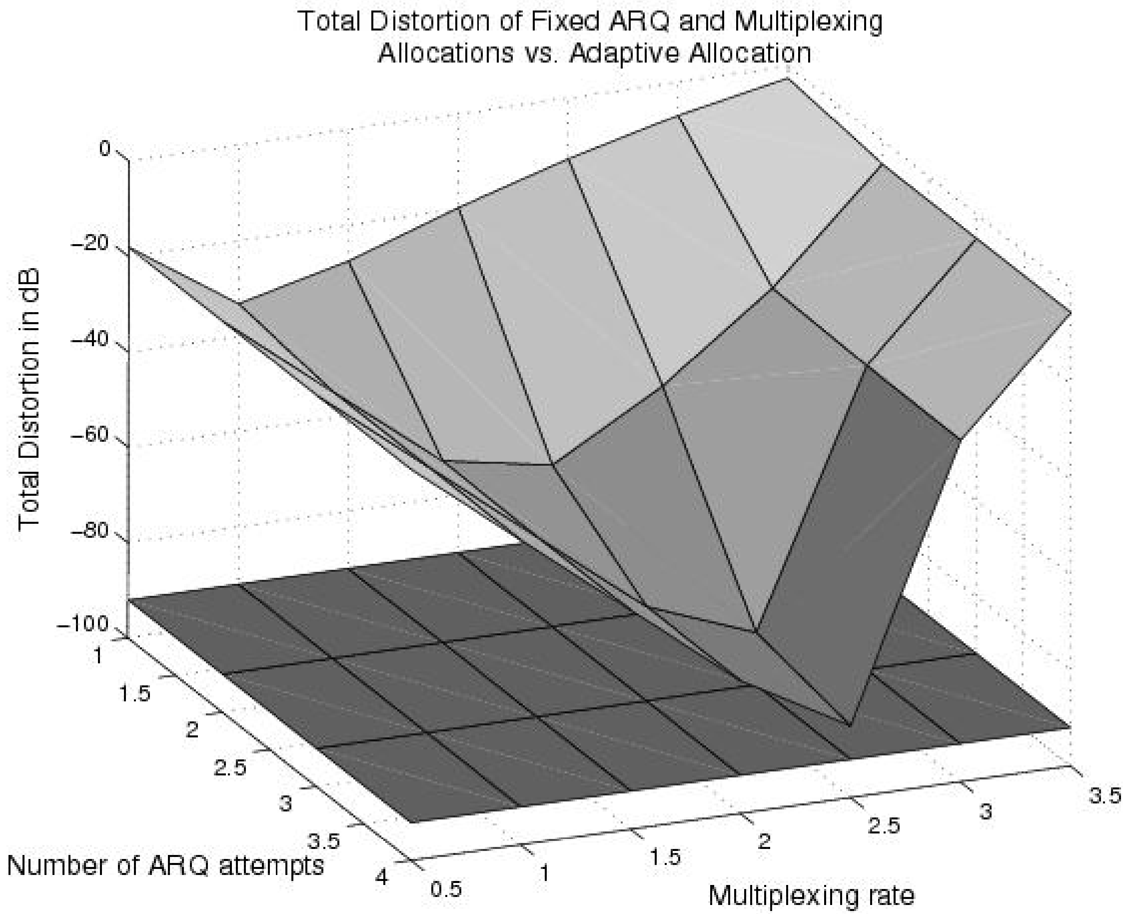
Multiplexing decreases as the queue size increases or as deadlines become shorter.

ARQ Window Size



We use longer ARQ windows with empty queues and long deadlines.

Distortion for Fixed and Adaptive Controls



Note the gap between the best fixed policy and the adaptive results.

MIMO-ARQ at Finite SNR

- When delay affects distortion we no longer have a separation theorem.
- The best fixed pair of rate and ARQ incurs more distortion than adaptation.
- Optimal fixed rate and ARQ assignment does not utilize all available ARQ.
- Therefore, the high SNR regime is a poor approximation in this case.

Summary

- Computed the end-to-end distortion exponent for a MIMO system
- The exponent is a point on the diversity-multiplexing tradeoff curve
- Results in a separation theorem for finite-block length codes
- Delay sensitivity precludes a separation theorem

Today's Talk: Three Topics
A Sampling of Ideas

- Energy Efficiency in Multiple-Access Networks
- Diversity-Multiplexing Tradeoffs in MIMO Systems
- Distributed Inference in Wireless Sensor Networks ✓

DISTRIBUTED INFERENCE IN WIRELESS SENSOR NETWORKS

[w. J. Predd & S. Kulkarni, *Proc. 2006 IEEE Info. Th'y Workshop*, Uruguay]

*Outline: Collaborative Regression**

- *Background*
- *A Model for Distributed Learning*
- *A Collaborative Training Algorithm*
- *Performance Guarantees*
- *Energy Considerations*

* [Related work w. Predd & Kulkarni, *IEEE Trans. Inform. Th'y*, Jan. 2006]

Classical (Supervised) Learning

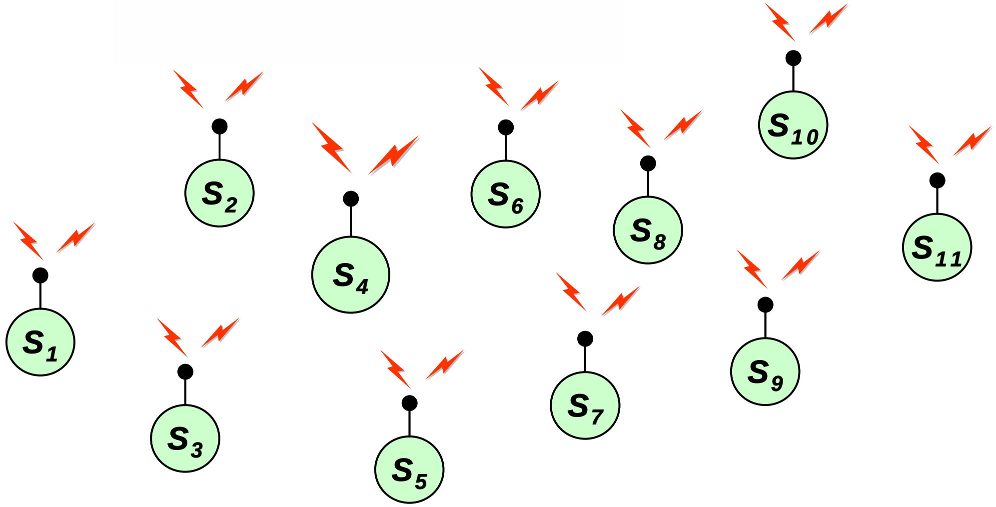
- Input space $\mathcal{X} = \mathcal{R}^d$, Output space $\mathcal{Y} = \mathcal{R}$
- (X, Y) is an $\mathcal{X} \times \mathcal{Y}$-valued r.v. with $(X, Y) \sim \mathbf{P}_{XY}$
- Design $g : \mathcal{X} \to \mathcal{Y}$ to predict outputs from inputs and minimize expected loss

$$\mathbf{E}\{|g(X) - Y|^2\}$$

- $\mathbf{P}_{XY}$ is unknown
- Given **training examples** $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^n \subset \mathcal{X} \times \mathcal{Y}$

A Model for Distributed Learning (in WSNs)

- Sensor *i* measures $S_i = \{(\mathbf{x}_j, y_j)\} \subseteq S$.



