



a||e **CTTC**
Centre
Tecnològic
de Telecomunicacions
de Catalunya

innovating communications

CAPACITY AND SCHEDULING IN MULTI-USER MIMO SYSTEMS

A cross-layer approach

Carles Antón-Haro, PhD

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)

ACE Course on *MIMO Communication Systems and Antennas*, KTH, Stockholm, Sept 5-9, 2005.

REFERENCES

The sources of this seminar are mainly:

- T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, John Wiley & Sons, New York, 1991.
- D. Tse, P. Wiswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005. Chaps 5,6,10.
- A. Goldsmith *et al*, *Capacity Limits of MIMO Channels*, IEEE Trans. on Selected Areas in Communications, Vol. 21, No. 5, June 2003.
- H. Boche and M. Wiczanowski, *Queueing Theoretic Optimal Scheduling for Multiple Input Multiple Output Multiple Access Channel*, ISSPIT 2003, Darmstadt, Germany.

MOTIVATION

- ❑ Best way for multiple users to transmit over a shared medium? Orthogonal access? Simultaneous?
- ❑ Differences between uplink (multiple access) and downlink (broadcast) channels?
- ❑ Impact of multiple transmit and/or multiple receive antennas?
- ❑ In multi-user systems, can we take advantage of fading?
- ❑ Can the scheduling process be enhanced with channel-related information?
- ❑ Combined use of queue and channel information for scheduling?
- ❑ Information theory approach: keep it general !!

3

OUTLINE

- ❑ Motivation
- ❑ A review of capacity issues in single-user systems
 - ✓ Definition, Capacity for MIMO systems.
- ❑ Capacity issues in multi-user systems:
 - ✓ Broadcast (BC) and Multiple Access (MAC) channels.
 - ✓ Capacity regions for SISO BC & MAC. Sum capacity. Symmetric capacity.
 - ✓ Multi-user diversity. Channel-aware scheduling.
 - ✓ Fairness issues: Proportional Fair Scheduling
 - ✓ Slow-fading channels: Opportunistic Beamforming
- ❑ Channel- and queue-aware scheduling
 - ✓ Motivation
- ❑ Q&A

4

OUTLINE

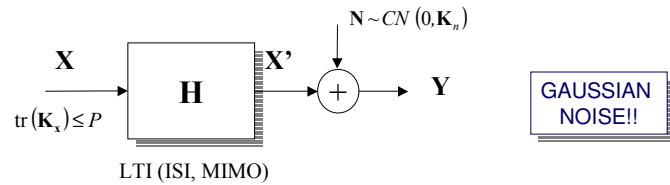
- Motivation
- **A review of capacity issues in single-user systems**
 - ✓ Definition, Capacity for MIMO systems.

5

A REVIEW OF CAPACITY ISSUES IN SINGLE-USER SYSTEMS

6

CAPACITY IN LINEAR TIME INVARIANT SYSTEMS



- Definition of mutual information

$$I(X; Y) = h(X) - h(X|Y) \quad \text{with} \quad h(X) = E_x(-\log f_x(x))$$

$$h(X|Y) = E_{xy}(-\log f_{x|y}(x|y))$$

- Information capacity of an AWGN channel with power constraint P :

$$C = \max_{f_x(x)} I(X; Y)$$

$$\text{s.t. } \text{tr}(\mathbf{K}_x) \leq P$$

- Mutual information maximized for GAUSSIAN input:

$$\mathbf{X} \sim \mathcal{CN}(0, \mathbf{K}_x)$$

7

CAPACITY IN LINEAR TIME INVARIANT SYSTEMS

- In these conditions, maximizing mutual information amounts to:

$$C = \max_{f_x(x)} I(X; Y) = \max_{\mathbf{K}_x} I(X; Y) = \max_{\mathbf{K}_x} \log \frac{|\mathbf{K}_n + \mathbf{H}\mathbf{K}_x\mathbf{H}^H|}{|\mathbf{K}_n|}$$

Noise + interference Signal s.t. $\text{tr}(\mathbf{K}_x) \leq P$

- **Remarks:**

- In general, \mathbf{K}_x depends on \mathbf{H} and what information is available @ Tx side (partial, full, none).
- Units: bits/s/Hz...when $\log = \log_2$
- Interpretation (Shannon's Channel Capacity Theorem): For every data rate R ...

