Cross-layer Scheduling and Resource Allocation in Wireless Communication Systems

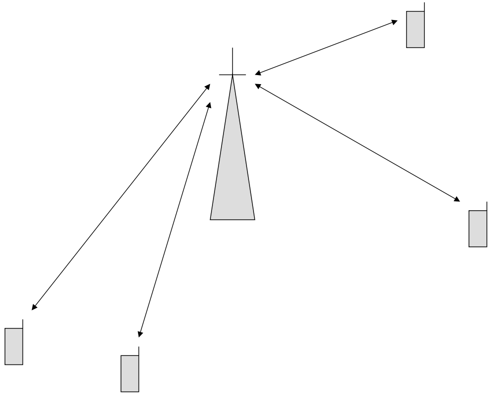
Srikrishna Bhashyam

Department of Electrical Engineering
Indian Institute of Technology Madras

2 July 2014

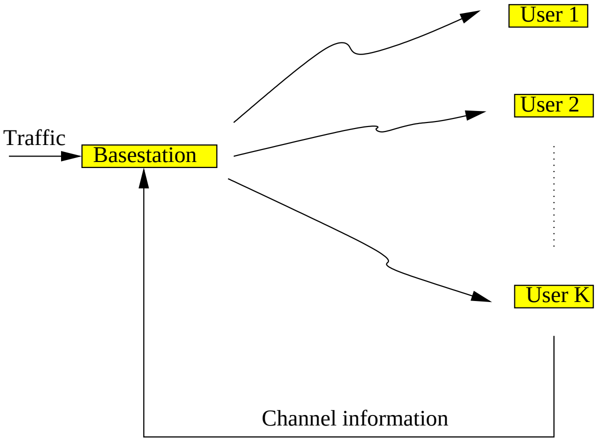
Wireless Systems

Cellular System



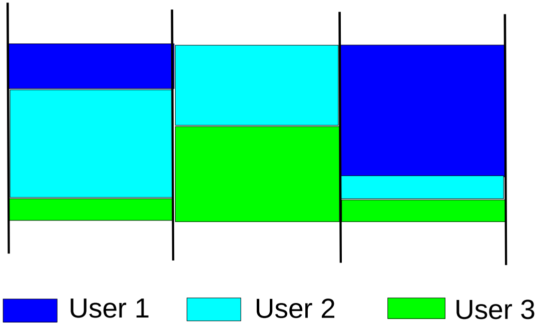
- Time-varying channel
- Resource sharing – Interference constraints

Downlink Resource Allocation Problem



- Physical resources: power and bandwidth
- Total transmit power constraint
- Maximize system throughput
- Fairness or Quality of Service (QoS) constraints

Dynamic Resource Allocation

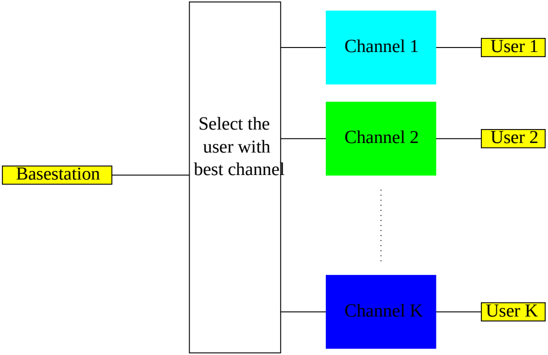


Periodic reallocation of resources

- Resources: Time, Bandwidth, Power
- Adaptation to channel and traffic conditions
- Dynamic resource allocation
  - Reallocation period of the order of a millisecond

Adapting to the Channel

Adapting to the Channel: Maximizing Capacity[1]

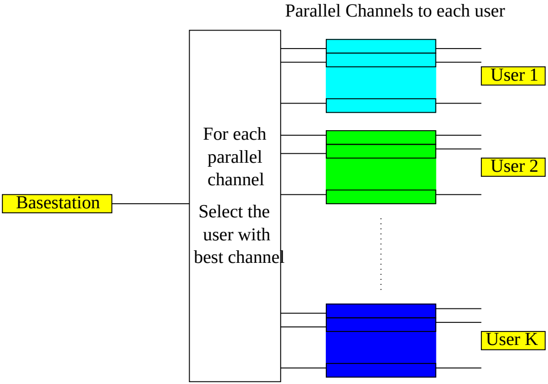


- Infinite backlog assumption
- All power and bandwidth resources to one user
- User with best achievable rate chosen: $i = \arg\max_k R_k$, where R_k is the rate that can be supported by user k.

---

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

Maximizing Capacity: Parallel Channels



- Bandwidth resources split to achieve parallel channels
- For each channel n, user with best channel conditions chosen:

$$i_n = \arg\max_k R_{k,n}.$$

- Water-filling power allocation

Fairness

Proportional Fairness[2] [3]

- $i = \arg\max_k \frac{R_k}{R_{k,av}}$,

  where $R_{k,av}$ is the average rate that can be supported by user k.

- $\max \sum_k \log\left(T_k\right)$,

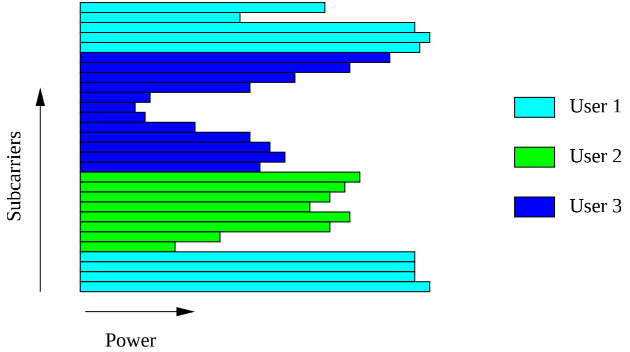
  where T_k is the average long-term throughput of user k.

---

[2] E. F. Chaponniere, P. Black, J. M. Holtzman, and D. Tse, "Transmitter directed multiple receiver system using path diversity to equitably maximize throughput," *U. S. Patent No. 6449490*, September 2002.

[3] P. Viswanath, D. N. C. Tse, R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277-1294, June 2002.

Parallel Channels: OFDM[4] [5]



- Available resources:
  - Subcarriers
  - Transmit power
- Channel is frequency-selective ⇒ subcarriers not identical.