- Example:
  - $\mathcal{X} = \mathcal{R}^3$ (coordinates in space-time),
  - $\mathcal{Y} = \mathcal{R}$ (temperature)
  - $(X, Y) \sim \mathbf{P}_{XY}$ models structure of a temperature field

"A distributed sampling device with a wireless interface"

The Centralized Approach

- Sensor *i* sends $S_i = \{(\mathbf{x}_j, y_j)\}$ to a centralized processor

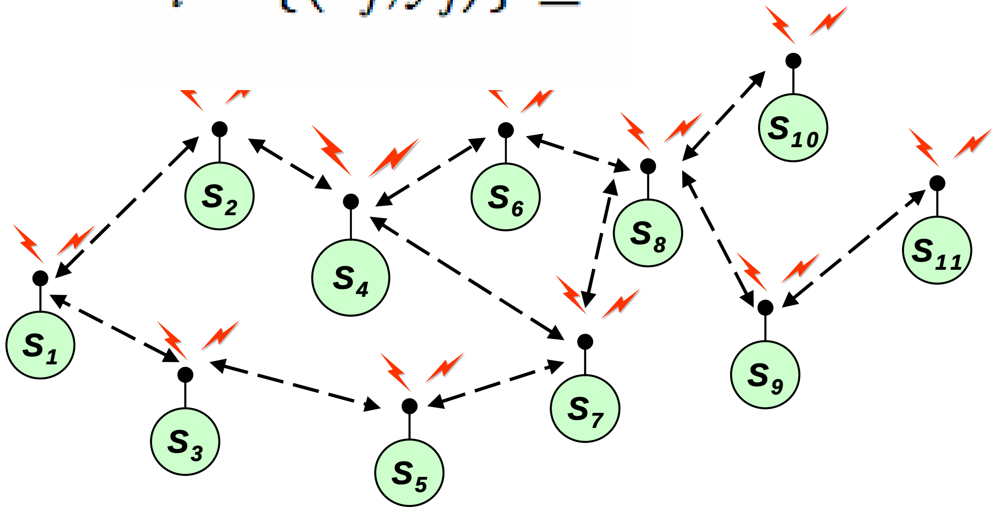
- "Learn" using (Reproducing) Kernel Methods:
  - For a positive semi-definite kernel **K(,)**:

$$f_\star = \arg\min_{f \in \mathcal{H}_K} \sum_{i=1}^{n} (f(\mathbf{x}_j) - y_i)^2 + \lambda \|f\|_{\mathcal{H}_K}^2$$

- Assumption: energy and bandwidth constraints preclude the sensors from sending $S_i = \{(\mathbf{x}_j, y_j)\} \subseteq S$ for centralized processing.

The Seed of a Model…

- Sensor i measures $S_i = \{(\mathbf{x}_j, y_j)\} \subseteq S$

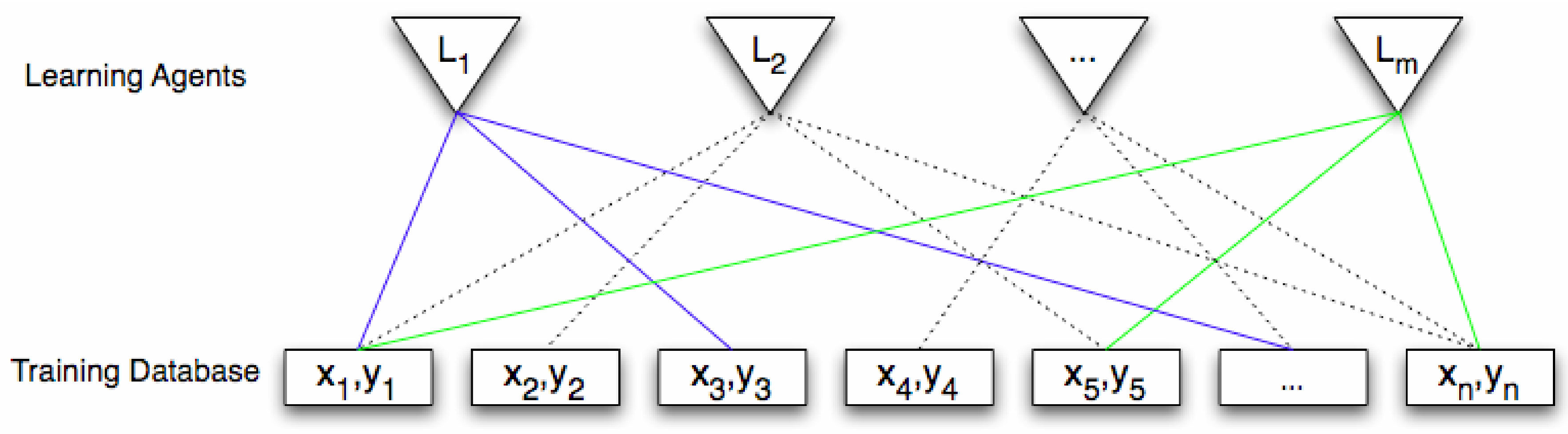


Assumption: Sensor i can access all neighboring sensors' measured data.

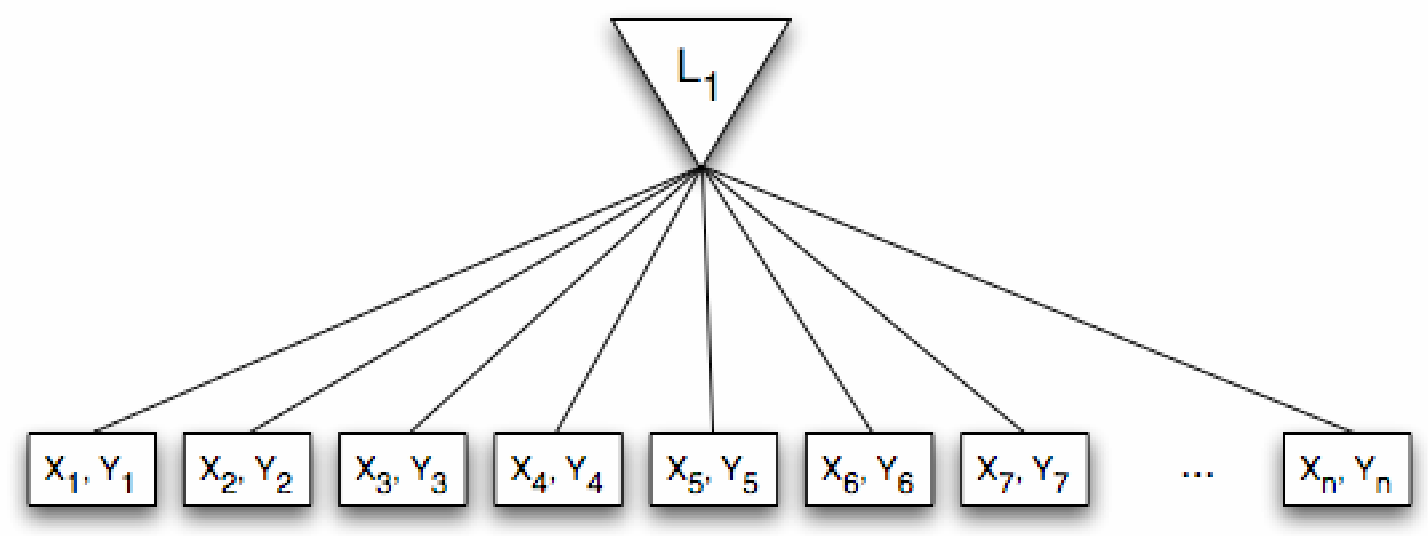
- **Informal justification**: local communication is efficient