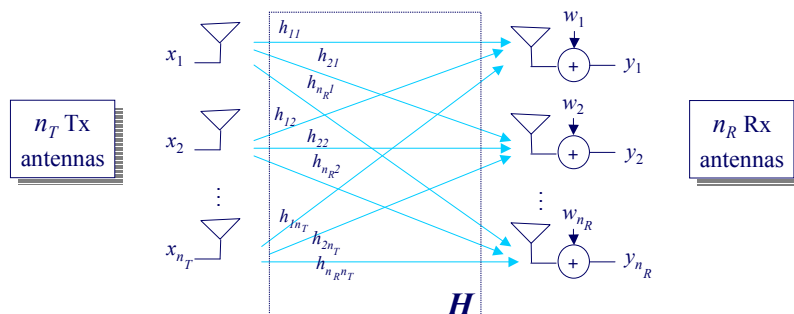
Information capacity (C) provides an upper bound of the achievable data rates (R)

- Assumptions: Gaussian input symbols & ideal channel coding (and decoding)
- Useful equivalence:

$$C = \max_{\mathbf{K}_x} \log \frac{|\mathbf{K}_n + \mathbf{H}\mathbf{K}_x\mathbf{H}^H|}{|\mathbf{K}_n|} = \max_{\mathbf{K}_x} \log |\mathbf{I} + \mathbf{K}_n^{-1}\mathbf{H}\mathbf{K}_x\mathbf{H}^H|$$

8

MIMO CHANNEL MODEL



- Simplest model:

- Channel: Flat fading (frequency), static / independent Rayleigh fading (time)
- Noise: Gaussian (spatially) white $\mathbf{N} \sim CN(0, \mathbf{K}_n) \rightarrow \mathbf{W} \sim CN(0, N_o \mathbf{I}_{n_R})$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_R} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n_T} \\ h_{21} & h_{22} & \cdots & h_{2n_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_R 1} & h_{n_R 2} & \cdots & h_{n_R n_T} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_T} \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n_R} \end{bmatrix} \quad \mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$$

9

CAPACITY OF MIMO SYSTEMS (LTI)

$$C = \max_{\mathbf{K}_x} \log \left| \mathbf{I} + \mathbf{K}_w^{-1} \mathbf{H} \mathbf{K}_x \mathbf{H}^H \right| \quad \mathbf{K}_w = N_o \mathbf{I}_{n_R}$$

- SISO**, Shannon Capacity

$$\mathbf{K}_x = P \quad (\mathbf{K}_w = N_o) \quad \Rightarrow \quad C = \log \left(1 + \frac{P|h|^2}{N_o} \right) = \log(1 + \text{SNR})$$

Asympt growth

LOG in power

i.e. 1 bits/s/Hz
every 3 dB

- MIMO, no CSI at Tx** – Isotropic transmission:

$$\mathbf{K}_x = \frac{P}{n_T} \mathbf{I}_{n_T} \quad \Rightarrow \quad C = \log \left| \mathbf{I} + \frac{P}{N_o n_T} \mathbf{H} \mathbf{H}^H \right|$$

$$C \approx n \log \frac{P}{N_o n_T} + \sum_{i=1}^n \log \lambda_i^2$$

Asympt growth

LOG in power
LIN in antennas

i.e. n bits/s/Hz
every 3 dB

- MIMO, full CSI at Tx** – Waterfilling over channel eigenmodes (SVD):

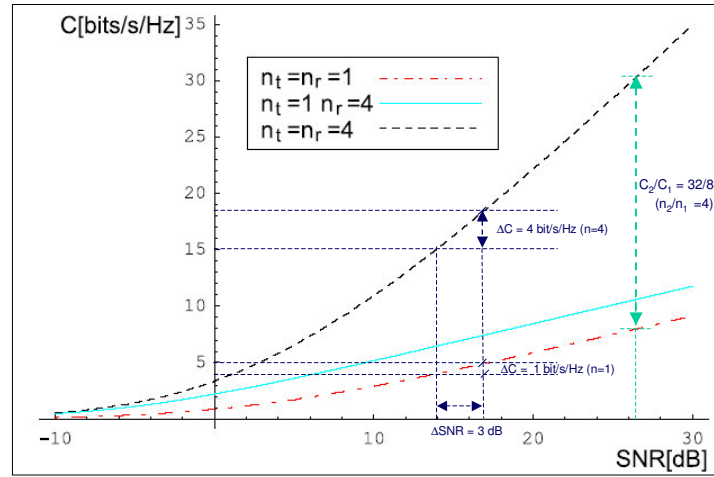
$$\mathbf{K}_x = \mathbf{V} \text{diag}(P_1 \dots P_n) \mathbf{V}^H \quad \Rightarrow \quad C = \sum_{i=1}^n \log \left[1 + N_o^{-1} \lambda_i^2 P_i \right]$$

$$\mathbf{H} = \mathbf{U} \text{diag}(\lambda_1 \dots \lambda_n) \mathbf{V}^H$$

Power allocation (Lagrange): $P_i(\lambda_i) = \left(\mu - \frac{N_o}{\lambda_i^2} \right)^+$ $i = 1 \dots n$ $\sum_{i=1}^n P_i = P$ $n = \min(n_T, n_R)$

