PHY-MAC Cross-layer: multiuser systems

Table of contents

1. Introduction to cross-layer Design

- Introduction
    - The wireless medium
    - Rational for cross-layer
    - An example: PHY-MAC design for an ad-hoc network
    - Cautionary cross-layer

- Multiaccess communications
    - Multiuser detection theory
    - Multiuser information theory
    - Communication networks: Interaction between physical and higher network layer
    - Case 1: SDMA for downlink
    - Case 2: CDMA multipacket acces in ad-hoc networks

- Tools for cross-layer
    - Modeling
    - Tools

Introduction: The wireless medium

- The success of the artificial layered architecture for wired networks has had a great impact on network design paradigms, but, is it appropriate for wireless**?**

- Different wireless networks have appeared: cellular, P2P, multihop, adhoc..

- The wireless medium is an interference-limited medium: collision redefinition in accessing. Multiaccess channel can increase throughput considerably in bursty packet transmission

- Multihop routing involving relays and power control in order to maximize capacity by spatial reuse of the resources (e.g. spectral frequencies)

- Two types of services: guaranteed service and best effort service (e-mail)

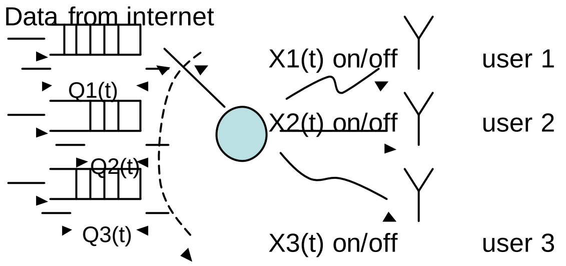
Introduction: Rational for cross-layer

- The ultimate goal is to match the instantaneous radio channel conditions and capacity needs with the traffic and congestion conditions: scarce radio resources and limited power can be optimized.

- As examples of PHY-MAC joint design in standards: In 3GPP, the enhancement High Speed Downlink Packet Access (HSDPA), BRAN HiperLAN2, IEEE802.11x protocols

- Motivating examples:
  - The current deployment of TCP protocol interprets all losses as being congestion related and thus decreases the packet transmission rate. A solution is link layer ARQ at a faster time-scale than that of TCP control loop
  - Data over Wireless in a Multi-user context: Channel state dependent scheduling algorithms

---



Data from internet

Q1(t)

Q2(t)

Q3(t)

X1(t) on/off

X2(t) on/off

X3(t) on/off

user 1

user 2

user 3

Multiuser diversity allows in this case to go from 1/6 pack/slot with round robin TDM to 7/24 pack/slot

How to design on-line algorithms? Support of divers QoS?

An example: PHY-MAC-Network design for an ad-hoc network

Ad-hoc networks

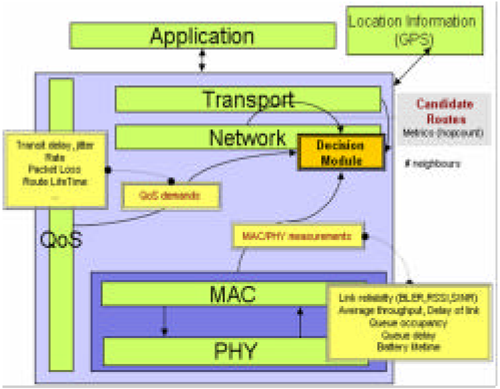
Desirable features are:

- self- configuration

CROSS-LAYER

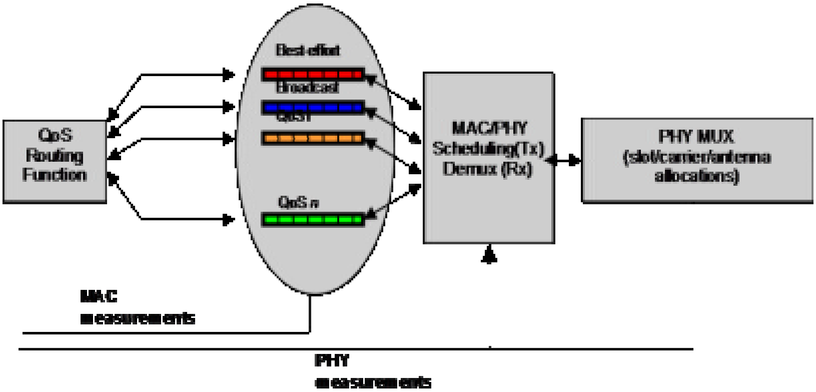
DESIGN

- self organisation
- power conciousness



PHY-MAC architectural design

An efficient MAC-PHY must take into account QoS requirements from applications as well as typical traffic scenarios.



Cross-layer signalling needs

An example: PHY-MAC-Network

MAC → Netw
- Link control
  - Fragmentation
  - Link adaptation
- MAC/PHY measure
  - Reliability (BER, FER, RSSI, SNIR)
  - Quality (throughput, delay, jitter)
  - Scheduling (queuing, delay, traffic)
  - Power capacity

MAC signalling
- Between layers: rx. Info, diversity, up/down CSI
- Between terminals: up/down CSI

QoS on MAC-PHY
- Traffic class
- Contention free/non-free
  - Centralized/Decentralized
  - Broadcast/Unicast/Multicast
- MultiAccess comm.
- Timing synchronization

On MAC performance
- Mobility
- Netw. Connectivity
- Traffic load and its spatial location
- Asymmetric links

PHY
- Radio freq. And regulation
- Channelization
- Power levels and emission masks
- Reconfigurability
- High speed efficiency

Introduction: Cautionary cross-layer

- The cooperation should be cautious though, because cross-layer design can lead to adverse unintended interactions, and adequate care is necessary not to forget the importance of a good architectural design

- Cross-layer design can create "loops" and stability can be broken

- Robustness is also an important issue

A little bit of history

- 1964 Kleinrock popularized in his Ph.D the OSI architecture
- 1973 Gallager connected networking and information theory: protocol surplus
- 1970 Abramson proposed the ALOHA protocol
- 1977 Gallager presented the routing problem (minimum delay in a packet switch) in a convex optimization way

Introduction: Multiaccess communications

•The main problem is the contention among the sources and the need to share the channel resource.

•Coding, detection, good channel modeling, source burstiness, and delay are all important issues. Multiuser information theory and multiuser detection theory focus on the first three. In the terminology of layers, the two latter topics are more at a multiaccess (MAC) layer or network layer (where packets are dealt with as “black boxes”), and the other topics are more at the physical layer. However, theh PHY and MAC cannot be cleanly separated, so that the areas of:

- multiuser detection or signal processing,
- multiuser information theory, and
- multiaccess networking issues

are best understood or developed in concert.

Table of contents

2. An insight into fairness at the Physical Layer

1.- Introduction: PHY layer views on fairness

-Fairness explicit in the cost function that solves a given resource allocation problem

- Fairness implicit in techniques without detailed attention of the designer

- Fairness with the addition of some QoS requirements to a resource allocation problem.

2.- The illustration of the fairness

3.- Example: Multi-antenna transmit processing

---

1. Introduction

The implementation of any resource sharing is different in wired or wireless networks

A. In explicit cost functions

1) Proportional fairness: This criterion is usually the preferred one in signal processing or information theory, because it yields the optimization of the global performance, see e.g*:* the sum rate maximization of the Gaussian MIMO Broadcast channel. Mathematically, a proportional fair allocation is such that the optimization problem can be expressed as

$$\max_{r_k} \sum_{k=1}^{K} g_k(r_k)$$

which incurs in a performance penalty caused to the worst user. In other words, this scheme gives advantage to the users with better channel conditions

$$\max_{r_k} \sum_{k=1}^{K} \frac{g_k(r_k) - g_k^*(r_k)}{g_k^*(r_k)} \leq 0.$$

However, it shall be noted that there are mechanisms to overcome this problem, e.g. the satisfaction obtained with the instantaneous resource can be divided by the average of the resource for the *k*th user obtained until a certain time spot.

