Università degli Studi di Padova
Bressanone Summer School
27 giugno 2005

Techniques for High Throughput Wireless Packet Data Access

Robert Elliott, David Mazzarese,
Robert Novak, and Witold Krzymień
University of Alberta / TRLabs
Edmonton, Alberta, Canada

Outline

1. **Scheduling Algorithms for Cellular Downlink Spectrally Efficient Single-Carrier SISO Packet Data Access**
2. **Spectrally Efficient High Throughput SISO Spread Spectrum OFDM Cellular Downlink for Packet Data**
3. **High Throughput Downlink Wireless Packet Data Access with Multiple Antennas and Multi-User Diversity**

Università degli Studi di Padova
Bressanone Summer School
27 giugno 2005

Scheduling Algorithms for Cellular Downlink Spectrally Efficient Single-Carrier SISO Packet Data Access

Robert Elliott & Witold Krzymień
University of Alberta / TRLabs
Edmonton, Alberta, Canada

Outline of Section 1

- Introduction
- Why scheduling?
- Scheduling algorithms
- Simulation set-up
- Selected simulation results
- Conclusions

Introduction (1)

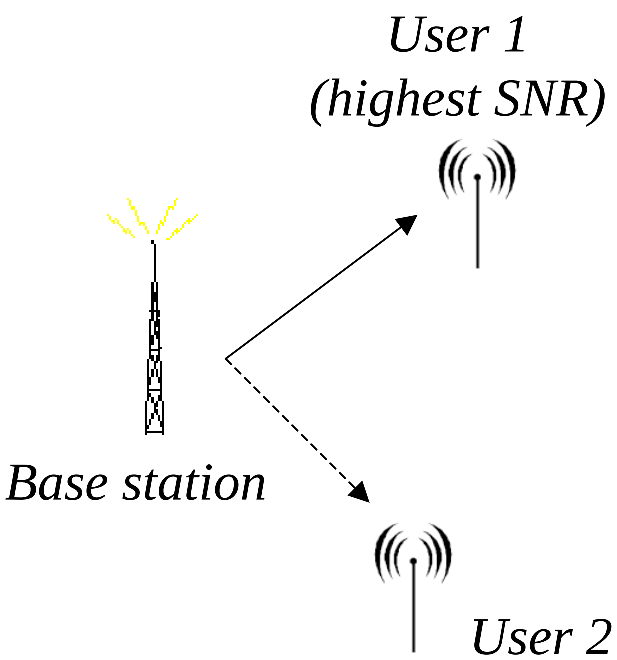
- Asymmetric data throughput requirements for reverse link and forward link: usually much higher requirements on the downlink
  - emphasis of this work on best effort, delay-tolerant services on the downlink
- 1st, 2nd and even 3rd generation cellular systems not designed to handle efficiently asymmetric data services.
- Several systems (e.g. 1xEV-DO, 1xEV-DV, HSDPA) have been proposed and standardized to enable high throughput packet data service (especially on the downlink) as an evolution of third generation cellular radio systems

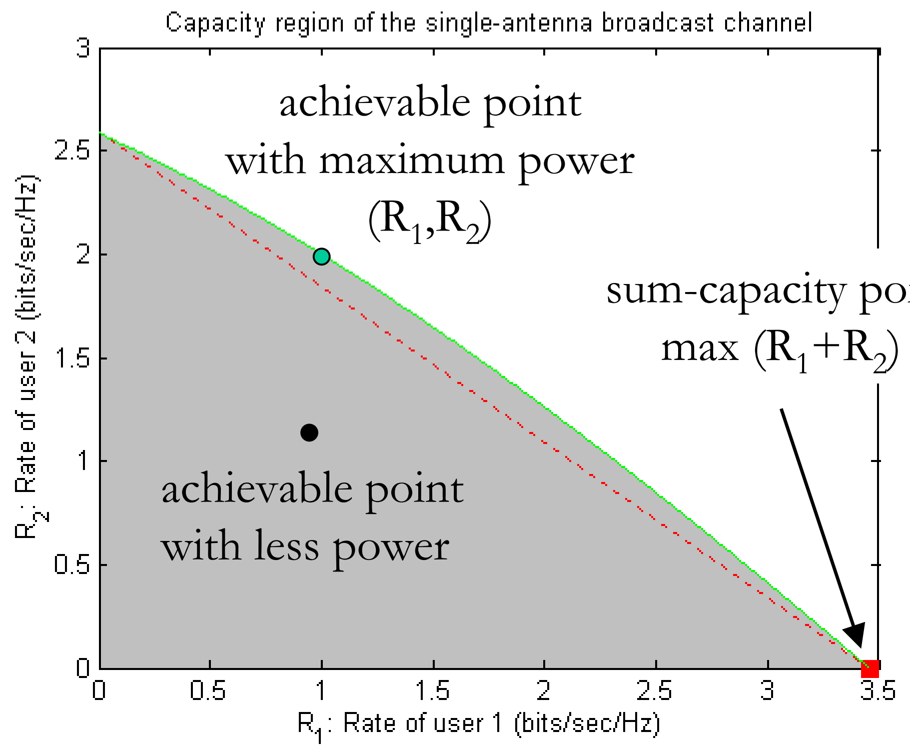
Introduction (2)

- The following techniques employed to enable high bit rates:
  - transmission of packets over very short (~ 1 ms) slots to one user at a time
  - frequent estimation & prediction of SINR
  - best sector selection (instead of hand-off)
  - reporting of predicted channel conditions or possible bit rates to the base station
  - adaptive modulation & coding (instead of power control – Tx always at full power)
  - scheduling of packet transmissions to exploit multi-user diversity
  - stop & wait type II hybrid ARQ (employing soft packet combining and incremental redundancy) with interlaced multi-slot packet transmissions (synchronous or asynchronous re-transmission of packets)

Why Scheduling? (1)

- Several papers ([Knopp/Humblet, 1995], [Grossglauser/Tse, 2001], [Bedekar *et al*., 1999]) show fading channel capacity in a single antenna (SISO) system is maximized when all Tx power is dedicated to a single user at time & users share the channel on time-slot basis (TDM)





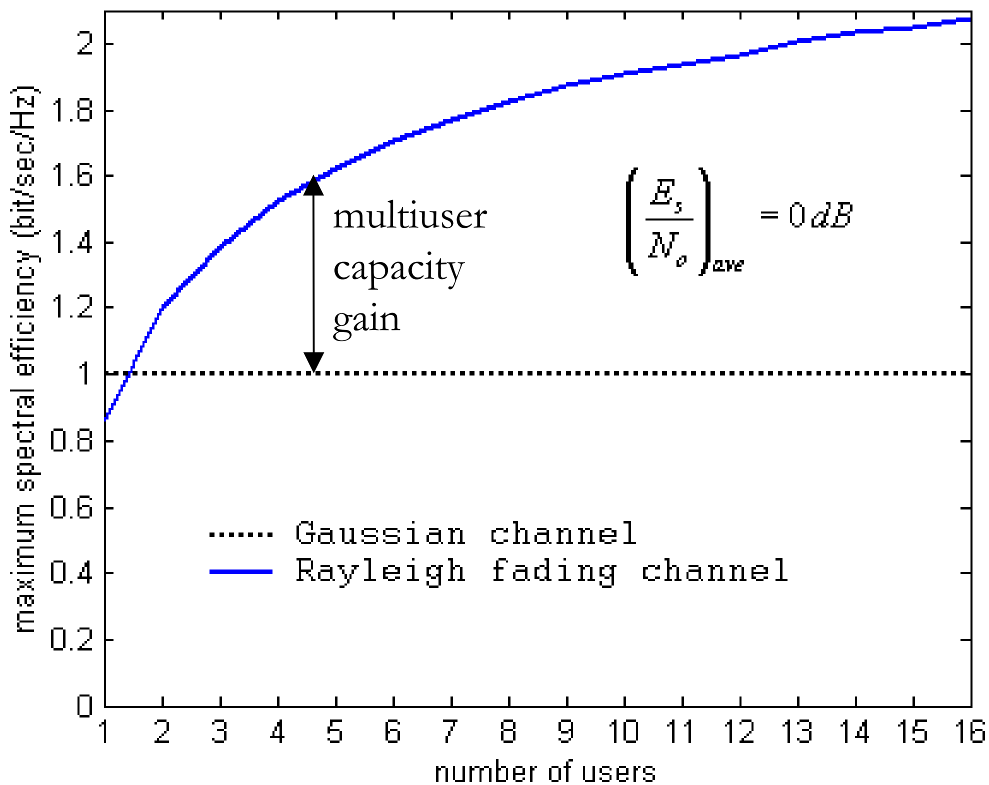
[Knopp/Humblet, 1995]: R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC'95*, Seattle, WA, June 1995, vol. 1, pp. 331-335.

[Grossglauser/Tse, 2001]: M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proc. IEEE INFOCOM 2001*, Anchorage, AK, Apr. 2001, vol. 3, pp. 1360-1369.

[Bedekar *et al*., 1999]: A. Bedekar, S. Borst, K. Ramanan, P. Whiting, and E. Yeh, "Downlink scheduling in CDMA data networks," in *Proc. IEEE GLOBECOM'99*, Rio de Janeireo, Brazil, Dec. 1999, vol. 5, pp. 2653-2657.

Why Scheduling? (2)

- Presence of many users within the system creates *multiuser diversity*



- Scheduler decides, which user receives a packet in each time slot

Scheduler must balance several conflicting goals

- Maximize system throughput
- Minimize delays seen by transmitted packets
- Minimize delays seen by each user requesting service
- Provide a fair amount of service to each user in the system
- Meet Quality of Service (QoS) requirements (i.e. a minimum throughput / maximum delay)

Scheduling Algorithms (1)

- **Round Robin (RR) –** Choose users in cyclic order, ignoring channel conditions
- **Maximum Requested Rate (maxD or "greedy" algorithm)**
  - Select user with the largest requested rate r_k, with ties broken at random

- Above algorithms provide lower and upper throughput bounds -

- **Proportionally Fair (PF A)** [Jalali *et al*., 2000]
  - Choose user with largest $r_k / \overline{R}_k$, where

$$\overline{R}_k(n) = \left(1 - \frac{1}{t_c}\right)\overline{R}_k(n-1) + \frac{1}{t_c} R_k(n-1)$$

t_c = 1000 slots

  - $R_k(n\text{-}1) = r_k$ if user k selected for transmission or re-transmission in slot n-1
  - $R_k(n\text{-}1) = 0$ otherwise

  - The way in which $R_k(n\text{-}1)$ is evaluated and included in the calculation of $\overline{R}_k(n)$ assumes all transmissions take the maximum number of slots, and hence the algorithm does not account for early terminations due to hybrid ARQ.

- **PF B -** Accounts for early terminations by evaluating $R_k(n\text{-}1)$ as follows:

$$R_k(n-1) = \begin{cases} r_{\max}^k(n-1) & \text{if transmission of a new packet to user } k \text{ has started in slot } n\text{-}1 \\ 0 & \text{if a re-transmission, or user } k \text{ not scheduled for transmission in slot } n\text{-}1 \end{cases}$$

$r_{\max}^k(n-1)$: The maximum bit rate possible for the physical layer packet size scheduled for user k

$r_{\max}^k$ is achieved if a packet is decoded correctly after one transmission attempt;

r_k is achieved if a packet is decoded correctly after using the maximum allowed number of re-transmissions

[Jalali *et al*., 2000] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Proc. IEEE Vehicular Technology Conf.* (*VTC 2000-Spring*), Tokyo, Japan, May 2000, vol. 3, pp. 1854-1858.

Scheduling Algorithms (2)

- Maximum Rate / Proportionally Fair Hybrid (maxD/PF)
  - Choose user *k* with largest value of:

$$a\frac{r_k}{\overline{R}_k}+(1-a)\frac{r_k}{\overline{R}_{av}}$$

  - r_k and $\overline{R}_k$ same as in for the PF algorithm
  - $\overline{R}_{av}$ is the average value of $\overline{R}_k$ across all users
  - a is a tunable parameter ($0 < a < 1$) that tunes throughput and delay characteristics between maxD and PF algorithms
    - a=0 ≡ maxD algorithm a=1 ≡ PF algorithm
  - Analyzed for a=0.25, a=0.5, and a=0.75

Scheduling Algorithms (3)

- Modified Longest Weighted Delay First (M-LWDF) [Andrews *et al*., 2001]
  - Choose user k with largest $\gamma_k r_k W_k$
  - r_k = requested rate, W_k = head-of-line packet delay
  - γ_k is a multiplier; it can be different for each user
    - $\gamma_k = 1$: maximum rate × delay hybrid ("M-LWDF 1")
    - $\gamma_k = 1/\overline{R}_k$: maximum relative rate (PF) × delay hybrid (M-LWDF 2")
- Exponential Rule (EXP) [Shakkottai/Stolyar, 2000]
  - Replace W_k above with $\exp\left(\dfrac{W_k}{1+\sqrt{\overline{W}}}\right)$
  - $\overline{W}$: average head-of-line delay over all users in sector
  - $\gamma_k = 1$: "EXP 1" – $\gamma_k = 1/\overline{R}_k$: "EXP 2"
  - Results vary depending on whether delays are in units of slots or seconds
    - Slots: "EXP1A", "EXP2A"; Seconds: "EXP1B", "EXP2B"
- maxD/PF/EXP Hybrid: multiply maxD/PF metric by exp(·) term above

[Andrews *et al*., 2001] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag*., vol. 39, no. 2, pp. 150-154, Feb. 2001.

[Shakkottai/Stolyar, 2000] S. Shakkottai and A. Stolyar, "A study of scheduling algorithms for a mixture of real and non-real time data in HDR," *Bell Labs Tech. Memo*, Oct. 2000.

Scheduling Algorithms (4)

- Slot-Wise Fair ("slot-fair") [Liu *et al.*, 2001 (1)]
  - Maximize throughput under the constraint of approximately equal distribution of slots among users
  - Choose user with highest $(U_k + v_k)$
    - U_k = (requested rate in kbps)/(38.4 kbps)
    - $v_k(n+1) = v_k(n) - a\left(\mathbf{1}_k(n) - \frac{1}{K}\right)$: updated every slot
      - $\mathbf{1}_k(n) = 1$ if user k receives data in slot n; $\mathbf{1}_k(n) = 0$ otherwise
      - $a = 0.25$
      - K = number of users in sector
- Packet-Wise Fair ("packet-fair")
  - Same as above, except $a = 0.5$, and only update v_k if a new packet is scheduled; approx. equal distribution of packets instead of slots

[Liu *et al.*, 2001 (1)] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 2053-2064, Oct. 2001.

Scheduling Algorithms (5)

- Minimum Performance ("minperf") [Liu *et al*., 2001 (2)]
  - Maximize throughput while attempting to provide a specified minimum rate C_k to each user k
  - Select user with largest $\alpha_k U_k$
  - U_k same as previous slide
  - $\alpha_k(n) = \max\left\{\alpha_k(n-1) - a\left[\bar{R}_k(n) - C_k\right], 1\right\}$
    - $\bar{R}_k(n)$ calculated the same as in the PF algorithm
    - $\bar{R}_k(n)$, C_k in bps
    - $a = 10^{-4}$
  - Analyzed for $C_k = \{5,10,15,20,25\}$ kbps for all users in sector

[Liu *et al*., 2001 (2)] X. Liu, E. K. P. Chong, and N. B. Shroff, "Transmission scheduling for efficient wireless resource utilization with minimum-performance guarantees," in *Proc. IEEE Vehicular Tech. Conf.* (*VTC 2001-Fall*), Atlantic City, NJ, Oct. 2001, vol. 2, pp. 824-828.

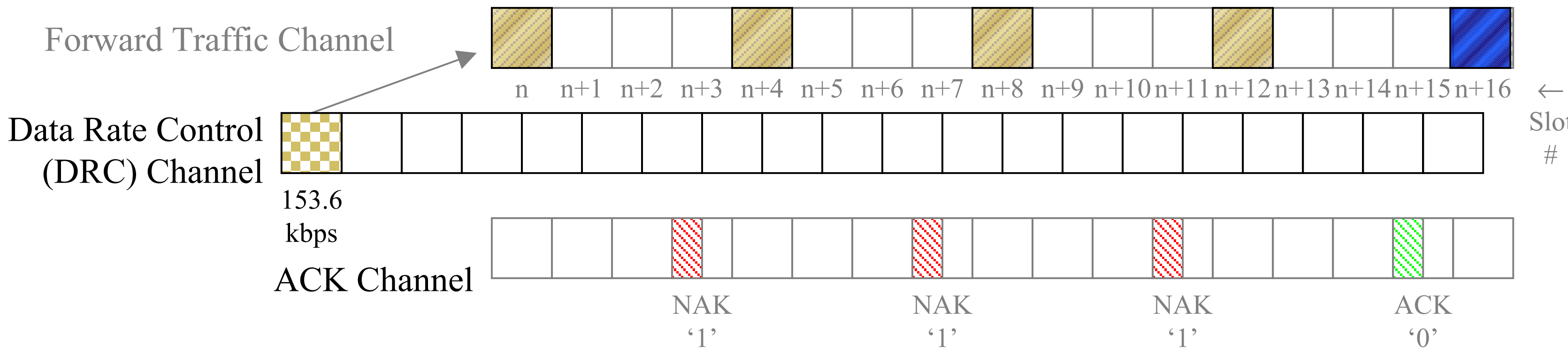
1xEV-DO

- Standardized as IS-856 [TIA, 2000]
- Designed to handle data traffic only
- 1 slot = 1.67 ms (5/3 ms exactly)
- Mobiles send transmission format requests to base station over a data request channel (DRC)
- Scheduler decides which user to transmit based on DRC requests
- Synchronous retransmissions occur 4 slots apart
- Uses S&W type II hybrid ARQ and acknowledgement channel to end retransmissions early if packet received correctly

Format (DRC request)	Possible Data Rates (kbps)	Max. TX Slots	Modulation	Effective Code Rate Range	Packet Size (bits)	Min. SIR (dB) for 1% PER (AWGN)
1	38.4-614.4	16	QPSK	1/5-8/9	1024	-13.5
2	76.8-614.4	8	QPSK	1/5-8/17	1024	-10.5
3	153.6-614.4	4	QPSK	1/5-8/21	1024	-7.4
4	307.2-614.4	2	QPSK	1/5-8/23	1024	-4.3
5	614.4	1	QPSK	1/3	1024	-1.0
6	307.2-1228.8	4	QPSK	1/3-16/23	2048	-4.2
7	614.4-1228.8	2	QPSK	1/3-2/3	2048	-1.2
8	1228.8	1	QPSK	2/3	2048	3.7
9	921.6-1843.2	2	8-PSK	1/3-2/3	3072	1.5
10	1843.2	1	8-PSK	2/3	3072	7.1
11	1228.8-2457.6	2	16-QAM	1/3-2/3	4096	3.4
12	2457.6	1	16-QAM	2/3	4096	9.2

[TIA, 2000] TIA/EIA/IS-856, *cdma2000 High Rate Packet Data Air Interface Specification*, Telecommunications Industry Association, Arlington, VA, Nov. 2000.

Format Request Determination (1)



– Delay of about 3 slots between SIR estimation, DRC determination and request, and user being scheduled
  - Prediction of channel desirable, especially at high Doppler
  - Prediction methods beyond the scope of this work

(a)
– Use current estimated SIR minus a margin:

$$\widehat{SIR}(n+3) = SIR_M(n) - \Delta_{1\%}$$

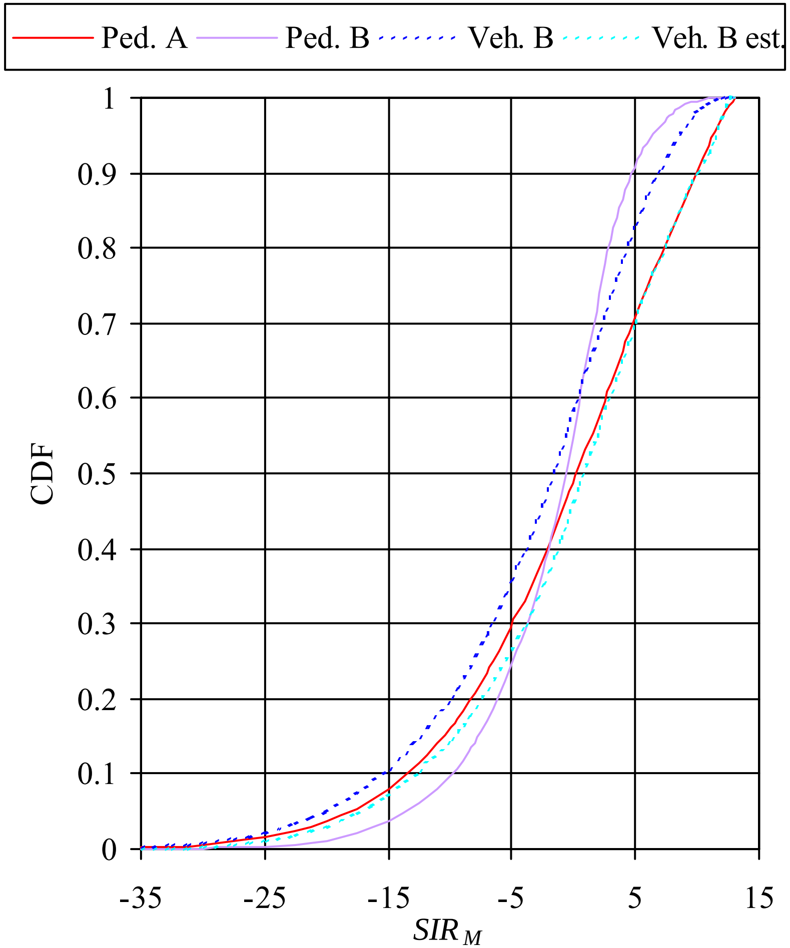
– Compare $\widehat{SIR}$ with AWGN values for 1% PER, select highest rate with E_S/N_0 (1%) < $\widehat{SIR}$

– Margins depend on format being tested, tuned to provide ~1% PER per format

(b)
– Perfect prediction: Mobile knows exactly the value of the SIR in future slots

- DRC channel is assumed to be error-free

Format Request Determination (2)



- SIR values reduced due to implementation imperfections
  - inter-chip interference, Tx non-linearity, non-ideal spreading waveform, ADC quantization noise, adjacent channel interference
- Effects combined into one parameter:

$$SIR_M = \frac{SIR}{1+\frac{SIR}{\alpha}}, \quad \alpha = 10^{1.3}\,(13\text{ dB})$$

- Mobile in Veh. B channel makes DRC requests based on path loss and shadowing components of the signal only when using margins

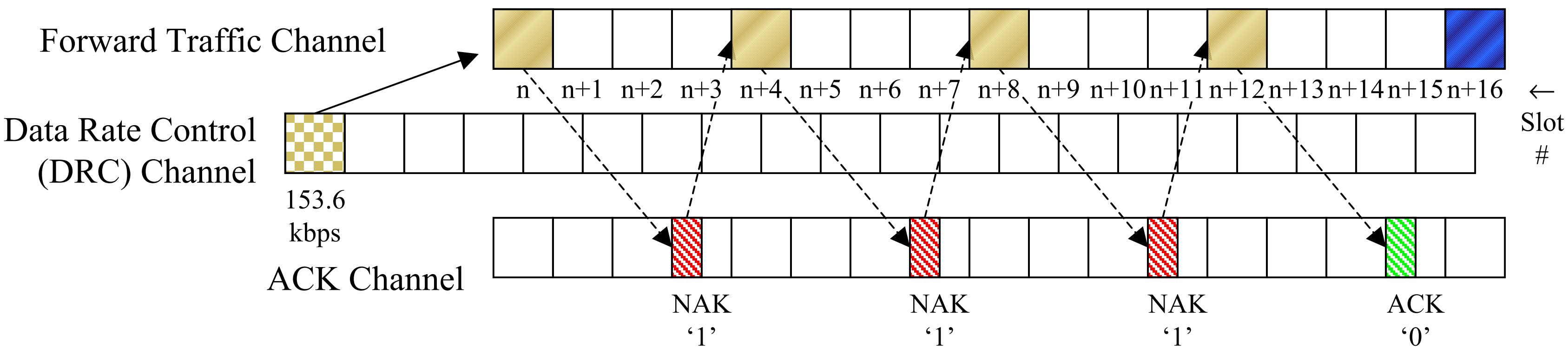
Obtaining Margins for Simulations

- Supportable format determinations based on comparing current SIR minus margins $\Delta_{1\%}$ with thresholds for 1% PER (see the 1xEV-DO format table)
  - Highest supportable rate chosen
  - When format rates are equal, for pedestrian channels, choose format that uses fewest slots; for vehicular channel choose format with more slots
- Run simulation starting with margin of 0 for all formats
- Check PER for each format
- If PER not in the range of 0.95% to 1.05%, adjust margin for that format
  - If PER>1.05%, increase margin; if PER<0.95%, decrease margin
- Re-run simulation with new margins
- Repeat until PER for all formats within desired range

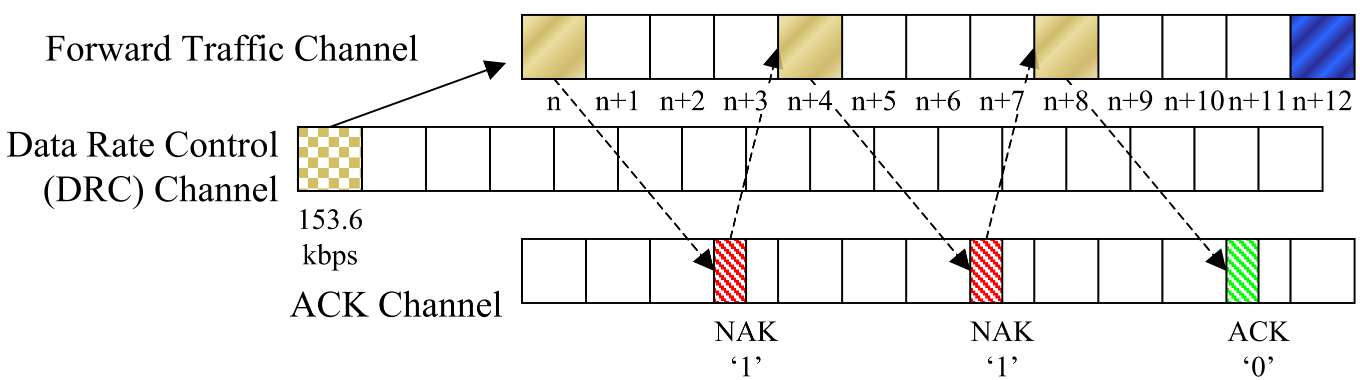
Note: Some formats may not be supportable at 1% in certain channels. In this event, the margin will increase until that format never gets selected, due to SIR-$\Delta_{1\%}$ always being smaller than the threshold for that format.

Re-Transmission / Early Termination Example

- Normal termination

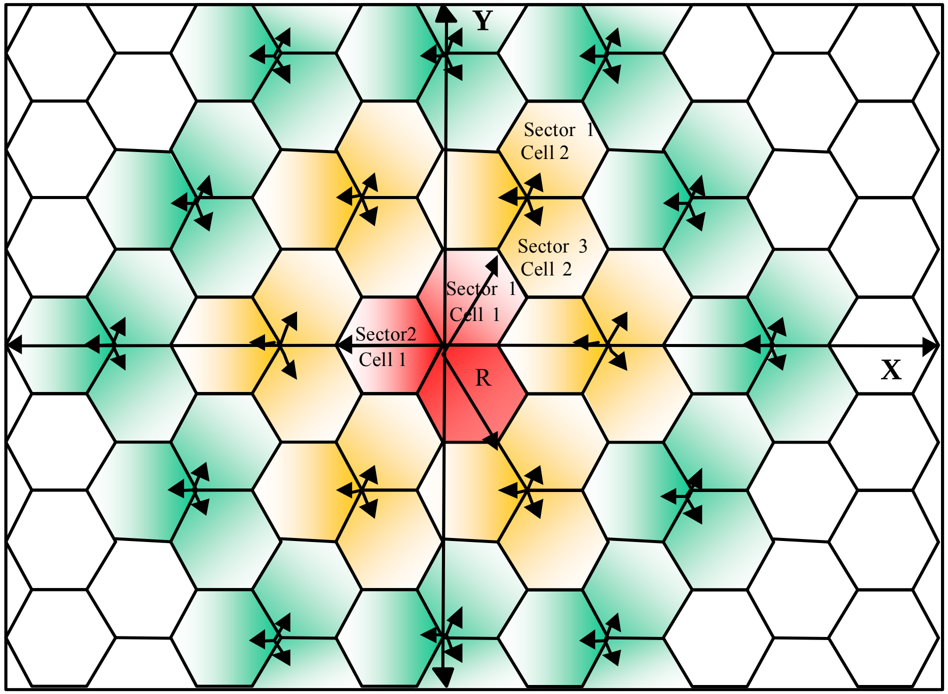


- Early termination



Actual Data Rate: 204.8 kbps

Simulation Set-Up (1)



- Data packets assumed to be always available for transmission to every user
- All users considered equal (i.e. best effort service with no QoS differences considered)

- 100 drops of 1, 2, 4, 8, 16, 32 users per drop in a sector of central cell, uniformly over area of sector
- 18000 slots (30 sec) per drop
- Surrounding sectors generate interference (two rings of interferers)
- SIR value available for each user, in every slot, for all drops
- Each signal undergoes COST-231 Walfish-Ikegami path loss, lognormal shadowing (σ=6.5 dB), and Rayleigh fading
- ITU pedestrian A and B, and vehicular B channels considered
- Cell layout according to [ETSI 1998]

[ETSI 1998] ETSI TR 101 112 *Universal Mobile Telecommunications System (UMTS): Selection procedures for the choice of radio transmission technologies of the UMTS*, 1998-04.

Simulation Set-Up (2): Channel Models

ITU-R Pedestrian A

Tap	Relative Delay (ns)	Avg. Power (dB)
1	0	0
2	110	-9.7
3	190	-19.2
4	410	-22.8
5	-	-
6	-	-

- Non-resolvable paths: flat Rayleigh fading
- 97% of transmitted power captured
- Max Doppler shift = 5.5 Hz (3 km/h at 2 GHz)

ITU-R Pedestrian B

Tap	Relative Delay (ns)	Avg. Power (dB)
1	0	0
2	200	-0.9
3	800	-4.9
4	1200	-8.0
5	2300	-7.8
6	3700	-23.9

- 2 combined paths and 1 single path
- RAKE receiver with maximal ratio combining
- Max Doppler shift = 5.5 Hz

ITU-R Vehicular B

Tap	Relative Delay (ns)	Avg. Power (dB)
1	0	-2.5
2	300	0
3	8900	-12.8
4	12900	-10.0
5	17100	-25.2
6	20000	-16.0

- 1 combined path and 2 single paths
- RAKE receiver with maximal ratio combining
- Max Doppler shift = 110 Hz (60 km/h at 2 GHz)

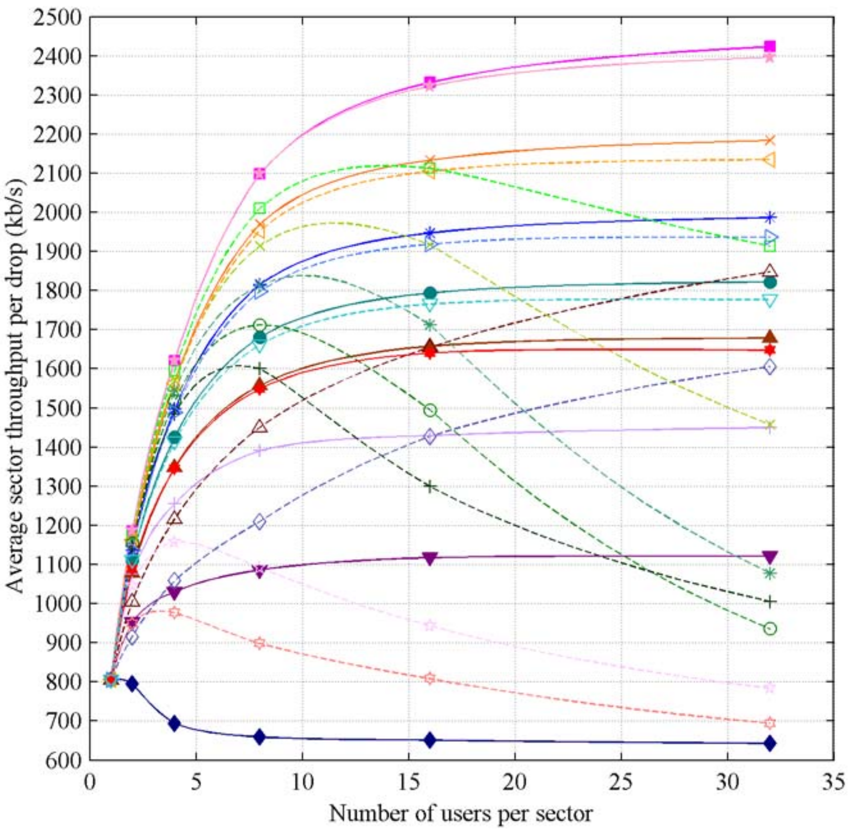
- Combined paths' power reduced proportional to time offset
- Remaining paths not captured, but contribute to interference
- Self-interference also accounted for

[ITU-R 1997] *Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000*, Recommendation ITU-R M.1225, 1997.