A General Model

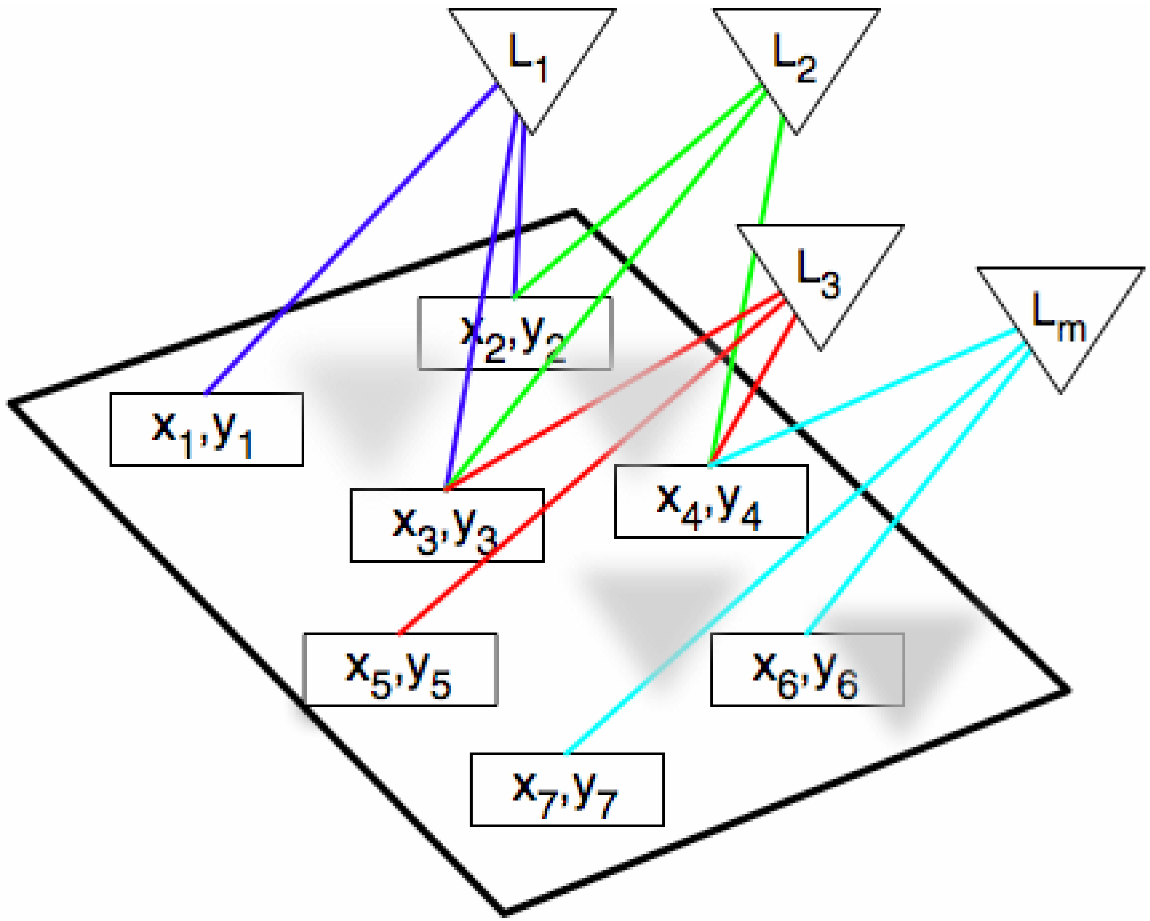
- *m* ***learning agents*** *(i.e., sensors)*
- *n* ***training examples*** $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$