2) Max-min fairness:

$$\max_{r_k} \min_{k} g_k(r_k)$$

3) Weighted fairness:

A Pareto-optimal point is such that no user can increase its own utility without decreasing the utility of some other user. Certainly, there is an intermediate point between the two extremes of the fairness axis (Pareto-optimal points) that have been described up to this point. In fact, both the proportional and the max-min fairness can be described using the weighted fairness

$$\max_{r_k} \sum_{k=1}^{K} c_k g_k(r_k) \qquad \max_{r_k} \min_{k} c_k g_k(r_k)$$

As a global picture, weighted fairness can be seen as the equilibrium point between the extreme positions, namely the max-min fairness and the proportional fairness. In fact, the weighted fairness might include all the intermediate operating points between the total worry about the global performance and the stringent concern on the individual needs.

Since the fairness criterion is totally subjective, it is still unclear which is the best option from a system point of view. Depending on the burstiness of the traffic, the number of users, the time scale in the system, etc, the scheduling procedure at the assigning entity might select one option or another. This choice (among others) determines the overall system (and service) performance, thus the price consumers are willing to pay. The selection of the best criterion is still an open issue.

B. In implicit cost functions

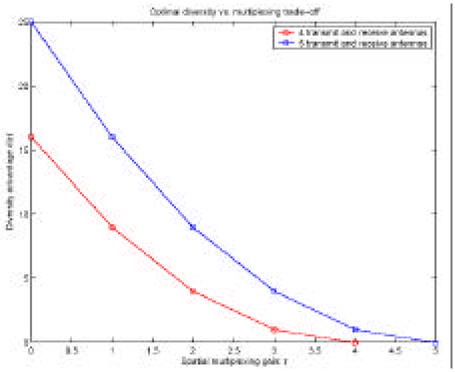
Besides fairness in explicit cost functions, it can also be found implicit in any resource allocation, which might not necessarily be described by a cost function. A clear example of that at the physical layer is the multi-antenna transmit processing case. In that paper, three well-known multi-antenna strategies for a multi-user MISO scenario are analyzed in terms of fairness. But from the study, it can be confirmed that a higher mean performance comes at the expense of an *unfair* distribution of the resources (a higher variance). This result is very interesting since it is made clear that fairness can appear an any stage of the definition of the system, without noticing, so the designer should carefully take it into account.

C. With minimum QoS requirements

Fairness also appears whenever there is a constraint on the performance metric for the users,e.g*:* in terms of rate or BER, which will impose some special characteristics in the performance of the system. Two examples of which are the capacity region and the diversity vs*:* multiplexing trade-off.

The **capacity region** can be described as the feasible combination of rates for a set of users in a system. Any operation point of the system is found within that region, since it shows the physical limits of the data rate transmission. Certainly, this is a physical limit, but these regions might be modified when a constraint on the rate is added. In general, the feasible region will be limited to a smaller area disregarding the operation points that are below the specified rate threshold. As a remark, the constraint could also be expressed according to any other metric denoting the QoS.

The **diversity-multiplexing trade-off** reflects the trade-off between the error probability and the data rate of a system, and it is shown in a practical scenario in The curves in the figure refer to the cases with 4 and 5 transmit and receive antennas, so that any practical strategy is contained in the region between them and the axes. Similarly to the capacity region, these curves also show the theoretical limits in the operation of the system, and any working point might be in the region between the curve and the axes.



Again, some minimum QoS requirements could be added into the system, so that the feasible region of operation is reduced. It is important to note that when multiple users come into play, the multiplexing gain should be seen as a multiuser gain, because the streams would be then intended for different users: the link efficiency is transformed into a network efficiency.

2. The illustration of the fairness

Now, a more deep review is made, which will prove that communications can benefit from the extension of some economical concepts that have been used in resource distribution for a long time.