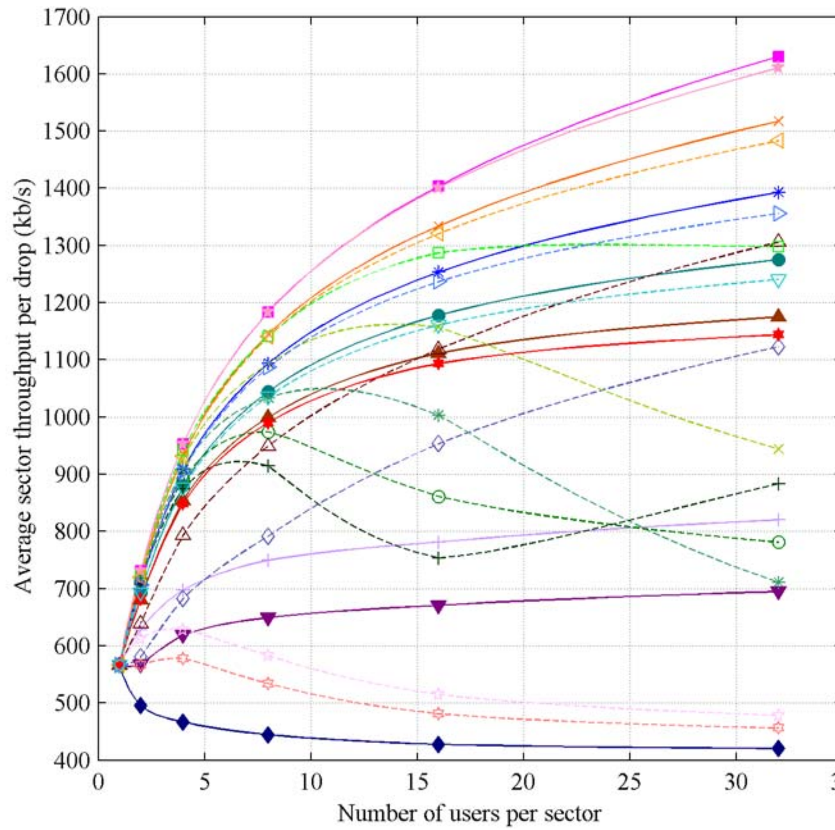
Selected Simulation Results

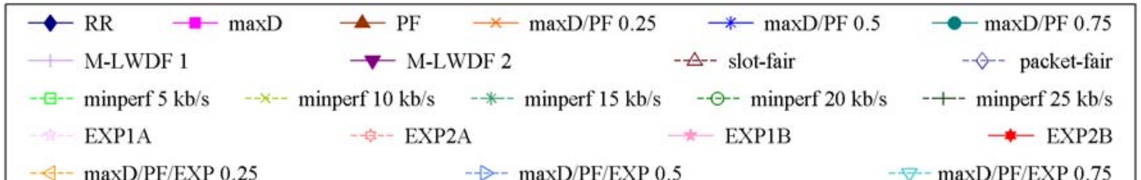
Average sector throughput per drop: Pedestrian channels with margins

ITU pedestrian A channel



ITU pedestrian B channel

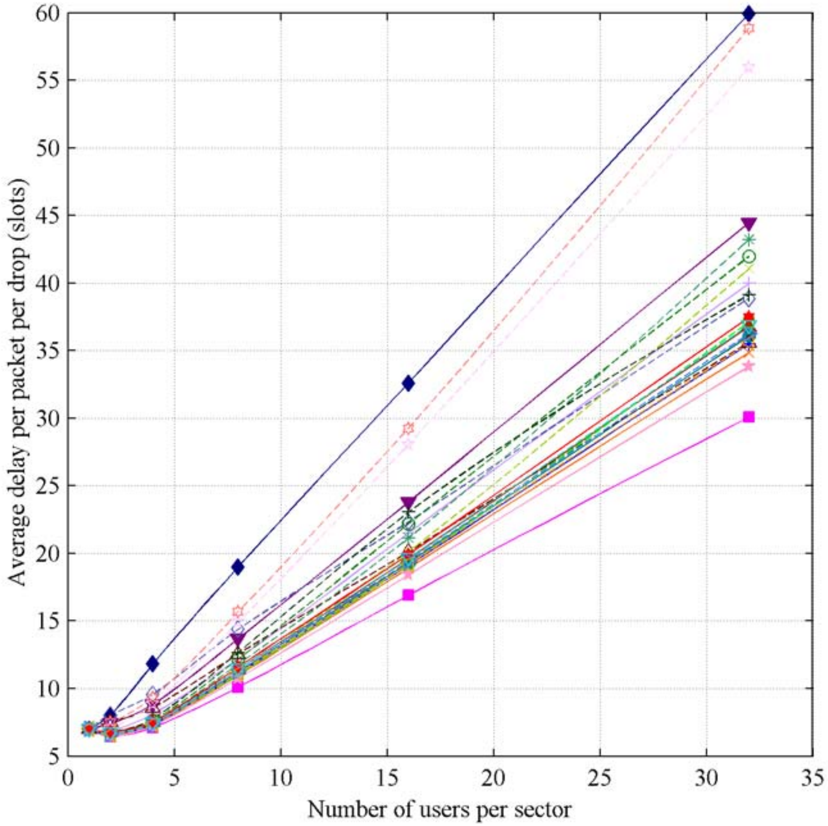




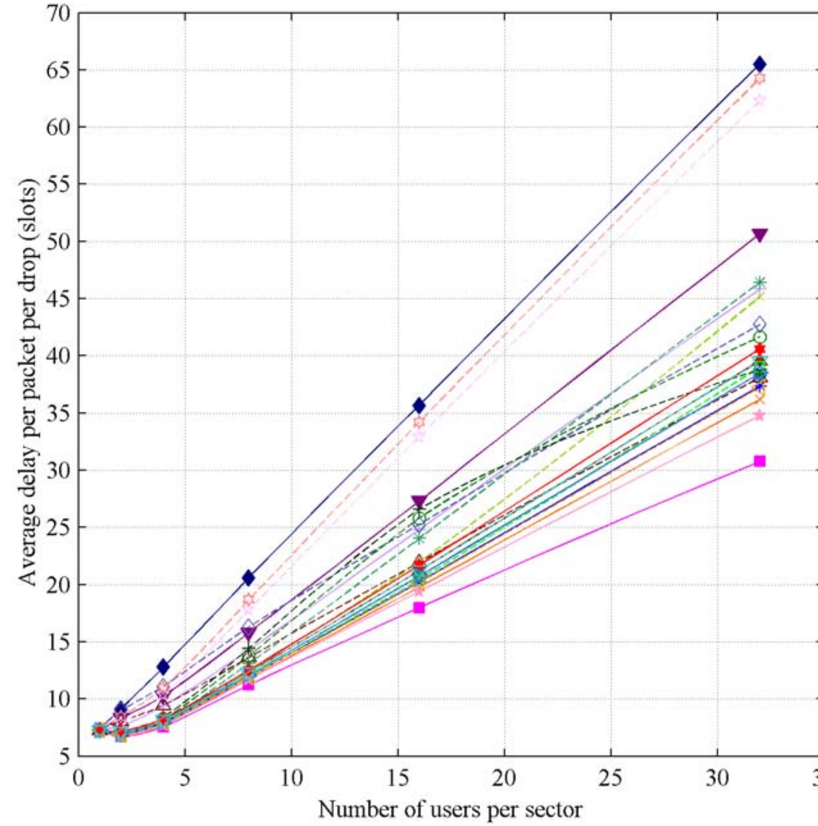
Selected Simulation Results

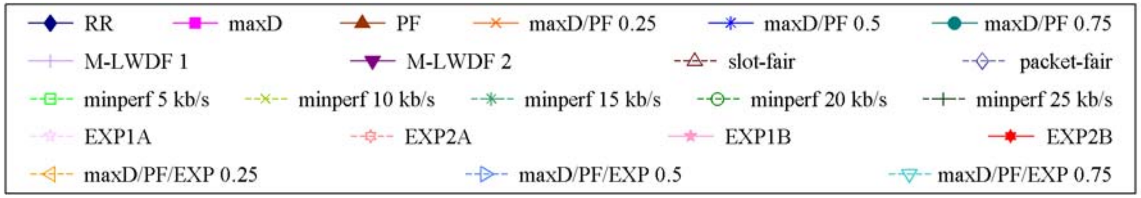
Average delay per packet per drop: Pedestrian channels with margins

ITU pedestrian A channel



ITU pedestrian B channel



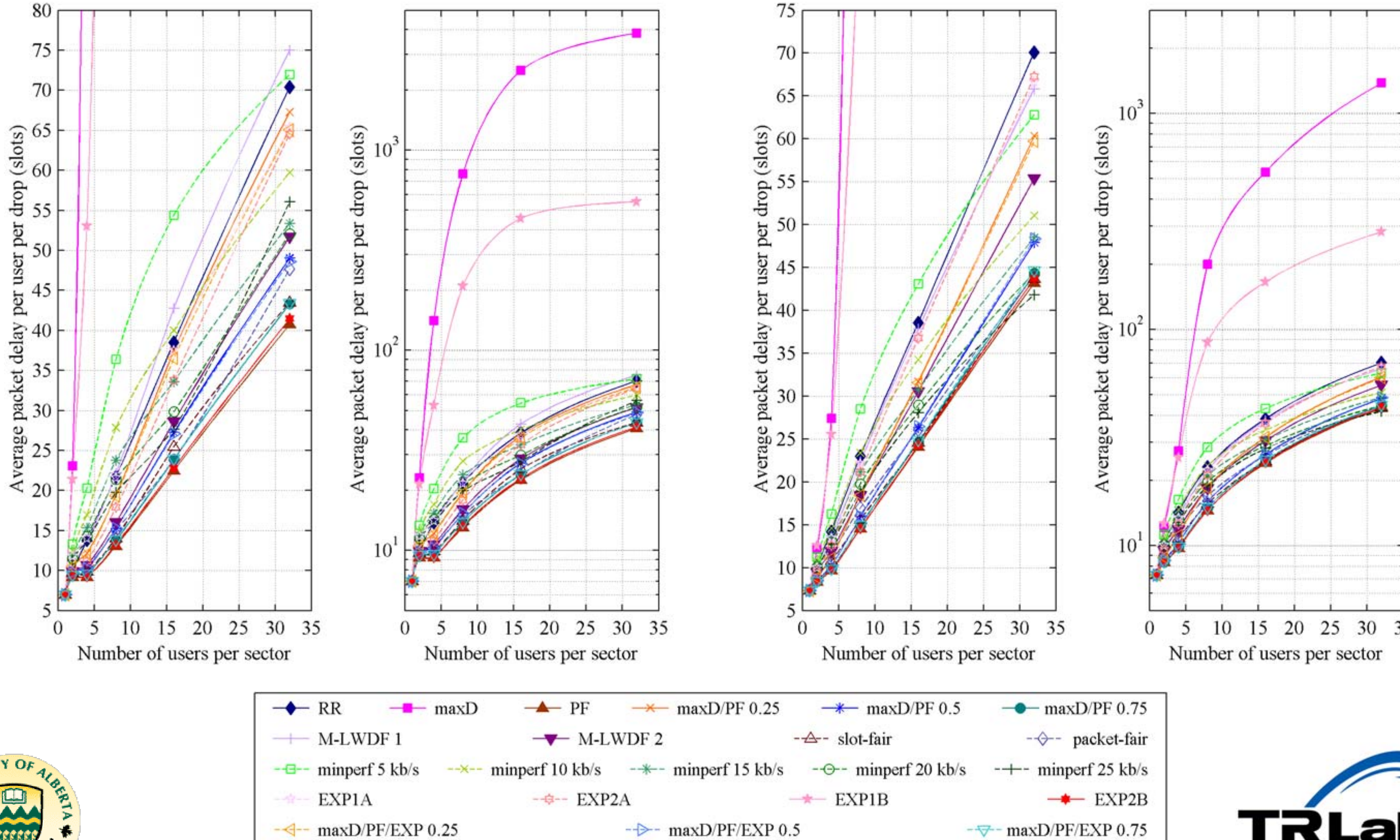


Selected Simulation Results

Average packet delay per user per drop: Pedestrian channels with margins

ITU pedestrian A channel

ITU pedestrian B channel

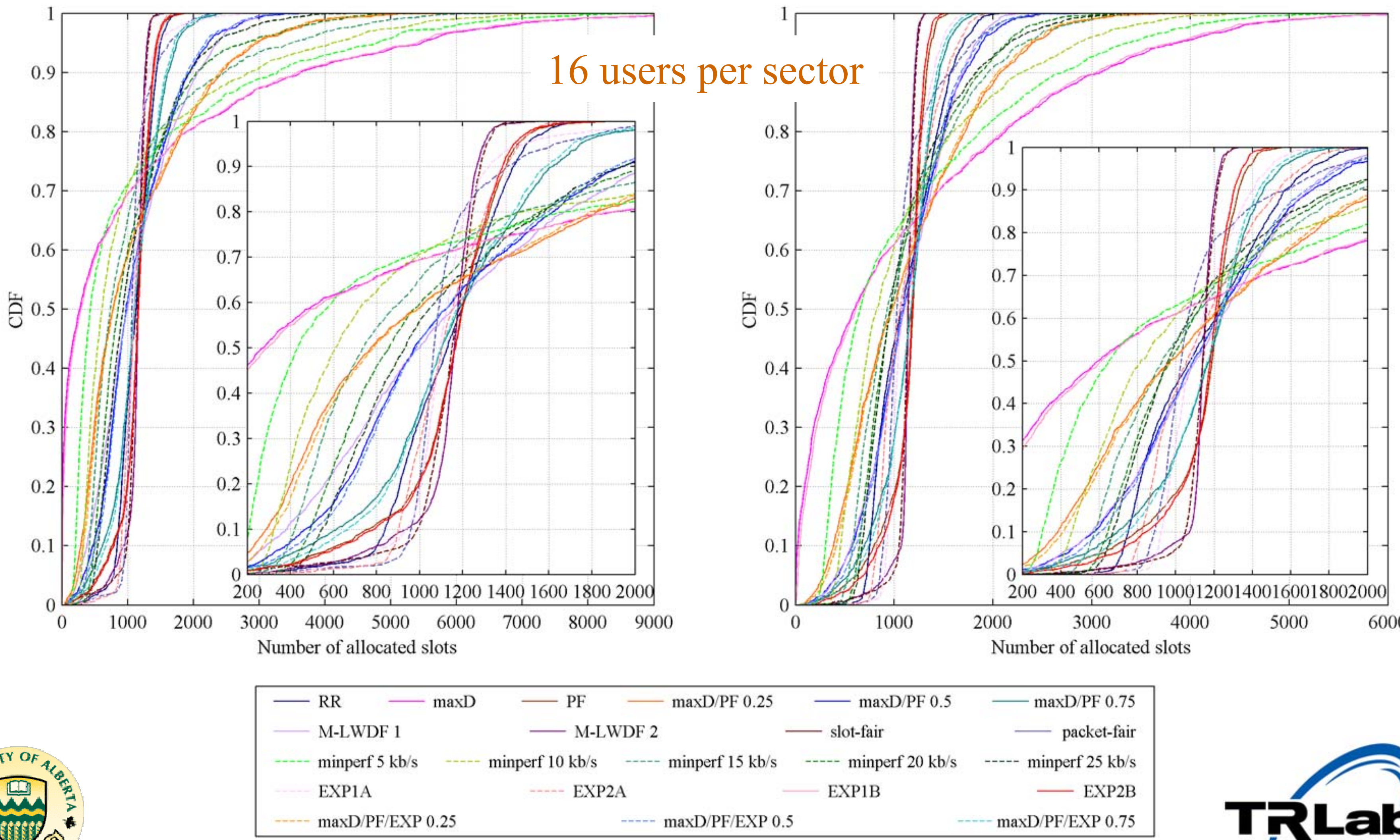


Selected Simulation Results : Slot Allocation

Distribution of allocated slots per user per drop: Pedestrian channels with margins

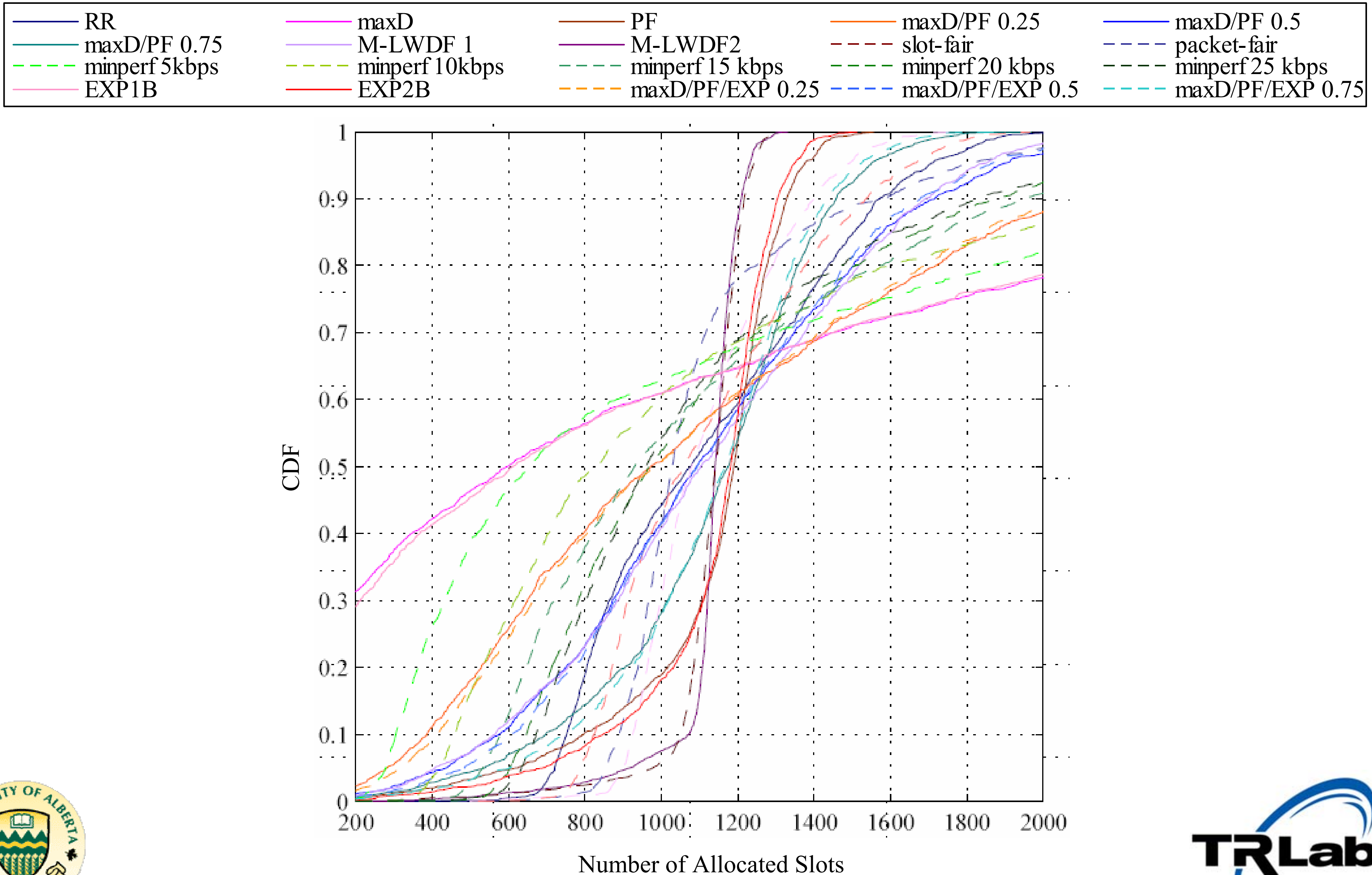
ITU pedestrian A channel

ITU pedestrian B channel



Selected Simulation Results: Slot Allocation

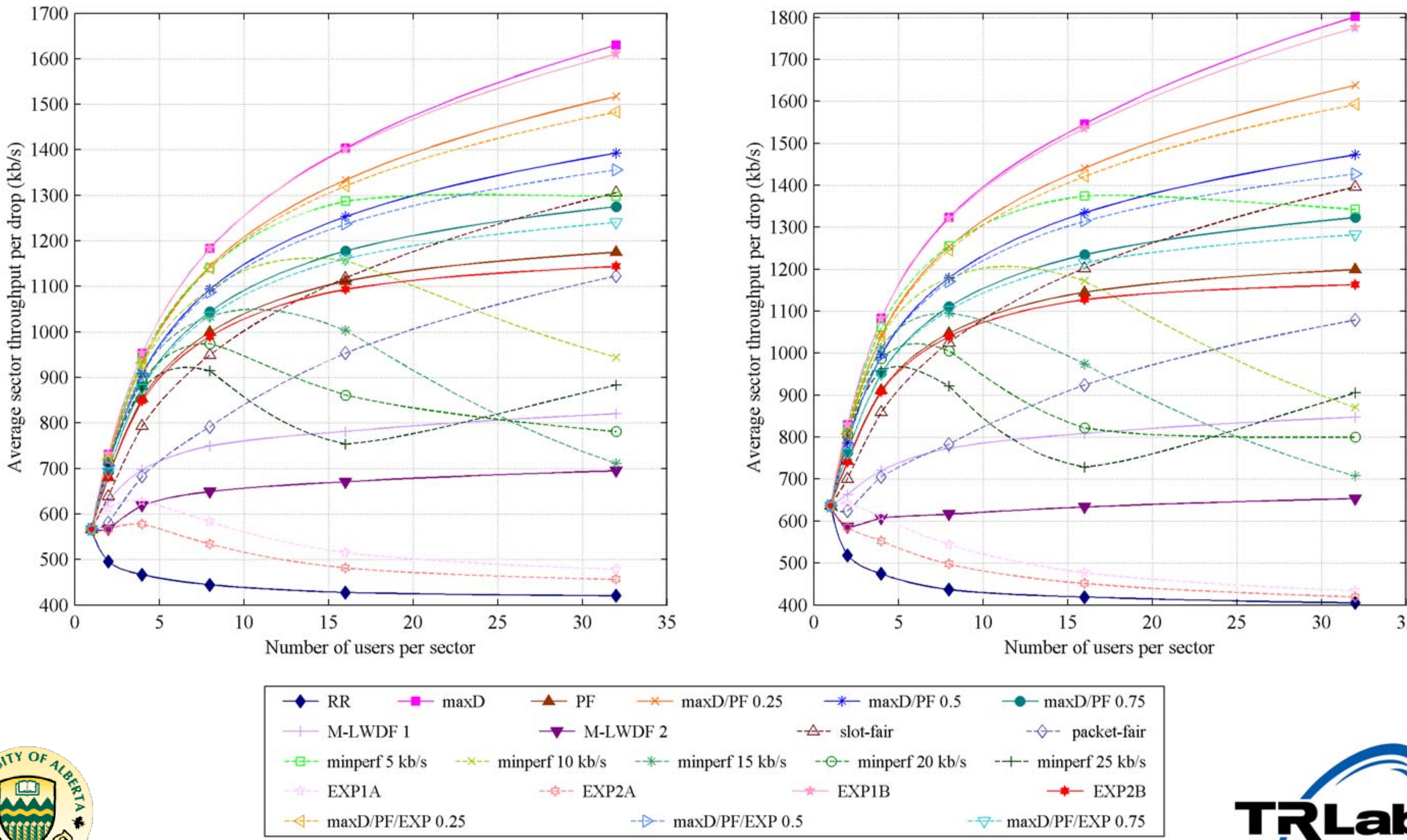
Distribution of Allocated Slots per User per Drop with 16 Users per Sector for ITU Pedestrian B Channel



Selected Simulation Results

Average sector throughput per drop: Margins vs. perfect prediction

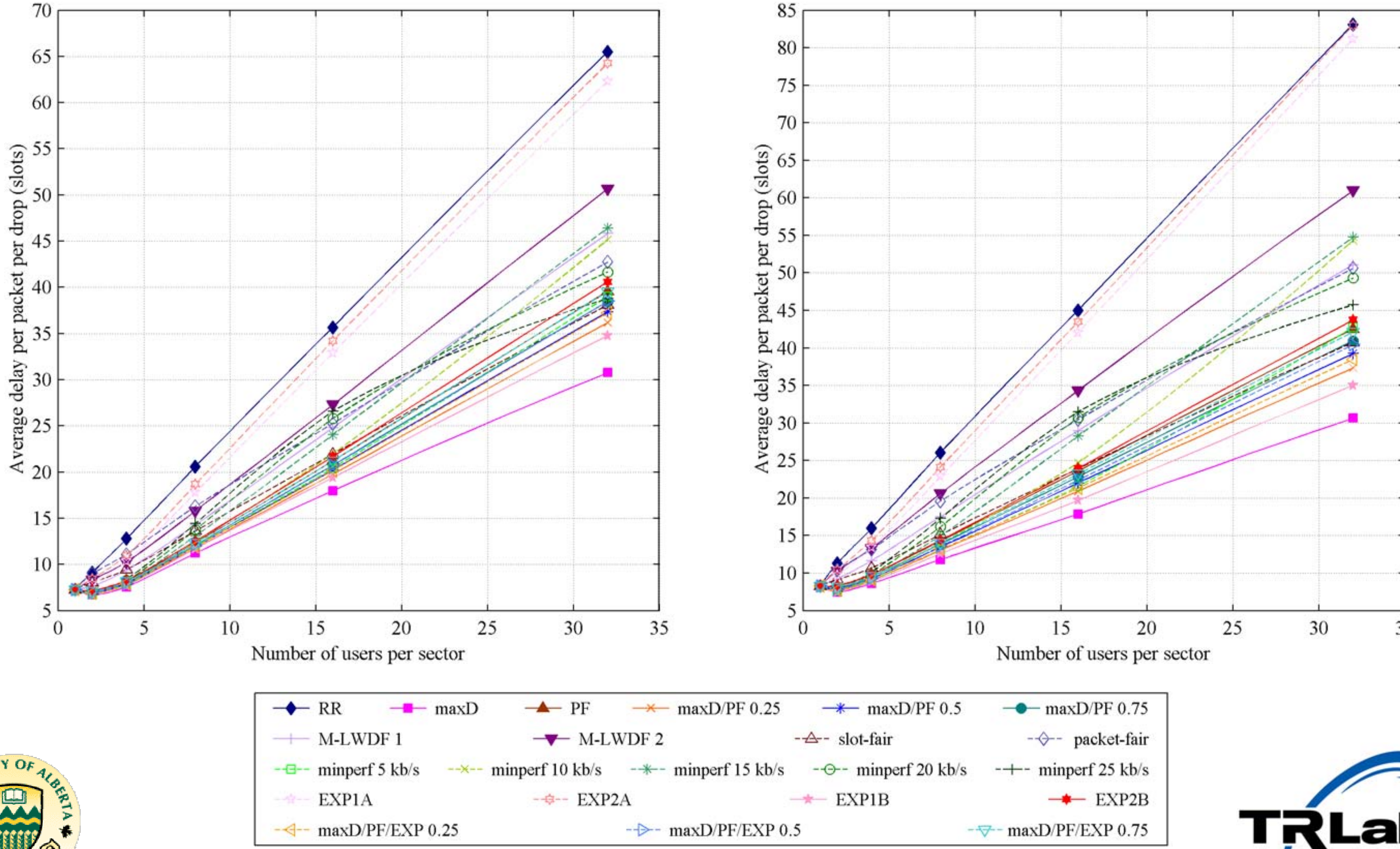
Margins ITU ped. B Perfect multislot prediction



Selected Simulation Results

Average delay per packet per drop: Margins vs. perfect prediction

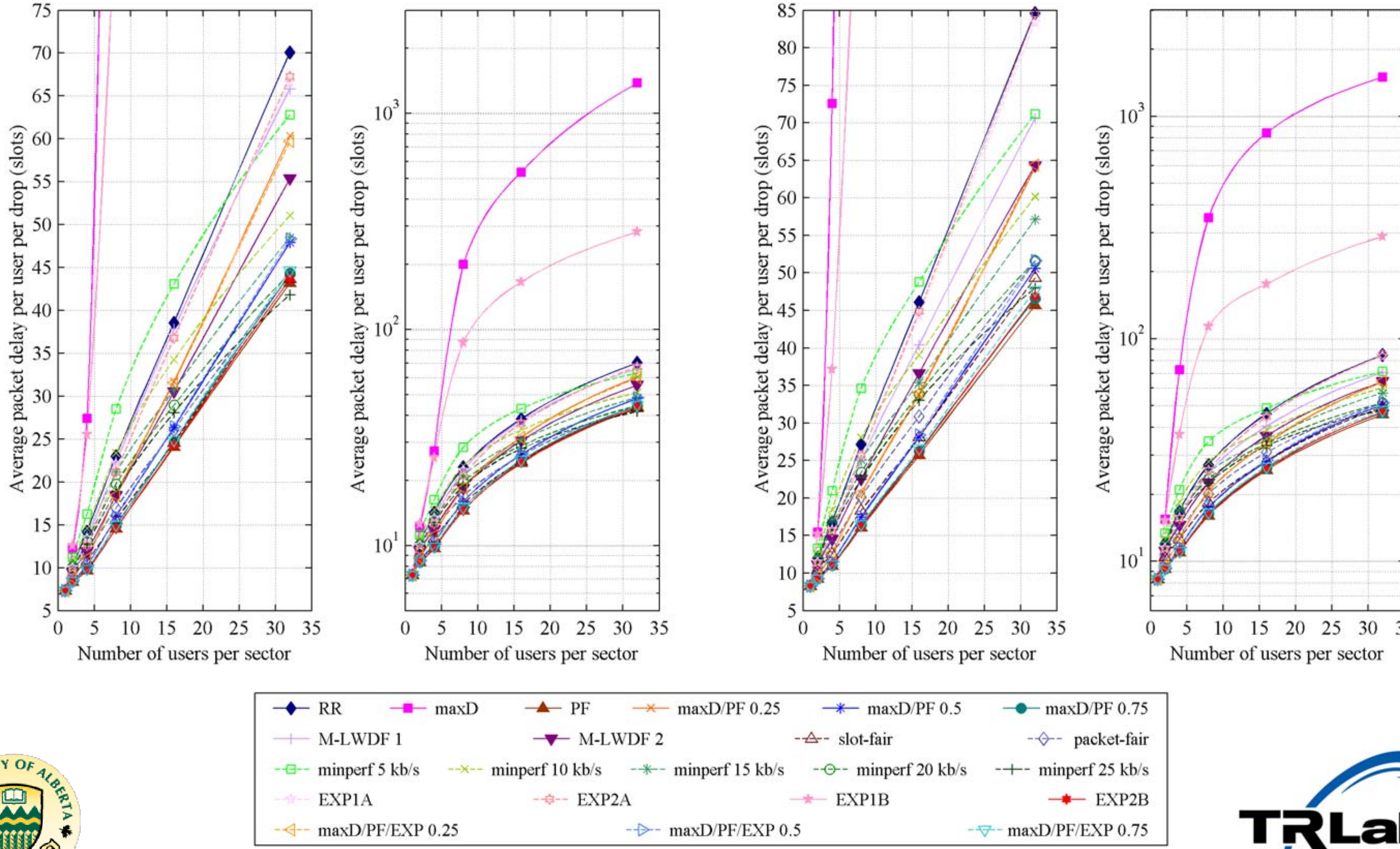
Margins ITU ped. B Perfect multislot prediction