Example
Centralized Learning

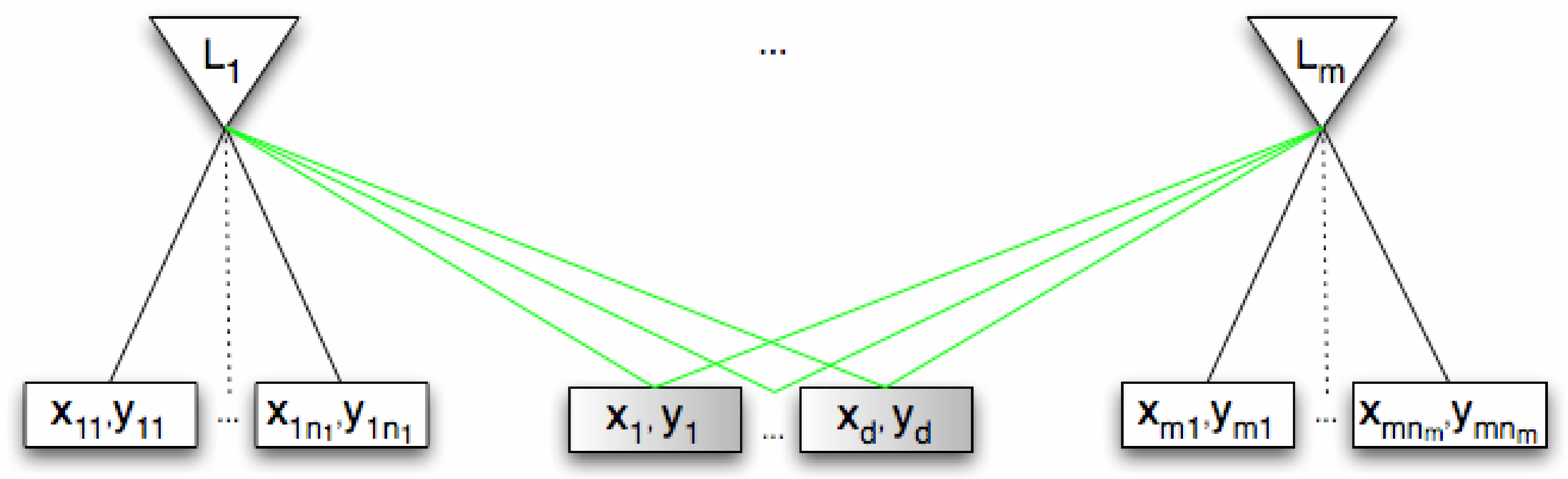


Example
Spatio-temporal Field Estimation in WSNs



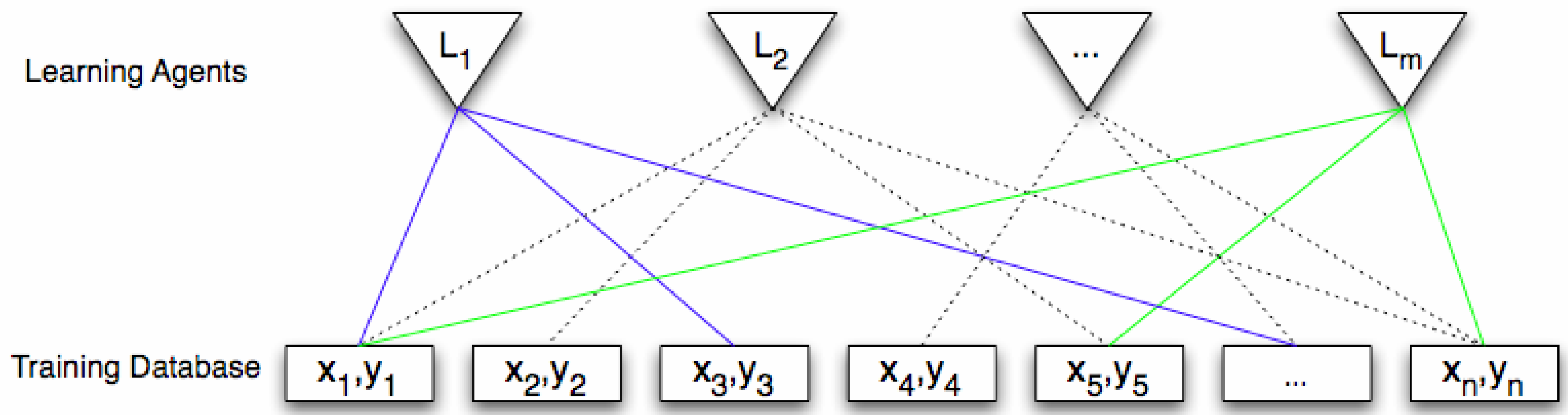
Collaborative Regression

Example
A public database

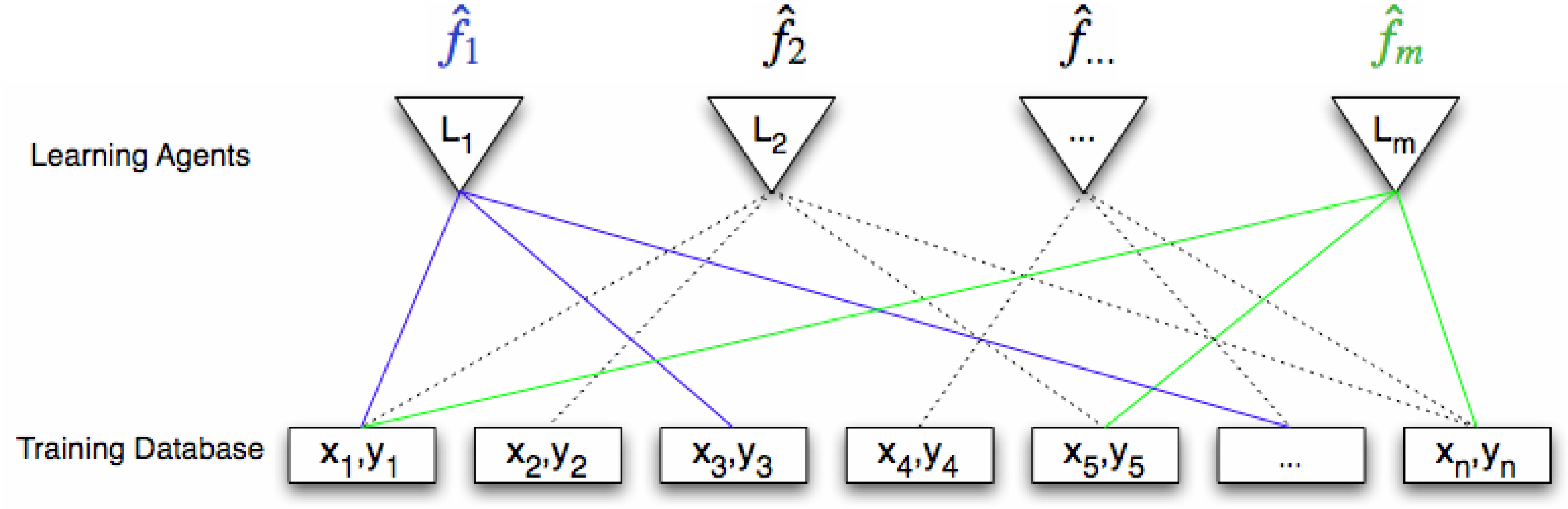


The General Case

- *m* ***learning agents*** *(i.e., sensors)*
- *n* ***training examples*** $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$



"Local" Learning
A Natural Approach

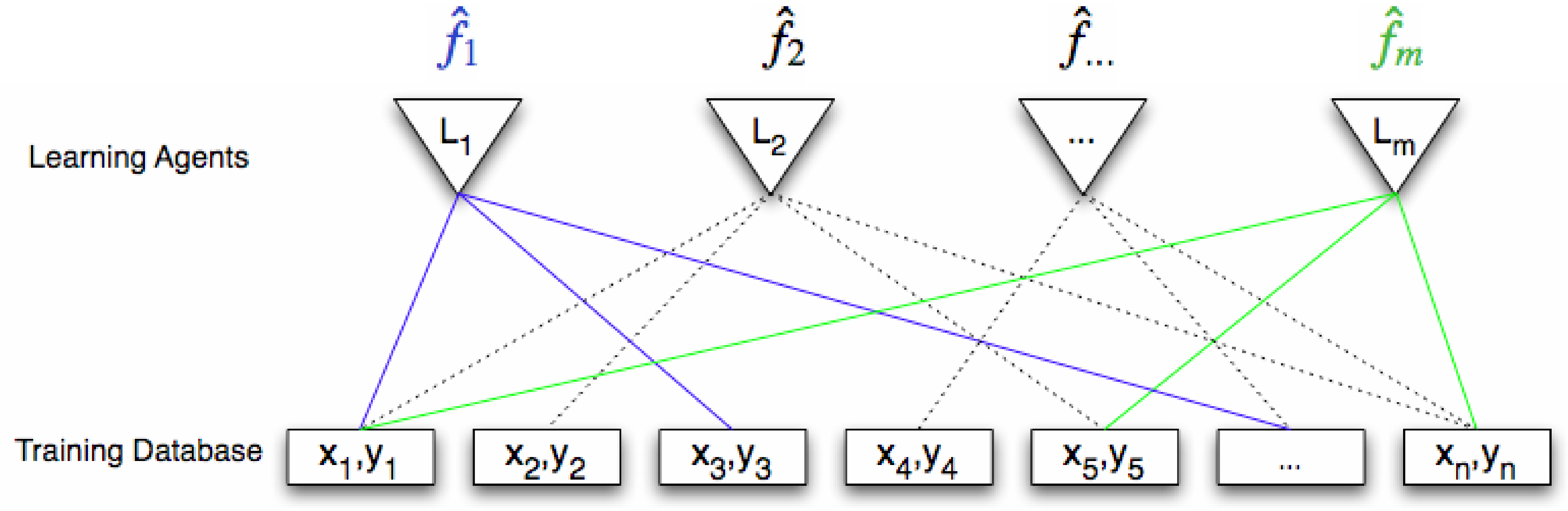


$$\hat{f}_1 = \arg\min_{f \in \mathcal{H}_K} \frac{1}{|N_1|} \sum_{j \in N_1} (f(\mathbf{x}_j) - y_j)^2 + \lambda_1 \|f\|^2_{\mathcal{H}_K}$$

$$\hat{f}_m = \arg\min_{f \in \mathcal{H}_K} \frac{1}{|N_m|} \sum_{j \in N_m} (f(\mathbf{x}_j) - y_j)^2 + \lambda_m \|f\|^2_{\mathcal{H}_K}$$