10

CAPACITY OF MIMO SYSTEMS (LTI)



D. Tse, P. Wiswanath, *Fundamentals of Wireless Communications*, Cambridge Univ. Press. 2005

SISO: LOG in power

i.e. 1 bits/s/Hz every 3 dB

MIMO: LOG in power, LIN in antennas

i.e. n bits/s/Hz every 3 dB

11

OUTLINE

- Motivation
- A review of capacity issues in single-user systems
 - ✓ Definition, Capacity for MIMO systems, time-varying systems.
- **Capacity issues in multi-user systems:**
 - ✓ **Broadcast (BC) and Multiple Access (MAC) channels.**
 - ✓ **Capacity regions for SISO BC & MAC. Sum capacity. Symmetric capacity.**
 - ✓ **Multi-user diversity. Channel-aware scheduling.**
 - ✓ **Fairness issues: Proportional Fair Scheduling**
 - ✓ **Slow-fading channels: Opportunistic Beamforming**
 - ✓ **Capacity regions for MIMO BC & MAC. Duality principle.**

12

CAPACITY ISSUES IN MULTI-USER SYSTEMS

13

BROADCAST AND MULTIPLE-ACCESS CHANNELS

Broadcast Channel (BC):

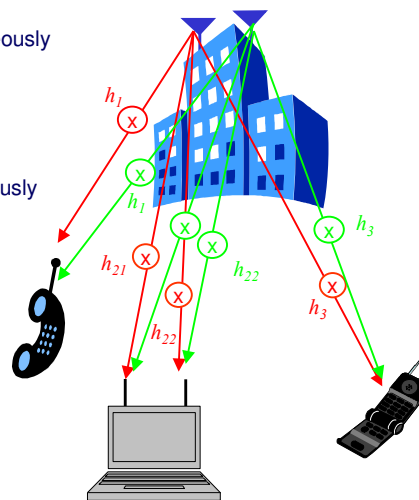
- Downlink
- One transmitter to many receivers simultaneously

Multiple Access Channel (MAC):

- Uplink
- Many transmitters to one receiver simultaneously

Remarks:

- Users can be regarded as an antenna array in a large area.
- Cooperation among antennas within the SAME location.
- Multiple antennas in one location enable Space Division Multiple Access or stream Multiplexing.

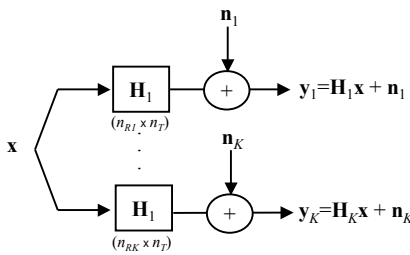


14

MIMO BC and MAC – CHANNEL MODEL

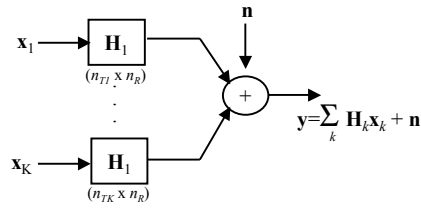
- One base station (BS) equipped with n_T (n_R) antennas
- K user equipments (UE) equipped with n_{Rk} (n_{Tk}) antennas each

BS Shared power constraint



Broadcast Channel (BC)

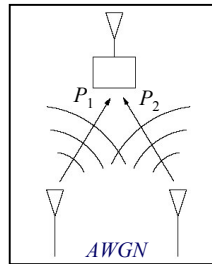
UE individual power constraints



Multiple Access Channel (MAC)