Selected Simulation Results

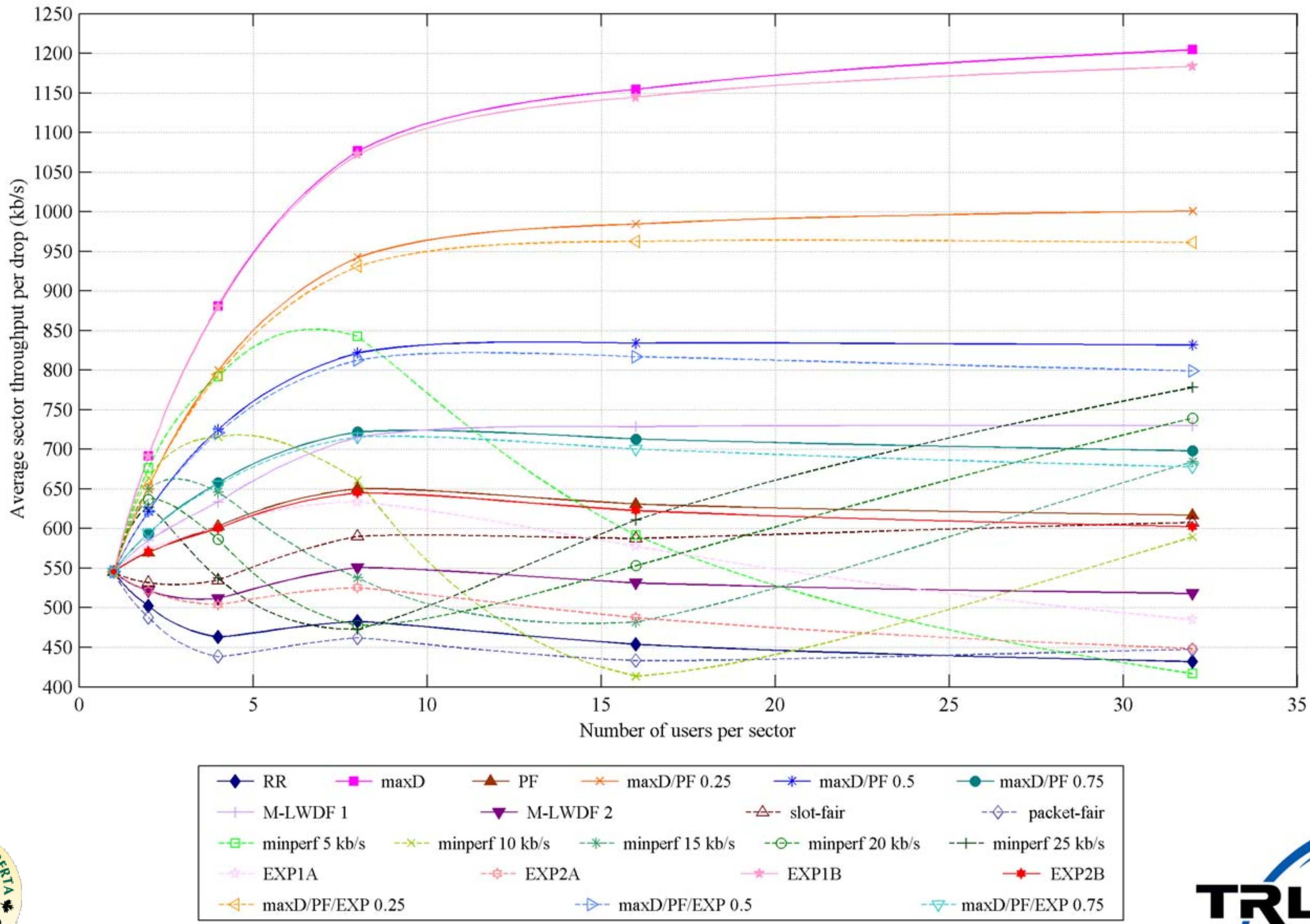
Average packet delay per user per drop: Margins vs. perfect prediction

Margins ITU ped. B Perfect multislot prediction



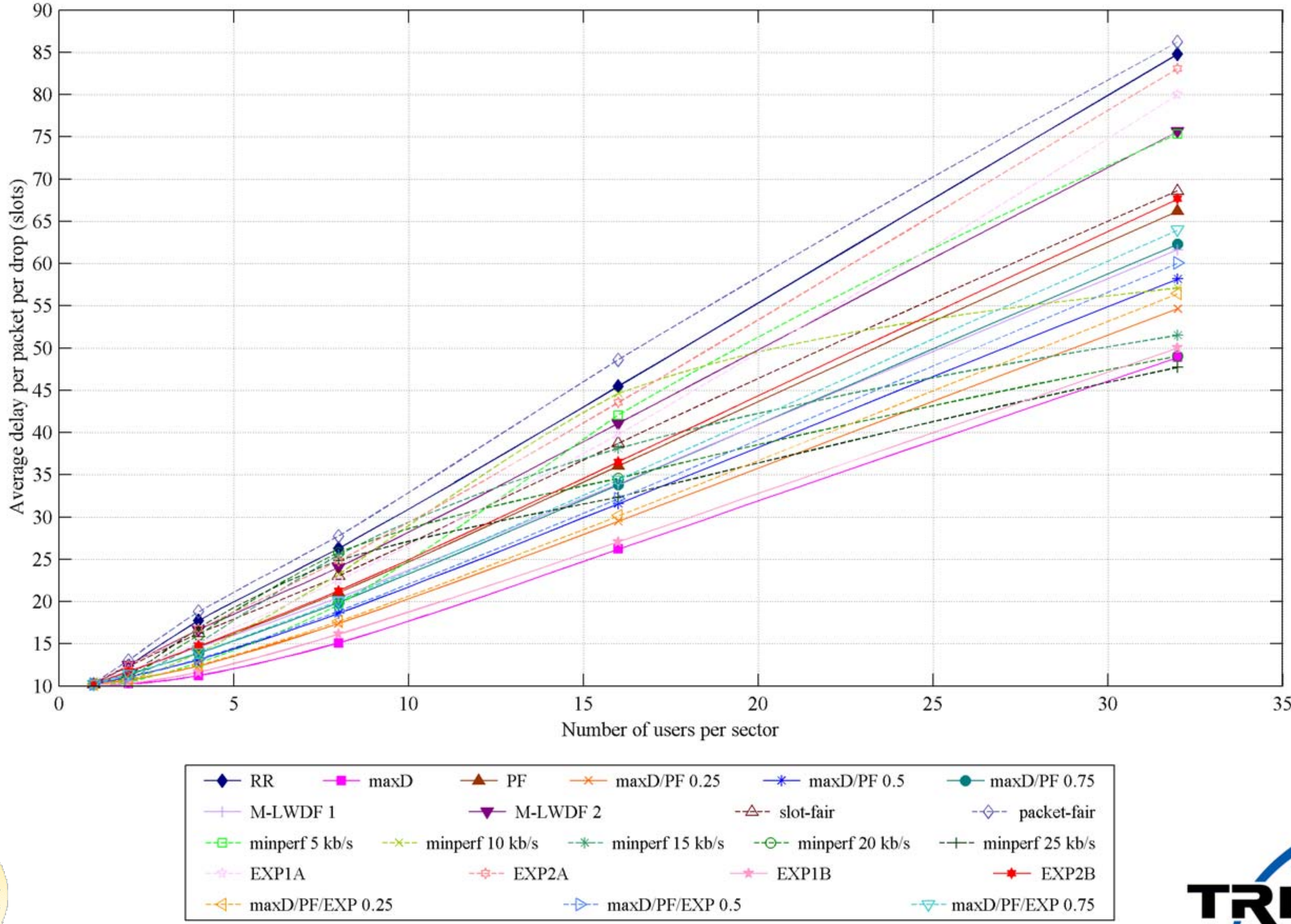
Selected Simulation Results

Average sector throughput per drop: ITU vehicular B channel with margins



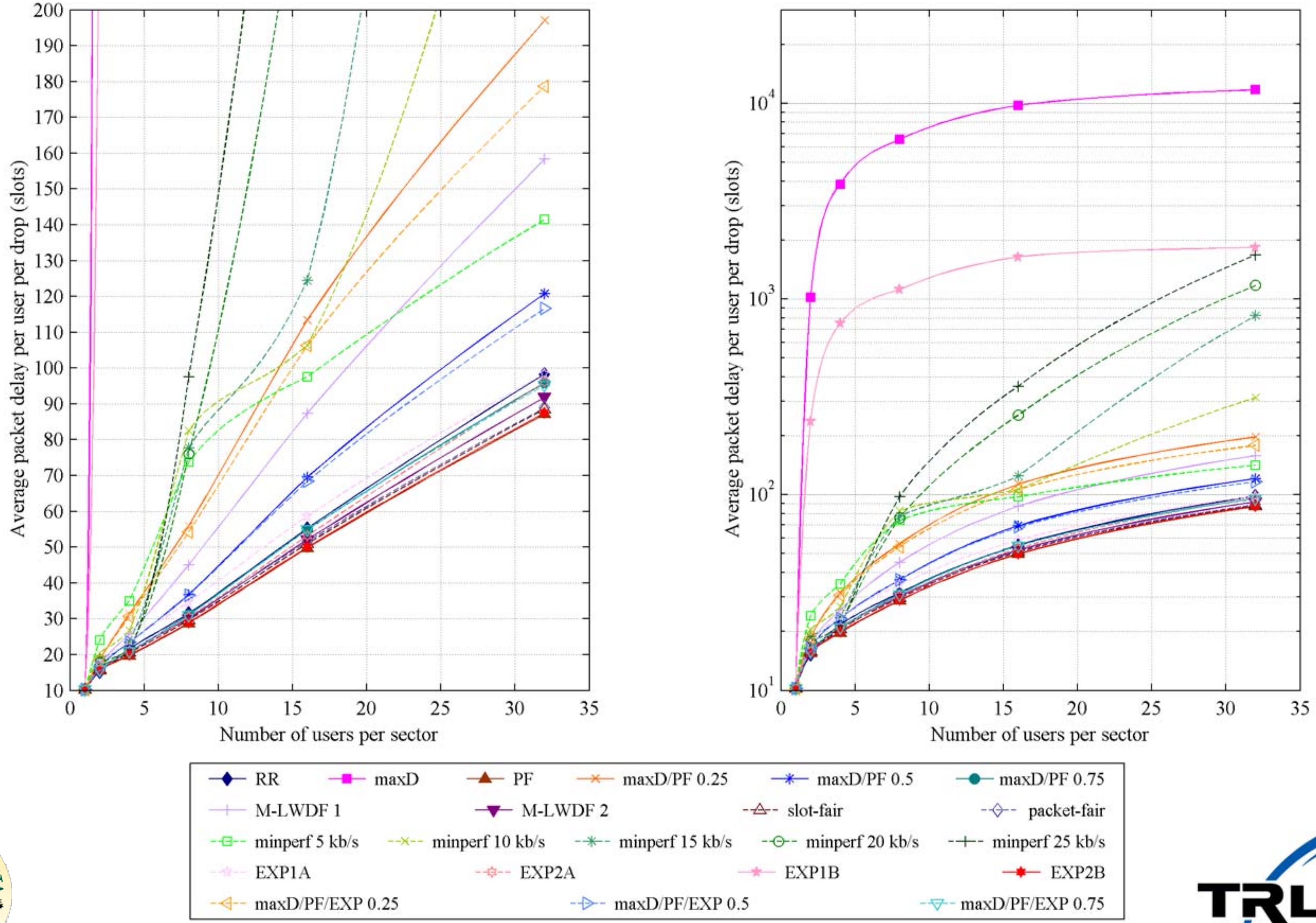
Selected Simulation Results

Average delay per packet per drop: ITU vehicular B channel with margins



Selected Simulation Results

Average packet delay per user per drop: Vehicular channel with margins



UNIVERSITY OF ALBERTA QUAECUMQUE VERA

TRLabs

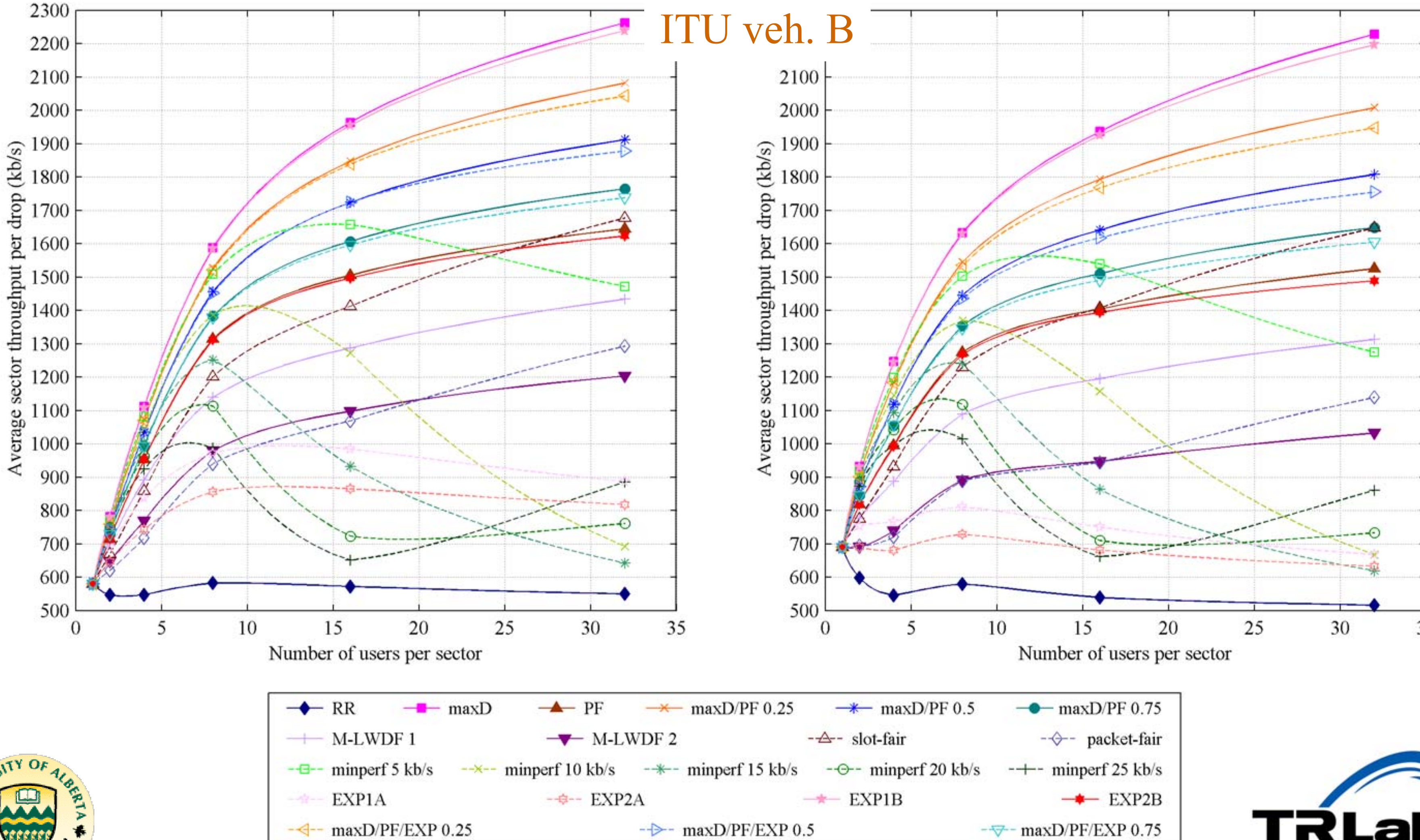
Selected Simulation Results

Average sector throughput per drop: Perfect prediction comparisons

Perfect 1-slot prediction + margins

Perfect multislot prediction

ITU veh. B



Conclusions for SISO Scheduling (1)

- Tradeoff exists between throughput and delay
  - Algorithms that give large average sector throughputs also display low average delay per packet, but yield high average delays per user, as well as large variations in throughputs, delays, and allocated slots across the set of users
  - Algorithms that control delay sacrifice throughput and result in higher average delays, but experience smaller variations in throughputs and delays across users, and have the smallest likelihood of having packets with extremely long delays
- PF, slot-fair, EXP2B algorithms provide best balance between throughputs, delays, slots, and fairness
- maxD, minperf, EXP1B algorithms show high throughputs, but low fairness
- RR, M-LWDF, EXP1A, EXP2A algorithms provide good fairness, but sacrifice throughput

Conclusions for SISO Scheduling (2)

- The use of outdated channel information plus margins can be an effective replacement for prediction, provided that the channel estimates are reasonably good
  - If estimates are bad, all performance measures suffer
  - If good, use of margins can even provide higher throughput than perfect prediction for some algorithms
- "PF A" and "PF B" methods of calculating the average rate in algorithm metrics in most cases did not change performance significantly
  - Exceptions: "minperf" algorithm in all cases, all algorithms in ITU veh. B channel with outdated channel information: increased throughput, decreased delay per packet, etc., when the "PF B" approach was used
- The algorithms can also be used for other similar TDM wireless packet services, with the same general trends & relations between them.

Some Related Publications

1. R.C. Elliott, W.A. Krzymień, “Scheduling algorithms for high throughput packet data service in cellular radio systems”, *Can. J. of Electrical & Computer Engineering*, Special Issue on Advances in Wireless Communications and Networking, vol. 29, no. 1/2, January/April 2004, pp. 117-127 .
2. R.C. Elliott, W.A. Krzymień, “Scheduling algorithms for the cdma2000 packet data evolution”, in the *Proc. 2002 IEEE Semi-Annual Vehicular Technology Conference (VTC2002-Fall),* Sept. 2002, Vancouver, BC, Canada, pp. 304-310.
3. C. H. Rentel, W. A. Krzymień, B. Darian, V. Vanghi, R. Elliott, "Comparative forward link data channel performance evaluation of HDR and 1XREME systems", in the *Proc. 2002 IEEE Semi-Annual Vehicular Technology Conference (VTC2002-Spring)*, Birmingham, Alabama, USA, May 2002, pp. 160-164.

Università degli Studi di Padova
Bressanone Summer School
27 giugno 2005

Spectrally Efficient High Throughput SISO Spread Spectrum OFDM Cellular Downlink for Packet Data

Robert Novak & Witold Krzymień
University of Alberta / TRLabs
Edmonton, Alberta, Canada

Outline of Section 2

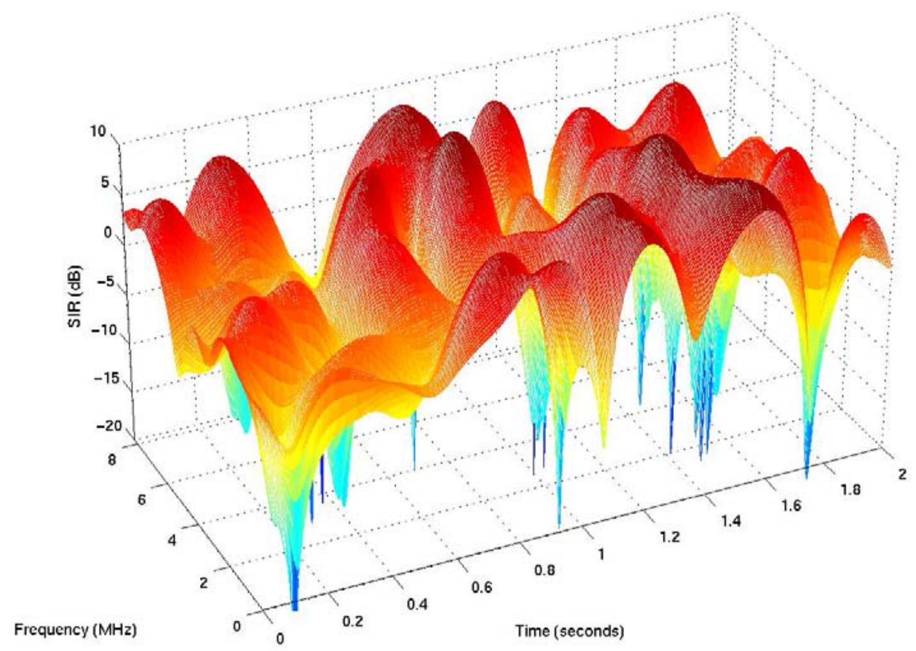
- Introduction
  - Frequency selectivity and temporal fading
- System description
  - Two-dimensional dynamic resource allocation
- System features and performance
  - Simulation parameters
  - Allocation of disjoint sub-bands
  - SIR quantization
  - Synchronous and asynchronous re-transmission
  - Maximum re-transmission interval constraint
- Conclusions
- References

Introduction

- Spectrally efficient packet data access will be one of the main features of future cellular systems.
- Internet access requires duplex transmission. However, usually throughput required on the downlink (forward link) is much higher than on the uplink (reverse link).
- Packet data transmission is delay tolerant.
- Several single carrier cellular packet data access systems using adaptive transmission techniques have been standardized and their deployment is under way.
- A downlink spread spectrum OFDM system with frequency and time domain allocation (SS-OFDM-F/TA) of radio resources to maximize throughput is described
  - Dynamic allocation based on propagation conditions of mobile users
  - Significant improvement on conventional *location-dependent radio resource allocation* for SS-OFDM systems (see slide 24)

Frequency Selective and Temporal Fading

- The received SINR varies in time and frequency.
- Small-scale fading for each user is independent.
- Packet data service is delay tolerant.
- SINR based allocation can be best achieved by dividing the time-frequency plane into small rectangles.

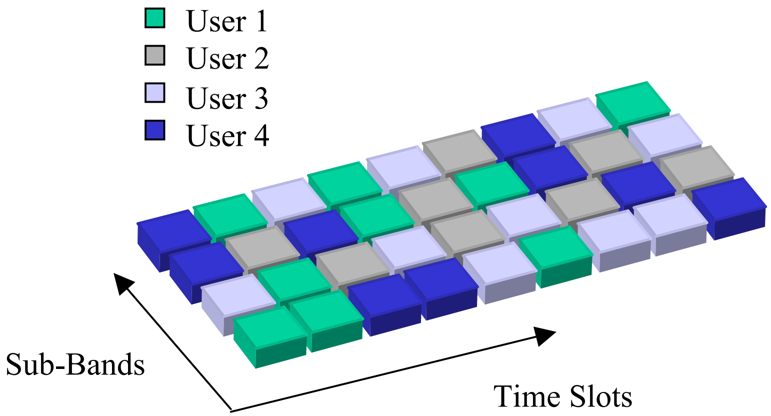


- The objective is to transmit to the mobile user with the highest SINR in a given frequency sub-band and time slot. In this manner, high channel gains are exploited while spectral nulls and deep fades are avoided.

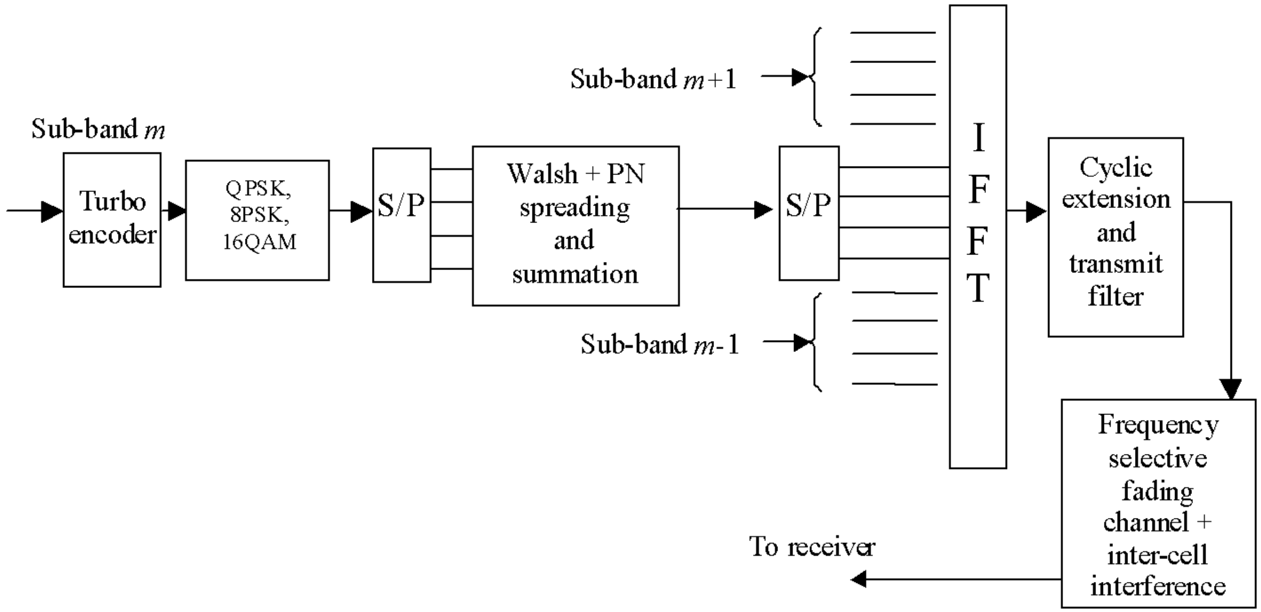
Two-Dimensional Resource Allocation

- Division of frequency band into *M* sub-bands, and division of time into time slots
  - Propagation condition dependent sub-band and time slot allocation to users
    - Proportionally fair scheduling
    - Allocation to a user with the highest ratio of supportable to average data rate
  - Due to correlation of adjacent subcarriers, allocation of groups of subcarriers (sub-bands) is sufficient
  - Any user may receive transmission in any sub-band during any time slot to exploit multi-user diversity

- One packet per sub-band(s)
  - One mobile may receive different packets over more than one sub-band
  - Transmission (and re-transmissions) of a packet over several disjoint sub-bands is also considered

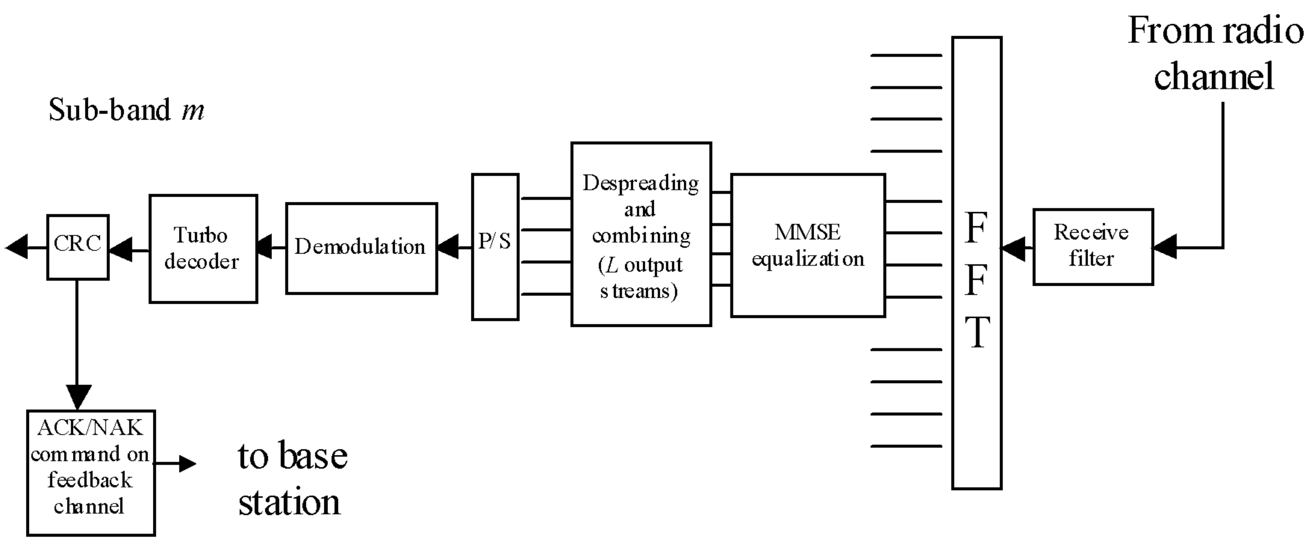


Transmitter



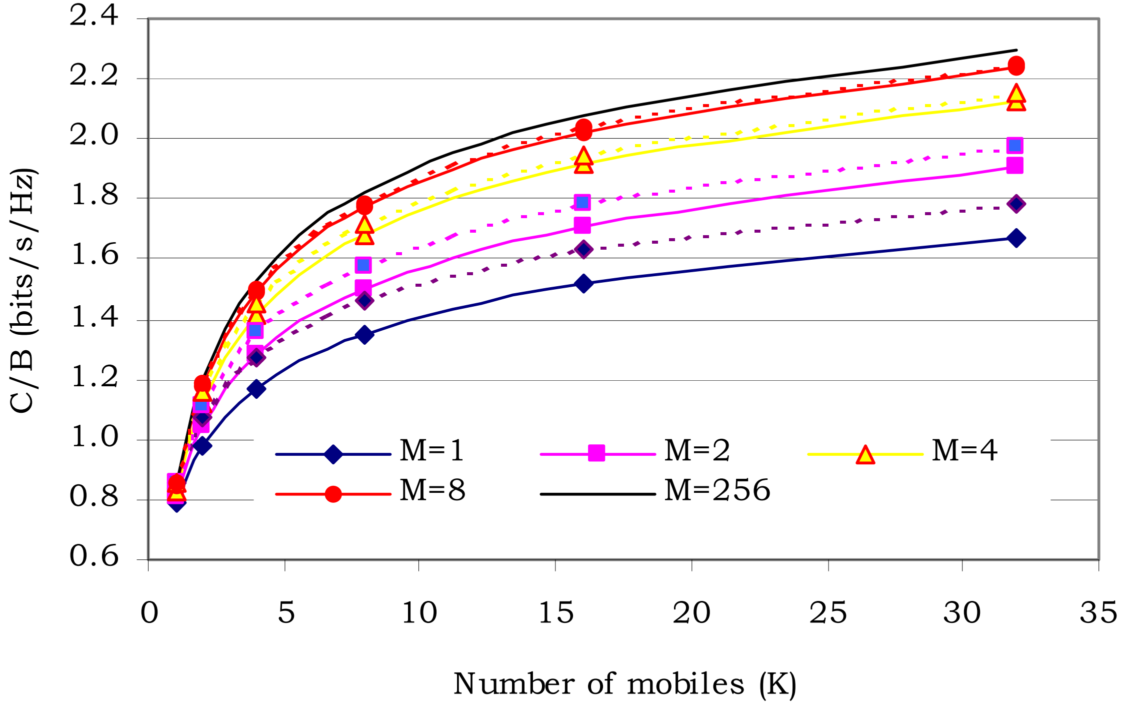
- Each sub-band may be used by a different user
- Adaptive coding and modulation of each packet transmission
- Walsh sequences are concatenated with sector specific PN sequences
- Spreading is primarily used for better inter-cell interference mitigation and base station identification

Receiver



- MMSE equalization is used before despreading
- User receives only sub-band(s) intended for it
- ACK/NAK feedback after each transmission allows for early termination of multi-slot packets.

Capacity of SS-OFDM Systems



Capacity of SS-OFDM (solid) and OFDM (dashed) systems.

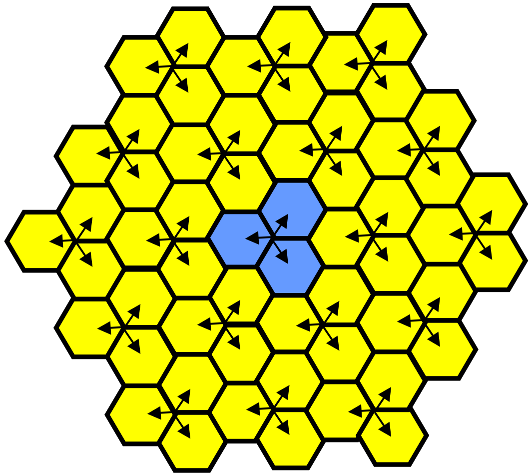
10 MHz bandwidth, ITU Indoor B channel, $N = 256$ subcarriers

Average Es/No = 0 dB.

- For M >4, OFDM and SS-OFDM systems have nearly the same capacity. SS-OFDM has some advantages for multi-cell applications, so it is considered.
- $M = 8$ instead of $M = 256$ sub-bands (subcarrier allocation) decreases the demand on channel state information feedback by a factor of 32, with negligible loss in capacity.
- Conventional SS-OFDM (no multi-user diversity and spreading over entire transmission bandwidth) is equivalent to point $M = 1$, $K = 1$.

System Features and Performance

Radio Channels and Simulation Parameters (1)



Indoor B		Pedestrian B	
Delay (ns)	Power (dB)	Delay (ns)	Power (dB)
0	0.0	0	0.0
100	-3.6	200	-0.9
200	-7.2	800	-4.9
300	-10.8	1200	-8.0
500	-18.0	2300	-7.8
700	-25.2	3700	-23.9

- Simulations organized into runs:
  - Simulation runs are 30 seconds long; 100 runs considered for each set of parameters. A different set of user locations in each run.
- Exponentially correlated log-normal shadowing process (decorrelation distance of 5 metres); standard deviation of 12 and 10 dB for Indoor and Pedestrian. Rayleigh small-scale fading. Correlation of antennas between different base stations = 0.5; between different sectors of the same BS = 1.0.
- An embedded sector in a layout of 19 3-sector cells simulated.