---

[4] C. Y. Wong, R. S. Cheng, K. B. Letaief, R. D. Murch, “Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation”, *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 10, pp. 1747-1758, October 1999.

[5] J. Jang, K. B. Lee, “Transmit power adaptation for multiuser OFDM systems”, *IEEE Journal on Selected Areas in Communications,* vol. 21, no. 2, pp. 171-178, February 2003.

Fairness: Joint Subchannel and Power Allocation

- Proportional rate subcarrier allocation[6]
- Proportional rate subcarrier allocation + power optimization[7]
- Joint subcarrier and power allocation[8]
- Utility Maximization[9]

---

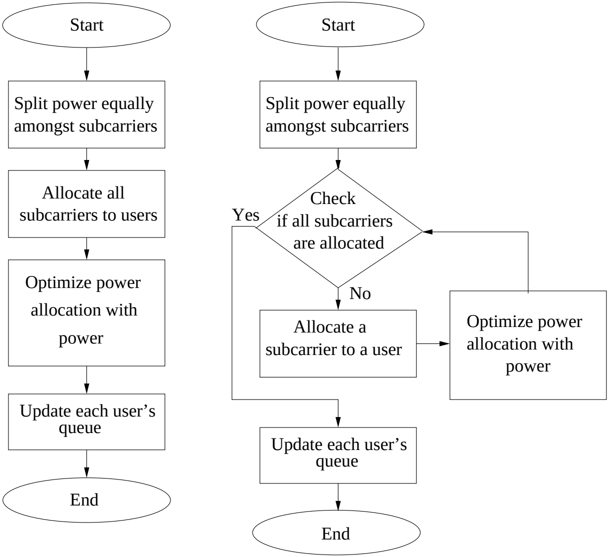
[6] W. Rhee, J. M. Cioffi, "Increase in Capacity of Multiuser OFDM System Using Dynamic Subchannel Allocation", *Proceedings of the 51st IEEE Vehicular Technology Conference, Tokyo,* vol. 2, pp. 1085-1089, Spring 2000.

[7] Z. Shen, J. G. Andrews, B. L. Evans, "Adaptive Resource Allocation in Multiuser OFDM Systems with Proportional Rate Constraints", *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726-2737, November 2005.

[8] C. Mohanram, S. Bhashyam, "A sub-optimal joint subcarrier and power allocation algorithm," *IEEE Communications Letters*, vol. 9, no. 8, pp. 685-687, August 2005.

[9] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, Downlink scheduling and resource allocation for OFDM systems, Wireless Communications, IEEE Transactions on, vol. 8, no. 1, pp. 288296, 2009.

Fairness: Joint Subchannel and Power Allocation



Gradient Algorithm

Stolyar[10]

- General utility functions
- Multiuser scheduling at the same time
- Proportional Fairness is a special case

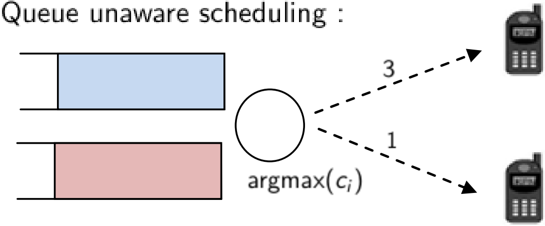
---

[10] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling for multi-user throughput allocation," *Operations Research*, vol. 53, no. 1, pp. 12-25, 2005.

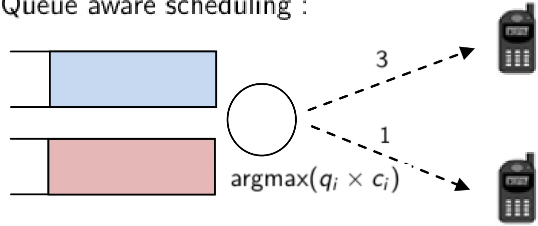
Adapting to the Channel and Traffic

Why Queue-aware Scheduling?

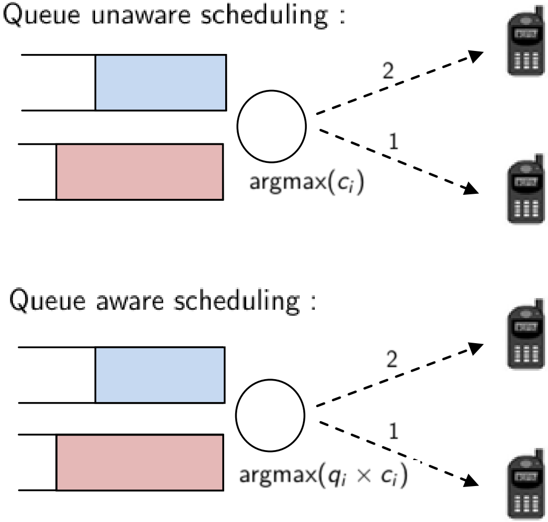
Queue unaware scheduling :



Queue aware scheduling :



Why Queue-aware Scheduling?



Queue unaware scheduling :

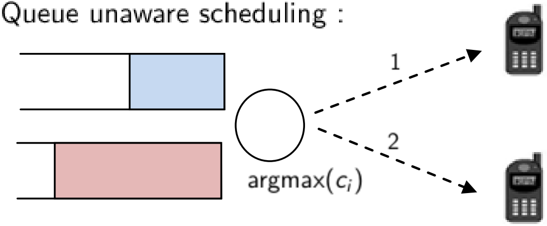
argmax(c_i)

Queue aware scheduling :

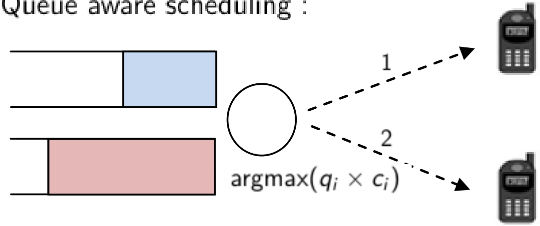
argmax($q_i \times c_i$)

Why Queue-aware Scheduling?