A. The index of fairness

$$IF = \frac{\left(\sum_k r_k\right)^2}{K \sum_k r_k^2}$$

A clear drawback of this type of indices is that they only measure relative performance, thus only serve to compare allocations, but not to design a system.

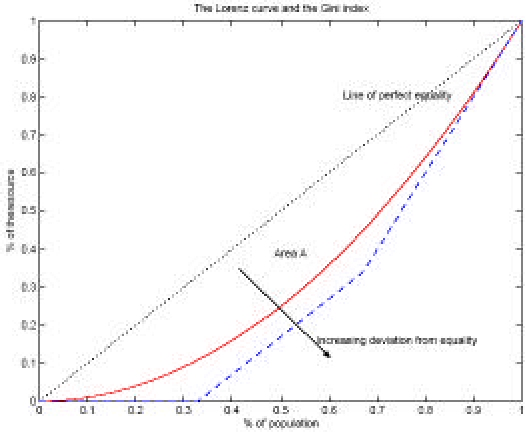
Clearly, the fairness index depends on the distribution of the resource *rk*, but does not provide an idea on how these resources are shared.

B. The Gini index as a measure of inequality

Essentially, the Gini index measures the degree of fairness/unfairness of a resource allocation $g_k(r_k)$ for the *K* users such that $g_1(r_1) \leq g_2(r_2) \leq \cdots \leq g_K(r_k)$.

To illustrate what the Gini index means, in the Figure the percentage of the resource is plotted as a function of the percentage of the population. If there is perfect equality in these quantiti $g_i(r_i) = g_j(r_j), \forall i, j.$ the Lorentz curve in the Figure will be the 45-degree line starting at the origin, i.e*:* for any percentage of the population (users) the resource is shared equally among all of them. This might be the distribution that is socially the most fair. As it is depicted with the continuous red line and the dashed line, other Lorentz curves might exist within this unit box.

The area between the perfect equality and any other Lorentz curve (area A in Figure) corresponds to the Gini index.



Note that this index is one of the most used indicators of social and economic conditions, since the more area among the curves, the more concentrated is the wealth.

One drawback of this method when applied to communications is that the degree of inequality disregards the fact that two distributions might have the same mean.

Another possible drawback of this method is that the Gini index might be difficult to compute at the physical layer

C. The mean vs: variance trade-off

To overcome the inherent problems of the IF or the Gini index, which measure only relative performance because they are scalar values and might not be sufficient to interpret fairness, a 2-D plot of two parameters might be well-suited.

A plot showing the individual behavior (denoted by the variance) and the global outcome (expressed by the mean) seems to be a good alternative.

---

3. An example

A Unified Fairness Framework in Multi-Antenna **Multi-User** Channels

From ICECS'04 (Tel Aviv, December 2004), Diego Bartolome and Ana Perez

- Introduction
- Problem statement
- Fairness analysis: Multi-antenna Multi-user techniques
- Simulations
- Conclusions

Introduction

- Fairness is usually studied at higher layers
- Trade-off (global performance vs. individual needs) at the physical layer
  - Max sum implies high differences among individuals
  - Max min penalizes global performance

- **Observed fact:**
  - **A better mean value comes at the expense of an uneven distribution of the resources**
- **OBJECTIVE: Measurement of this trade-off**

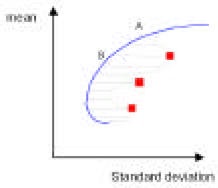
Introduction

- Traditional index of fairness is not sufficient

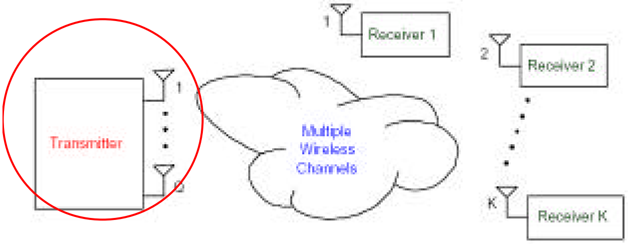
$$IF = \frac{\left(\sum_k r_k\right)^2}{K \sum_k r_k^2}$$