15

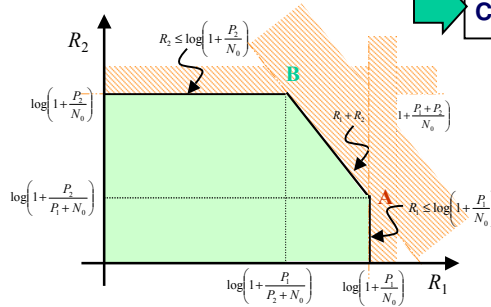
CAPACITY REGION FOR MAC-AWGN



- SISO, MAC, AWGN channel, $K=2$ users:

$$y[m] = x_1[m] + x_2[m] + w[m]$$

- **Single user:** Rate R achievable (with arbitrarily low error rate) iff $R < C \rightarrow C$ upper bound on performance
- **Multi-user:** UEs communicate with BS in a shared bandwidth \rightarrow trade-offs turning up!!
 - Set of achievable rates (R_1, R_2) with simultaneous communication??



CAPACITY REGION, C !!

- Characterizes *optimal* trade-off achievable by *any* MA scheme.
- User 2 gets $R_2 > 0$ while user 1 attains single-user bound (A) !!
- HOW? *Successive interference Cancellation* (SIC).
- Reversing detection order leads to different rate split (B) - fairness

16

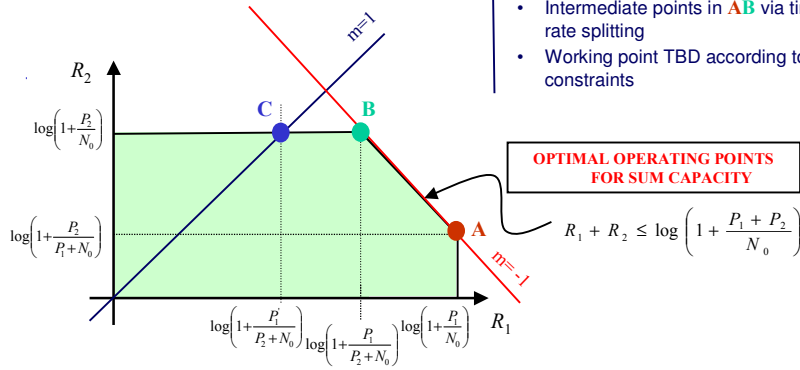
MEASURES OF INTEREST

- Some performance measures (scalars) for a capacity region:

- Sum capacity

$$C_{\text{sum}} := \max_{(R_1, R_2) \in C} R_1 + R_2$$

- Reached at **A****B** segment (ANY point)
- Points **A**, **B** achievable via SIC
- Intermediate points in **A****B** via time sharing or rate splitting
- Working point TBD according to fairness constraints



- Symmetric capacity

$$C_{\text{sym}} := \max_{(R, R) \in C} R$$

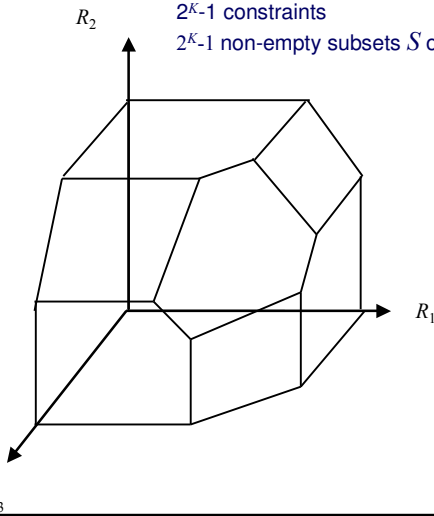
- Reached @ boundary (near/far) - C

17

GENERAL CASE: MAC with K users

$$\sum_{k \in S} R_k \leq \log \left(1 + \frac{\sum_{k \in S} P_k}{N_0} \right)$$

$2^K - 1$ constraints
 $2^K - 1$ non-empty subsets S of users



$$R_1 \leq \log \left(1 + \frac{P_1}{N_0} \right)$$

$$R_2 \leq \log \left(1 + \frac{P_2}{N_0} \right)$$

$$R_3 \leq \log \left(1 + \frac{P_3}{N_0} \right)$$

$$R_1 + R_2 \leq \log \left(1 + \frac{P_1 + P_2}{N_0} \right)$$

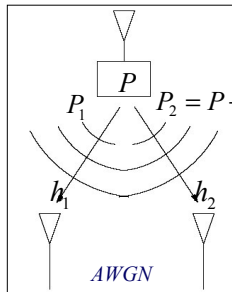
$$R_2 + R_3 \leq \log \left(1 + \frac{P_2 + P_3}{N_0} \right)$$

$$R_1 + R_3 \leq \log \left(1 + \frac{P_1 + P_3}{N_0} \right)$$

$$R_1 + R_2 + R_3 \leq \log \left(1 + \frac{P_1 + P_2 + P_3}{N_0} \right)$$

18

CAPACITY REGION FOR BC-AWGN



- SISO, BC, AWGN channel, $K=2$ users:

$$y_1[m] = h_1 x[m] + w_1[m] \quad y_2[m] = h_2 x[m] + w_2[m]$$

- BS communicates with UE in a shared bandwidth & shared power (P) \rightarrow trade-offs turning up!!