Queue unaware scheduling :

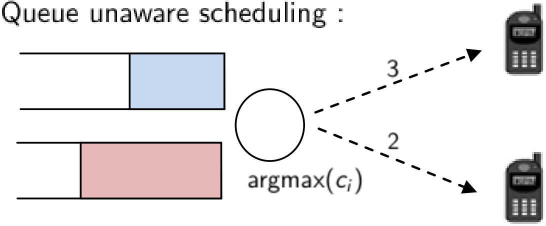


Queue aware scheduling :

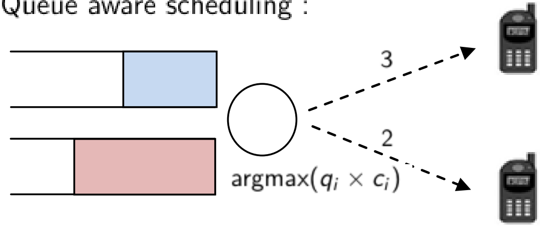


Why Queue-aware Scheduling?

Queue unaware scheduling :

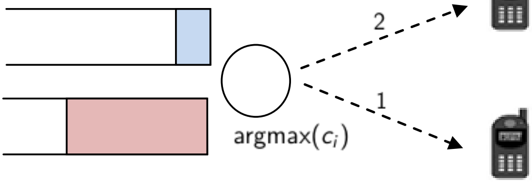


Queue aware scheduling :



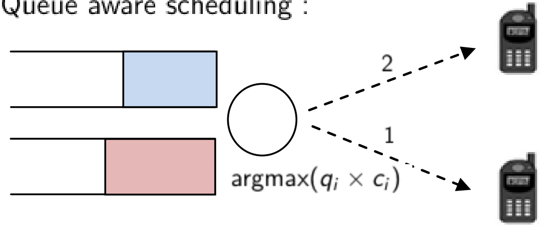
Why Queue-aware Scheduling?

Queue unaware scheduling :



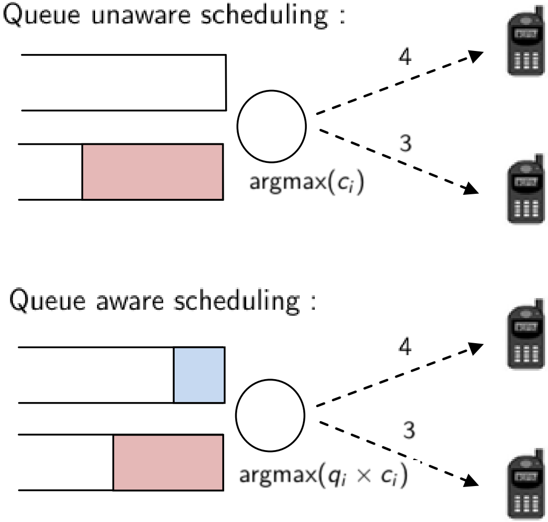
argmax(c_i)

Queue aware scheduling :

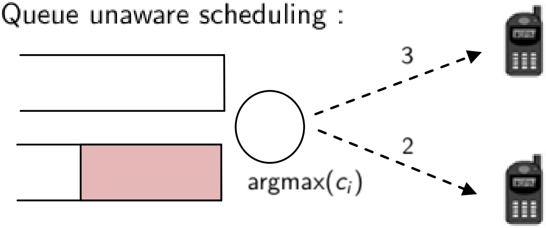


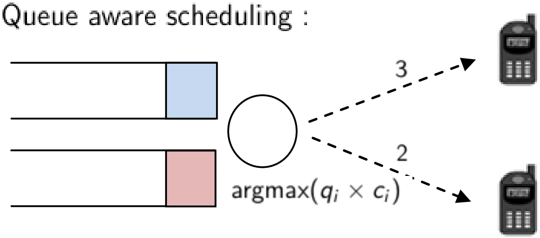
argmax($q_i \times c_i$)

Why Queue-aware Scheduling?

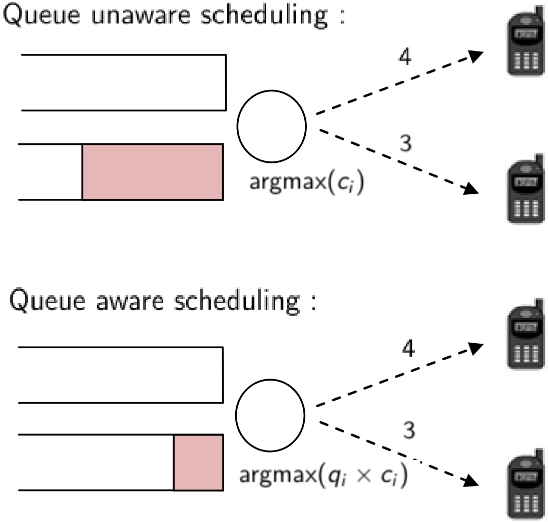


Why Queue-aware Scheduling?



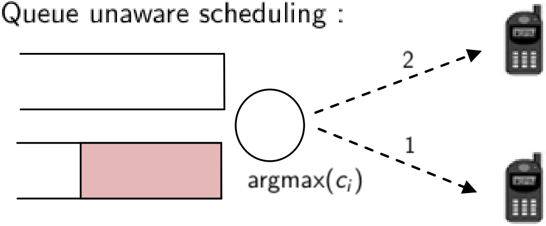


Why Queue-aware Scheduling?

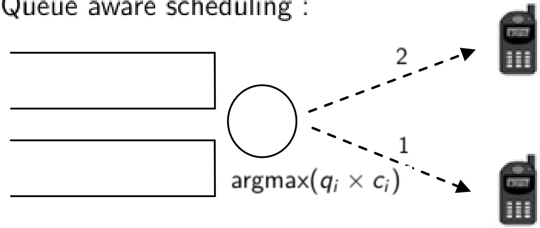


Why Queue-aware Scheduling?