Radio Channels and Simulation Parameters (2)

	Indoor	*Pedestrian*
N (subcarriers)	256	512
M (sub-bands)	8	4
Bandwidth	10 MHz	5 MHz
Main lobe bandwidth	9.4787 MHz	4.7393 MHz
T_s (SS-OFDM symbol duration)	27.008 μs	108.032 μs
T_g (cyclic prefix)	1 μs	4 μs
Time slot length	1.7925/1.344 ms	1.7925/1.344 ms
Data symbols per sub-band/time slot	1600/1200	1600/1200
Pilot & MAC symbols per transmission	448/336 (21.875%)	448/336 (21.875%)

In general, the system in the pedestrian channel has 32 sub-bands, allocated in sets of 8 sub-bands.

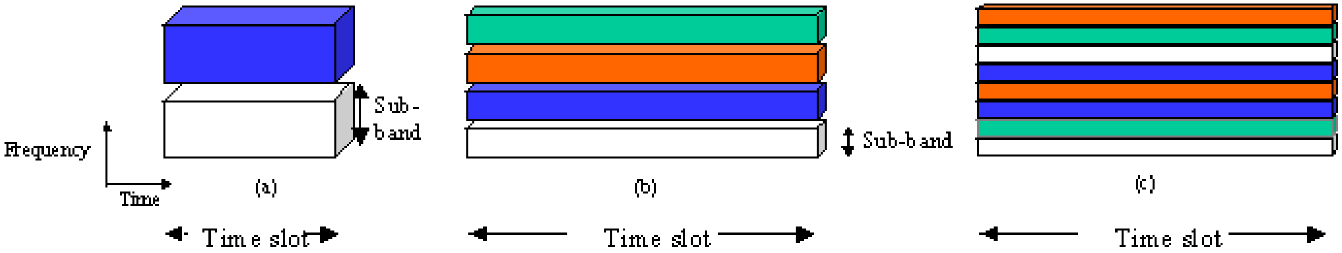
- Outdoor pedestrian channel users are moving at low speeds
  - But, urban environment with buses or cars
  - We consider this environment with maximum Doppler shifts from 5 to 100 Hz.
  - ITU Vehicular A channel is very similar to ITU Pedestrian B channel
- Allocation is based on out-dated channel estimates (delay of 3 slots)
  - Also consider perfect prediction (p.p.) with 100 Hz Doppler shift

Design Challenges

- Efficient 2-dimensional resource allocation requires rectangular segments on the time-frequency plane where the channel is effectively constant, and a good estimate of the channel gain

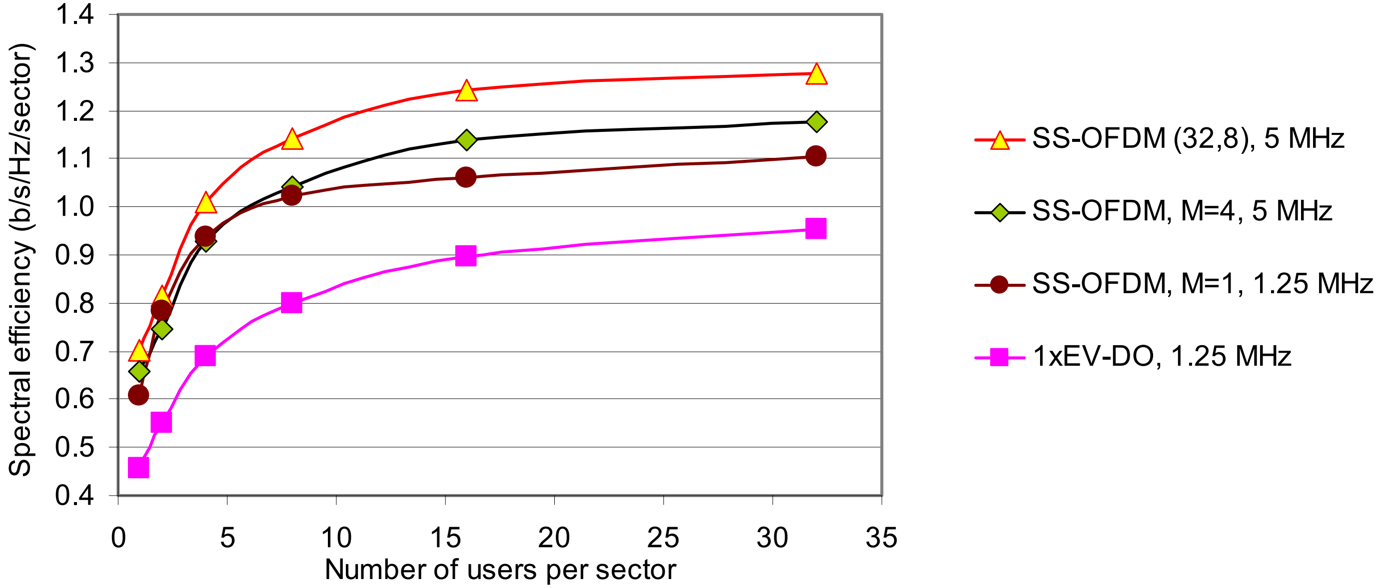
- How do we do this if …
  - Highly-frequency selective channel
  - High mobile velocity
  - Need to maintain sufficiently large packet sizes
    - Good turbo code performance
    - Maximize payload per allocation in order to minimize overhead
  - Limited amount of channel state information (CSI) feedback

Allocation of Disjoint Sub-Bands



- Sub-band grouping
  - The formation and allocation of groups of sub-bands is adaptive
  - SINR estimates in every sub-band are sent on the feedback channel.
  - Allocation is based on the supportable rate in the M_s best sub-bands for each user
    - The remaining M-M_s sub-bands are allocated in the same way to other users/packets
- Advantages
  - Allocate several disjoint sub-bands (dynamically) together for transmission of a single packet without lengthening the time slot (increases throughput)
  - Throughput always as good or better than with decreased packet size, while requiring less forward link signalling
  - Simple, "conflict-free" allocation algorithm

Comparison to Single Carrier Systems



- Average spectral efficiency per sector of SS-OFDM-F/TA in comparison to 1xEV-DO[1] in an ITU Pedestrian B channel; identical packet structures, σ_{shad} = 6.5 dB.
- Allocating groups of sub-bands increases throughput
  - Users with poorer channel conditions benefit the most
- Sub-band format makes the system easily scalable in frequency
- Maximizes efficiency when system load (i.e. number of users per MHz) is low

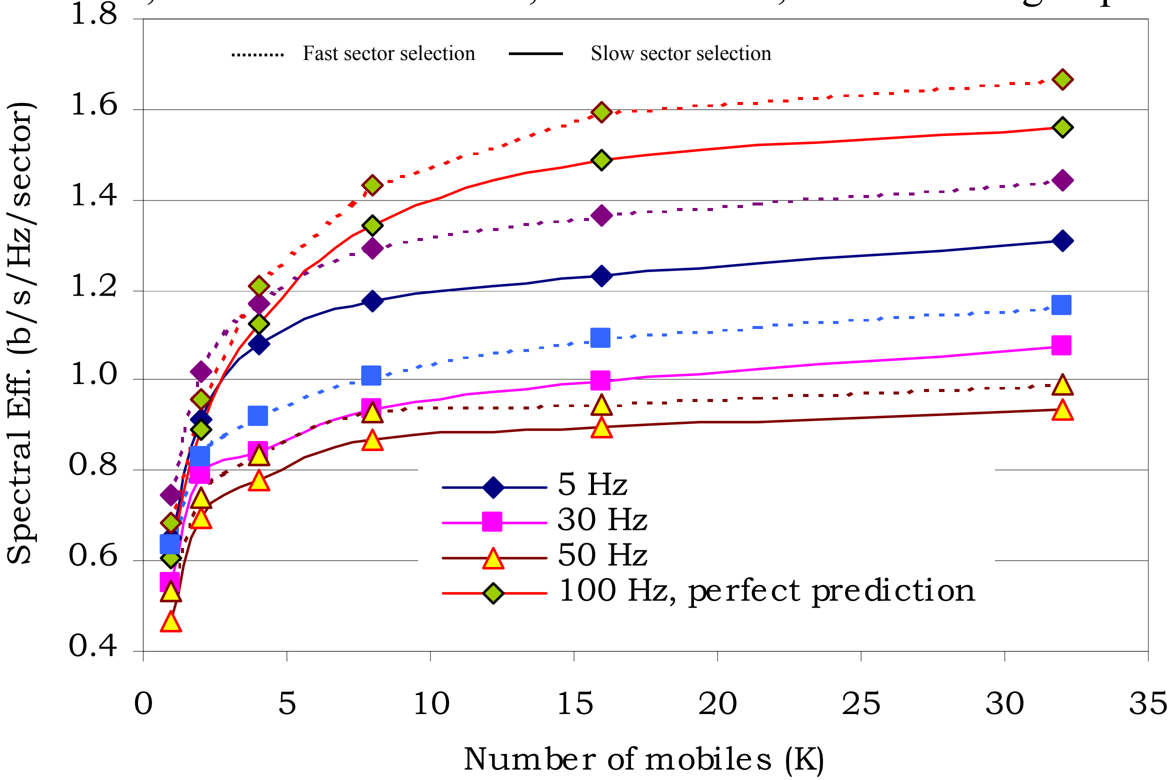
[1]Estimated from: R. Elliott, "*Scheduling Algorithms for High Throughput Packet Data Service in Cellular Radio Systems*", M.Sc. Thesis, Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB., Fall 2003.

Best Sector Selection

- Slow sector selection
  - Selection every 30 seconds
  - Simple to implement
- Fast sector selection
  - Selection every second (1 Hz)
    - Adequate to average out small scale fading
    - Shorter than correlation time of shadowing process (~6 seconds at 5 Hz maximum Doppler shift)
  - Network signalling is much more demanding

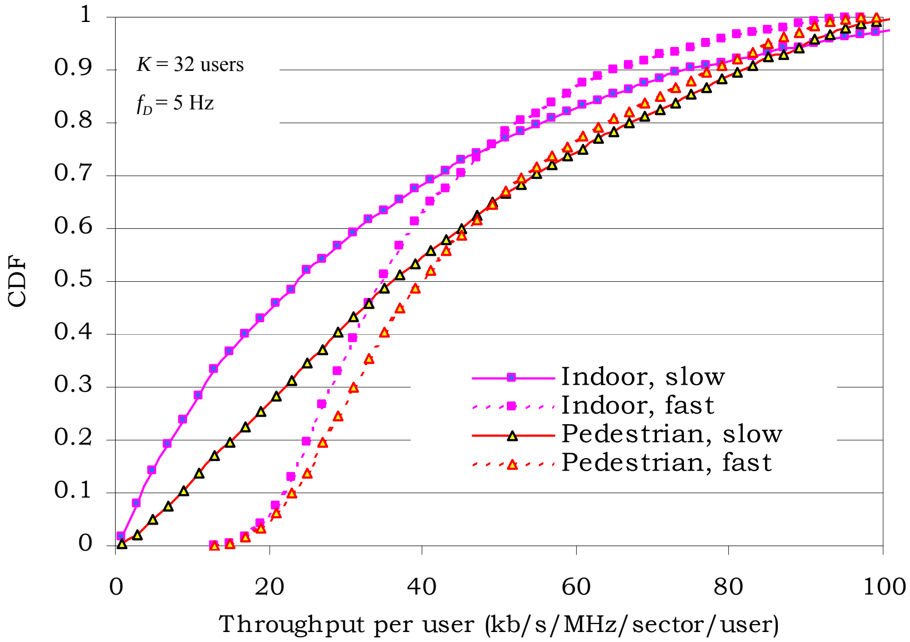
Throughput at Various Mobile Speeds

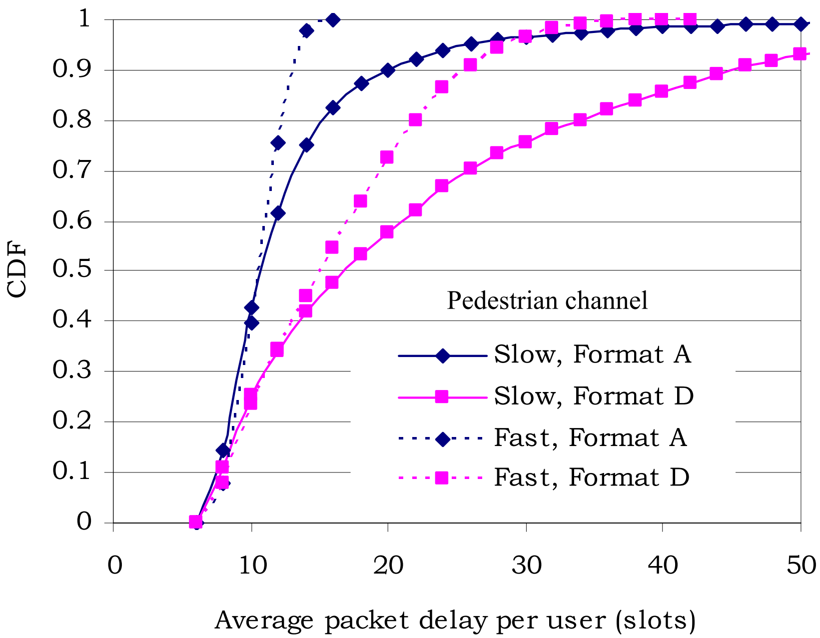
5 MHz bandwidth, Pedestrian B channel, 32 sub-bands; allocation in groups of 8 (1.34 ms slots)



- Throughput decreases with increased mobile speed (CSI outdated by 3 slots); still significant multi-user diversity gain at > 50 Hz
- With perfect prediction of CSI, high Doppler shift provides significant time diversity increasing throughput
- Fast sector selection increases aggregate throughput; increases the throughput of the 'worst' mobiles.

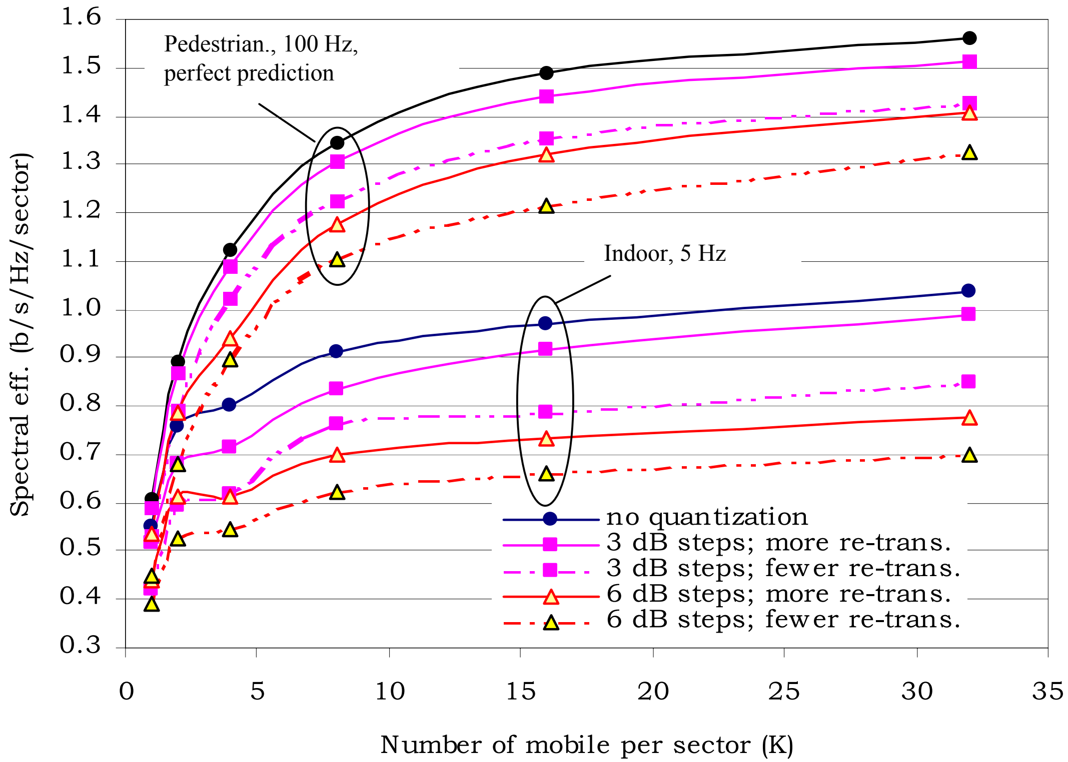
CDF of Throughput and Delay





- Fast sector selection at 1Hz improves throughput and fairness:
  - Increase in throughput highest for users with worst channels
  - Lower delays; prevents long-term degradation of user's conditions
- Probability of actual sector switch is 7 - 12%.

Throughput with Quantized SIR Feedback



In the Pedestrian B channel, sub-band grouping is used (32,8).

Allocation and formation of groups is based on quantized CSI only.

- Can still exploit multi-user diversity with coarse CSI feedback
  - Throughput at 100 Hz Doppler with prefect prediction and 6 dB CSI steps is higher than with continuous CSI and 5 Hz Doppler shift.

Type II Hybrid ARQ: Packet Re-Transmissions

- Synchronous re-transmissions
  - If NACK is received, re-transmission occurs every 4 slots in the same sub-band
  - Re-transmissions occur every 4 slots regardless of the channel conditions until the packet is successfully delivered or the maximum allowed number of re-transmissions is reached.

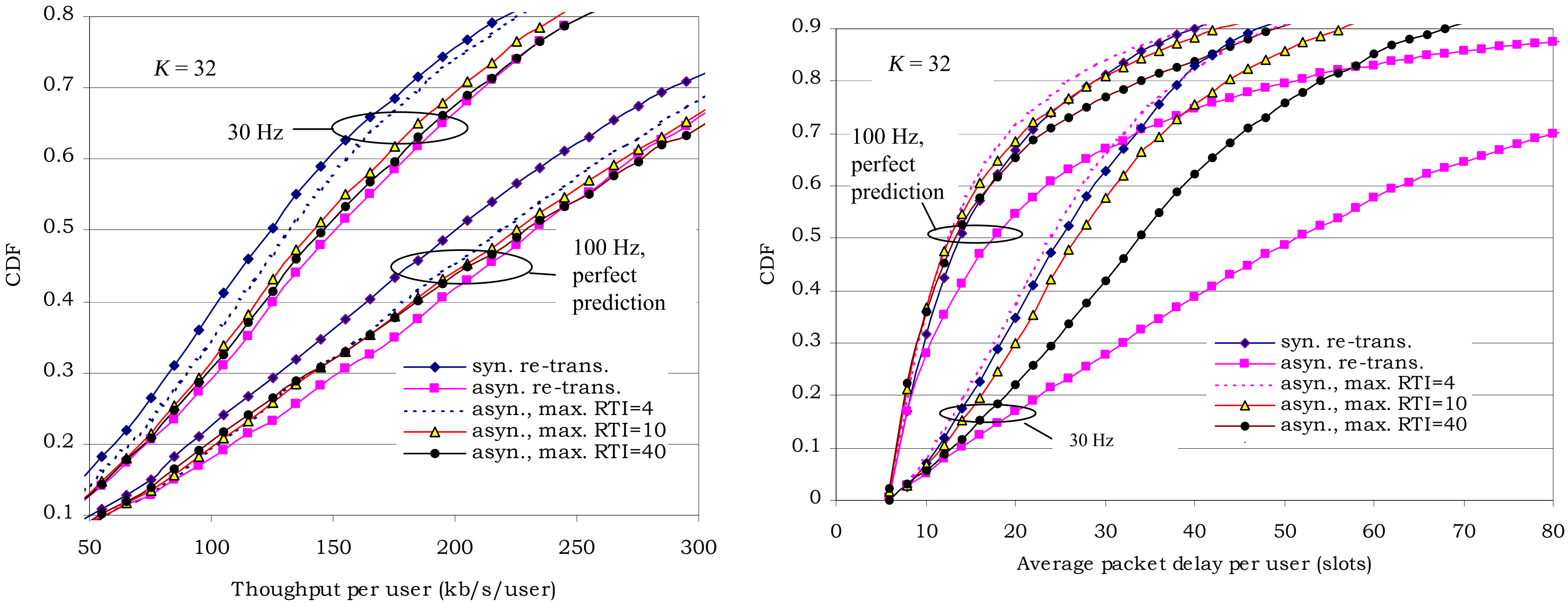
- Asynchronous re-transmissions
  - Re-transmit when scheduler chooses the user again. Several algorithms for selecting:
    - Start a new packet
    - Re-transmit a packet
      - which packet, as there are up to 4*M*
  - Another option: re-transmit when channel conditions are as good as or better than during the original transmission

Asynchronous Algorithms and Max. RTI (1)

- Best results from minimum delay with re-transmit priority and parallel allocation algorithm (**MD-RP-PA)**
  - Determine M_k out of M sub-bands for user k to receive transmissions
  - Find the M_k packets that have been waiting the longest for re-transmission intended for user k
  - Arrange packets so that the lowest transmission format (e.g. QPSK, rate 1/5) is used on the sub-band with the lowest SIR, and so on.
- Low delay (in comparison to other asynchronous algorithms), high throughput
  - However, packet delays are much longer than with synchronous transmission
- Constrained maximum re-transmission interval (RTI) in order to minimize delays associated with asynchronous re-transmission
  - A packet is selected for transmission if it has been waiting too long
  - Prioritized over all other transmissions
  - In sub-band grouping, sub-bands re-allocated for this transmission

Asynchronous Algorithms and Max. RTI (2)

5 MHz bandwidth, Pedestrian B channel, 32 sub-bands; allocation in groups of 8 (1.34 ms slots)



- Asynchronous re-transmissions with constrained maximum RTI
  - Higher throughput compared to synchronous
  - Delay comparable to synchronous re-transmissions
  - Users with average channel conditions benefit the most from increased throughput

Conclusions

- The SS-OFDM-F/TA system can achieve high throughputs
  - nearly 1.7 b/s/Hz/sector with prediction and fast sector selection using ITU recommended channel models.
- Allocation of groups of sub-bands is a simple, and effective means of exploiting highly frequency-selective channels
  - maintains reasonable packets sizes, low signalling overhead
- Several asynchronous algorithms have also been proposed.
  - typically, increased throughput at a cost of increased delay (minimally in case of min RTI algorithms) has been achieved.
- The system has also been shown to perform well even with coarse quantization of the SIR estimates.
- Constellations greater than 16 QAM have not been considered to avoid the need for complex receiver structures.

Selected References

[1] R. Novak, W. A. Krzymień, "A downlink SS-OFDM-F/TA packet data system employing multi-user diversity", *in Proc. of the 3rd Intl. Workshop on Multi-Carrier Spread-Spectrum (MCSS 2001)*, Oberpfaffenhofen, Germany, pp. 181-190, Sept. 2001.

[2] R. Novak, W. A. Krzymień, "An adaptive downlink spread spectrum OFDM packet data system with two-dimensional radio resource allocation: performance in low-mobility environments", *in Proc. WPMC'02*, Honolulu, HI, Oct. 2002.

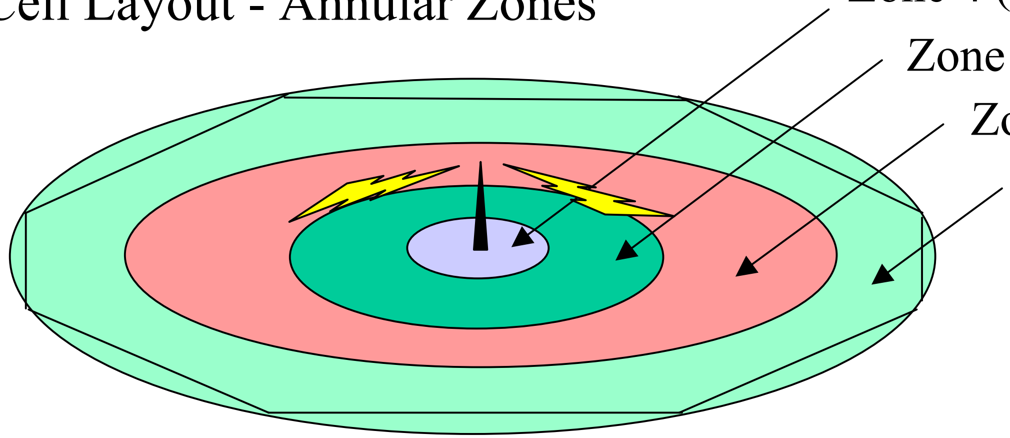
[3] R. Novak, W. A. Krzymień, "SS-OFDM-F/TA system packet size and structure for high mobility cellular environments", *in Proc. VTC'03-Spring*, Jeju, Korea, April 2003.

[4] R. Novak, W. A. Krzymień, "Packet re-transmission options for the SS-OFDM-F/TA system", *in Proc. of the 4rd Intl. Workshop on Multi-Carrier Spread-Spectrum (MCSS 2003)*, Oberpfaffenhofen, Germany, pp. 89-100, Sept. 2003

[5] R. Novak, W. A. Krzymień, "Efficient packet data service in a spread spectrum OFDM cellular system with 2-dimensional radio resource allocation", *European Transactions on Telecommunications*, vol. 15, no. 3, pp. 185-199, May-June 2004.

Location Dependent Allocation (addnl. slide)

Cell Layout - Annular Zones



The maximum data rate per user in a given zone is fixed

- Serving users near the cell boundary is costly
  - High power transmission: creates large intra-cell and inter-cell interference
  - Low throughput due to poor large-scale channel conditions, limits cell throughput
- Location Dependent Radio Resource Allocation (resource = code channels)
  - SS-OFDM, coded QPSK, zero-delay constraint, multi-code system, power control
  - Allocate (reserve) more resources to users near transmitter, less to distant users
  - Increases throughput, reduces transmit power: inter- and intra-cell inference is lower
  - Distant users still served according to some minimum rate guarantee (limiting factor in cellular throughput)
  - No channel state information at transmitter; only power control "up/down" feedback and mobile location (or equivalently, large-scale channel conditions)

Transmission Formats (additional slide)

Max. no. of slots	Modulation	Code rate (approx.)	Number of complete trans.	Data rate (Kb/s/sub-band)	Packet size (bits)	SIR for 1% PER	Max. no. of slots	Modulation	Code rate (approx.)	Number of complete trans.	Data rate (Kb/s/sub-band)	Packet size (bits)	SIR for 1% PER
8	QPSK	1/5	9.4-1.2	37.9-303.5	408	-13.1	8	8 PSK	1/5-11/17	2.5-1	216.5-1731.6	2328	-5.0
4	QPSK	1/5	4.7-1.2	75.9-303.5	408	-10.0	4	8 PSK	1/5-11/17	1.2-1	432.9-1731.6	2328	-1.8
2	QPSK	1/5	2.4-1.2	151.7-303.5	408	-6.9	2	8 PSK	1/3-11/17	1	865.8-1731.6	2328	1.7
1	QPSK	1/5	1.2	303.5	408	-3.7	1	8 PSK	11/17	1	1731.6	2328	8.3
8	QPSK	1/5-1/3	4.8-1	73.6-589.1	792	-10.5	8	16 QAM	1/5-2/3	2.5-1	287.9-2302.9	3096	-3.0
4	QPSK	1/5-1/3	2.4-1	147.3-589.1	792	-7.3	4	16 QAM	1/5-2/3	1.2-1	575.7-2302.9	3096	0.3
2	QPSK	1/5-1/3	1.2-1	294.6-589.1	792	-4.0	2	16 QAM	1/3-2/3	1	1151.5-2302.9	3096	3.7
1	QPSK	1/3	1	589.1	792	-0.7	1	16 QAM	2/3	1	2302.9	3096	10.5
8	QPSK	1/5-13/20	2.5-1	145-1160.4	1560	-7.5	8	16 QAM	1/5-4/5	2-1	359.3-2874.2	3864	-2.0
4	QPSK	1/5-13/20	1.2-1	290.1-1160.4	1560	-4.3	4	16 QAM	1/5-4/5	1	718.5-2874.2	3864	1.0
2	QPSK	1/3-13/20	1	580.2-1160.4	1560	-1.0	2	16 QAM	2/5-4/5	1	1437.1-2874.2	3864	5.9
1	QPSK	13/20	1	1160.4	1560	5.0	1	16 QAM	4/5	1	2874.2	3864	13.7

- 24 Possible Raw Formats
  - QPSK, 8 PSK, and 16 QAM, turbo coding with 1/5 mother rate
  - 1536 symbols per sub-band/time slot; 1200 data, and 336 MAC and pilot symbols
  - Preamble assumed to be part of MAC & pilot overhead
  - Type II Hybrid ARQ with 1 to 8 possible re-transmissions

Packet Format Sets (additional slide)

*Packet format set **A***			*Packet format set **B***			*Packet format set **C***			*Packet format set **D***		
Max. no. of slots	*Packet size (bits)*	*SIR at 1% PER (dB)*	*Max. no. of slots*	*Packet size (bits)*	*SIR at 1% PER (dB)*	*Max. no. of slots*	*Packet size (bits)*	*SIR at 1% PER (dB)*	*Max. no. of slots*	*Packet size (bits)*	*SIR at 1% PER (dB)*
8	408	-13.1	8	408	-13.1	8	408	-13.1	8	408	-13.1
4	408	-10.0	4	408	-10.0	4	408	-10.0	8	792	-10.5
2	408	-6.9	2	408	-6.9	4	792	-7.3	8	1560	-7.5
1	408	-3.7	2	792	-4.0	4	1560	-4.3	8	2328	-5.0
1	792	-0.7	2	1560	-1.0	4	3096	0.3	8	3864	-2.0
2	2328	1.7	2	2328	1.7	2	3096	3.7	4	3096	0.3
1	1560	5.0	2	3096	3.7	2	3864	5.9	2	3096	3.7
1	2328	8.3	2	3864	5.9	1	2328	8.3	2	3864	5.9
1	3096	10.5	1	2328	8.3	1	3096	10.5	1	2328	8.3
1	3864	13.7	1	3096	10.5	1	3864	13.7	1	3096	10.5
			1	3864	13.7				1	3864	13.7

- 4 sets of transmission formats (selected from formats on last slide)
  - A given system design will use one of these 4
  - Different number of maximum re-transmissions allowed;
    - ***A*** allows fewest (low delay, low adaptation) and ***D*** allows the most (higher delay, high adaptation)
  - PER curves for each format, and number of retransmissions, used in system level simulations

Università degli Studi di Padova
Bressanone Summer School
27 giugno 2005

High Throughput Downlink Wireless Packet Data Access with Multiple Antennas and Multi-User Diversity

David Mazzarese & Witold Krzymień
University of Alberta / TRLabs
Edmonton, Alberta, Canada

Outline of Section 3

I. Introduction and background
- The MIMO broadcast channel (BC) model
- The SISO broadcast channel and multiuser diversity
- Capacity of multiuser MIMO channels
- A brief introduction to dirty-paper coding (DPC)

II. Throughput maximization in Gaussian channels
- Asymptotically optimal solution with single-antenna receivers
- Asymptotically optimal solution with multiple-antenna receivers

III. Maximum-throughput scheduling algorithms for MIMO broadcast fading channels
- Low-complexity scheduling algorithms
- Relation to receive antenna selection algorithms

IV. Spatial multiplexing by linear processing
- Throughput maximization: constraints, algorithms and performance
- Proportionally-fair scheduling: impact on delay and complexity

V. Conclusions

I. INTRODUCTION

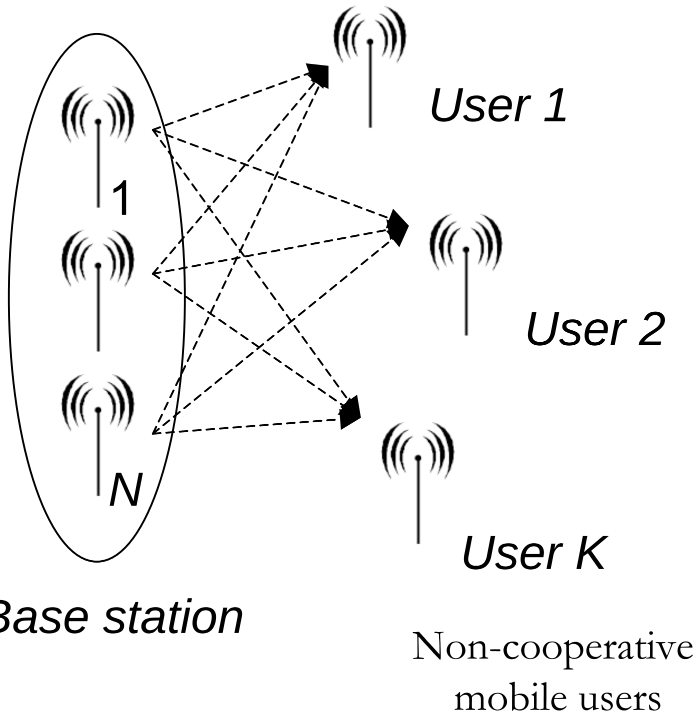
I.A. The MIMO Broadcast Channel Model
I.B. The SISO Broadcast Channel and Multiuser Diversity
I.C. Multiuser MIMO Channels
- Single-user transmission strategies
- Performance limits with optimal signalling (Dirty Paper Coding)

I.D. A brief introduction to Dirty-Paper Coding

I.A. Introduction

System model and notation

MIMO Broadcast Channel



(N, M_k, K) MIMO BC

- N - number of transmit antennas
- M_k - number of receive antennas of user k
- K - number of mobile users
- P - total transmit power constraint

Both the transmitter and the receivers know the channel gains.