Collaborative Regression

Local learning is "locally incoherent"



Local incoherence:

Sensor ***1*** *and sensor* ***m*** *both train with* $(\mathbf{x}_1, y_1)$

but $\hat{f}_1(\mathbf{x}_1) \neq \hat{f}_m(\mathbf{x}_1)$

Collaborative Regression

...And local incoherence is **undesirable**

Define distance measure

$$\Delta_i(f,g) = \frac{1}{|N_i|}\sum_{j\in N_i}(f(\mathbf{x}_j) - g(\mathbf{x}_j))^2 + \lambda_i \|f-g\|^2_{\mathcal{H}_K}$$

...And local incoherence is **undesirable**

Define distance measure

$$\Delta_i(f,g) = \frac{1}{|N_i|}\sum_{j \in N_i}(f(\mathbf{x}_j) - g(\mathbf{x}_j))^2 + \lambda_i \|f - g\|_{\mathcal{H}_K}^2$$

Lemma: For any locally incoherent rules $\{\hat{f}_i\}_{i=1}^m$

there exists a set of locally coherent rules $\{\hat{g}_i\}_{i=1}^m$

such that

$$\frac{1}{m}\sum_{i=1}^m \Delta_i(\hat{g}_i, f_\star) \leq \frac{1}{m}\sum_{i=1}^m \Delta_i(\hat{f}_i, f_\star)$$

Summary

- *In this model for distributed learning, "local learning"* ***requires only local communication****.*

- *However, it leads to* ***local incoherence****, which* ***is provably "undesirable"****.*

- **Can agents (i.e., sensors) collaborate to gain the "optimality" of coherence, while retaining efficiency locality?**

A Collaborative Training Algorithm
Intuition

- Local learning as a building block.

Iterate over sensors **s = 1,…, m**

sensor **s**

Computes using local data:

$$\hat{f}_s = \arg\min_{f \in \mathcal{H}_K} \sum_{j \in N_s} (f(\mathbf{x}_j) - y_j)^2 + \lambda_s \|f\|^2_{\mathcal{H}_K}$$

Updates labels of local data:

$$\{\mathbf{x}_j, \hat{f}_s(\mathbf{x}_j)\}_{j \in N_s} \rightarrow \{(\mathbf{x}_j, y_j)\}_{j \in N_s}$$

end

A Collaborative Training Algorithm
Intuition (cont'd)

- Need multiple passes + inertia term

Initialize: $\hat{f}_{s,0} = 0 \in \mathcal{H}_K$

for **t=1,…., T**

Iterate over sensors **s = 1,…, m**

sensor **s**

Computes using local data:

$$\hat{f}_{s,t} = \arg\min_{f \in \mathcal{H}_K} \sum_{j \in N_s} (f(\mathbf{x}_j) - y_j)^2 + \lambda_s \|f - f_{s,t-1}\|^2_{\mathcal{H}_K}$$

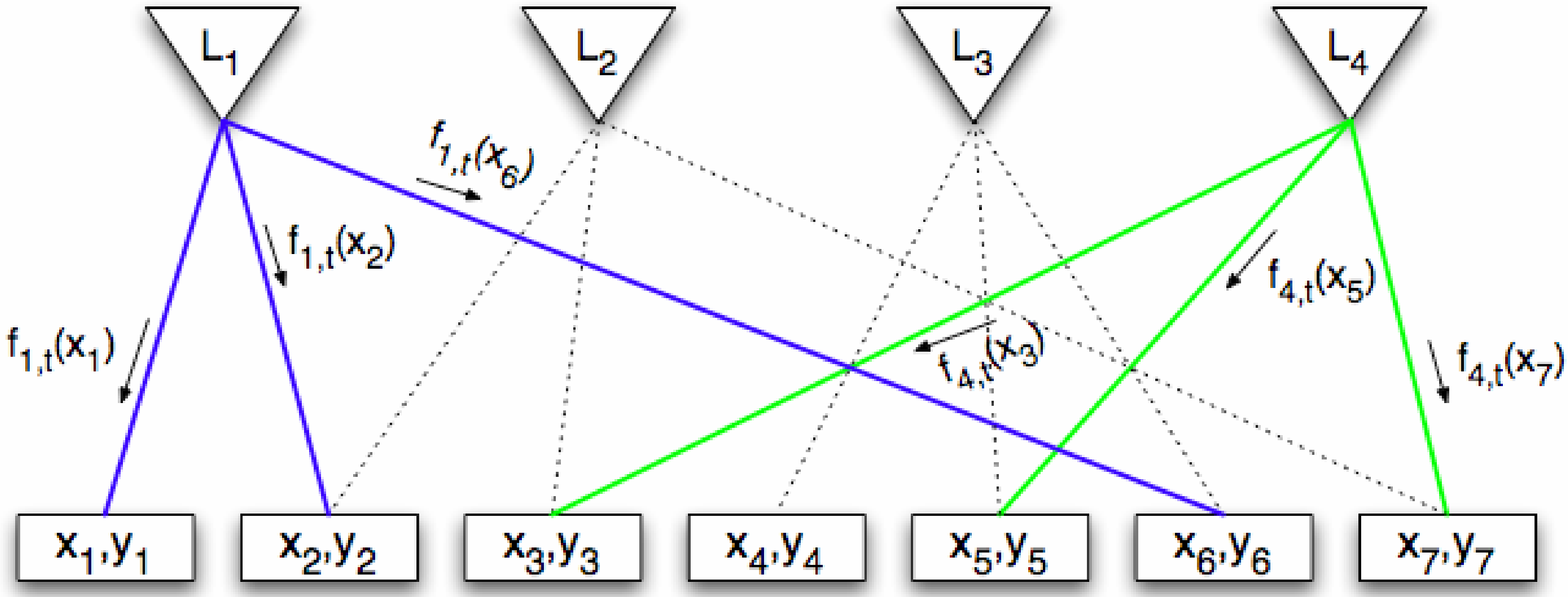
Updates labels of local data:

$$\{\mathbf{x}_j, \hat{f}_{s,t}(\mathbf{x}_j)\}_{j \in N_s} \rightarrow \{(\mathbf{x}_j, y_j)\}_{j \in N_s}$$