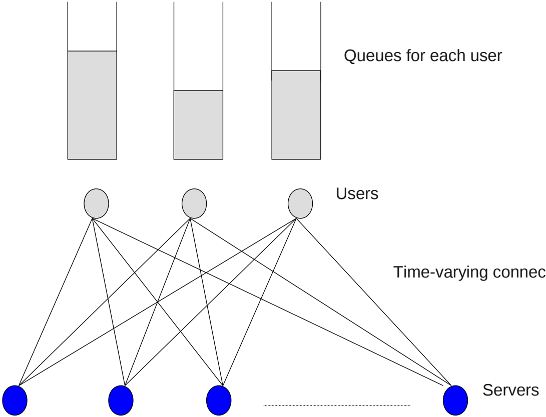
Queue unaware scheduling :



Queue aware scheduling :



Adapting to the Channel and Traffic[11]



- Multi-Queue Multi-Server Model for each time slot
- Server: Subcarrier/Group of subcarriers/Spreading code

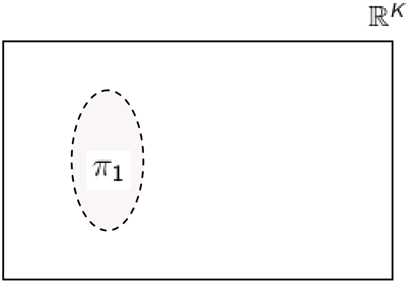
---

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

Resource Allocation/Cross-layer Scheduling Goals

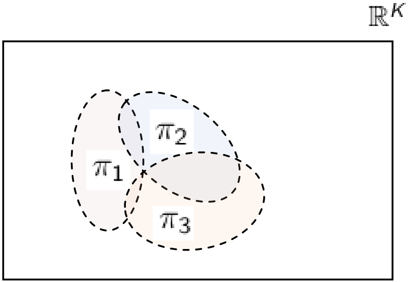
- Scheduling Goals
    - Stability and throughput optimality
        - Stability: Average queue length finite
    - Packet delay constraints
    - Fairness

Stability Region



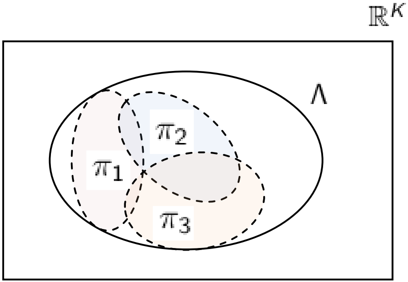
π_1 is a policy in $\mathcal{P}$.

Stability Region



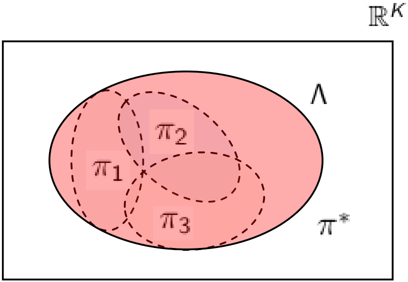
π_1, π_2 and π_3 are policies in $\mathcal{P}$.

Stability Region



π_1, π_2 and π_3 are policies in $\mathcal{P}$.

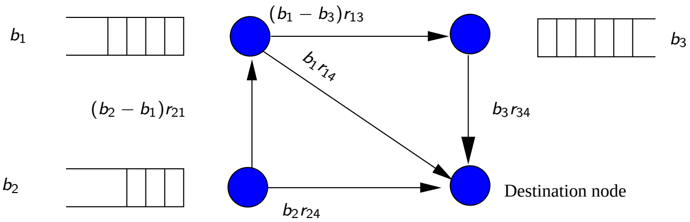
Stability Region



π^* is a throughput optimal policy in $\mathcal{P}$.

Stability in a general wireless network

- Dynamic backpressure policy[12] [13]



- Interference model: Only certain links can be activated simultaneously
- Scheduling problem: Which links will you activate?
- Solution: Activate those links such that the sum of their weights is maximum.

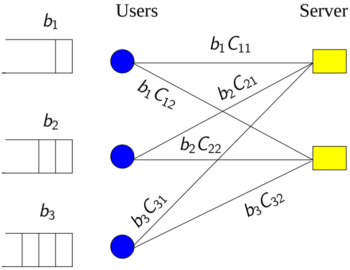
---

[12] L. Tassiulas, A. Ephremides, "Stability properties of constrained queueing systems and scheduling for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control,* vol. 37, no. 12, pp. 1936-1949, December 1992.

[13] L. Georgiadis, M. J. Neely, L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 1-144, 2006.

Dynamic back-pressure policy for our setting

Max-Weight Scheduling



- Only one link per server to be activated. Which links to activate?
- *Solution*:
    - Make the servers as destination nodes.
    - Assign the weights for each link as in back-pressure policy.
    - Activate those links such that the sum of their weights is maximum.

$$\max \sum_k b_n C_{nk}$$

b_n: Backlog of user n, C_{nk}: Capacity of user n on server k

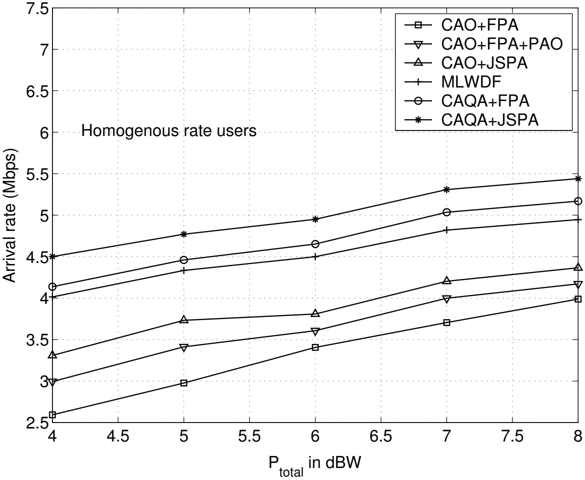
Joint Server and Power Allocation

- Finite number of power levels
  - Max-weight scheduling
- Joint subcarrier and power allocation
  - Joint optimization
  - Sub-optimal solutions[14]

---

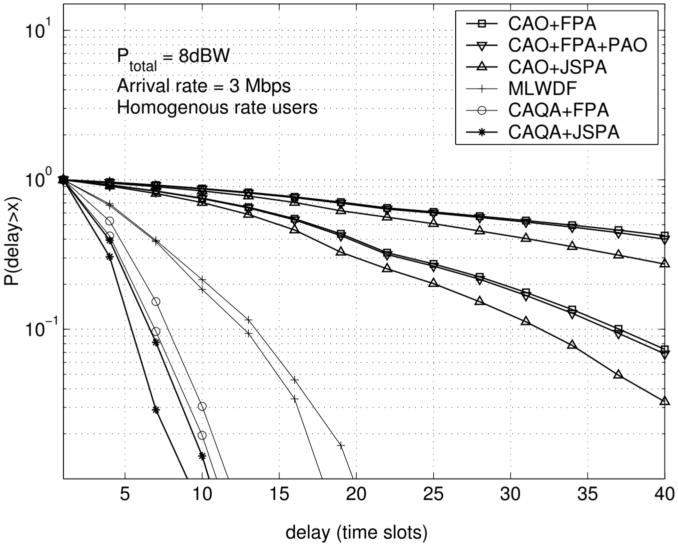
[14] C. Mohanram, S. Bhashyam, "Joint subcarrier and power allocation in channel-aware queue-aware scheduling for multiuser OFDM," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, September 2007.