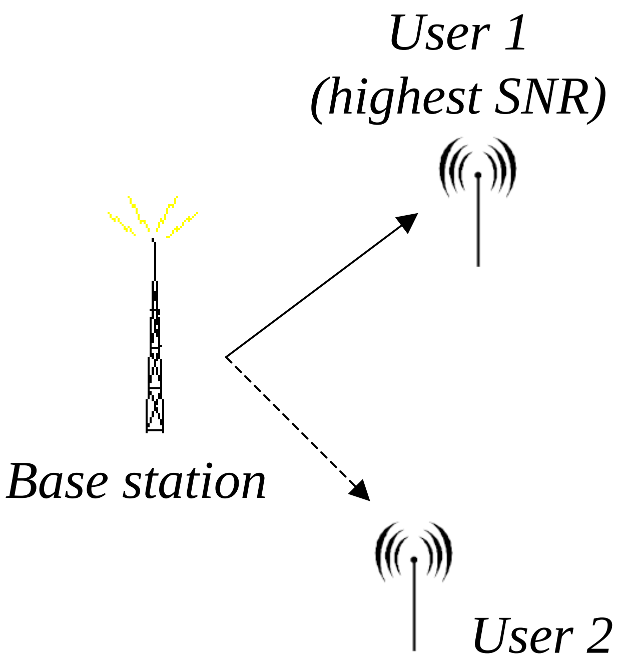
- Channel matrices remain constant for one time slot (quasi-static fading).
- Average throughput simulations assume independent fading in space and time

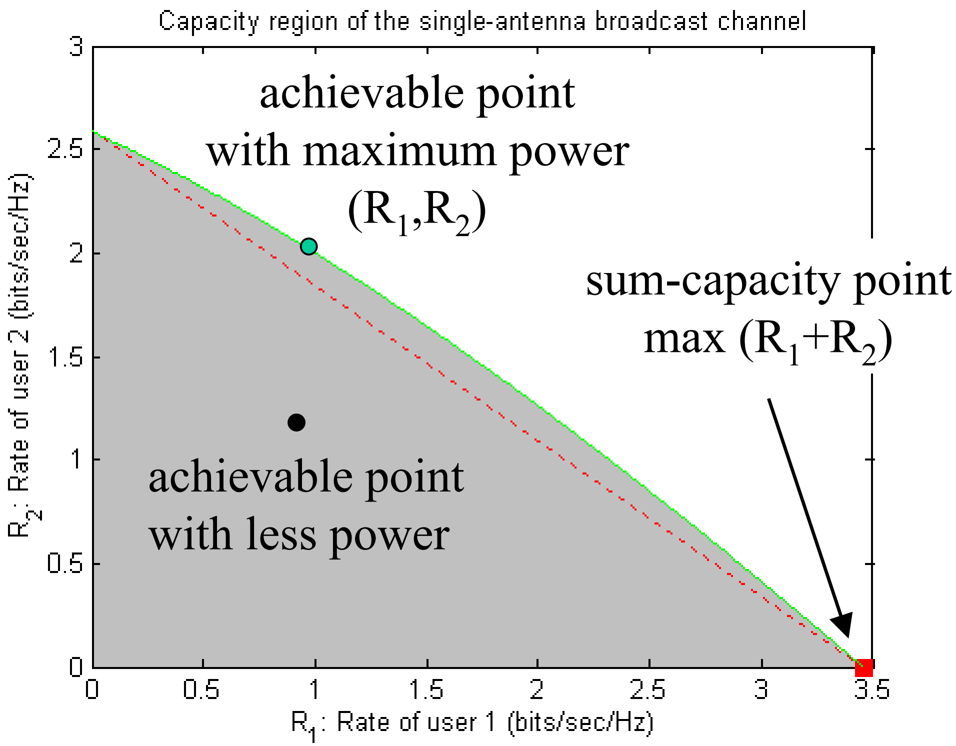
I.B. Introduction

SISO multiuser broadcast channel: 2-user capacity region

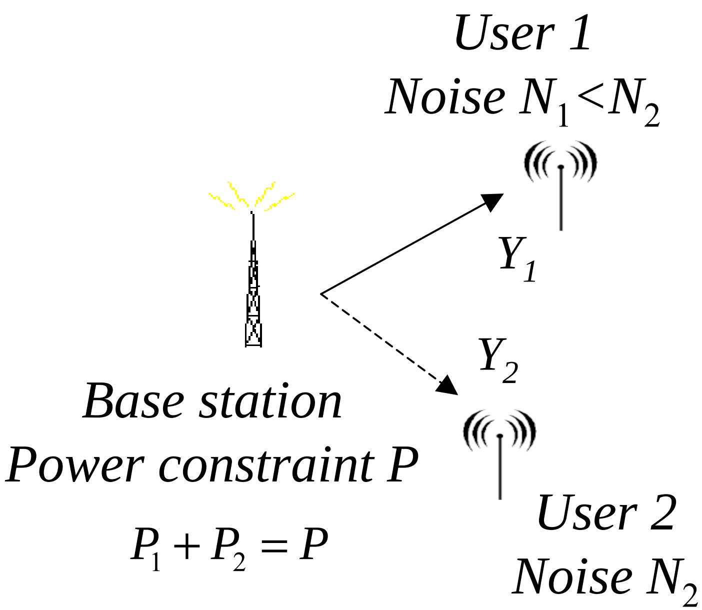
Why transmit to a single user at a time with maximum power? [1]

(1,1,2) BC

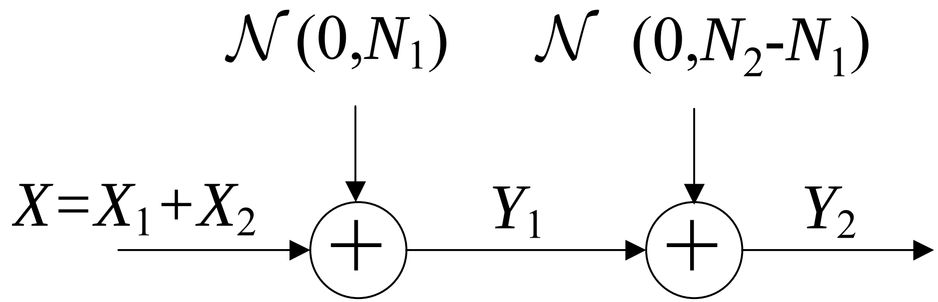




Optimal signalling: 2-user broadcast channel (single-antenna case)



This is a *degraded* broadcast channel [2]



Achievable rates

$$R_1 = \frac{1}{2}\log\left(1+\frac{P_1}{N_1}\right) \quad \& \quad R_2 = \frac{1}{2}\log\left(1+\frac{P_2}{N_2+P_1}\right)$$

Coding

- The capacity-achieving coding scheme is called superposition coding.
- The two users use independent codebooks of rate R_1 and R_2
- The two codewords are added together.

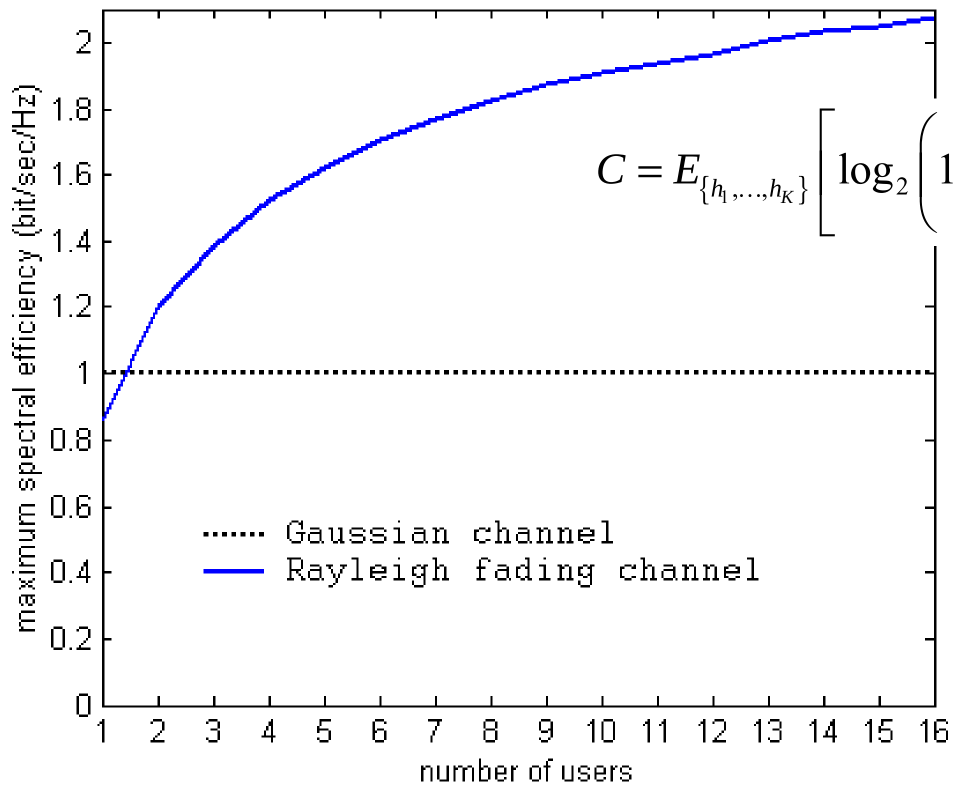
Successive Decoding

- User 1:
    - first decode X_2 from Y_1
    - subtract X_2 from Y_1
    - then decode X_1
- User 2:
    - decode X_2 by treating X_1 as interference

Introduction

SISO multiuser fading channels: multiuser diversity

Capacity gain on multiuser single antenna fading channels
with maximum throughput scheduling



$$C = E_{\{h_1,\ldots,h_K\}}\left[\log_2\left(1+\frac{P}{N_o}\max_{1\le k\le K}\left(\|h_k\|^2\right)\right)\right]$$

$$\frac{P}{N_o} = 0\ dB$$

The scheduling algorithm chooses to transmit with full power to the user with the largest SNR in each fading state.

Exploiting multiuser diversity in practice

High-throughput packet-data access for 3G cellular: A brief overview

- channel estimation and prediction of SIR in every time slot by each mobile user
- CQI: feedback of SIR to the base station (or data rate request)
- scheduling of packet transmission to users by the base station with adaptive modulation and coding
- transmission over very short time slots (channel fading gains remain constant) one user at a time at maximum power
- packet retransmissions with hybrid ARQ employing soft packet combining and incremental redundancy

User 1 packet 1 Slot 1	User 3 packet 1 Slot 2	User 1 packet 2 Slot 3	…	User 1 packet 1 Slot 5	…

SIR: Signal-to-Interference Ratio
ARQ: Automatic Repeat Request
CQI: Channel Quality Indicator

I.C. MULTIUSER MIMO CHANNELS

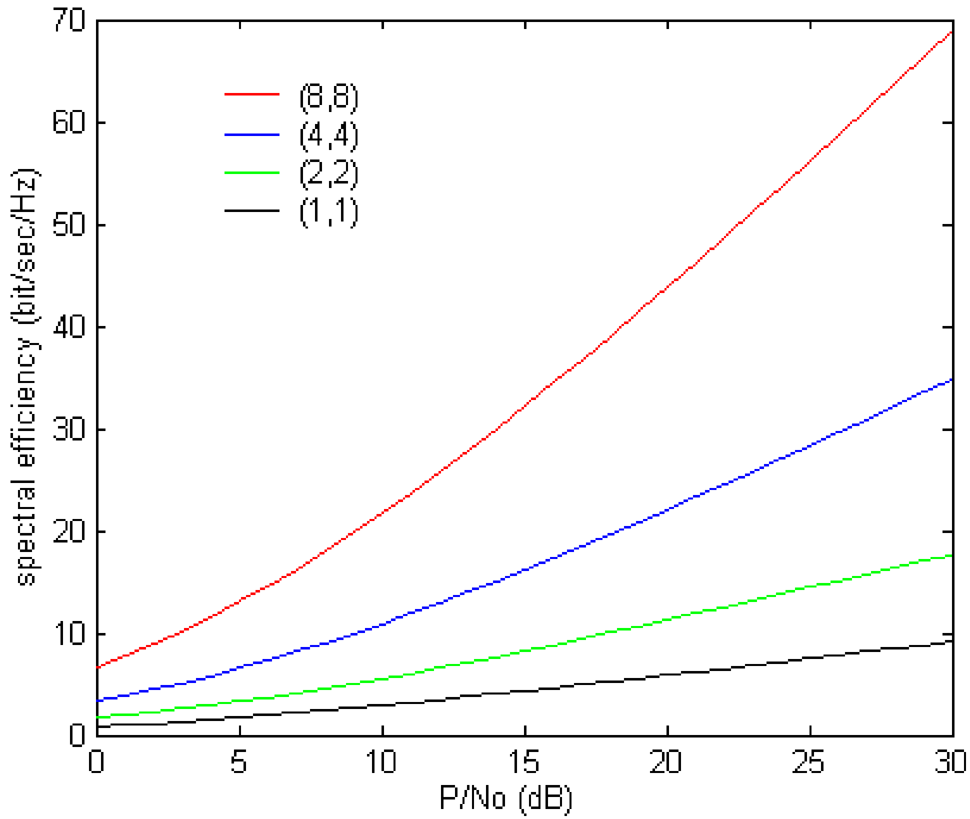
- **Single-user transmission strategies**
- **Performance limits with optimal signalling**

Objective: Spatial multiplexing gain

MIMO channels: point-to-point link [3,4]

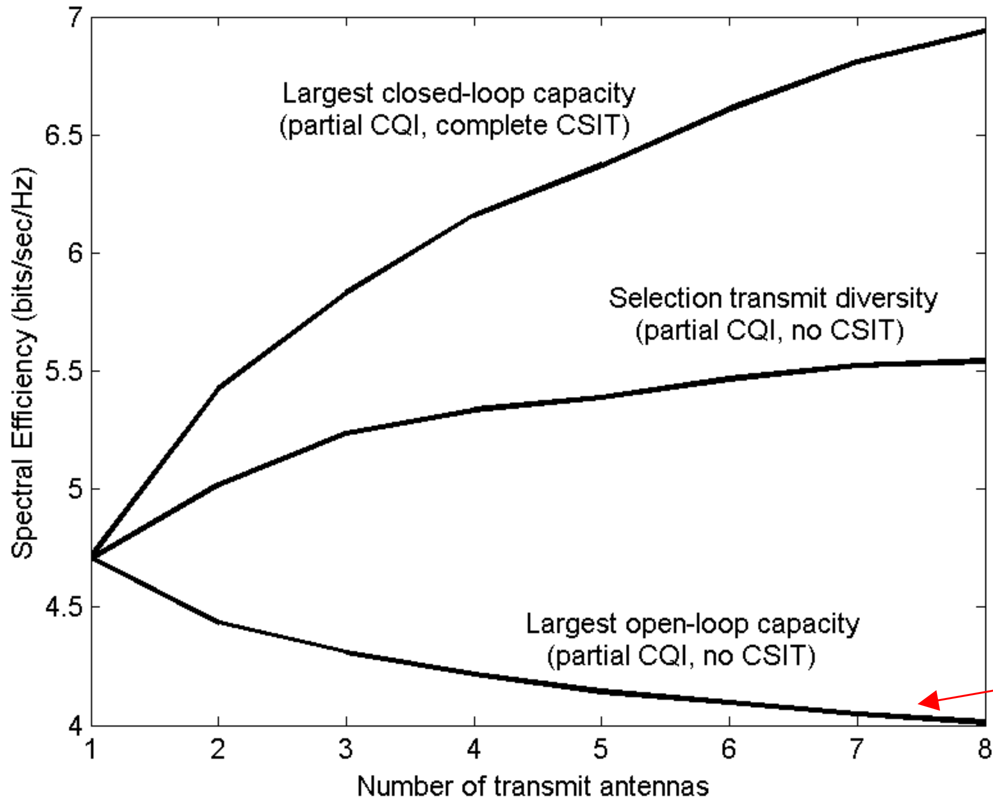
Single user MIMO Rayleigh fading channel

N_T transmit and N_R receive antennas: (N_T, N_R) MIMO channel



High SNR slope min (N_T, N_R)

Channel hardening and its impact on multiuser diversity with *single-user* transmission strategies [5,6]



(N,1,8) MIMO BC

CQI: channel quality indicator

- feedback by all users
- used for scheduling

CSIT: channel state information at the transmitter

- feedback only by the selected users
- used for spatial processing

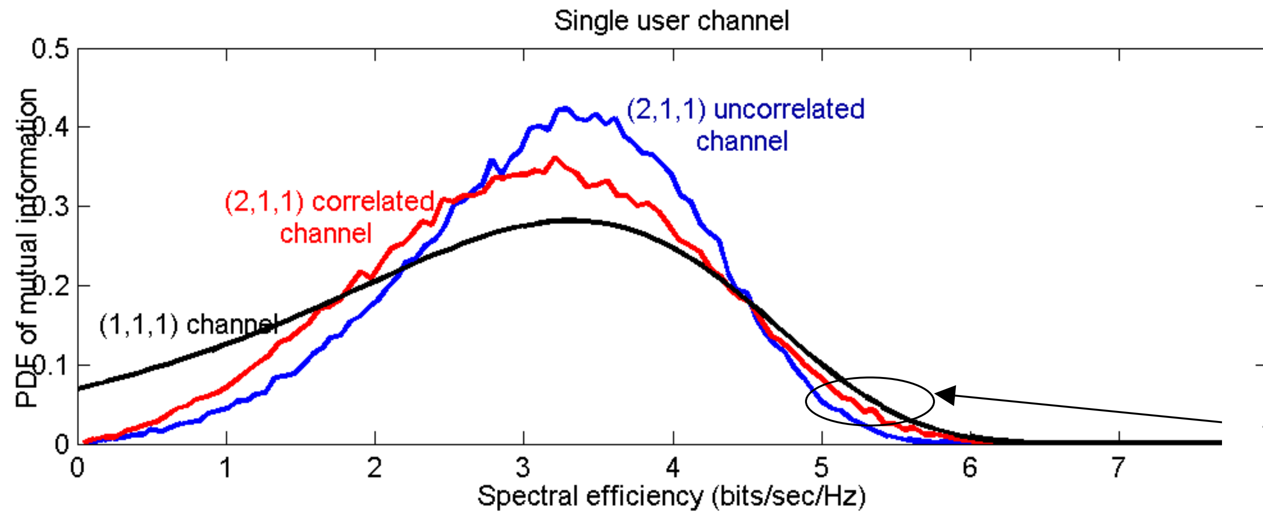
Channel hardening effect

Channel hardening of open-loop MIMO channel mutual information: spatial diversity reduces the randomness in the channel.

This effect is contrary to the advantage of multiuser diversity.

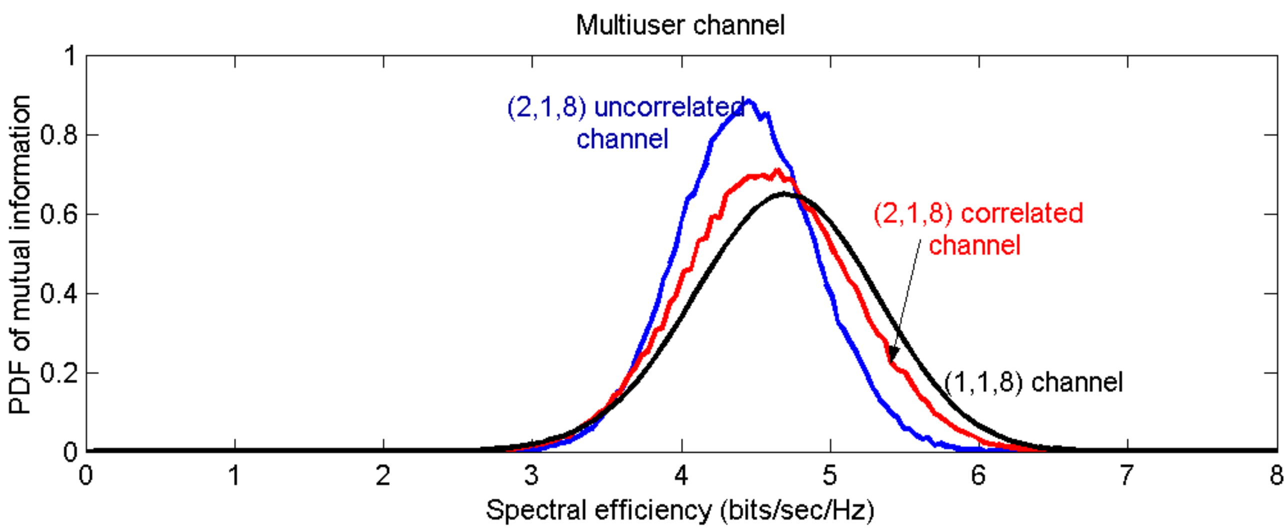
None of these solutions achieve spatial multiplexing gain.

Effect of transmit antenna correlations on multiuser diversity with single-user transmission strategies (to the user with the largest mutual information)



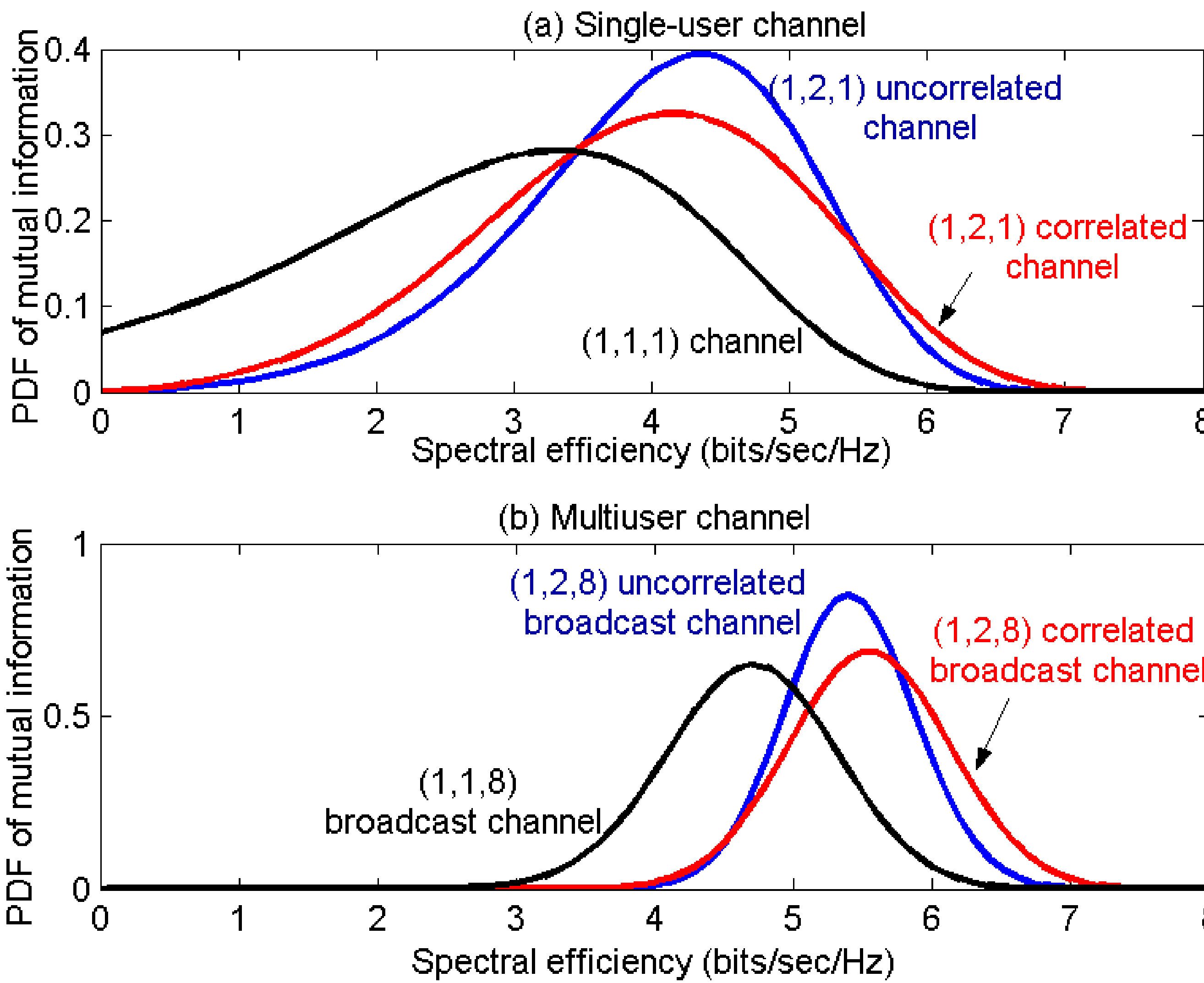
Multiple transmit antenna scenarios:
CQI = open-loop mutual information

Tail of PDF is what matters



Open-loop capacity is not a good choice of CQI

Effect of receive antenna correlations on multiuser diversity with single-user transmission strategies [7]



Multiple receive antenna scenarios:

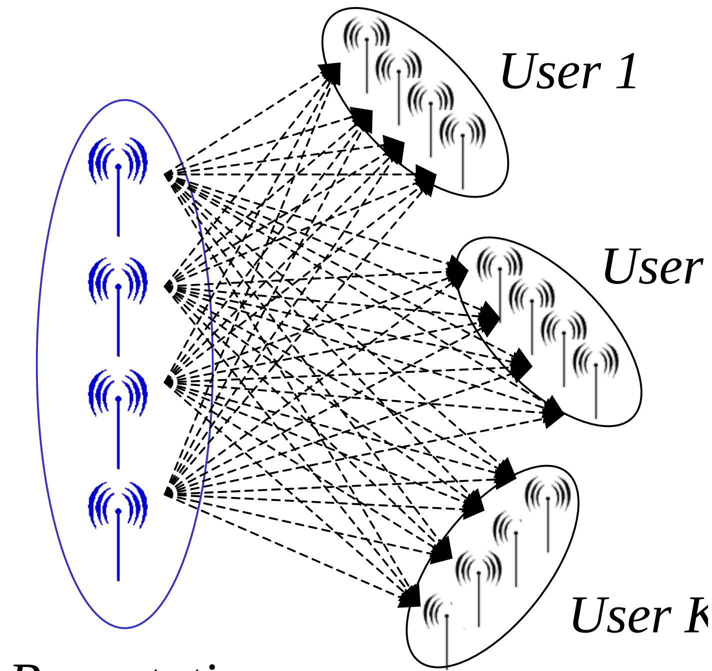
CQI = MRC mutual information

MRC = Maximal Ratio Combining

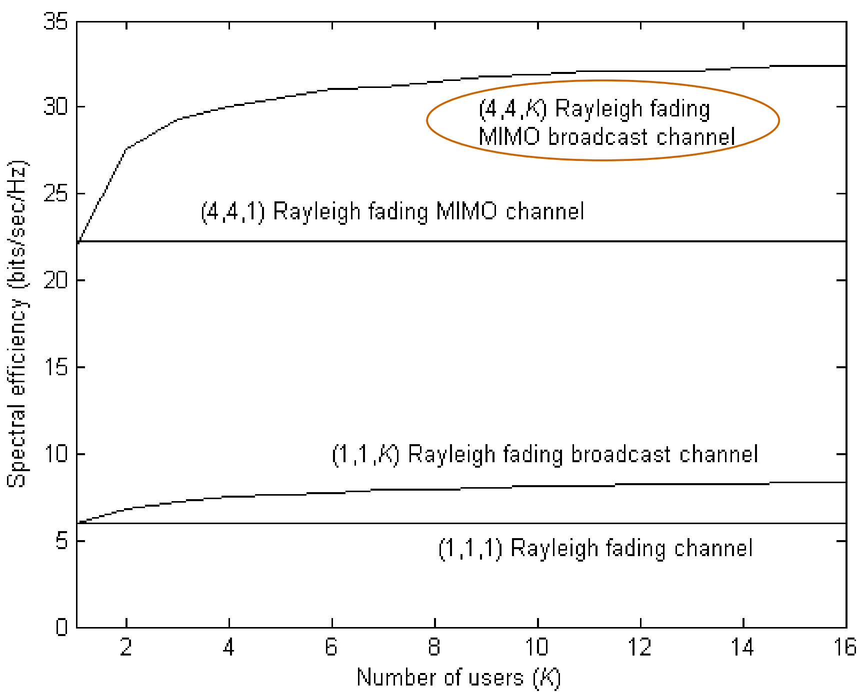
MIMO channels and multiuser systems

Numerical results

(4,4,*K*) MIMO BC



Downlink maximum throughput (sum-capacity) in Rayleigh fading



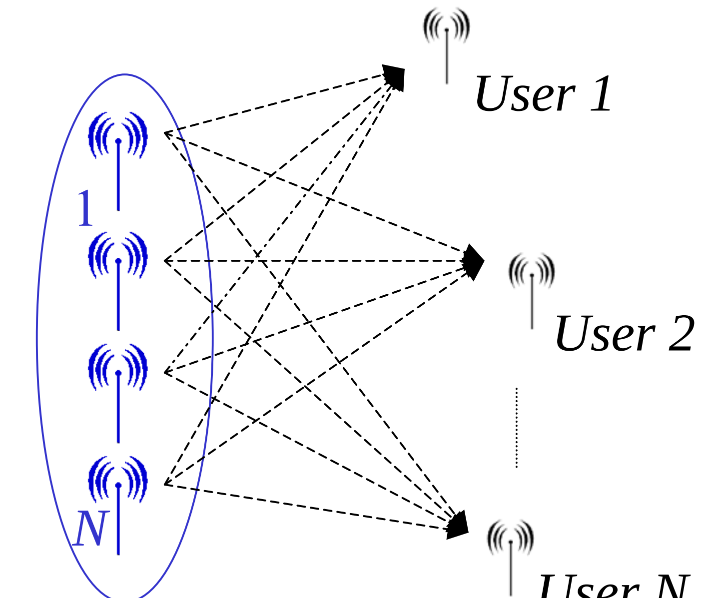
$$\frac{P}{N_o} = 20\ dB$$

The capacity gains of MIMO systems and multiuser diversity can be achieved simultaneously

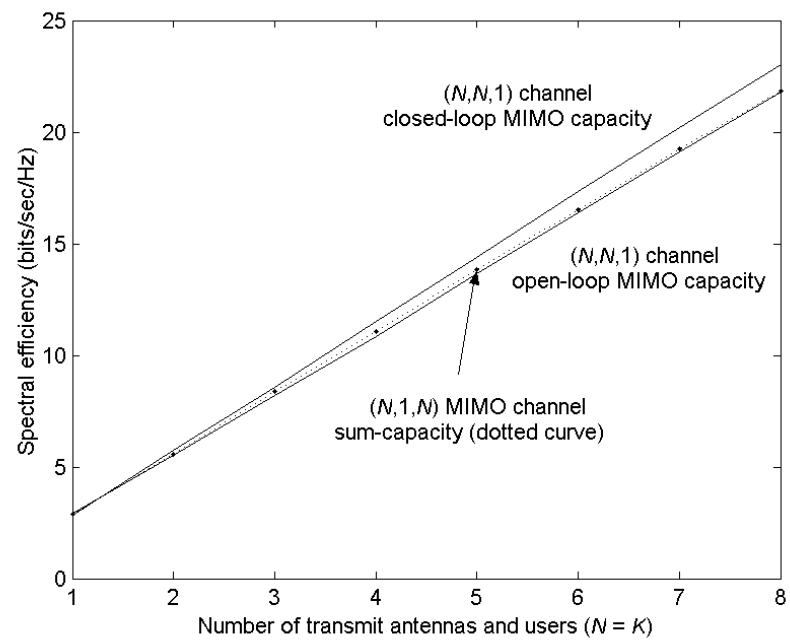
MIMO channels and multiuser systems

Numerical results

(*N*,1,*N*) MIMO BC



Downlink maximum throughput (sum-capacity) in Rayleigh fading



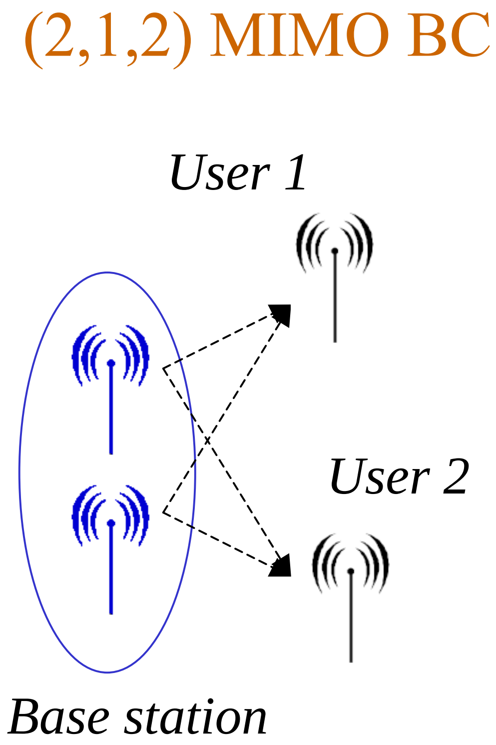
A MIMO channel can be created even though the mobile users have only one receive antenna and cannot cooperate at the receiver side