  - It measures only a relative performance
  - It might mask the effects of a better technique
- This ratio appears in many situations, for example in spectral estimation
- OBJECTIVE: Find how to show fairness

Particular Issues



- Inspired by portfolio selection
  - Higher risk -> higher expected profit
- Particularized framework:
  - Multi-antenna broadcast channel
  - Zero Forcing, Dirty Paper, Cooperative



Problem Statement

- K Single antenna terminals
  - Not higher than number of antennas (Q)
- Signal model assuming a tx-rx processing

$$\mathbf{y} = \mathbf{AHBs} + \mathbf{w} \in \mathbb{C}^{K \times 1}$$

- Instantaneous power constraint

$$tr(\mathbf{B}^H\mathbf{B}) \leq P_T$$

  - Adaptation of the (unitary) beamforming
  - A Uniform Power Allocation is assumed in this paper
- Three well-known techniques are analyzed
  - Cooperative, Zero Forcing, Dirty Paper

Mean vs. variance analysis

- Signal Model
  - Cooperative $y_k^{CO} = \lambda_k \sqrt{p_k} s_k + w_k$ transmit-receive processing
  - Zero Forcing $y_k^{ZF} = \alpha_k \sqrt{p_k} s_k + w_k$
  - Dirty Paper $y_k^{DP} = d_k \sqrt{p_k} s_k + w_k$ interference pre-subtraction
- Mean and variance analysis
  - Cooperative $E_{\mathbf{H},K}(l_k^2) = Q/K$
  - Zero Forcing $E_{\mathbf{H},K}(a_k^2) = (Q-K+1)/K$
  - Dirty Paper $E_{\mathbf{H},K}(d_k^2) = (2Q-K+1)/2K$

variance comes next

Mean and variance analysis

- **Cooperative**

$$s_{\mathbf{H},k}(l_k^2) = \sqrt{Q/K}$$

- **Zero Forcing**

$$s_{\mathbf{H},k}(a_k^2) = \frac{\sqrt{Q-K+1}}{K}$$

- **Dirty Paper**

$$s_{\mathbf{H},k}(d_k^2) = \frac{\sqrt{Q+\frac{1}{12}(K-5)(K-1)}}{K}$$

Traditional Index of Fairness (IF)

- Signal Model
  - Cooperative

$$IF^{CO} = \frac{\mathrm{E}^2_{\mathbf{H},K}(\lambda_k^2/K)}{\mathrm{E}^2_{\mathbf{H},K}(\lambda_k^2/K) + \sigma^2_{\mathbf{H},K}(\lambda_k^2/K)} = \frac{Q}{Q+K} = \boxed{\frac{1}{1+\xi}}$$

  - Zero Forcing

$$IF^{ZF} = \frac{\mathrm{E}^2_{\mathbf{H},K}(\alpha_k^2/K)}{\mathrm{E}^2_{\mathbf{H},K}(\alpha_k^2/K) + \sigma^2_{\mathbf{H},K}(\alpha_k^2/K)} = \boxed{\frac{1-\xi+1/Q}{1-\xi+2/Q}}$$

  - Dirty Paper

$$IF^{DP} = \frac{\mathrm{E}^2_{\mathbf{H},K}(d_k^2/K)}{\mathrm{E}^2_{\mathbf{H},K}(d_k^2/K) + \sigma^2_{\mathbf{H},K}(d_k^2/K)} = \boxed{\frac{(2-\xi+1/Q)^2}{(2-\xi+1/Q)^2+4/Q+(\xi-5/Q)(\xi-1/Q)/}}$$