Results: Max. Arrival Rate vs. Transmit Power



- Max. arrival rate for less than 0.5% packets dropped
- CAO: Channel-aware only, CAQA: Channel-aware Queue-aware
- FPA: Fixed power allocation, JSPA: Joint subcarrier and power

Results: Delay Performance



- Best and worst delay performance among users plotted

Fairness and Utility Maximization[15]

- Arrival rate vector outside stability region
  - Support a fraction of the traffic
  - Optimize utility based on long term throughput
  - Flow control to get stabilizable rates + stabilizing policy
  - Fairness based on choice of utility function
    - Proportional fairness

---

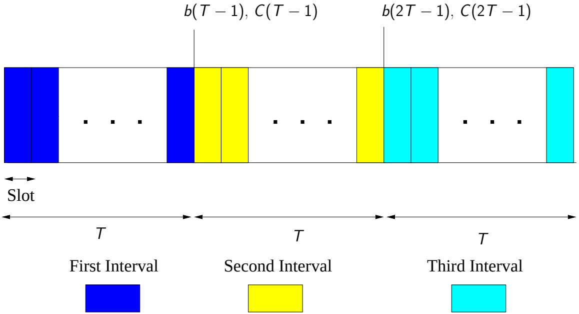
[15] L. Georgiadis, M. J. Neely, L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 1-144, 2006.

Adapting with Partial Information

Infrequent measurements

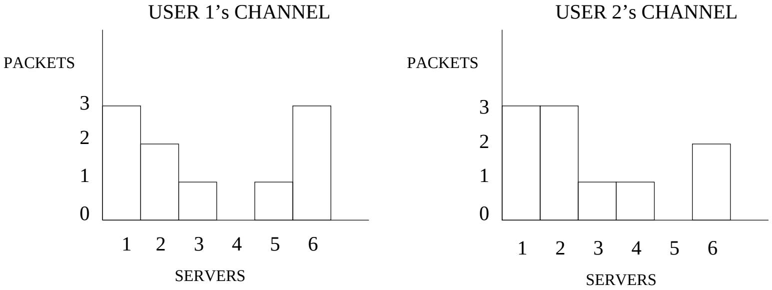
C. Manikandan, S. Bhashyam, R. Sundaresan, "Cross-layer scheduling with infrequent channel and queue measurements," IEEE Transactions on Wireless Communications, vol. 8, no. 12, pp. 5737-5742, December 2009.

Using Delayed Information



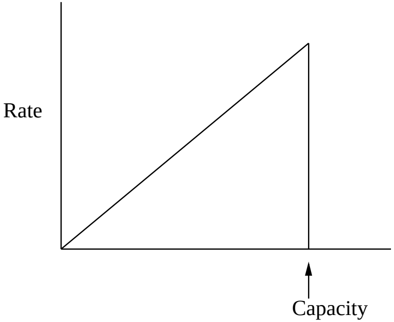
- Time-slots are grouped into intervals
- Channel and queue information available only once in T slots

Channel model



- C_{nk}: channel capacity of user n on server k.
- $C_{nk} \in \{0, 1, 2, 3\}$.

Loss model



Packets sent

R_{nk}	0
$R_{nk} <= C_{nk}$	$R_{nk} > C_{nk}$

- R_{nk}: number of packets user n transmits on server k.
- $C_{nk}(lT-1)$: channel information available at the start of l^{th} interval.

Scheduling with infrequent measurements

- Retain throughput optimality of dynamic backpressure policy
- Two policies: Policy 1 and Policy 2[17]
- Comparison with KLS policy[18]
- Delayed network-state information[19]

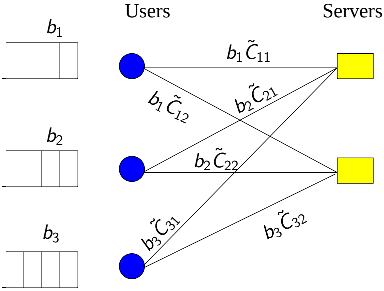
---

[17] C. Manikandan, S. Bhashyam, R. Sundaresan, "Cross-layer scheduling with infrequent channel and queue measurements," *IEEE Transactions on Wireless Communications,* vol. 8, no. 12, pp. 5737-5742, December 2009.

[18] K. Kar, X. Luo, S. Sarkar, "Throughput-optimal scheduling in multichannel access point networks under infrequent channel measurements," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2619-2629, July 2008.

[19] L. Ying and S. Shakkottai, On throughput optimality with delayed network-state information, IEEE Transactions on Information Theory, vol. 57, no. 8, pp. 5116–5132, 2011.

Policy 1 & Policy 2



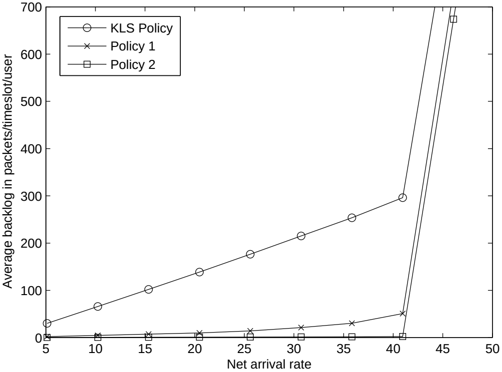
$$\text{Define} \quad \tilde{C}_{nk} = \max E\left[T_{nk}(t) | C_{nk}(lT-1)\right]$$
$$= \max_r r \Pr\{r \leq C_{nk} \mid C_{nk}(lT-1)\}$$

- Policy 1 is the dynamic back pressure policy for our setting
- Assignment changes every slot
- Policy 2: Update queue information after each server is scheduled