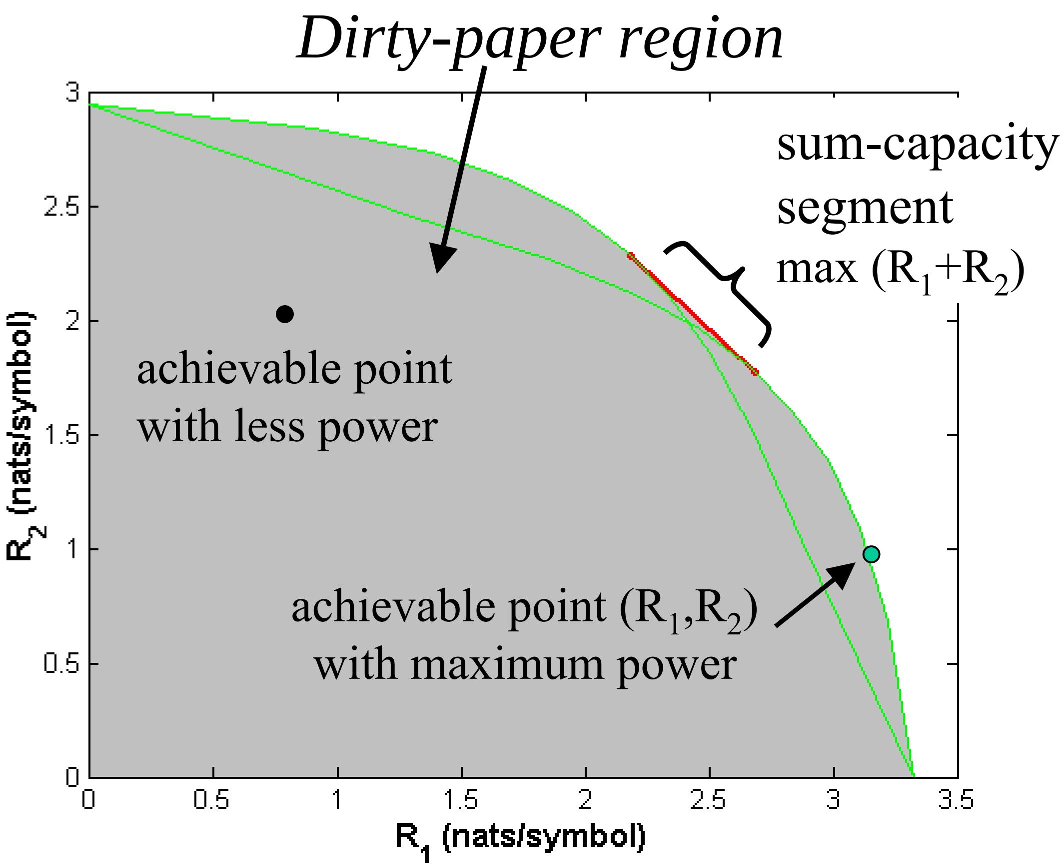
$$\frac{P}{N_o} = 0 \ dB$$

MIMO channels and multiuser systems

2-user capacity region [8]

Is it still optimal to transmit to a single user at a time? Answer: NO

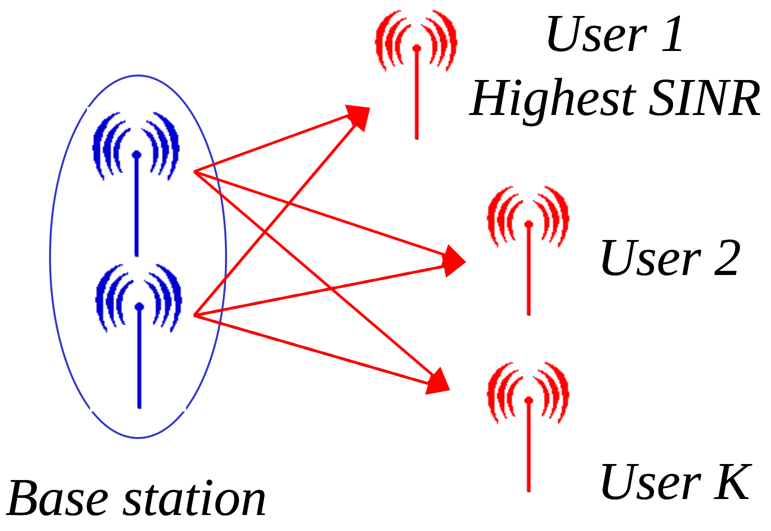




MIMO Channels and Multiuser Systems

The fundamentally different nature of the channel [8]

Perfect channel knowledge at the transmitter
Perfect channel knowledge at all receivers
Transmit power constraint



One transmitter antenna
degraded broadcast channel sum-capacity achieved by transmitting to one user at a time.

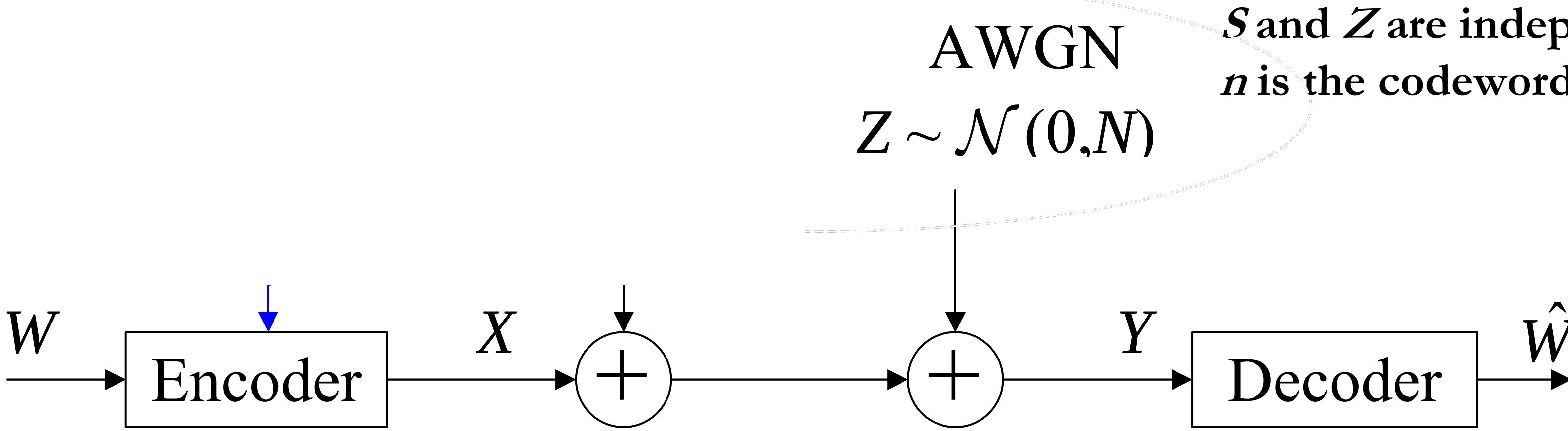
Add one transmitter antenna
Non-degraded broadcast channel sum-capacity achieved by dirty-paper coding and simultaneous transmission to several users.

On some channels it is even possible that the optimal number of active users be larger than the number of transmit antennas

I.D. A brief introduction to Dirty-Paper Coding

Principle of Dirty-Paper Coding [9]

S and ***Z*** are independent
n is the codeword length



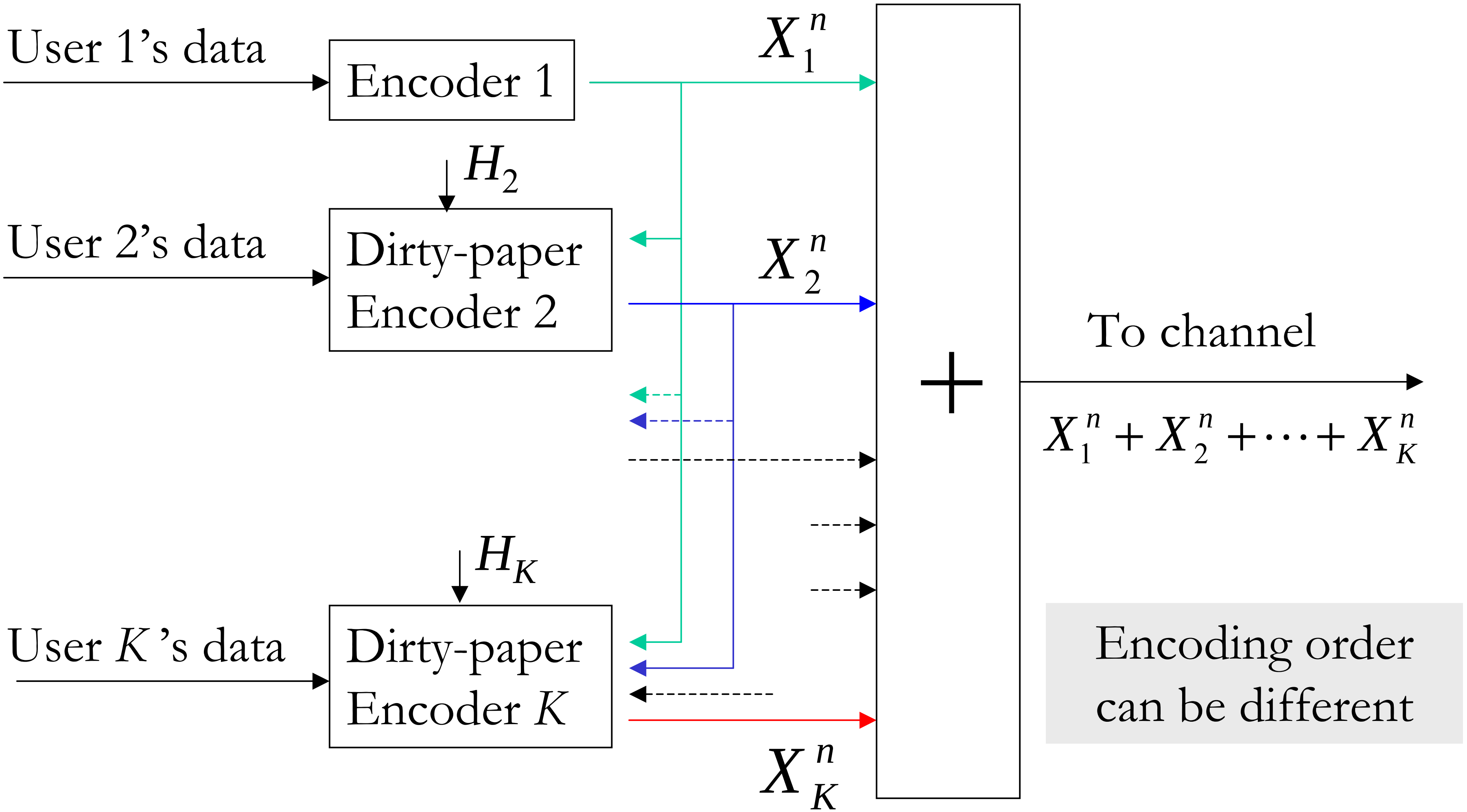
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \le P$$

$$Y = X + S + Z$$

Capacity of the channel with interference ***S*** **known non-causally at the transmitter** ***but unknown at the receiver*** **is the same as if** ***S*** **was not present**

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

MIMO BC: successive encoding [8]



Reference: Syed Ali Jafar, PhD Thesis, Fundamental Capacity Limits of Multiple Antenna Wireless Systems, Stanford, August 2003.

MIMO BC: successive encoding [8]

$$y_i = H_i\left(x_i + \sum_{j=1}^{i-1} x_j + \sum_{j=i+1}^{K} x_j\right) + n_i$$ Signal received by user i

$$s_i = H_i \sum_{j=1}^{i-1} x_j$$ is the known-interference, its effect has been removed

At the receivers:

- User K sees a channel as if there was no interference from other users
- User K - 1 sees interference from user K only
- User 1 sees interference from all users

MIMO BC: achievable rate vector

The achievable rate vector $(R_1, \ldots, R_K)$ is given by

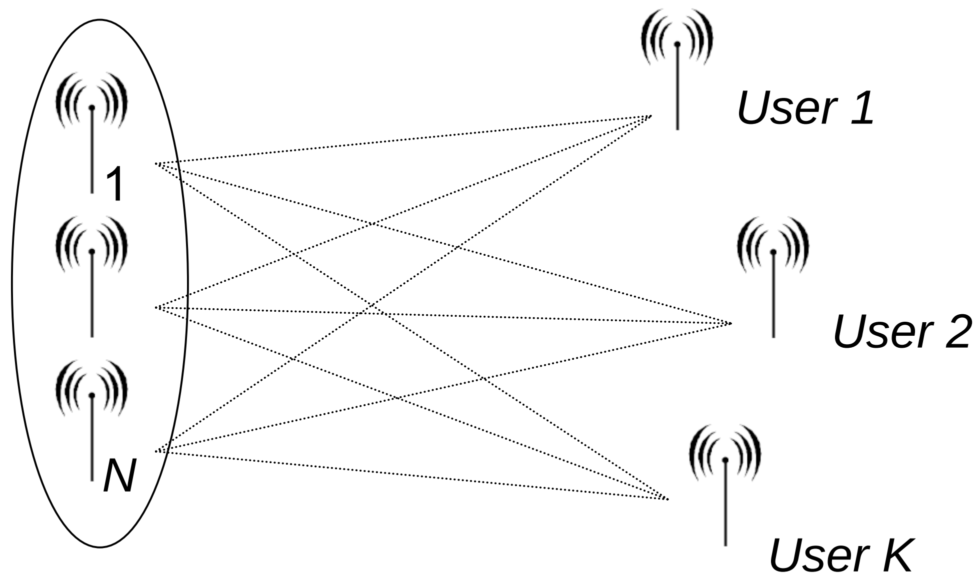
$$R_i = \frac{1}{2}\log\frac{\left|I + H_i\left(\sum_{j\geq i}\Sigma_j\right)H_i^{\dagger}\right|}{\left|I + H_i\left(\sum_{j>i}\Sigma_j\right)H_i^{\dagger}\right|} \qquad i = 1,\ldots,K$$

Covariance matrix of x_i is $\Sigma_i = E\left[x_i x_i^{\dagger}\right]$

Duality: MIMO BC and Sum-Power MIMO MAC [10]

Same capacity region

Same sum-capacity



MIMO BC	Transmitter Power constraint P	Receivers Non-cooperating
Sum-power MIMO MAC	Receiver	Transmitters Non-cooperating Sum-power constraint P

Hermitian channel matrices

Sum-power MIMO MAC introduced for solving the sum-capacity problem more easily

II. MULTIUSER MIMO CHANNELS
Novel results on the sum-capacity

What is the optimal number of active users?
What is the optimal power allocation?

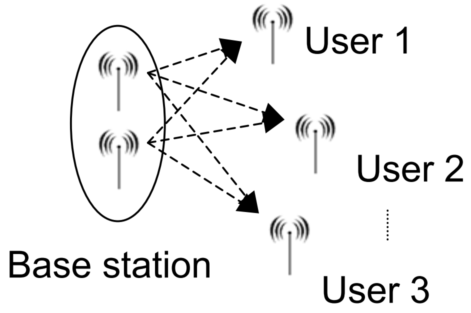
MIMO channels and multiuser systems

Sum-capacity analysis for the (2,1,K) MIMO BC [11]

Using the dual sum-power multiple-access channel formulation of the sum-capacity maximization and Lagrange duality theory for the optimization:

$$C = \max_{\{p_1,\ldots,p_K\}} \log\det\left(I + \sum_{i=1}^{K} p_i \boldsymbol{h}_i^* \boldsymbol{h}_i \right)$$

$$s.t. \sum_{i=1}^{K} p_i = P,\ p_i \geq 0,\ i = 1,\ldots,K$$



Closed-form expression of the sum-capacity: $C = \log\left(1 - 1/2\, \boldsymbol{b}^* \boldsymbol{\Phi}^{-1} \boldsymbol{b}\right)$

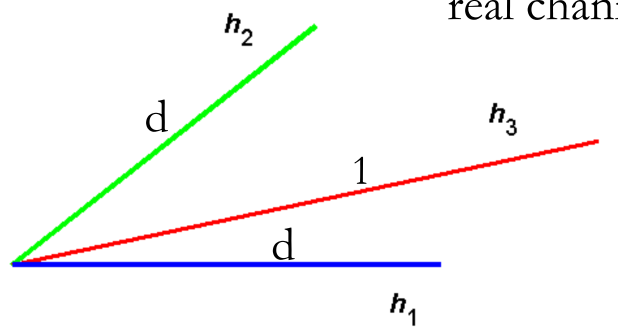
($\boldsymbol{b}$ and $\boldsymbol{\Phi}$ depend only on the channel gains and on the total power constraint P)

The optimal number of active users ($p_i > 0$) can be determined exactly with this analysis.

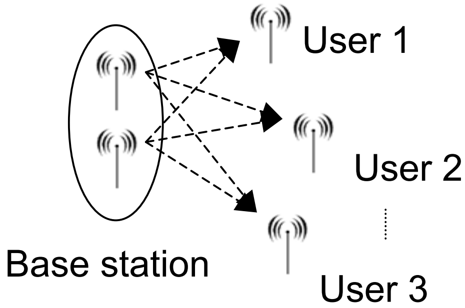
Optimal number of active users vs. transmit power on the dual sum-power MIMO MAC at the sum-capacity

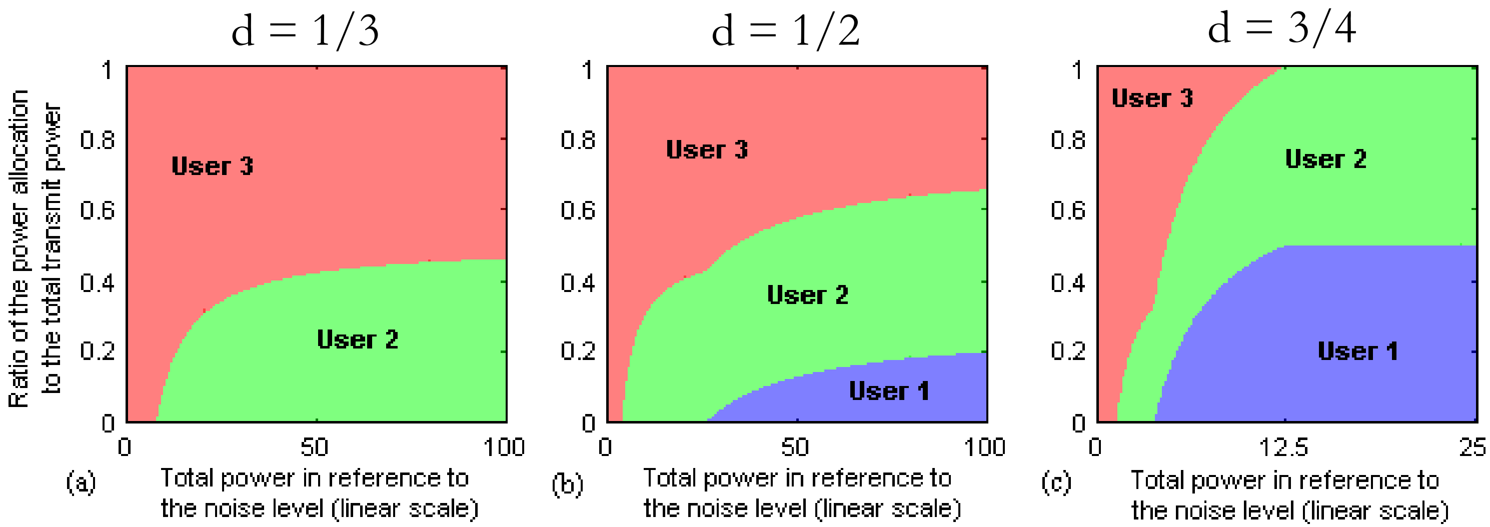
Example 1: (2,1,3) MIMO BC

real channel vectors



$$H = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$





The optimal number of active users can be

- larger than the number of transmit antennas
- a non-monotonic function of the total transmit power !

MAC to BC covariance transformations [10]

- Dirty-paper encoding order: K to 1
- Optimal BC covariance matrices: $\boldsymbol{\Sigma}_1 \ldots \boldsymbol{\Sigma}_K$
- Optimal MAC covariance matrices: $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_K$

$$\Rightarrow \quad \boldsymbol{\Sigma}_j = \boldsymbol{B}_j^{-1/2} \overline{\boldsymbol{A}_j^{1/2} \boldsymbol{P}_j \boldsymbol{A}_j^{1/2}} \boldsymbol{B}_j^{-1/2}$$

$\boldsymbol{A}_j$ interference experienced by user j on the BC

$\boldsymbol{B}_j$ interference experienced by user j on the MAC

These transformations relate the optimal covariance matrices for the MAC and the BC that achieve the same rate vector $(R_1, \ldots, R_K)$ on both channels.

**Optimal MAC covariance matrices are *independent* of the decoding order.
Optimal BC covariance matrices *depend* on the encoding order (for DPC).**

Asymptotic optimality of *N*-user scheduling with Dirty Paper Coding on the (*N*,1,*K*) MIMO BC [12]

- Dirty-paper encoding order: K to 1
- Optimal BC covariance matrices: $\boldsymbol{\Sigma}_1 \ldots \boldsymbol{\Sigma}_K$
- Asymptotic fraction of power allocated to user i on the dual MAC: r_i
- Assume $r_i > 0,\ i = K,\ldots,K-N+1$ asymptotically

$$\lim_{P\to\infty} \frac{Tr\left(\boldsymbol{\Sigma}_i\right)}{P} = 0 \text{ if } i \le K-N$$

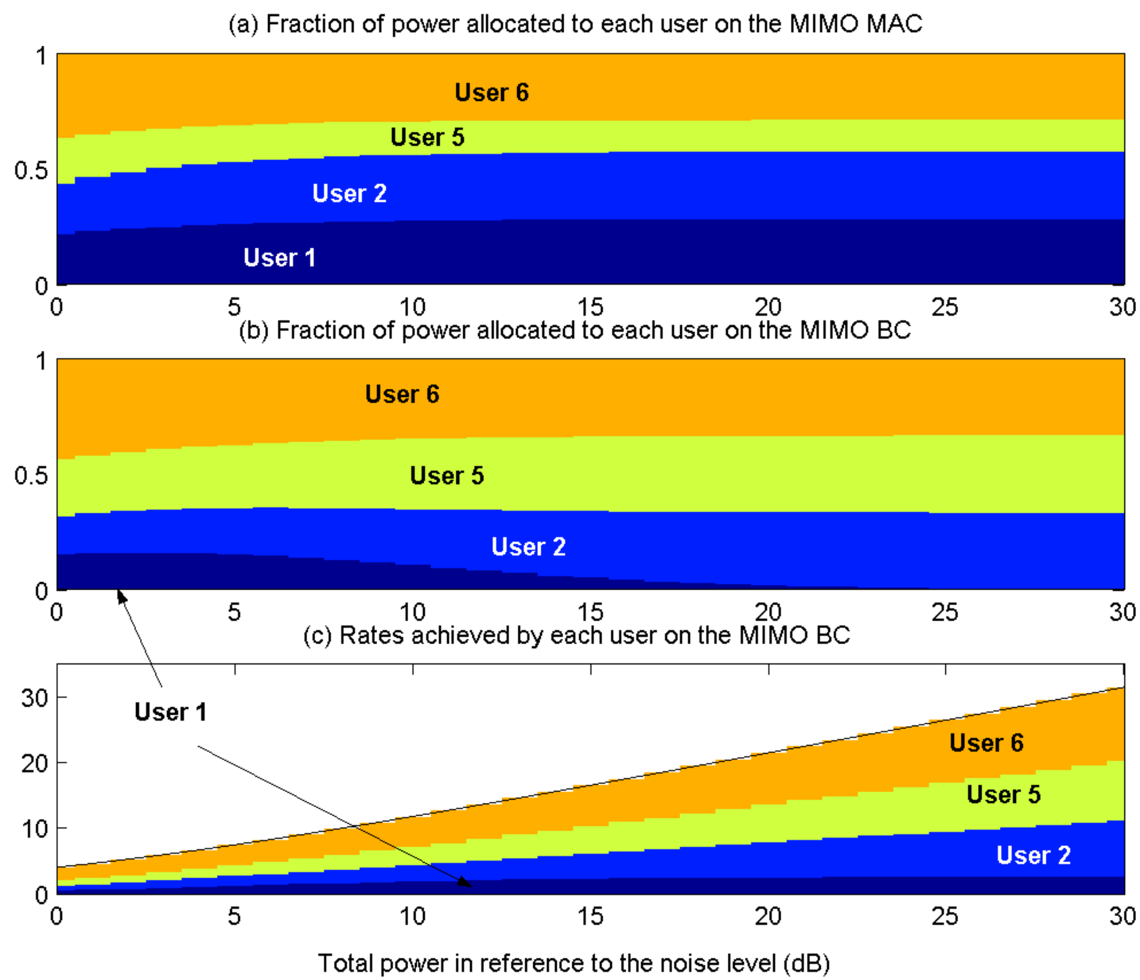
$$\lim_{P\to\infty} \frac{Tr\left(\boldsymbol{\Sigma}_i\right)}{P} > r_i \text{ if } i > K-N$$

Only N users are required to optimally exploit the N dimensions offered by the MIMO BC in the high power region. These N users are not unique in general.

Beamforming vector for user i is orthogonal to channel vector of user $j > i > K - N$

The joint effect of dirty-paper coding and beamforming completely orthogonalizes the channels between the N users in the high power region. Uniform power allocation becomes asymptotically optimal.

Asymptotic optimality of *N*-user scheduling with Dirty Paper Coding on the (*N*,1,*K*) MIMO BC [12]



Example 2

(3,1,8) MIMO BC

Single antenna receivers

Fixed channel matrix realization

Let the transmit power increase

Dirty-paper encoding order:

8 to 1

Asymptotic optimality of single-user scheduling with Dirty Paper Coding on the (N,N,K) MIMO BC [12]

- Dirty-paper encoding order: K to 1
- Optimal BC covariance matrices: $\boldsymbol{\Sigma}_1 \ldots \boldsymbol{\Sigma}_K$
- Asymptotic fraction of power allocated to user i on the dual MAC: r_i
- Assume $r_K > 0$

$$\Rightarrow \quad \lim_{P\to\infty} \frac{Tr\left(\boldsymbol{\Sigma}_i\right)}{P} = 0 \text{ if } i < K$$

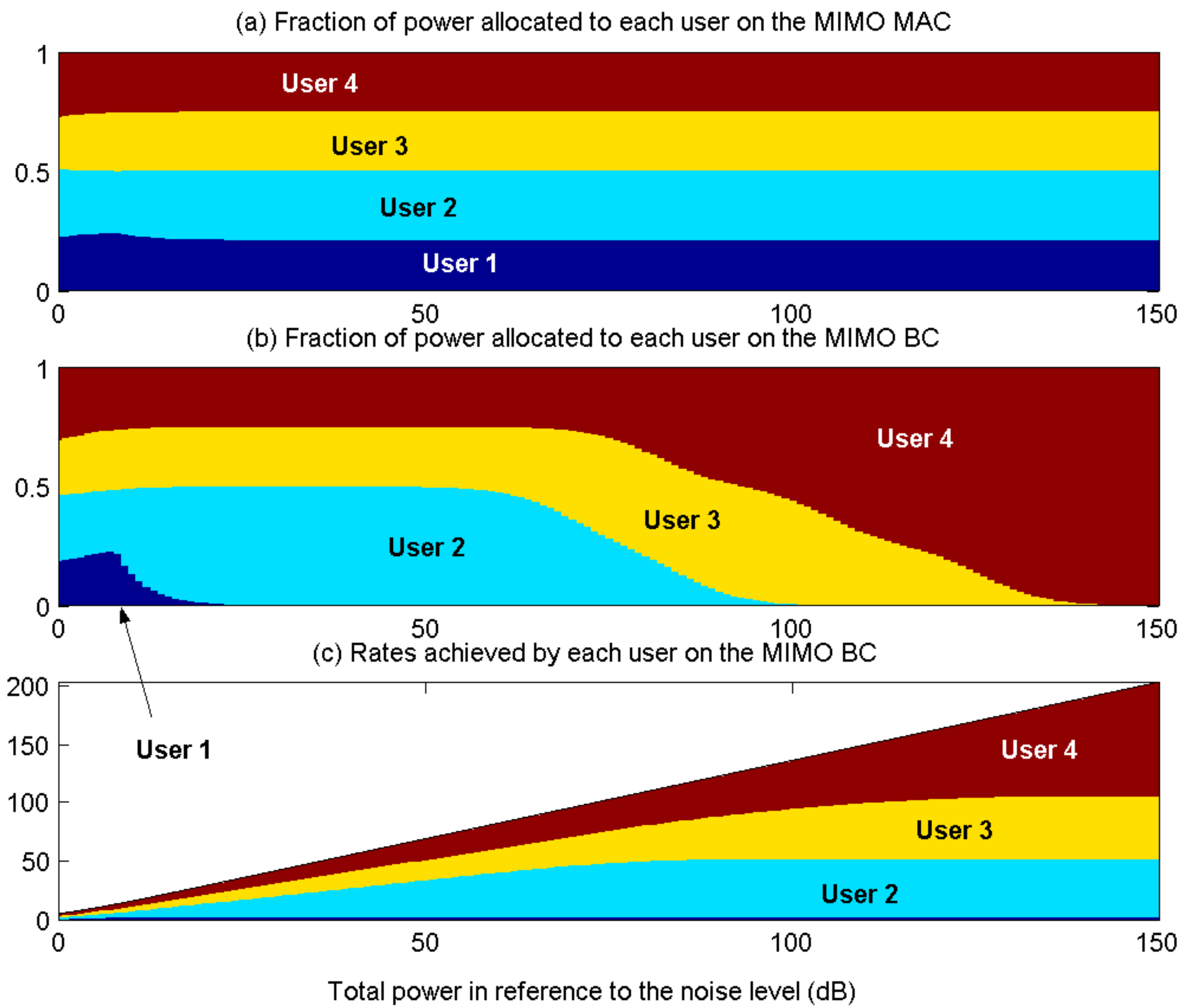
$$\lim_{P\to\infty} \frac{Tr\left(\boldsymbol{\Sigma}_K\right)}{P} > 0$$

Only 1 user is allocated a non-vanishing fraction of the total transmit power on the MIMO BC in the high power region.
This user is in general not unique.