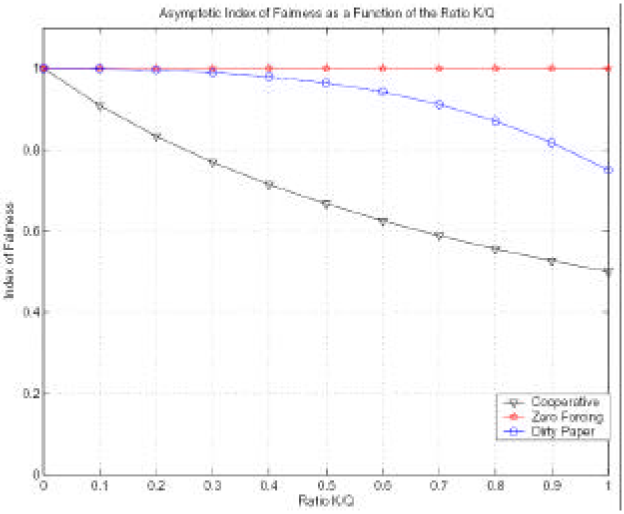
-

◆SUMMARY

Technique	Gain	Mean	Standard Deviation	Asymptotic IF
Cooperative	λ_k^2/K	Q/K	$\sqrt{Q/K}$	$1/(1+\xi)$
Dirty Paper	d_k^2/K	$(2Q-K+1)/2K$	$\sqrt{Q+\frac{1}{12}(K-5)(K-1)}/K$	$(2-\xi)^2/\left[(2-\xi)^2+\xi^2/3\right]$
Zero Forcing	α_k^2/K	$(Q-K+1)/K$	$\sqrt{Q-K+1}/K$	1

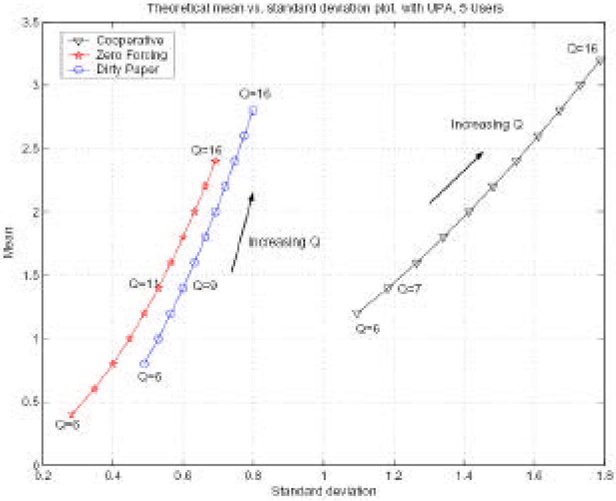
A plot on traditional view …

- Zero Forcing is the most fair
- Dirty paper comes next and behavior is not constant
- The best technique is the most unfair
- How can we detect the best technique?
- The variance among users is also important



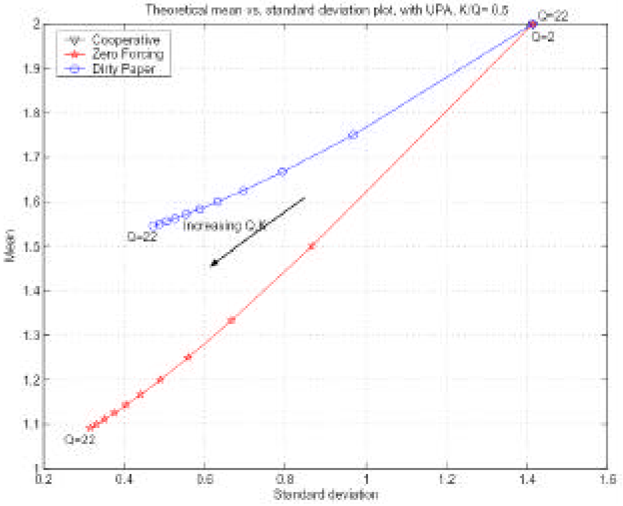
A new plot !