Simulation setup

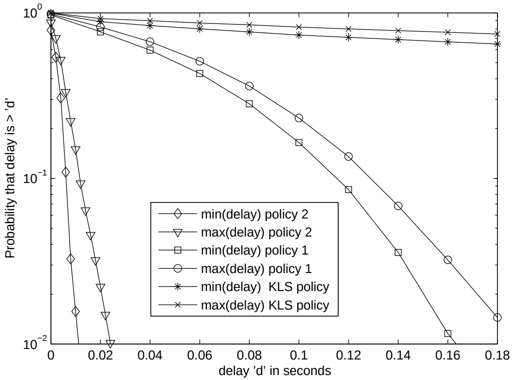
- Truncated Poisson arrivals
- 128 users and 16 servers
- Markov fading channel with probability transition matrix
- Backlog and delay are used as metrics for comparison
- Simulations for both symmetric and asymmetric arrivals
    - Symmetric case shown here

Average backlog comparison: Slow fading, $T = 8$



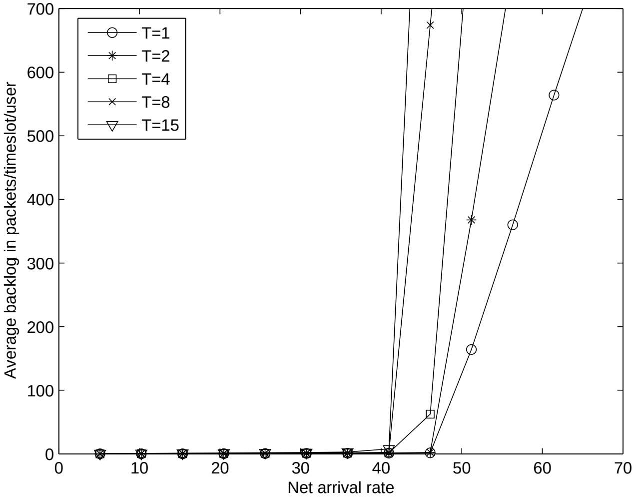
- All the policies have similar stability region.
- At low traffic, proposed policies outperform KLS policy.

Delay comparison

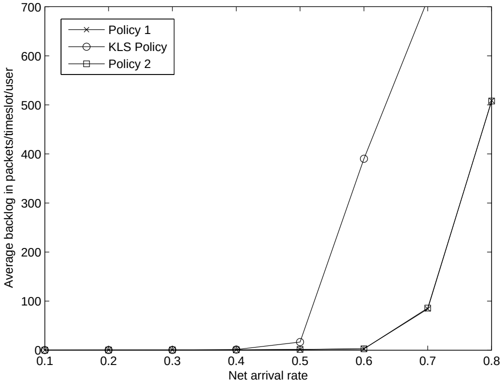


- Net arrival rate $= 25.6$, $T = 4$

Average backlog comparison vs T for Policy 2



Comparison of stability regions: Fast fading



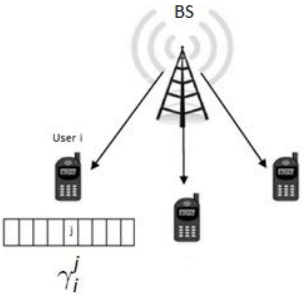
- 2 queues, 1 server, $T = 2$, states are $\{0, 1\}$
- Probability transition matrix: $\begin{bmatrix} \delta & 1-\delta \\ 1-\delta & \delta \end{bmatrix}$, $\delta = 0.1$

Adapting with Partial Information

Best M sub-band feedback in LTE

H. Ahmed, K. Jagannathan, S. Bhashyam, "Queue-Aware Optimal Resource Allocation for the LTE Downlink," Proceedings of IEEE GLOBECOM 2013, Atlanta, GA, USA, Dec. 2013.

Resource Allocation for the LTE Downlink



Resources

- OFDM with hundreds of sub-carriers (512, 1024, 2048)
- Group of 12 sub-carriers - Resource Block (RB)
- Sub-band - one to three RBs
- K users, N sub-bands, γ_i^j - SNR for i^{th} user in j^{th} sub-band

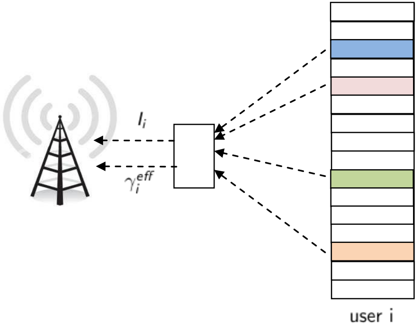
Components of Resource Allocation

- Sub-band Assignment
- Rate allocation
- Power allocation

- For Optimal allocation, perfect CQI is needed at the BS:
  - N bands for each of the K users
  - HUGE amount of feedback!

UE-selected sub-band feedback mode (3GPP)

- Limited feedback



- **Index set I_i :**
  The indices of the M best sub-bands of the i^{th} user

$$I_i = \{i_1, i_2, \ldots, i_M\}$$

- **Effective Exponential SNR Mapping (EESM) γ_i^{eff} :**

$$\gamma_i^{eff} = -\eta \ln \left(\frac{1}{M} \sum_{j=1}^{M} e^{-\frac{\gamma_i^{i_j}}{\eta}} \right)$$

Say $N = 43$ and $M = 4$.
Huge reduction in feedback overhead

Problem Setup

Goal: Find a resource allocation policy that maximizes throughput while keeping all the queues stable given the following limited information:

- The EESMs $\underline{\gamma}^{eff} = [\gamma_1^{eff}, \gamma_2^{eff}, \ldots, \gamma_K^{eff}]$
- The index sets $\underline{I} = [I_1, I_2, \ldots, I_K]$
- The queue length vector $\underline{Q} = [Q_1, Q_2, \ldots, Q_K]$.

Let $\mathcal{P}$ be the family of all policies which allocate equal power to all scheduled sub-bands, and have access only to the parameters $\underline{\gamma}^{eff}, \underline{I}$, and $\underline{Q}$.