- How to MUX data for both users at the BS? $x[m] = ??$
- Set of achievable rates (R_1, R_2) with simultaneous comms.??

- Assume: User 2 is the "strongest" ($|h_2| \geq |h_1|$) and superposition coding $x[m] = x_1[m] + x_2[m]$
- If x_1 decodable at UE₁ (weakest) in the presence of x_2 , so is at UE₂ (strongest) for all power splits P_1, P_2 (not possible if reversed order)

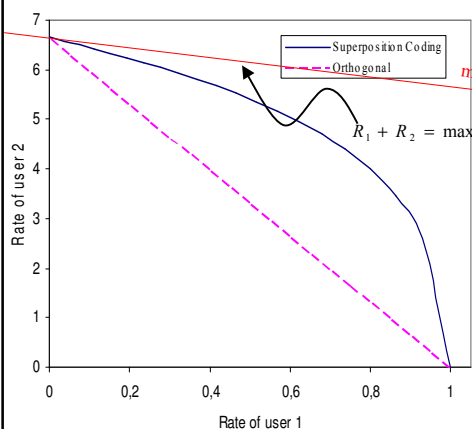
$$\text{SNIR}_{x_1 @ \text{UE}_1} = \frac{P_1 |h_1|^2}{(1-P_1) |h_1|^2 + N_0} \leq \frac{P_1 |h_2|^2}{(1-P_1) |h_2|^2 + N_0} = \text{SNIR}_{x_1 @ \text{UE}_2}$$

- So apply SIC at the strongest (UE₂) and

$$R_1 = \log \left(1 + \frac{P_1 |h_1|^2}{(P-P_1) |h_1|^2 + N_0} \right) \quad R_2 = \log \left(1 + \frac{(P-P_1) |h_2|^2}{N_0} \right)$$

19

CAPACITY REGION FOR BC-LTI (cont'd)



$$R_1 = \alpha \log \left(1 + \frac{P_1 |h_1|^2}{\alpha N_0} \right) \quad R_2 = (1-\alpha) \log \left(1 + \frac{(1-P_1) |h_2|^2}{(1-\alpha) N_0} \right)$$

$$\alpha = \alpha_{\text{opt}} = \frac{P_1}{P_1 + P_2} = \frac{P_1}{P}$$

Orthogonal multiple access

- SC boundary given by all P_1/P_2 splits
- Sum-rate: allocate ALL power to strongest user (UE₂)
...at the expense of delays!!

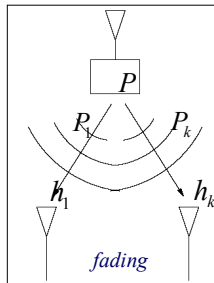
Best policy in BC-SISO: ONE user at a time

(vs. MAC-SISO: ALL users simultaneously)

- Orthogonal multiple access strictly suboptimal for all power splits!!!
 - SC: low power for strong user (UE₂) is efficiently exploited (x_1 removed) and low interference to weaker (UE₁)
- Remarks:
 - Strong assumption: DEGRADED BC
 - MIMO is non degraded.
 - Degradation not needed in UL (centralized Rx & CSI).

20

BC CHANNEL WITH FADING



- SISO, BC, fading channel, K users:

$$y_k[m] = h_k[m]x[m] + w_k[m]$$

- Assumptions:

- Fading processes ($\{h_k[m]\}$): Independent and identically distributed (symmetric case).