Collaborative Regression

A Collaborative Training Algorithm
Intuition (cont'd)

$$f_{1,t} = \arg\min_{f \in \mathcal{H}_K} \sum_{j \in \{1,2,6\}} (f(\mathbf{x}_j) - y_j)^2 + \lambda_1 \|f - f_{1,t-1}\|^2_{\mathcal{H}_K}$$



$$f_{4,t} = \arg\min_{f \in \mathcal{H}_K} \sum_{j \in \{3,5,7\}} (f(\mathbf{x}_j) - y_j)^2 + \lambda_4 \|f - f_{4,t-1}\|^2_{\mathcal{H}_K}$$

Collaborative Regression

Define

$$\Delta_s(f,g) = \sum_{j \in N_s} w_j (f(\mathbf{x}_j) - g(\mathbf{x}_j))^2 + \lambda_s \|f - g\|^2_{\mathcal{H}_K}$$

Theorem:

1. $\hat{f}_{s,T}$ converges (in norm) to a "relaxation" of centralized estimate.
2. $\{\lim_{T\to\infty} \hat{f}_{s,T}\}_{s=1}^m$ is locally coherent and satisfies

$$\lim_{T\to\infty} \hat{f}_{s,T} \in \mathrm{span}\{K(\cdot,\mathbf{x}_j)\}_{j \in N_s}$$

3. Moreover, $\{\hat{f}_{s,t}\}_{s=1}^m$ "improves" with every update

$$\frac{1}{m}\sum_{s=1}^m \Delta_s(\hat{f}_{s,t}, f_\star) \le \frac{1}{m}\sum_{s=1}^m \Delta_s(\hat{f}_{s,t-1}, f_\star)$$

Other Observations

- *Each sensor **locally** computes a **global** estimate.*

- ***Computation**: Sensor i solves $|\mathbf{N}_i|$ x $|\mathbf{N}_i|$ linear system*
- ***Storage**: $O(|\mathbf{N}_i|^2)$ (coefficients + kernel matrix)*
- ***Parallelism**: Non-sequential updates possible*
- ***Robustness** to ordering*

Why?

The Algorithm Can Be Derived in 3 Steps

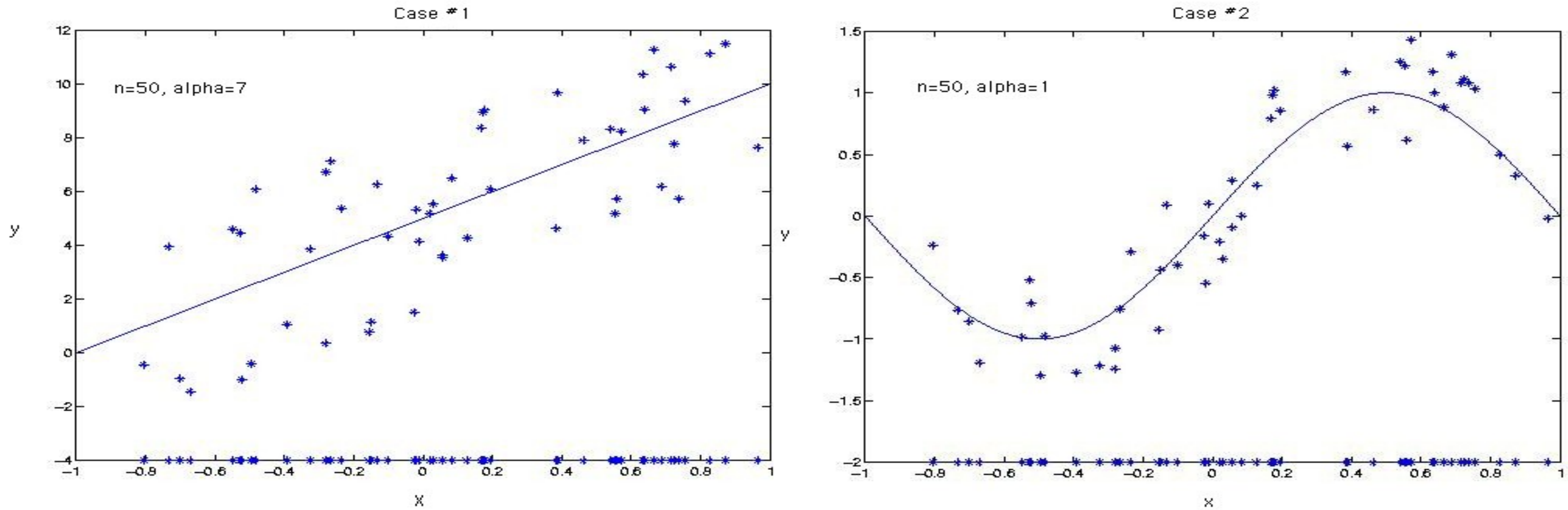
- ***Step 1:*** ***Interpret*** *classical (centralized) estimator as projection onto a Hilbert space.*

- ***Step 2:*** **Relax** *projection by "respecting" network topology*

- ***Step 3:*** ***Alternating projection algorithms*** *imply the distributed training algorithm.*