Impact of Limited Feedback - Outage

Outage Probability $P_{i,j}()$:
Outage occurs when the allocated rate exceeds the capacity

$$P_{i,j}(r_{i,j}) = \mathbb{P}\{C_{i,j} < r_{i,j} | \gamma_i^{eff}, I_i\}$$

Goodput $G_{i,j}()$:
Average successfully transmitted amount of data for i^{th} user in j^{th} sub-band

$$G_{i,j}(r_{i,j}) = r_{i,j}\left(1 - P_{i,j}(r_{i,j})\right) + 0 \times P_{i,j}(r_{i,j})$$

Methodology Used

- Lyapunov stability analysis
  - Minimizing the Lyapunov drift
  - Formulated this as a convex optimization problem
  - Solution using KKT conditions
- Calculation of Outage probability
  - Using a weak limit theorem on order statistics of sub-band SNRs[21]

---

[21] T. Ferguson, A Course in Large Sample Theory: Texts in Statistical Science, Chapman & Hall/CRC, 1996.

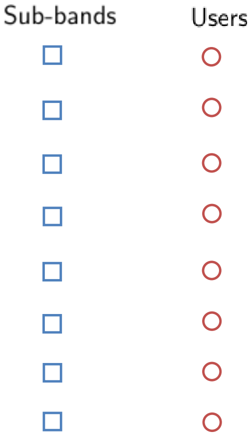
Throughput Optimal Resource Allocation Policy

During each time slot, the scheduler at the BS observes $\underline{\gamma}^{eff}, \underline{I}$, and $\underline{Q}$, and implements the following steps :

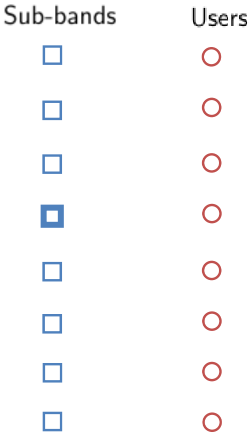
Rate and User allocation for a sub-band :

- Find the users which reported this sub-band
- Calculate the optimum rate which maximizes the goodput for each of these users
- Pick the user with the maximum queue-length goodput product
- Assign the sub-band to this user and transmit at its optimum rate

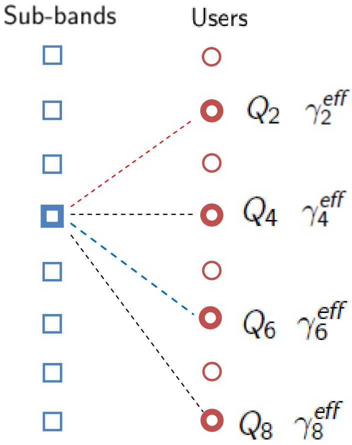
Throughput Optimal Resource Allocation Policy



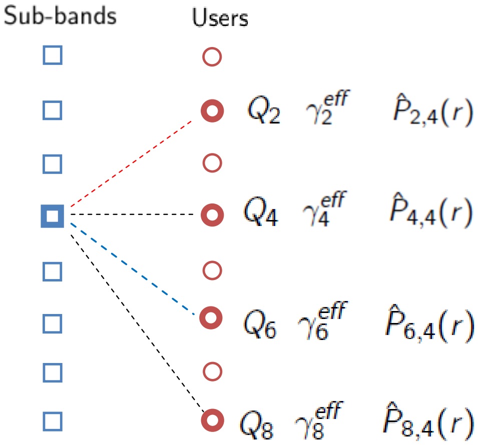
Throughput Optimal Resource Allocation Policy



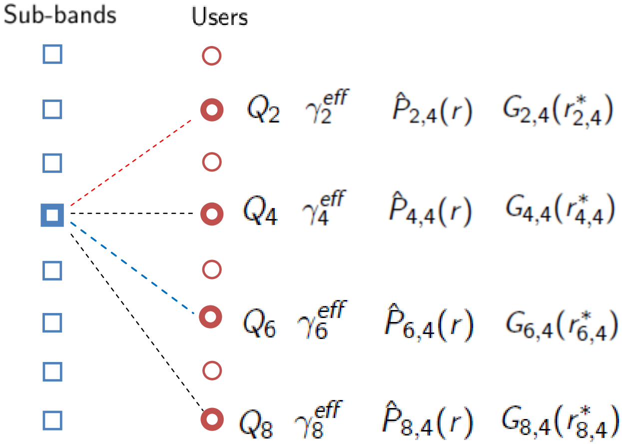
Throughput Optimal Resource Allocation Policy



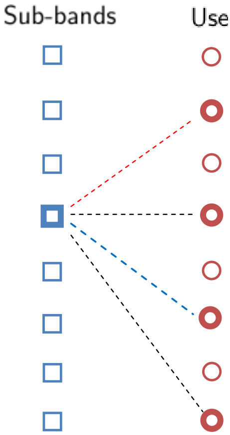
Throughput Optimal Resource Allocation Policy



Throughput Optimal Resource Allocation Policy



Throughput Optimal Resource Allocation Policy



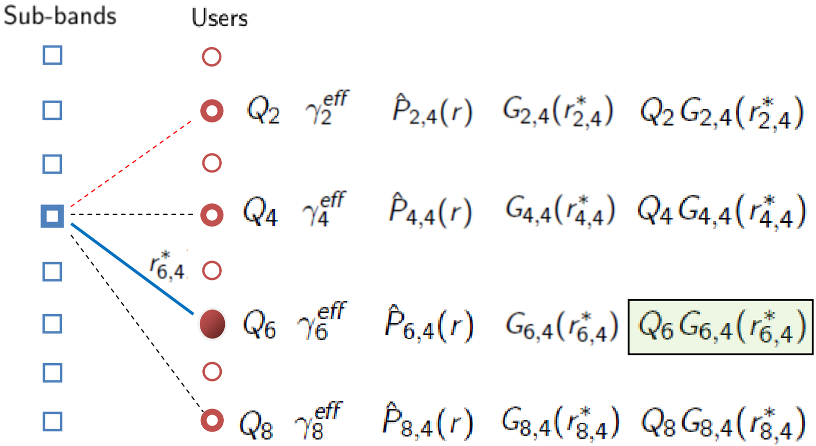
$Q_2 \quad \gamma_2^{eff} \quad \hat{P}_{2,4}(r) \quad G_{2,4}(r^*_{2,4}) \quad Q_2\, G_{2,4}(r^*_{2,4})$