- Both the mean and the variance are shown
- Trade-off can be clearly identified
  - Global performance
  - vs. individual needs
- Example
  - Number of antennas
  - to attain a mean



Another plot

- Very similar to the previous one
- Trade-off can be clearly identified
  - Global performance
  - vs. individual needs
- Example
  - It might not be interesting to use above a certain number of antennas



- **It could be applied to other degrees of freedom in multi-user communications**
  - Power Allocation
  - Bit Allocation (in systems with Link Adaptation)

Power management I

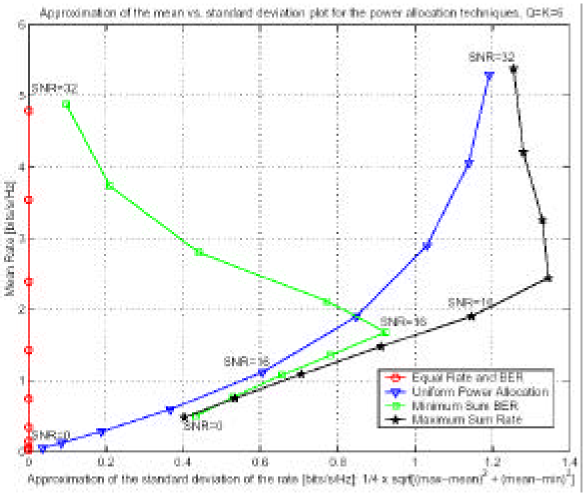
ZF transmitter

Mean vs. Approx of the stdev (rate)

The performance variation and the equivalences are clearer now

Not only the mean is important, also how the resources are distributed among users

Interesting issue: behavior of MSB



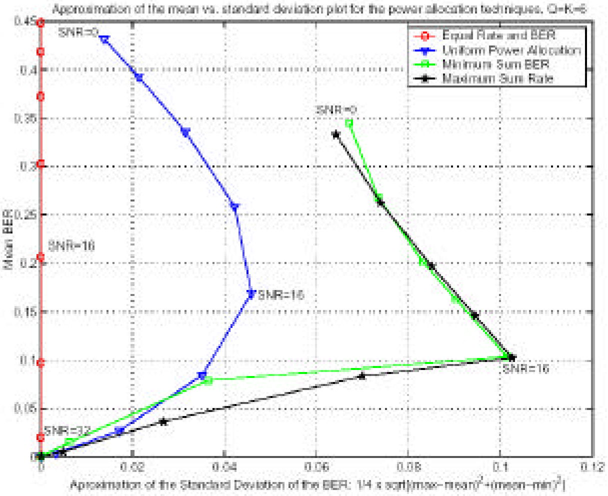
Power management II

Mean vs. Approx of the stdev (BER)

The performance variation and the equivalences are clearer now

As in the previous figure, the ERB is always a vertical line (equal behavior for all the users)

Good behavior of UPA

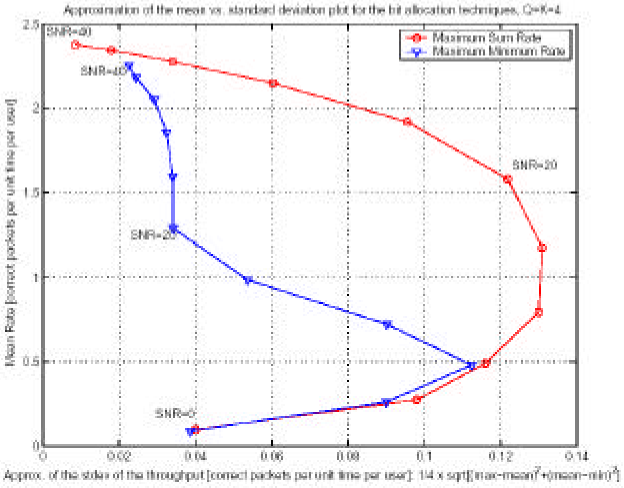


Bit allocation

Mean vs. Stdev for the bit allocation techniques

it is important to note that some users might be left out for the sake of the collective revenue

more mean comes at the expense of a higher variance



Conclusions

- A new view on fairness has been proposed
  - It evaluates the mean vs. the variance in a plot
    - Inspired by portfolio selection
  - It refines the traditional views
  - It is useful for the system design
    - Fairness is not only explicit in cost functions, but also implicit in the multi-antenna techniques
  - It could be applied to analyze other degrees of freedom in multi-user communications