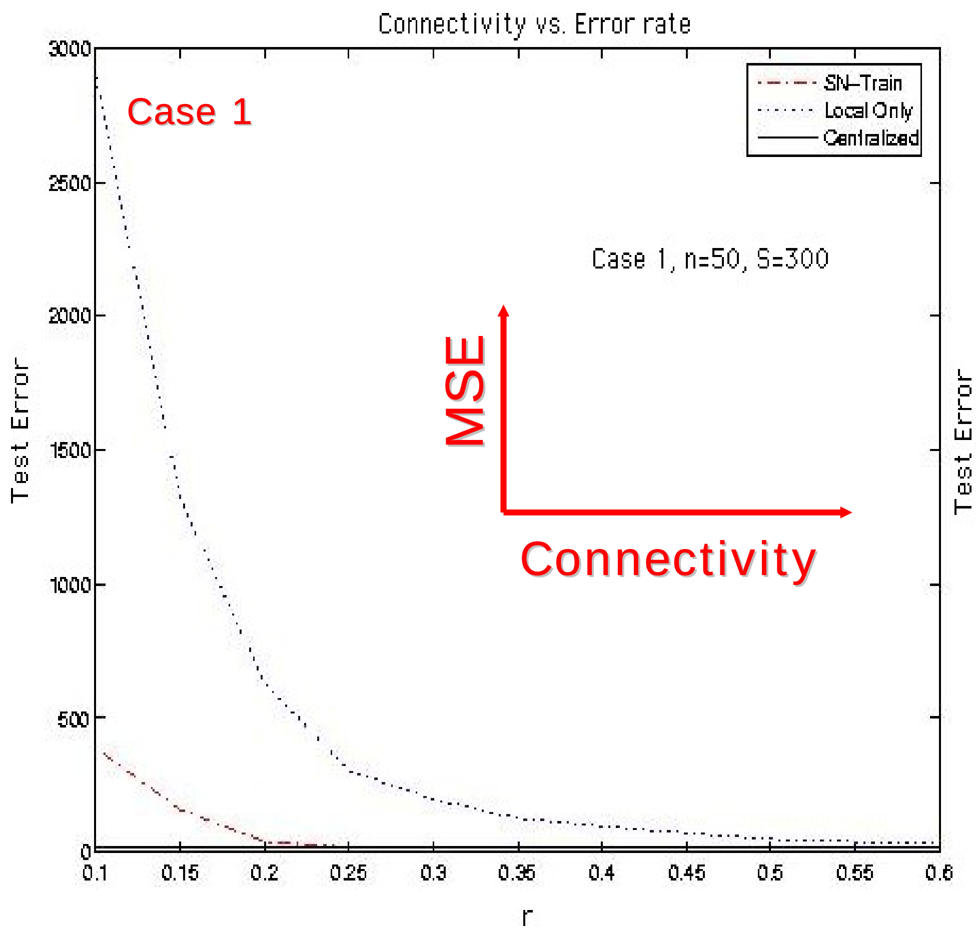
Experiments

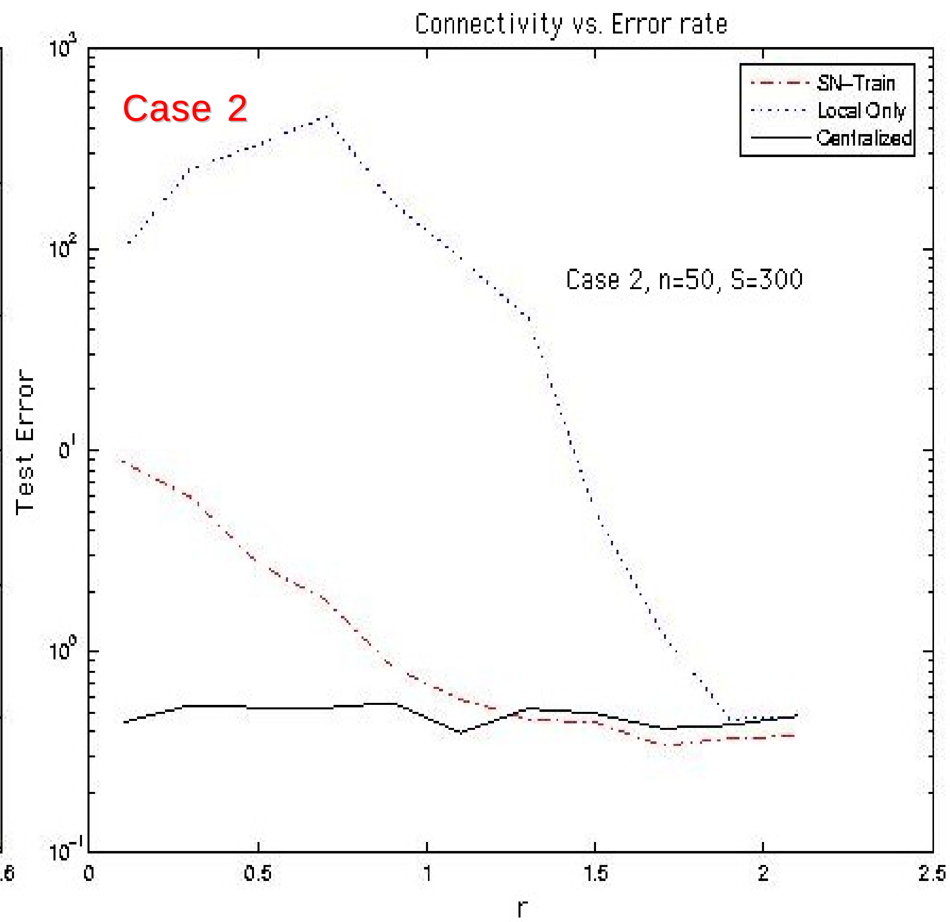
- n=50 sensors uniformly distributed about [-1, 1]
- Sensor i observes $y_i = \mathbf{f}(x_i) + n_i$
  - $\{n_i\}$ is i.i.d. standard normal
  - regression function **f** is linear (Case 1) or sinusoidal (Case 2)
- Sensors i and j are neighbors iff $|x_i - x_j| < r$
- Sensors employ linear (Case 1) or Gaussian (Case 2) kernel



Case 1 Case 2

How does collaboration effect generalization error?