- Power constraint (pooled power): $E_H \left[\sum_{k=1}^K P_k[m] \right] = P$

- Take the case with CSIT (i.e. power allocation possible):

- AWGN: Sum capacity maximized by transmitting to the BEST user
- Fading: Schedule the BEST user at EACH time (*greedy* approach). Equivalent point-to-point channel

$$|h|_{\text{eq}}^2 = \max_{k=1..K} |h_k|^2$$

- How to allocate power? Temporal waterfilling for the equivalent P2P channel

$$P^*(\mathbf{h}) = \left(\frac{1}{\lambda} - \frac{N_0}{\max_{k=1..K} |h_k|^2} \right)^+ \quad \Rightarrow \quad C_{\text{sum}} = E_h \left[\log \left(1 + \frac{P^*(\mathbf{h}) (\max_{k=1..K} |h_k|^2)}{N_0} \right) \right]$$

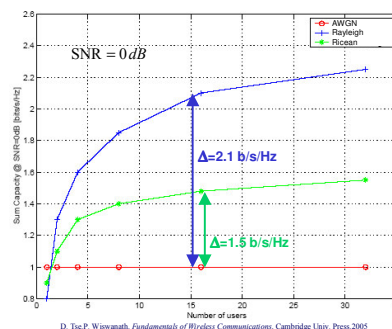
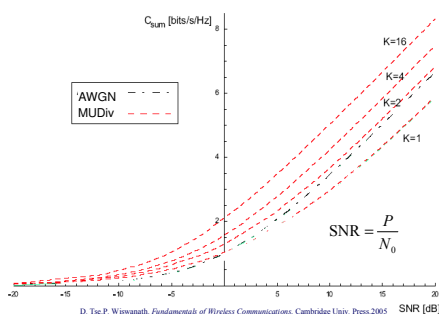
21

MULTI-USER DIVERSITY (MUDiv) GAIN

- With K users *FADING INDEPENDENTLY* and *OPPORTUNISTIC (DYNAMIC) SCHEDULING*, channel gain improves

$$|h_1|^2 \rightarrow |h|_{\text{eq}}^2 = \max_{k=1..K} |h_k|^2$$

Higher gain means higher (sum) rate!!



- Gain wrt AWGN for $K > 1$ (mid-high SNR)
- The amount of MUDiv increases with pdfs' tails: Rayleigh > Rice ($\kappa=5$, LOS, less "random")
- MUDiv gain increases with nr. of users (K): the stronger is the strongest channel

22

MULTI-USER vs. CLASSICAL DIVERSITY

- **Purpose:**
 - Classical (time/frequency/space): Increase link *reliability* (slow fading)
 - MUDiv: Increase *average* cell *throughput* (fast fading)
...but no rate guarantees in *specific* fading states
- **Means:**
 - Classical: *Counteract* adverse fading effects.
 - MUDiv: *Exploit* independent fading (capture strongest user)
- **Scope:**
 - Classical: Works at the *link* level
 - MUDiv: *System-wide* (active users)

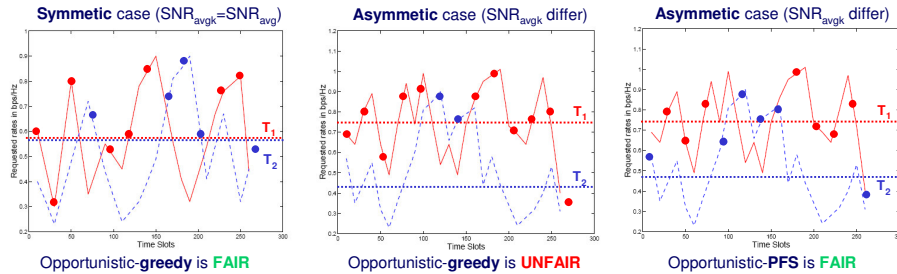
23

REMARKS ON MUDiv

- **Signalling:**
 - UEs: Track their link quality (common pilot)
 - BS: Access to quality measurements (delay-free feedback channel)
- **Delay in the feedback channel** (ass.: delay&error free)
 - Mismatch actual channel-measured channel
 - FIX: ↓ scheduling slots ⇒ ↑ signalling overhead ⇒ selective MUDiv (f/b iff above threshold)
- **Fairness & delay:**
 - Non-homogeneous user set in real-world networks (assumed so far)
 - Different statistics (Rayleigh, Rice,...) average SNRs (near-far).. RESOURCE ALLOCATION ??
 - **FIX: Proportional Fair Scheduler (PFS)**

24

PROPORTIONAL FAIR SCHEDULING (PFS)



- **Proportional Fair Scheduler:** Schedule user with *peak rate with respect to its average rate*

$$k^*[m] = \max_k \frac{R_k[m]}{T_k[m]} \quad T_k[m] = \begin{cases} (1 - 1/t_c)T_k[m] + (1/t_c)R_k[m] & k = k^* \\ (1 - 1/t_c)T_k[m] & k \neq k^* \end{cases}$$