However…

Rates achieved by other users only become negligible at very large values of *P*

Asymptotic optimality of single-user scheduling with Dirty Paper Coding on the (*N*,*N*,*K*) MIMO BC [12]



Example 3

(4,4,4) MIMO BC

Multiple antenna receivers

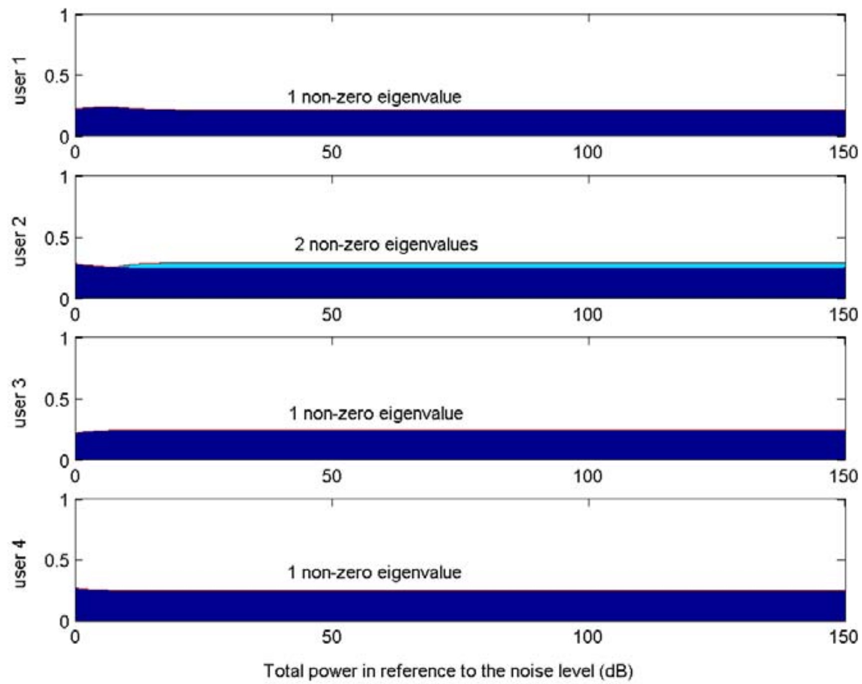
Dirty-paper encoding order:
4 to 1

Asymptotic optimality of single-user scheduling with Dirty Paper Coding on the (*N*,*N*,*K*) MIMO BC [12]

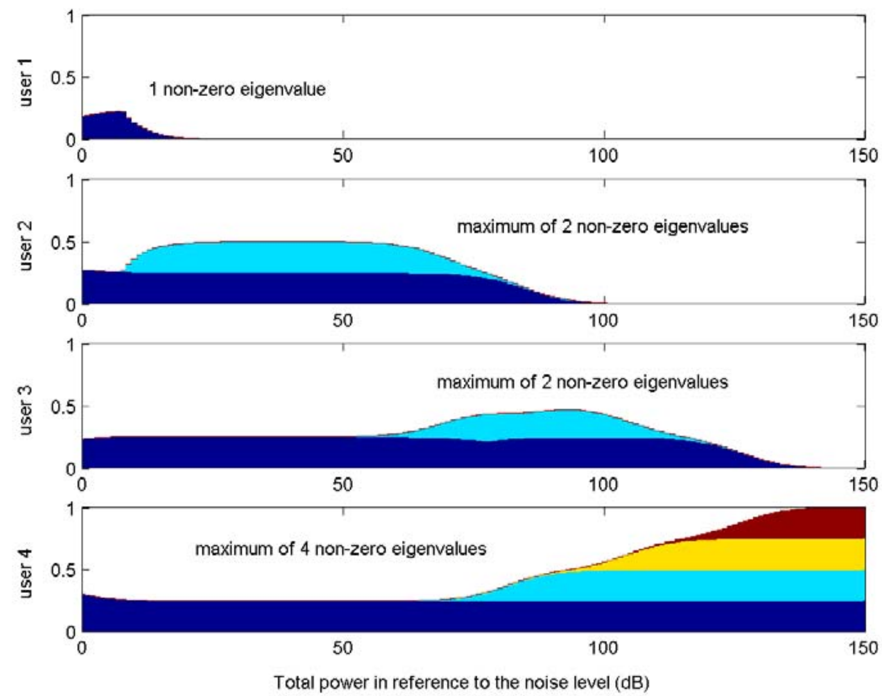
Example 3 (continued)
(4,4,4) MIMO BC

Dirty-paper encoding order: 4 to 1

Eigenvalues of optimal MAC covariance matrices

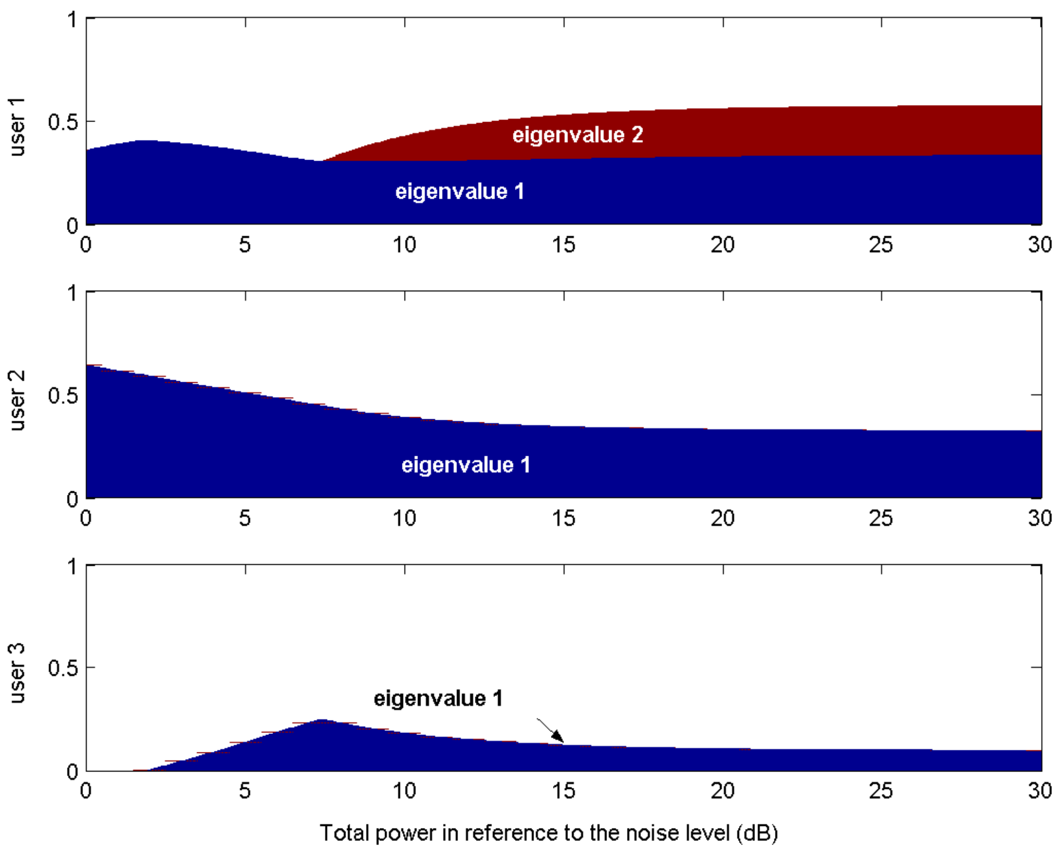


Eigenvalues of optimal BC covariance matrices

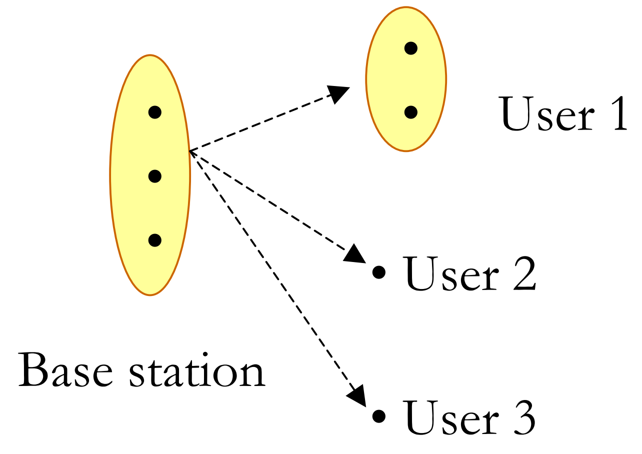


General case: the (*N*,*M*,*K*) MIMO BC [7]

Eigenvalues of optimal MAC covariance matrices



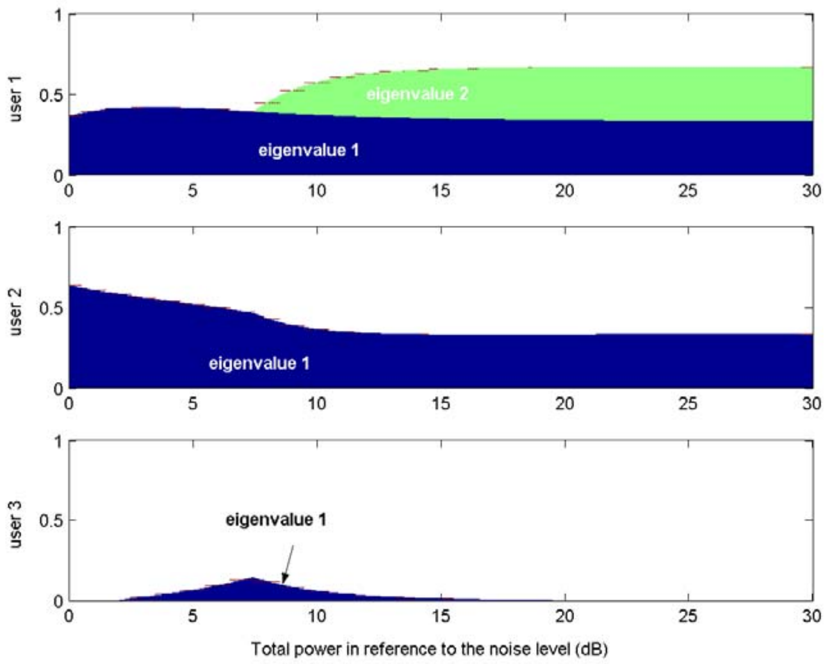
Example 4
(3,[2 1 1],3) MIMO BC



General case: the (N,M,K) MIMO BC [7]

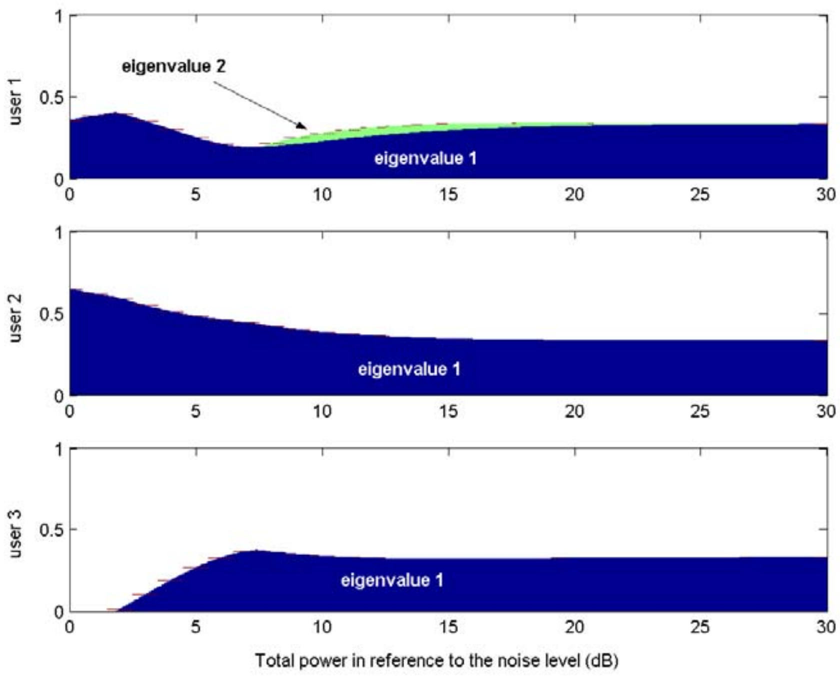
Example 4: (3,[2 1 1],3) MIMO BC

Eigenvalues of optimal BC covariance matrices



Dirty-paper encoding order: 1 to 3

Eigenvalues of optimal BC covariance matrices



Dirty-paper encoding order: 3 to 1

In both cases, the sum rate is equal to the sum-capacity of the MIMO broadcast channel at any value of the power, but the rates are distributed differently among the 3 users.

Relation to MIMO channels with co-channel interference

- **3 different problems**
  - MIMO channel with co-channel interference

    With M receive antennas, a user can only support M-1 infinite-power one-dimensional interferers before its achievable rate goes to 0
  - MIMO BC with self interference, e.g. (N,1,K) MIMO BC

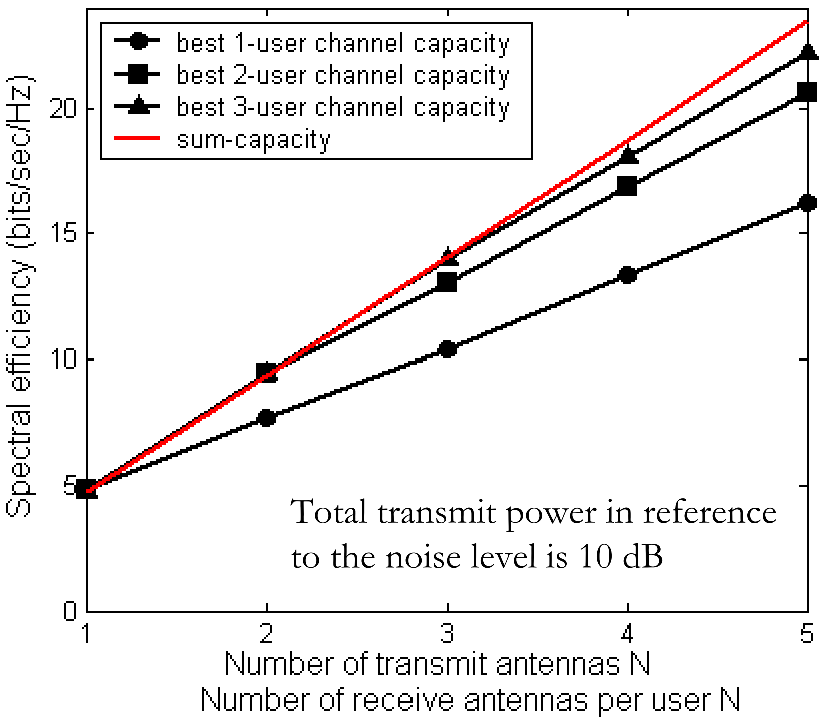
    Users with 1 receive antenna can support N-1 other infinite power users!
  - Dual MIMO MAC with sum-power constraint

    With N receive antennas a user can support more than N infinite power users!
- **Commonalities**
  - Infinite-power interferers reduce the rank of the channel
- **Differences: cooperation**
  - MIMO BC: dirty-paper coding and spatial processing completely orthogonalize the channels between the N infinite power users
  - Dual MIMO MAC: all users have infinite power but after successive decoding only N will have a rate proportional to log(P)

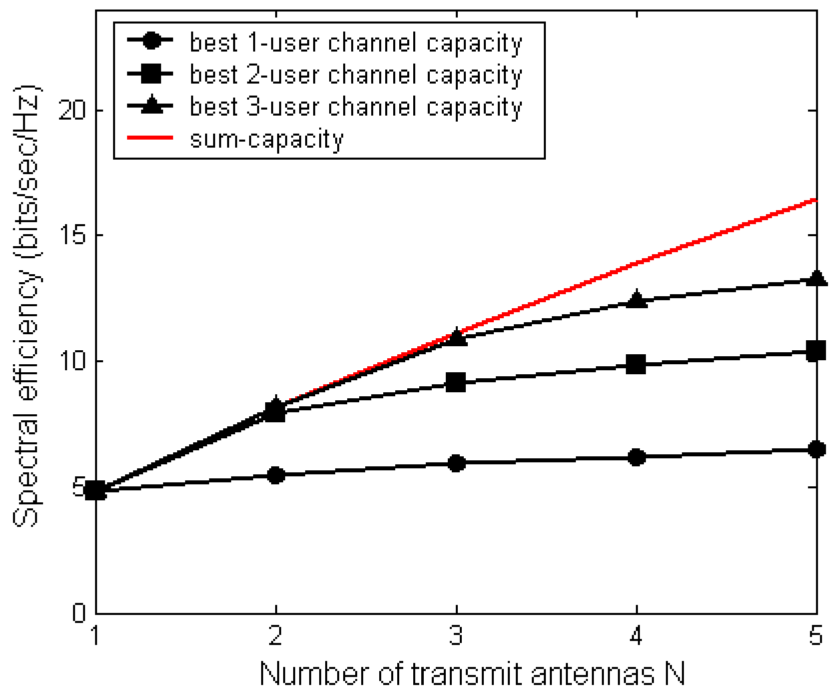
III. MAXIMUM-THROUGHPUT SCHEDULING ALGORITHMS FOR MIMO BROADCAST FADING CHANNELS [7][13]

Near optimality of *N*-user scheduling with optimal signaling on fading channels in medium power region : average sum-capacity [11]

(*N*,*N*,10) MIMO BC



(*N*,1,10) MIMO BC



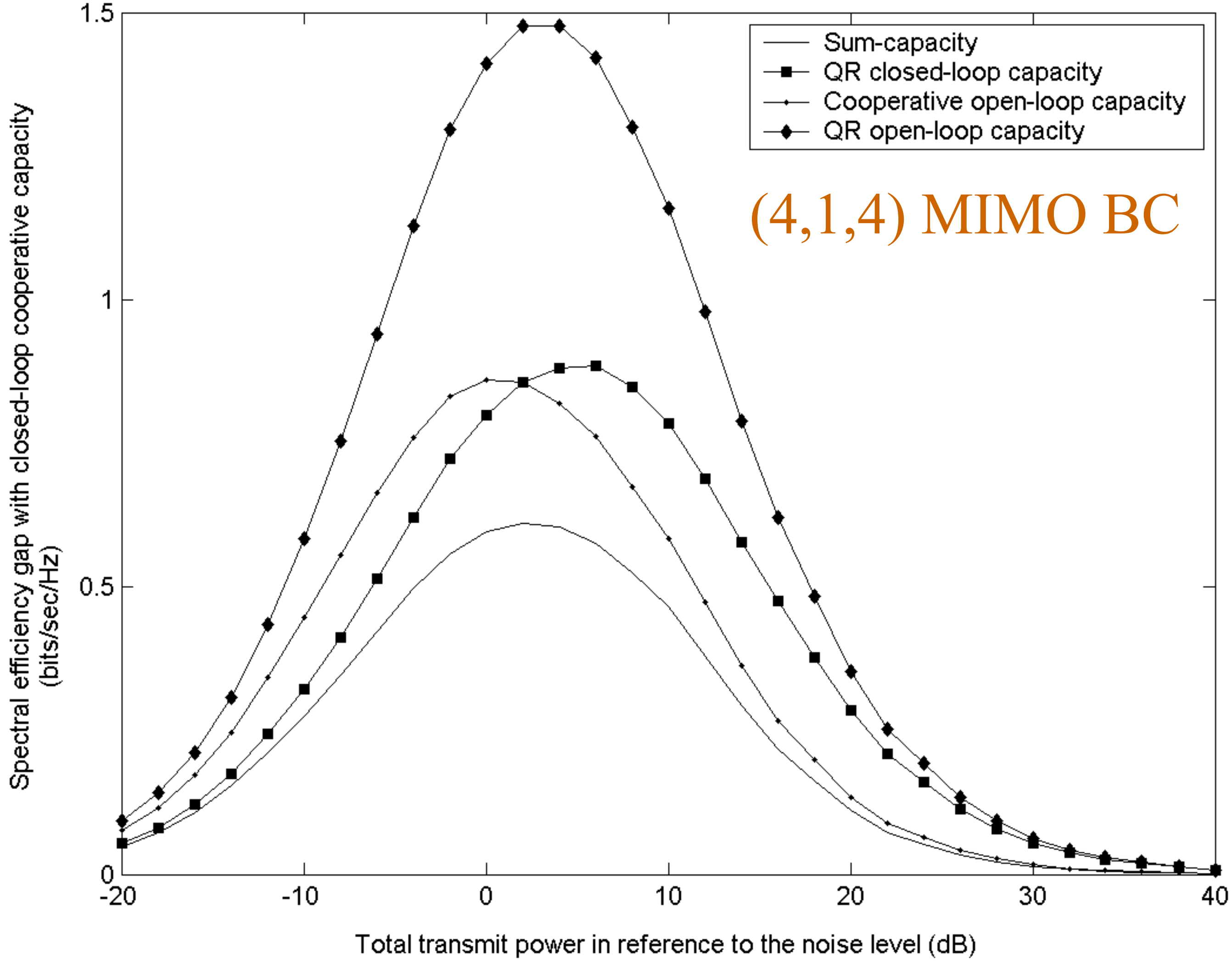
<u>*Rules*</u>

1. Need to be capable to transmit to at least N users

2. Need not transmit to more than N users

New role of scheduling algorithm: schedule jointly good users

Approximations of the sum-capacity of the (N,1,N) MIMO BC



QR capacity: first-order asymptotic approximation of the sum-capacity of the (N,1,N) MIMO BC in the high power region (closed-loop: with waterfilling power allocation) [7]

Relation to receive antenna selection algorithms

- **Motivation for using receive antenna selection algorithms**

  Sum-capacity ~= MIMO open-loop capacity at high power

- **Drawbacks of existing receive antenna selection algorithms**
  - Only designed for the high power region
  - Only designed for uniform power allocation at the transmitter
  - What is the performance in the medium power region?

- **Advantages of existing receive antenna selection algorithms**
  - Existing literature
  - Low computational complexity
  - Interference-avoidance properties

Low-complexity max-throughput scheduling
Receive antennas selection algorithms – example (Gorokhov, [14])

Algorithm II: Decremental selection

The set of receive antennas indexes is obtained as follows:

Set $\boldsymbol{A} = \left((E_s / N_o)^{-1} \boldsymbol{I}_N + \boldsymbol{H}^* \boldsymbol{H} \right)^{-1}$

$p = \arg\min_{1 \le l \le M} \boldsymbol{H}_l \boldsymbol{A} \boldsymbol{H}_l^*$

$r = \{1, \ldots, p-1, p+1, \ldots, M\}$

For $n = 1$ to $(M - N - 1)$

Update $\boldsymbol{A} = \boldsymbol{A} + \boldsymbol{A}\boldsymbol{H}_p^* \left(1 - \boldsymbol{H}_p \boldsymbol{A} \boldsymbol{H}_p^*\right)^{-1} \boldsymbol{H}_p \boldsymbol{A}$

Compute $p = \arg\min_{l \in r} \boldsymbol{H}_l \boldsymbol{A} \boldsymbol{H}_l^*$

Remove p from the set r

End

At each step, determine the receive antenna that can be removed such that the capacity decrease is minimized.

Low-complexity max-throughput scheduling
N transmit antennas, *K* single-antenna users [7,13]

Successive projections scheduling

Choose users successively such that their channel vectors are as much linearly independent as possible at each stage. At stage *k*:

$$u_k = \underset{1\le i\le K,\, i\notin S_{k-1}}{\arg\max}\left[\min_{(\alpha_1,\cdots,\alpha_{k-1})\in\mathbb{C}^{k-1}}\left\|\sum_{j\in S_{k-1}}\alpha_j \boldsymbol{h}_j - \boldsymbol{h}_i\right\|^2\right] \quad \text{with} \quad S_{k-1}=\{u_1,\ldots,u_{k-1}\}$$

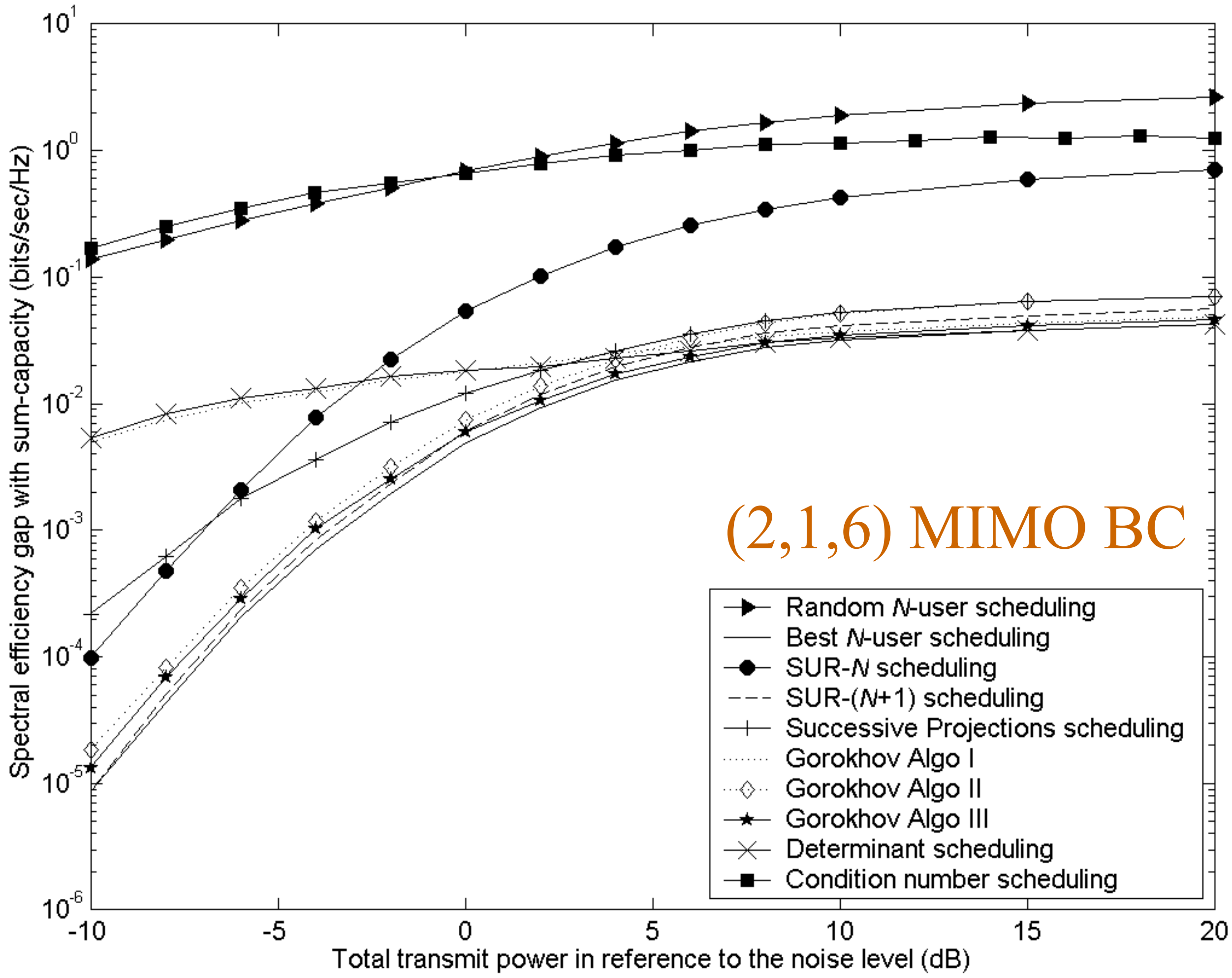
Equivalently:

$$u_k = \underset{1\le i\le K,\, i\notin S_{k-1}}{\arg\max}\left\|\boldsymbol{h}_i\left[\tilde{\boldsymbol{H}}^*_{S_{k-1}}\left(\tilde{\boldsymbol{H}}_{S_{k-1}}\tilde{\boldsymbol{H}}^*_{S_{k-1}}\right)^{-1}\tilde{\boldsymbol{H}}_{S_{k-1}} - \boldsymbol{I}_N\right]\right\|^2 \text{ with } \tilde{\boldsymbol{H}}_{S_{k-1}} = \begin{bmatrix}\boldsymbol{h}_{u_1}\\ \vdots \\ \boldsymbol{h}_{u_{k-1}}\end{bmatrix}$$

Low-complexity scheduling by successive projections finds the best *N*-user set almost surely with a complexity proportional to *KN*-*N*(*N*-1)/2. It also allows to:

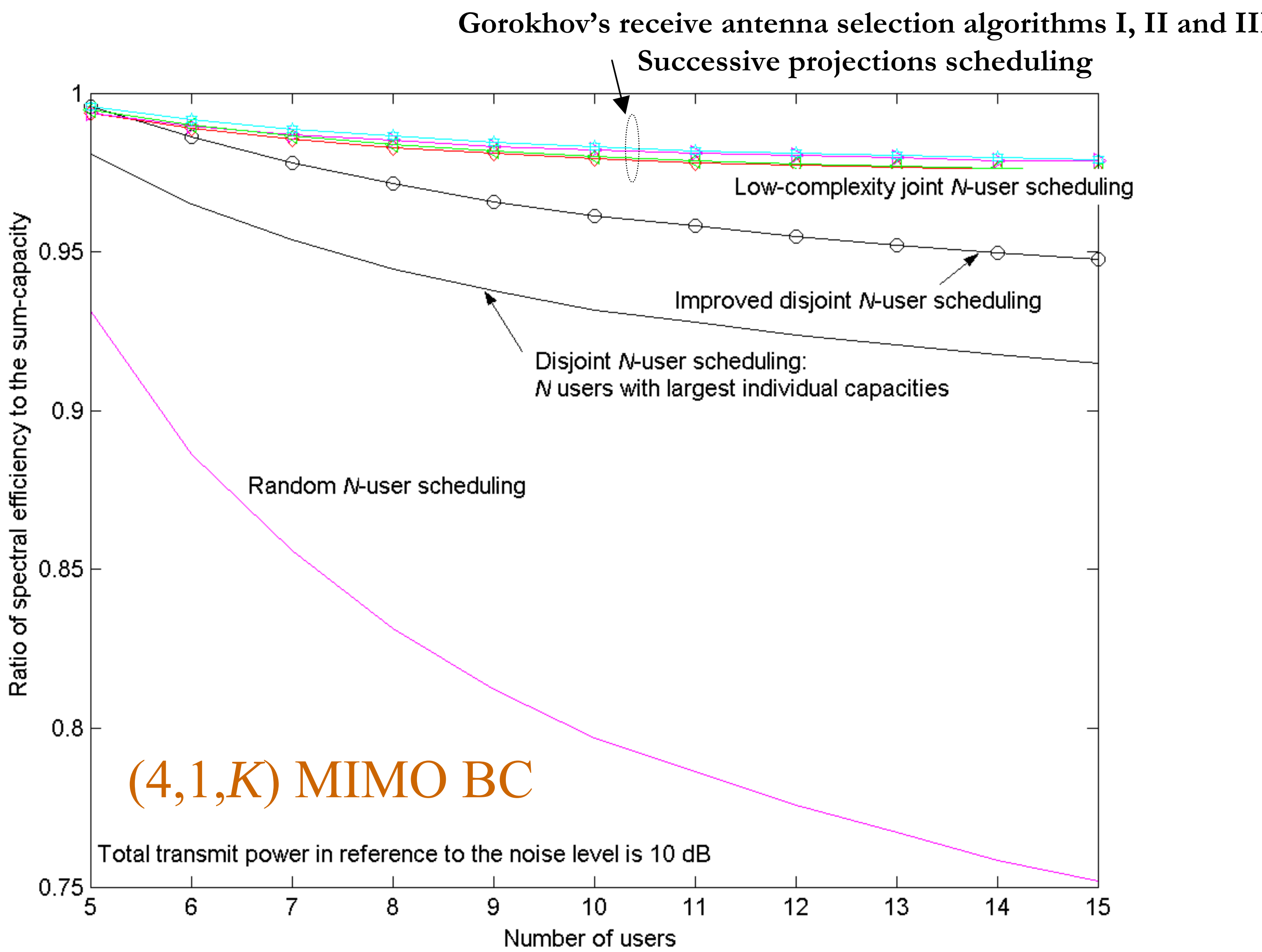
- Schedule users independently of the transmission scheme.
- Schedule up to *N* users in an interference avoidance way.
- Select *N*-*m* users that are spatially compatible with *m* already active users.

Scheduling algorithms for Dirty-Paper Coding



Best *N*-user scheduling: exhaustive search of all sets of *N* users among *K*.
SUR-(*N*+1): selection of *N*+1 user with largest individual capacity, and then exhaustive search.
Gorokhov's algorithm III performs best among the low-complexity algorithms.