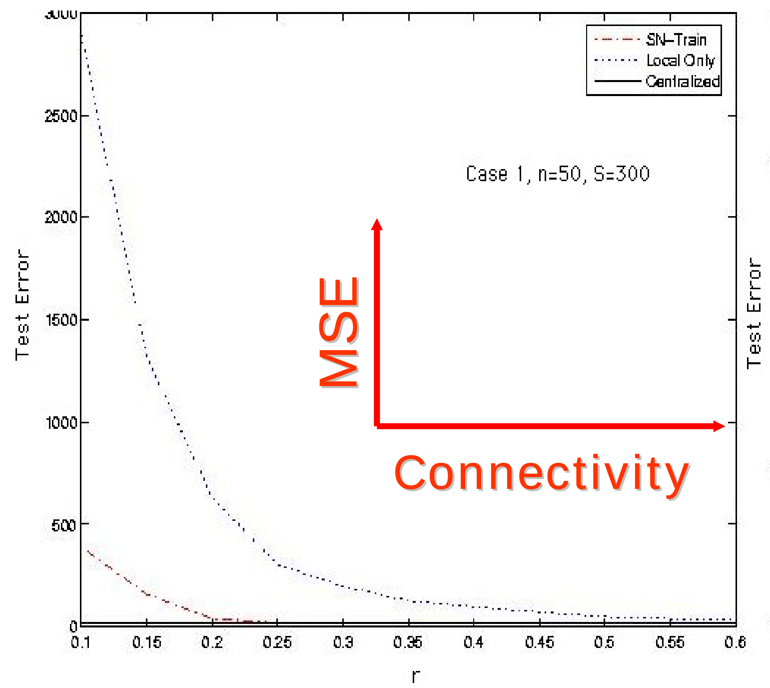




Collaborative Regression

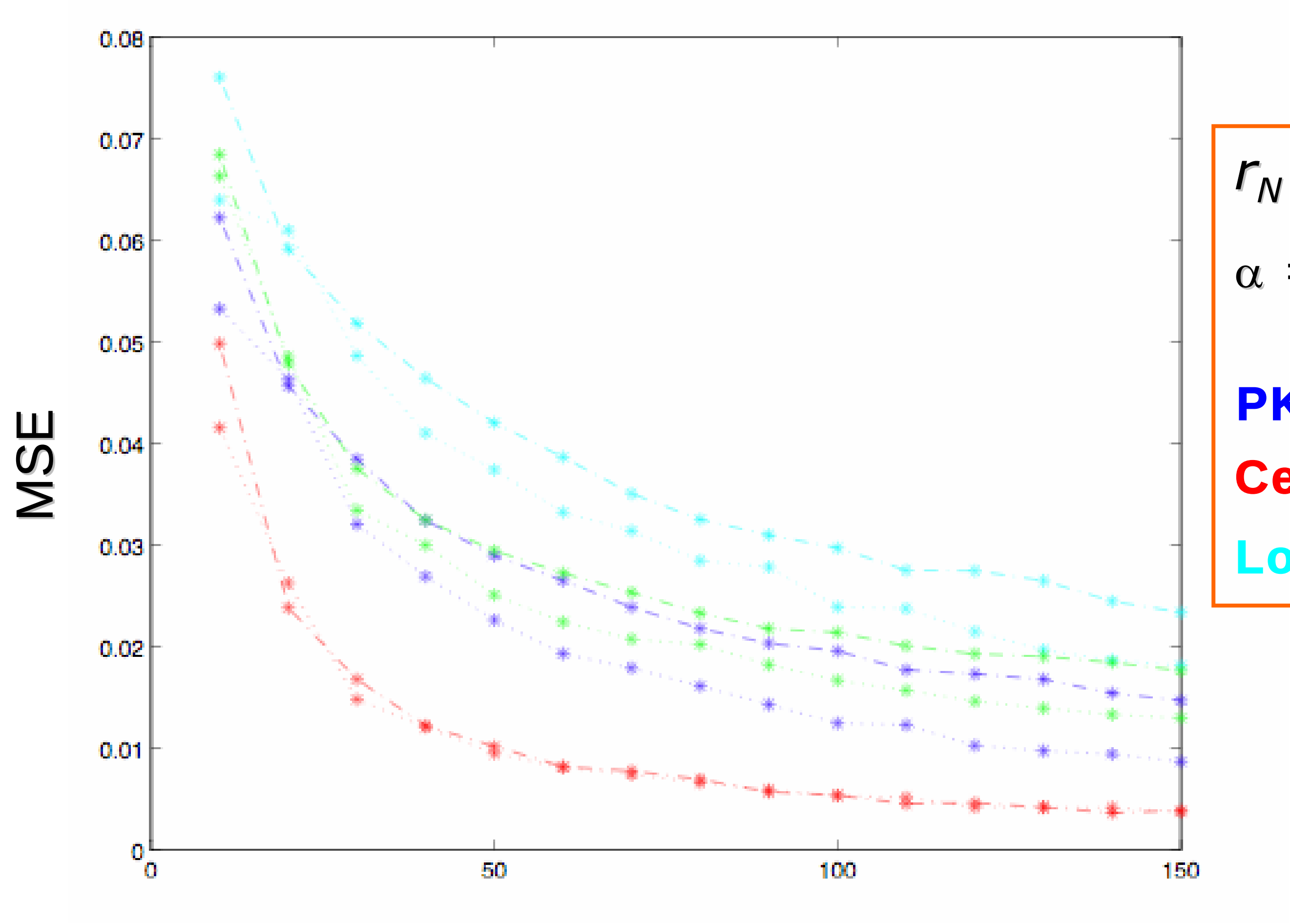
Energy Efficiency

- Overall error decreases with size of the neighborhoods.
- But, energy consumed by message-passing increases with neighborhood size.



- Question: What are the trade-offs?

Mean-Square Error vs. N



N (number of sensors)

$r_N = N^{\alpha}$

$\alpha = \{.30, .35, .40, .45\}$

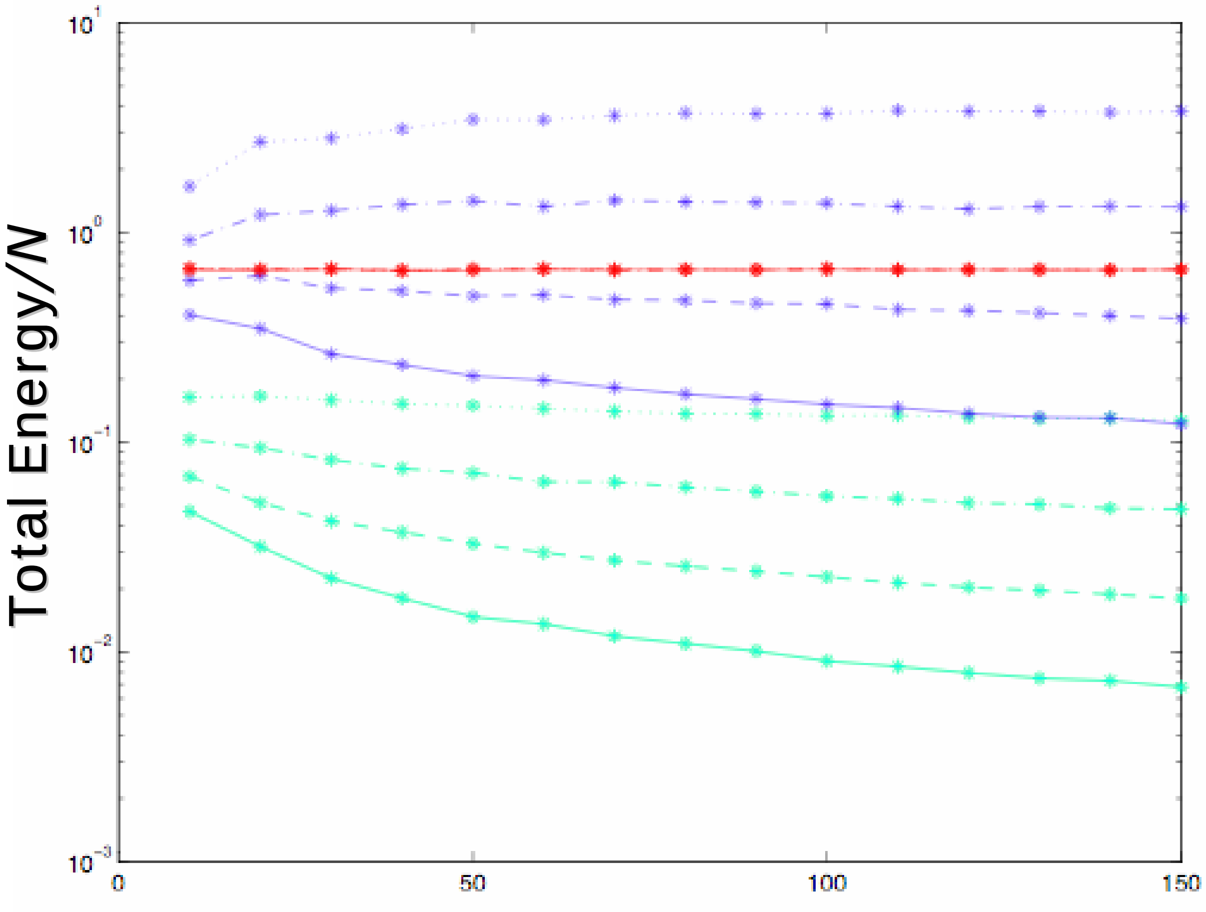
PKP

Centralized

Local Averaging

Collaborative Regression

Energy-per-Sensor vs. N



N (number of sensors)

$r_N = N^\alpha$

$\alpha = \{.30, .35, .40, .45\}$

PKP

Centralized

Local Averaging

Summary/Conclusions

- Model and algorithm for collaborative regression
- Convergence properties
- Cautionary note on energy
- Rich area for further work

Today's Talk: Three Topics
A Sampling of Ideas

- Energy Efficiency in Multiple-Access Networks ✓
- Diversity-Multiplexing Tradeoffs in MIMO Systems ✓
- Distributed Inference in Wireless Sensor Networks ✓

Thank You!