- **PFS vs. greedy** opportunistic schedulers:
 - Both channel-dependent (vs. round-robin, vs. queue-based). PFS implemented in IS-856.
 - Greedy: No *short-term* fairness, captures MUDiv, maximizes *average* sum-rate.
 - PFS: No *short-term* fairness, *long-term* fairness (same # access), captures some MUDiv, loss in average sum-rate.
- Latency time scale (t_c), a design parameter: if larger, larger averaging period, higher latency (schedule when hitting a really high peak)

D. Tse, P. Wornath, *Fundamentals of Wireless Communications*, Cambridge Univ. Press 2005

25

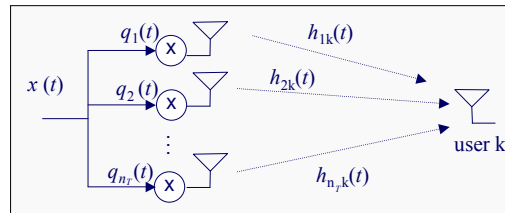
REMARKS ON MUDiv

- **Signalling:**
 - UEs: Track their link quality (common pilot)
 - BS: Access to quality measurements (delay-free feedback channel)
- **Delay in the feedback channel** (ass.: delay & error free)
 - Mismatch actual channel-measured channel
 - FIX: \downarrow scheduling slots $\Rightarrow \uparrow$ signalling overhead \Rightarrow selective MUDiv (f/b iff above threshold)
- **Fairness & delay:**
 - Non-homogeneous user set in real-world networks (assumed so far)
 - Different statistics (Rayleigh, Rice,...) average SNRs (near-far).. RESOURCE ALLOCATION ??
 - FIX: **Proportional Fair Scheduler (PFS)**
- **Limited and slow fluctuations** (ass: high & fast)
 - Limited: poor scattering/LOS – Slow : low mobility environment
 - Result: low cell throughput (peaks) - Delay requirements not met.
 - **FIX: Opportunistic beamforming.**

26

OPPORTUNISTIC BEAMFORMING

- Slow fading hurts: If all users fade slow \Rightarrow like $K=1$ user \Rightarrow no MUDlv
- Limited fluctuation hurts: lower peak rates



$$\mathbf{q}[m] = [q_1[m] \dots q_{n_t}[m]]^T$$

$$\mathbf{h}_k[m] = [h_{1k}[m] \dots h_{n_tk}[m]]^T$$

with

$$\|\mathbf{q}[m]\|^2 = 1$$

- **TRICK (MISO):** Induce fast and high fluctuations by transmit beamforming with a time-varying common set of random weights (e.g circularly symmetric Gaussian):

$$y_k[m] = (\mathbf{h}_k^T[m] \mathbf{q}[m]) x[m] + w_k[m]$$

Random weights

$$\text{SNR}_k[m] = \frac{|\mathbf{h}_k^T[m] \mathbf{q}[m]|^2}{N_0}$$

measure at UE_k
feedback to BS

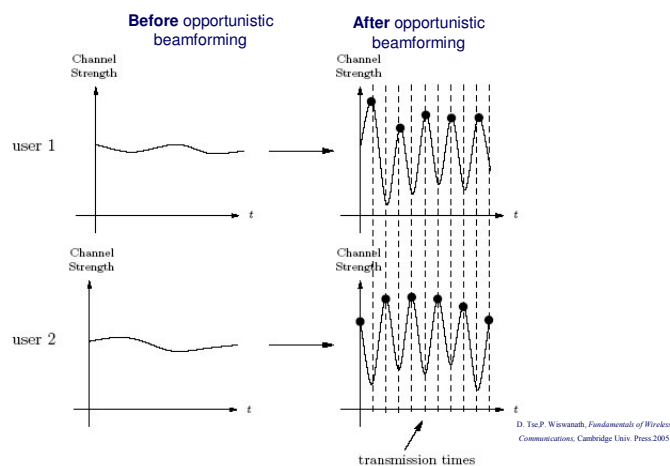
- **When are SNR peaks reached?:** When beam “points” at user k

$$\mathbf{q}[m] \propto \mathbf{h}_k^*[m]$$

“OPPORTUNISTIC BEAMFORMING”

27

OPPORTUNISTIC BEAMFORMING (cont'd)



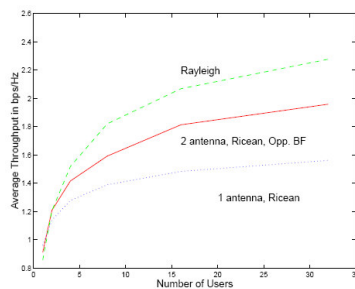
- **How fast should $\mathbf{q}[n]$ change?:** Design parameter:
 - Fast enough to induce fast fading
 - Slow enough for reliable channel estimation, timely feedback, stable loop.