$Q_4 \quad \gamma_4^{eff} \quad \hat{P}_{4,4}(r) \quad G_{4,4}(r^*_{4,4}) \quad Q_4\, G_{4,4}(r^*_{4,4})$

$Q_6 \quad \gamma_6^{eff} \quad \hat{P}_{6,4}(r) \quad G_{6,4}(r^*_{6,4}) \quad Q_6\, G_{6,4}(r^*_{6,4})$

$Q_8 \quad \gamma_8^{eff} \quad \hat{P}_{8,4}(r) \quad G_{8,4}(r^*_{8,4}) \quad Q_8\, G_{8,4}(r^*_{8,4})$

Throughput Optimal Resource Allocation Policy



Q_2 γ_2^{eff} $\hat{P}_{2,4}(r)$ $G_{2,4}(r^*_{2,4})$ $Q_2 G_{2,4}(r^*_{2,4})$

Q_4 γ_4^{eff} $\hat{P}_{4,4}(r)$ $G_{4,4}(r^*_{4,4})$ $Q_4 G_{4,4}(r^*_{4,4})$

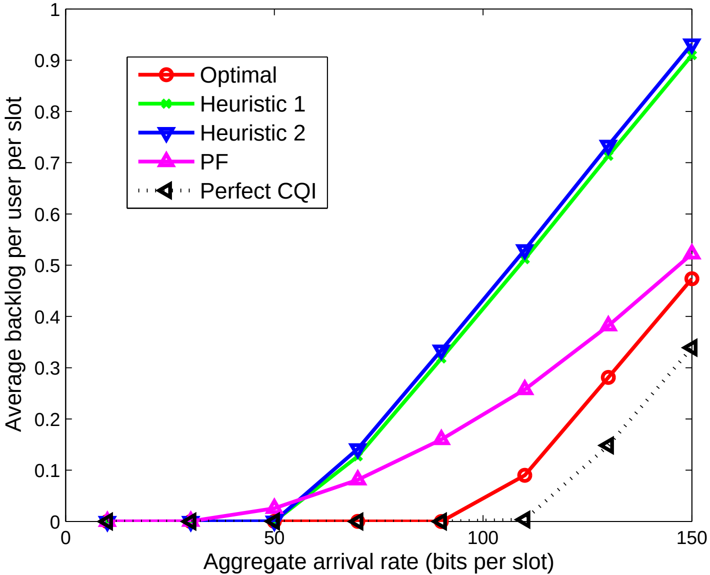
Q_6 γ_6^{eff} $\hat{P}_{6,4}(r)$ $G_{6,4}(r^*_{6,4})$ $\boxed{Q_6 G_{6,4}(r^*_{6,4})}$

Q_8 γ_8^{eff} $\hat{P}_{8,4}(r)$ $G_{8,4}(r^*_{8,4})$ $Q_8 G_{8,4}(r^*_{8,4})$

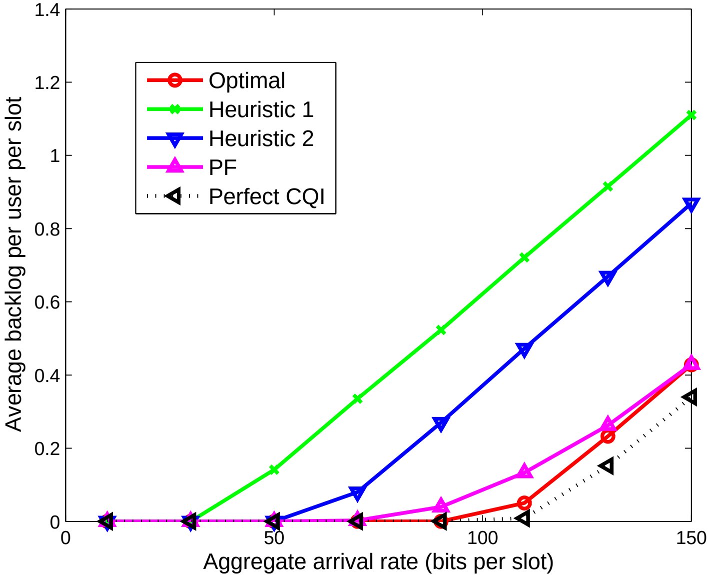
Simulation Results

- Validity of our limiting approximation in case of both i.i.d. and correlated sub-band SNRs
- Comparison of average backlog for various policies
  - Optimal - Max (Queue-length × Goodput)
  - Heuristic 1 - Max (Queue-length × EESM)
  - Heuristic 2 - Max (Queue-length × Estimate of CQI given EESM)
  - PF - Max (Goodput/average rate)
  - Perfect CQI - Max (Queue-length × CQI)

Comparison of various policies (M=3, i.i.d.)



Comparison of various policies (M=3, non i.i.d.)



Observations

- Policy naturally decouples for each sub-band (does not need solving any computationally intensive matching problems)
- Throughput optimality using Lyapunov stability framework
- Novel statistical model for EESM using a weak limit theorem
- Model for EESM valid for a larger class of sub-band SNR distribution (those which lie within the Gumbel domain of attraction)

Summary

Summary

- Adapting to the channel
- Adapting to the channel and traffic
  - Max-weight Scheduling
- Adapting to partial information
  - Conditional expected rate
  - Outage

Other Work

- Countering strategic behavior[22]
- Advanced physical layer options
    - Scheduling for cooperative base-stations[23]
    - Interference avoidance vs. Interference processing
- Distributed scheduling
    - Using local information only

http://www.ee.iitm.ac.in/∼skrishna/research.html

---

[22] A. K. Chorppath, S. Bhashyam, R. Sundaresan, "A convex optimization framework for almost budget balanced allocation of a divisible good," IEEE Transactions on Automation Science and Engineering, vol.8, no.3, pp.520-531, July 2011.

[23] M. R. Ramesh Kumar, S. Bhashyam, D. Jalihal, "Downlink Performance of 2-Cell Cooperation Schemes in a Multi-Cell Environment," Proceedings of WPMC 2008, Lapland, Finland, September 2008.

Acknowledgements

- Rajesh Sundaresan
- Krishna Jagannathan

- Chandrashekar Mohanram (MS)
- C. Manikandan (MS)
- Hussam Ahmed (MTech)
- Parimal Parag (DD-MTech)

- Department of Science and Technology