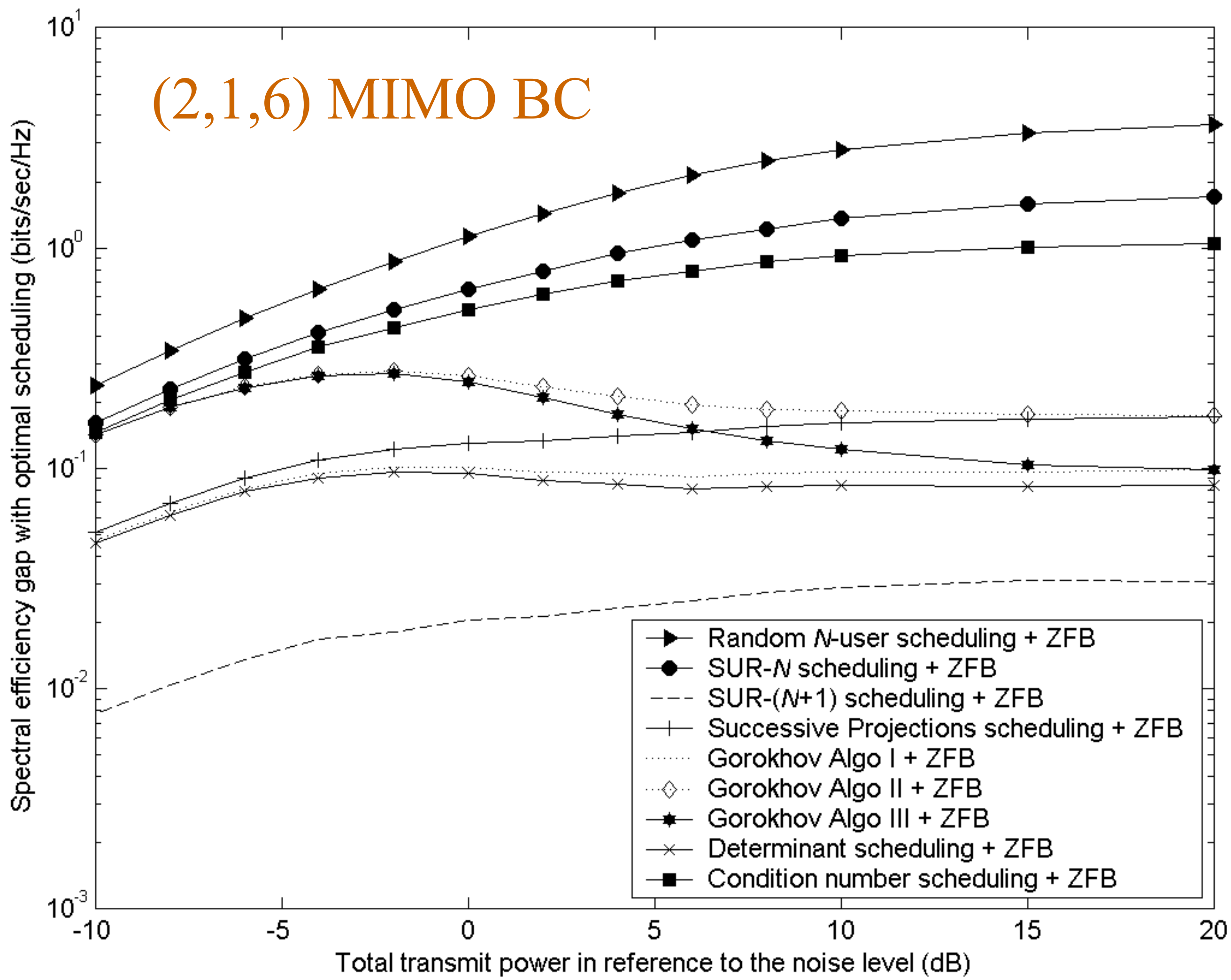
Low-Complexity Maximum-Throughput Scheduling Algorithms for Dirty-Paper Coding



IV. SPATIAL MULTIPLEXING BY LINEAR PROCESSING

Throughput Maximization
Proportionally-Fair Scheduling

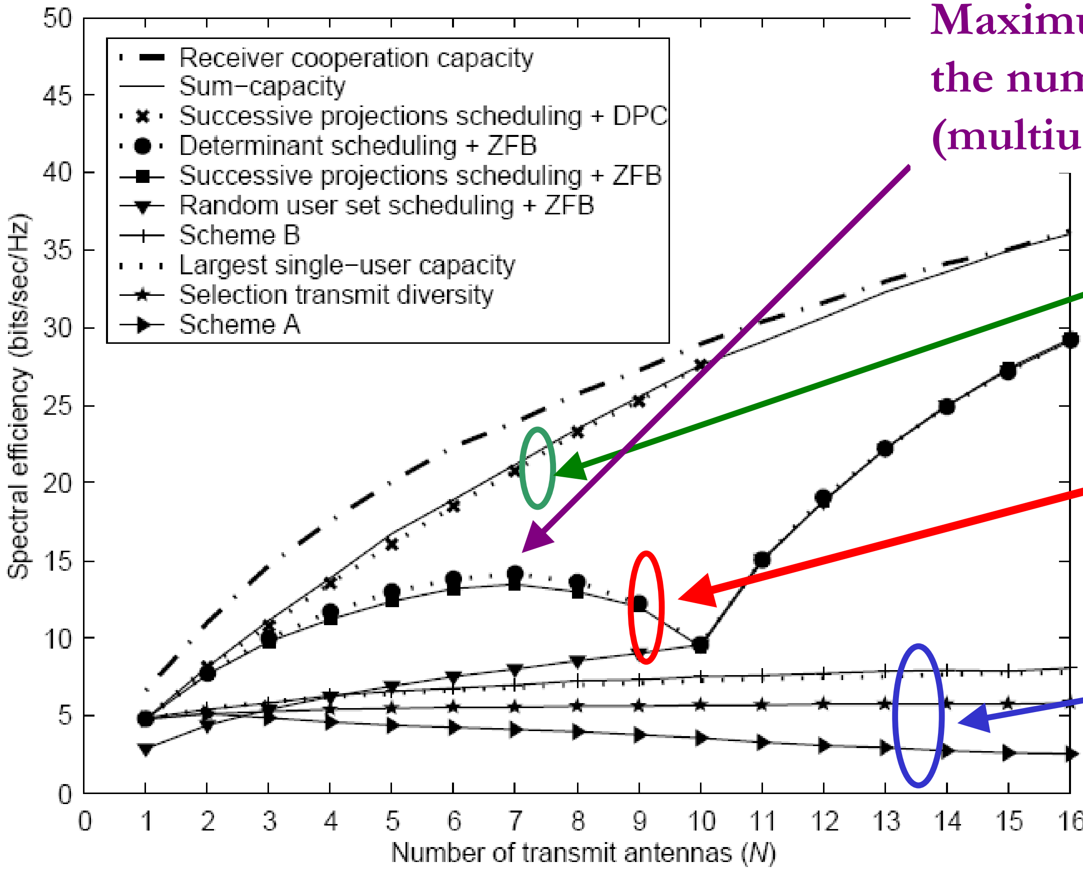
Scheduling algorithms for Zero-Forcing Beamforming (ZFB)



Relative performance of receive antenna selection algorithms is reversed when compared to DPC: Gorokhov's algorithm I is better with ZFB.

Average throughput: DPC vs. ZFB vs. single-user TDMA [15]

(*N*,1,10) MIMO BC with quasi-static Rayleigh fading



Maximum increases with the number of users (multiuser diversity)

Dirty-paper coding

Linear transmitter processing scheme (MIMO gain somewhat exploited)

Partial CSIT, or single-user transmission (no MIMO gain)

Total transmit power in reference to the noise level is 10 dB
ZFB: zero-forcing beamforming (channel inversion at transmitter)
DPC: dirty-paper coding
CSIT: channel state information at the transmitter

Effects of
- **channel state information**
- **scheduling algorithms**
- **linear processing**

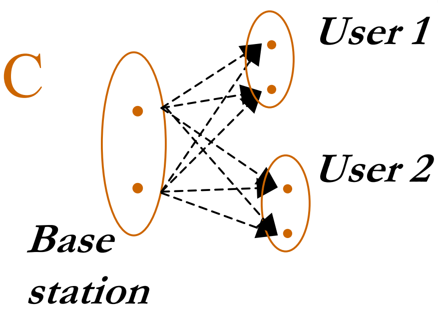
Spatial Multiplexing with Multiple Receive Antennas

- Overview of the problem
  - Base station and users have multiple antennas
  - Transmit channel inversion is no longer possible
  - Jointly optimized transmit and receive filters are needed
  - Problems arise when users do not know all channel matrices
- Previously proposed solutions
  - Group zero-forcing beamforming [16]
  - Joint transmit-receive filters with orthogonality property [17-19]

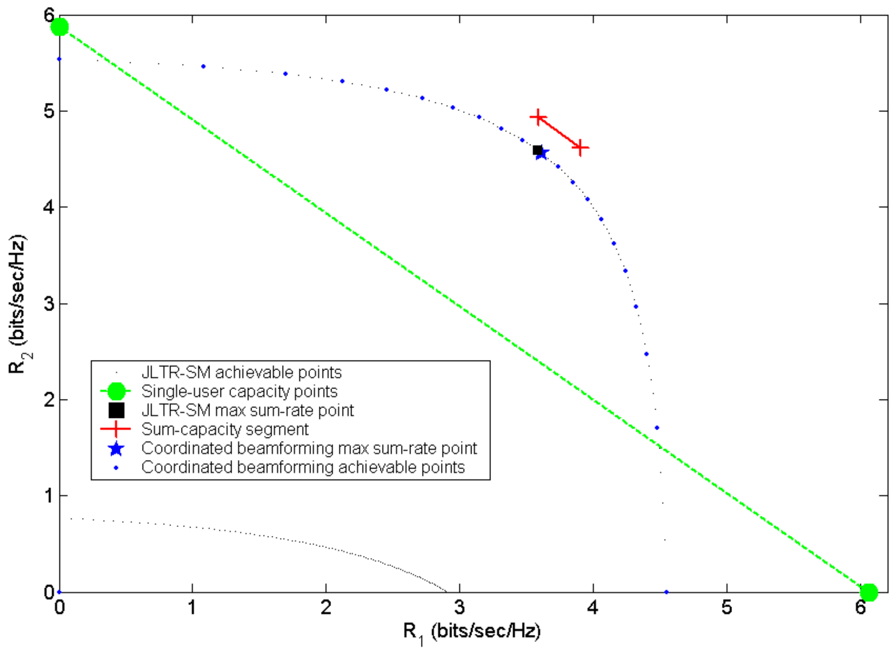
    → Constraints on the number of transmit and receive antennas
- Objectives
  - A scheme with no constraints on the number of antennas and users
  - A scheme that uses only local CSI but still has good performance

Achievable Rate Regions

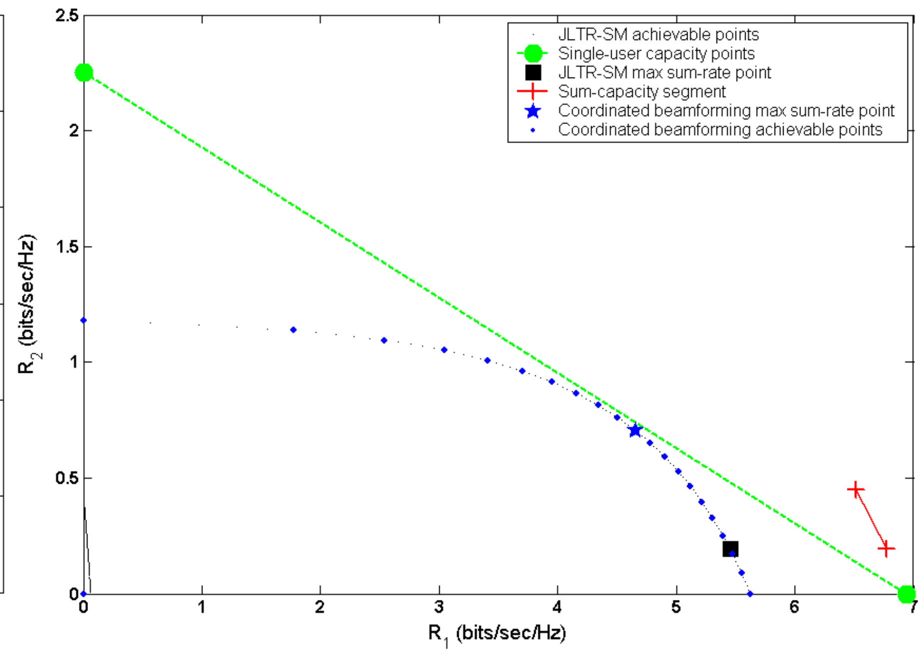
(2,2,2) GBC



Coordinated beamforming, recently proposed by Farhang et al [20] has no constraint on the number of antennas and allows transmitting to N users.



JLTR-SM [21] and coordinated beamforming outperform the single user capacities and approach the sum-capacity for some channel realization.



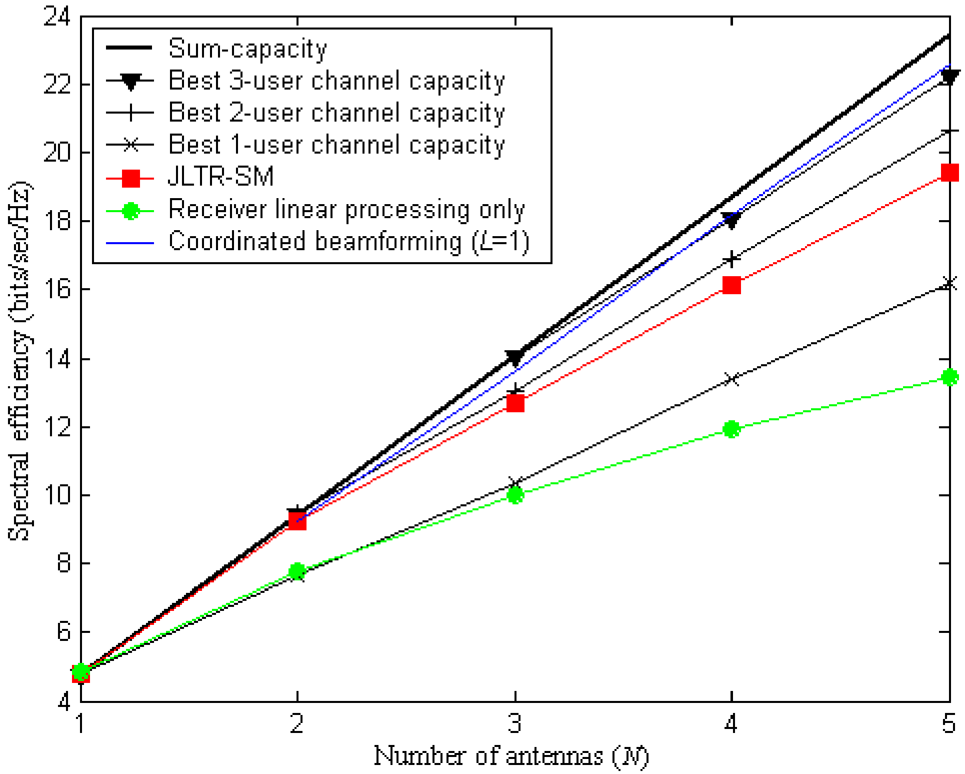
Both JLTR-SM and Coordinated Beamforming can be worse than single-user transmission and far from the sum-capacity for some other channel realization.

Coordinated Beamforming outperforms JLTR-SM when $N > 2$.

JLTR-SM = Joint Linear Transmit & Receive Spatial Multiplexing

Average Sum-Rate Simulations

Average spectral efficiency as a function of the number of transmit and receive antennas N.



(N,N,10) GBC

- The total number of users is $K = 10$.
- Each user has N receive antennas.
- The total power in reference to the noise level is 10 dB.
- L=1: one layer to each of the N users.
- Maximum throughput scheduling

- Receiver processing alone requires only local CSI, but it incurs a large loss in spectral efficiency
- JLTR-SM is limited to transmitting to 2 users simultaneously.
- Coordinated beamforming performs better.
- Gap to the sum-capacity increases as the number of antennas increases.

Spatial Multiplexing with Multiple Receive Antennas

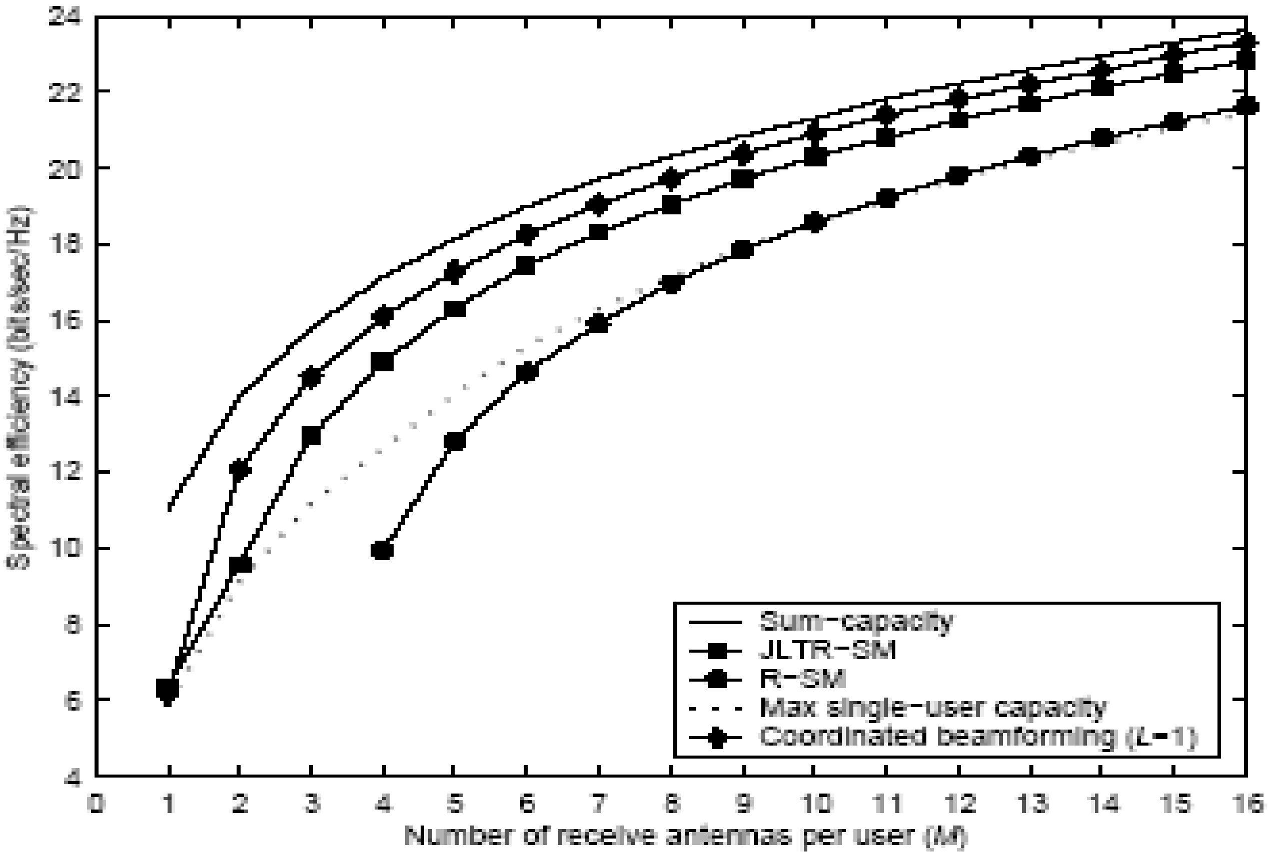


Figure 4-14 Average spectral efficiency on the (4,*M*,4) BC as a function of the number of receive antennas *M* per user. The total power in reference to the noise level is 10 dB.

Proportionally-fair scheduling algorithms

Joint proportionally-fair scheduling algorithm
Weighted sum-rate criterion [16]:

Choose the K_a users that achieve max $\sum_{k=1}^{K_a} \frac{R_k}{\overline{R}_k}$

- $\overline{R}_k$: average throughput of user k

- R_k : rate achievable by user k with some chosen transmission scheme and with K_a-1 other simultaneous users

The active users and their rates are determined jointly

→ **Combinatorial scheduling complexity**

Proportionally-fair scheduling algorithms (cont.)

Disjoint scheduling algorithm

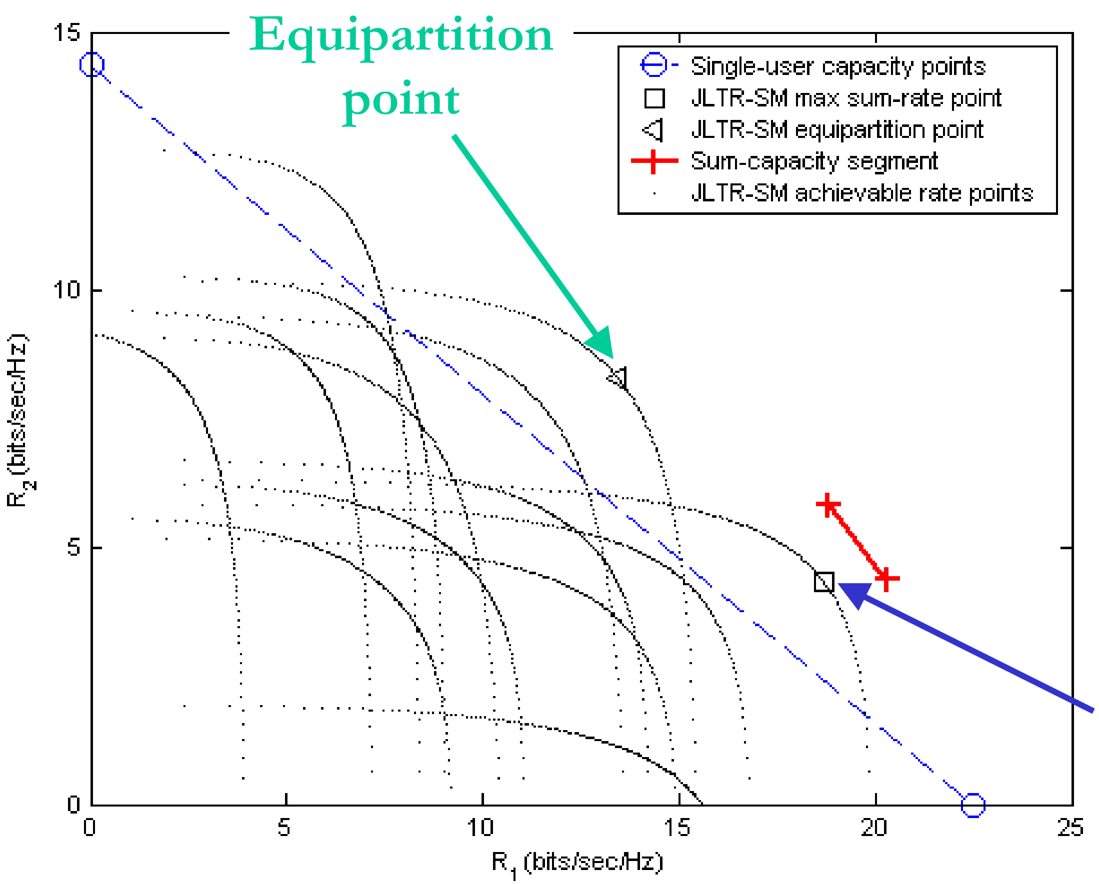
1. Single-user weighted capacities: choose the K_a users with the largest individual weighted capacities $C_k / \overline{R}_k$ [16] .
   C_k : capacity to user k in the absence of other users

2. Fairness in the transmission scheme: choose the rates
   - equipartition JLTR-SM [22]
   - equipartition coordinated beamforming. [22]

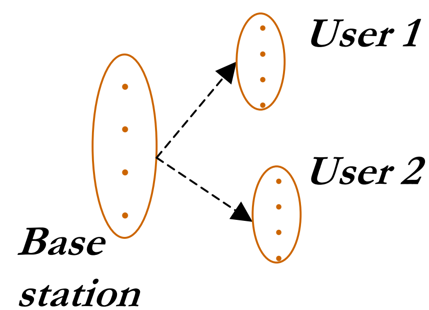
- Transmit to the K_a users simultaneously by equally allocating spatial channels and power: Each of the K_a users is allocated its strongest N/K_a layers and P/K_a of the total transmit power.

- Waterfilling power allocation independently for each active user across its own layers.

Achievable rate regions & equipartition point



4×4:2 GBC



2 layers to each user

Equipartition point provides more fairness in throughput between the two users than the maximum sum-rate point with only a small decrease in sum-rate.

V. Conclusions (1)

Sum-capacity of the MIMO broadcast channel

The nature of the multiuser MIMO broadcast channel requires to revisit transmission strategies for packet-data access systems

✓ It is no longer optimal to transmit to a single user at a time on the $(N,1,N)$ MIMO BC

✓ It is in general not sufficient to transmit to a single user at a time on the (N,N,K) MIMO BC unless the transmit power is very large

✓ Nature of the problem closely related to co-channel interference in MIMO channels

Linear spatial multiplexing

✓ Single-antenna receivers: transmitter channel inversion is not good enough, non-linear precoding is required (lattice-reduction techniques)

✓ Multiple-antenna receivers: main difficulty in obtaining channel state information to perform joint optimization of transmit and receive filters

Conclusions (2)

Low-complexity *N*-user scheduling

- *N*-user scheduling is required if dirty-paper coding is not feasible

- Receive antenna selection algorithms used as scheduling algorithms are efficient for throughput maximization in the single-receive antenna case

- Disjoint scheduling algorithms are efficient with path loss and shadow fading because then the SNR difference between users is more important than the spatial structure of the channel

- With multiple receive antennas, joint transmit/receive filtering allows to decrease the complexity of scheduling algorithms by taking care of the spatial interference

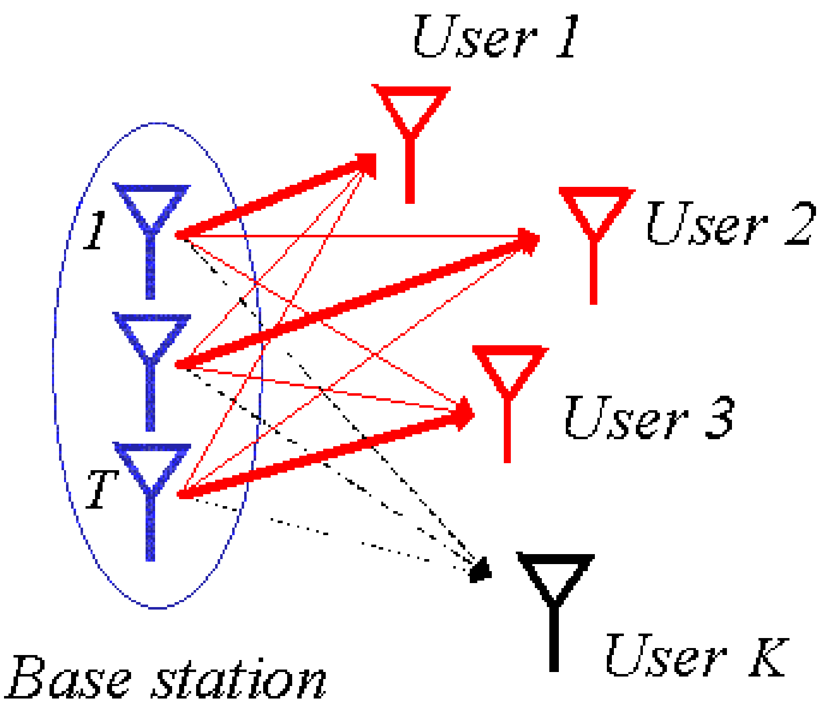
- Not optimizing the layers and power allocation (waterfilling) can be a good thing with proportionally-fair scheduling

- Main impact of multiple antennas: tight control of average delay and increased aggregate throughput

Appendix 1: Scheduling Algorithm with Partial CSIT: SINR Feedback [15]

Algorithm A (all transmit antennas used)

- Base station sends pilot signals from each antenna
- Users estimate SINR for each transmit antenna + feedback to base station
- Scheduler chooses T active users based on largest SINR for each Tx antenna
- BS transmits to all users simultaneously with weight vector $[0\ldots0\ 1\ 0\ldots0]^T$ and capacity-achieving codes
- Users decode their signals by treating other signals as interference



Algorithm B: Adaptive version selects a subset of transmit antennas to eliminate the interference limitation at high SNR

Appendix 2: Scheduling Algorithms with Complete CSIT [7,13]

- Random N-user scheduling: N users are randomly selected.
- Best N-user scheduling: for each possible set of N users, the sum-capacity of the N-user broadcast channel is computed. The best N-user set is the one that achieves the largest N-user sum-capacity. This scheduling algorithm has combinatorial complexity in "K choose N"
- Single-User Rates (SUR-N) scheduling: the N users with the largest individual capacities are selected.
- SUR-(N+1) scheduling: first select the N+1 users with the largest individual capacities, then the N users that offer the largest sum-capacity by exhaustive search with combinatorial complexity in "$N+1$ choose N".
- Successive Projections Scheduling chooses N users such that their channel vectors are as little linearly dependent as possible.
- Gorokhov's receive antenna selection algorithms [14]: by treating each user as a different antenna as in a single receiver with multiple antennas, Gorokhov's low-complexity algorithms allow to select N receive antennas out of K and to limit the capacity loss. Algorithm II is based on decremental selection. Algorithm III is based on incremental selection. Algorithm I is based on a different criterion to minimize the capacity loss. It is the only one of the 3 algorithms that does not take into account the value of the total transmit power constraint. It has been found as a solution in the high power region as P goes to infinity.
- Determinant Scheduling chooses the N user such that the N by N channel matrix has the largest determinant.
- Condition Number Scheduling chooses the N user such that the N by N channel matrix has the smallest condition number (ratio of largest to smallest eigenvalue).

References (1)

1. D.N.C. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," in *Proc. IEEE ISIT'97*, Ulm, Germany, 29 June – 4 July 1997, p. 27.
2. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New-York: Wiley, 1991.
3. G.J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.
4. I.E. Telatar, "Capacity of multi-antenna Gaussian channels" *European Transactions on. Telecommunications*, vol. 10, no. 6, pp. 585-595, Nov/Dec 1999.
5. A.G. Kogiantis, N. Joshi and O. Sunay, "On transmit diversity and scheduling in wireless packet data," in *Proc. IEEE ICC'01*, Helsinki, Finland, June 2001, pp. 2433 – 2437.
6. B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Information Theory*, vol. 50, no. 9, pp. 1893 -1909, September 2004.
7. D.J. Mazzarese, *High Throughput Downlink Wireless Packet Data Access with Multiple Antennas and Multi-User Diversity*, Ph.D. Thesis, University of Alberta, 2005.
8. G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.
9. M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
10. S. Vishwanath, N. Jindal and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Information Theory*, vol. 49, no. 10, pp. 2658–2668, October 2003.
11. D.J. Mazzarese and W.A. Krzymień, "Throughput maximization and optimal number of active users on the two transmit antenna downlink of a cellular system," in *Proc. IEEE PACRIM'03*, Victoria, Canada, August 2003, pp. 498 – 501.
12. D.J. Mazzarese and W.A. Krzymień, "Design rules for efficient scheduling of packet data on multiple antenna downlink," in *Proc. 14th IST Mobile and Wireless Communications Summit,* Dresden, Germany, June 2005.

References (2)

13. D.J. Mazzarese and W.A. Krzymień, "Scheduling algorithms and throughput maximization for the downlink of packet-data cellular systems with multiple antennas at the base station," submitted to *Wireless Personal Communications (Springer)*.
14. A. Gorokhov, D. Gore and A. Paulraj, "Receive antenna selection for MIMO flat-fading channels: theory and algorithms," *IEEE Trans. Information Theory*, vol. 49, no. 10, pp. 2687–2697, October 2003.
15. D.J. Mazzarese and W.A. Krzymień, "High throughput downlink cellular packet data access with multiple antennas and multiuser diversity," in *Proc. IEEE VTC'03-Spring*, Jeju, Korea, April 2003, pp. 1079–1083.
16. H. Viswanathan, S. Venkatesan and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas in Commun.*, vol. 21, no. 5, pp. 802–811, June 2003.
17. K. K.Wong, R. D. Murch, and K. B. Letaief, "Performance enhancement of multiuser MIMO wireless communications systems," *IEEE Trans. Communications*, vol. 50, no. 12, pp. 1960–1970, December 2002.
18. K. K. Wong, R. D. Murch, and K. B. Letaief, "A joint channel diagonalization for multiuser MIMO antenna systems," *IEEE Trans. Wireless Communications*, vol. 2, pp. 773–786, July 2003.
19. L.-U. Choi and R.D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Communications*, vol. 3, no. 1, pp. 20 - 24, January 2004.
20. B. Farhang-Boroujeny, Q. Spencer and L. Swindlehurst, "Layering techniques for space-time communication in multi-user networks," in *Proc. IEEE VTC'03-Fall*, Orlando, FL, USA, October 2003, pp. 1339-1343.
21. D.J. Mazzarese and W.A. Krzymień, "Linear space-time transmitter and receiver processing and scheduling for the MIMO broadcast channel", in *Proc. IEEE VTC'04-Spring*, Milan, Italy, May 2004.
22. D.J. Mazzarese and W.A. Krzymień, "Space-time linear processing and scheduling for the cellular downlink," in *Proc. 13th IST Mobile and Wireless Communications Summit,* Lyon, France, pp. 85-90, June 2004.

Grazie mille!