28

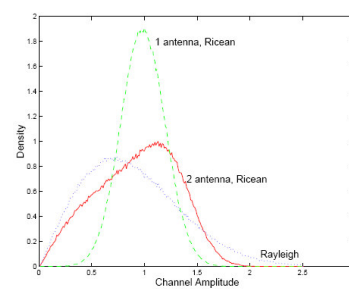
OPPORTUNISTIC BEAMFORMING (cont'd)

DOES OPPORTUNISTIC BEAMFORMING ALWAYS HELP?

- Slow fading $h_k[m] = h_k$: Constant $\rightarrow |h_k^* \mathbf{q}[m]|^2$: Fast & high fluctuation **YES**
- Fast Rayleigh fading: $h_k[m]$: i.i.d. Gaussian $\rightarrow |h_k^* \mathbf{q}[m]|^2$: i.i.d. Gaussian **NO**
i.e. identical distribution for ANY distribution of \mathbf{q}
- Fast Ricean Fading: $h_k[m] = h_k + h_{kw}[m]$ $\rightarrow |h_k^* \mathbf{q}[m]|^2 = |h_k^* \mathbf{q}[m] + h_{kw}^* \mathbf{q}[m]|^2$ **YES**
Additional power for FAST fluctuations \rightarrow No additional fluctuations



D. Tse, P. Wornath, *Fundamentals of Wireless Communications*, Cambridge Univ. Press 2005

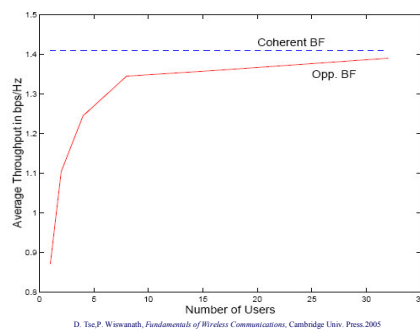


D. Tse, P. Wornath, *Fundamentals of Wireless Communications*, Cambridge Univ. Press 2005

29

REMARKS ON OPPORTUNISTIC BEAMFORMING

Opportunistic vs. coherent beamforming:



D. Tse, P. Wornath, *Fundamentals of Wireless Communications*, Cambridge Univ. Press 2005

- Performance: Comparable for high K (always a user to point at)
- CSIT needs:
 - Opp.: SNR only (Opp.)!!!
 - Coherent: full CSI

Multiple transmit antennas just for inducing fluctuations? Can we do better?

YES

**MULTIPLE ORTHOGONAL
RANDOM BEAMS**

- + Still inducing fast fading
- + Additional spatial multiplexing gain (SDMA)
- Extra overhead for SNR measurements & feedback

30

OUTLINE

- Motivation
- A review of capacity issues in single-user systems
 - ✓ Definition, Capacity for MIMO systems.
- Capacity issues in multi-user systems:
 - ✓ Broadcast (BC) and Multiple Access (MAC) channels.
 - ✓ Capacity regions for SISO BC & MAC. Sum capacity. Symmetric capacity.
 - ✓ Multi-user diversity. Channel-aware scheduling.
 - ✓ Fairness issues: Proportional Fair Scheduling
 - ✓ Slow-fading channels: Opportunistic Beamforming
- **Channel- and queue-aware scheduling**
 - ✓ **Motivation.**
- Q&A

31

CHANNEL- AND QUEUE-AWARE SCHEDULING

32

ASSUMPTIONS REVISITED

- **Implicit assumptions so far...**
 - Ass. 1: *Infinite* transmit buffer size:
 - Users can be delayed without bound (to maximize sum-rate).
 - Did not care much about packet arrival rates.
 - Ass.2 : Scheduled user(s) always have data to transmit
- **BUT in realistic scenarios...**
 - *Finite* buffer size:
 - When close to buffer overflow, user should be scheduled regardless of channel conditions.
 - If too many packets arrive, buffer bound to explode.
 - Traffic is *bursty*: no point in scheduling a user with empty buffer!
- **CONCLUSION:** *Channel* and *queue* (buffer) information must be jointly considered in the scheduling process (i.e. cross-layer)

33

QUESTIONS ?

34