Minimizing Feedback Load for Nested Scheduling Algorithms

COST289 MCM Barcelona, October 28 and 29, 2004

Vegard Hassel, Mohamed-Slim Alouini, David Gesbert, Geir E. Øien

Outline

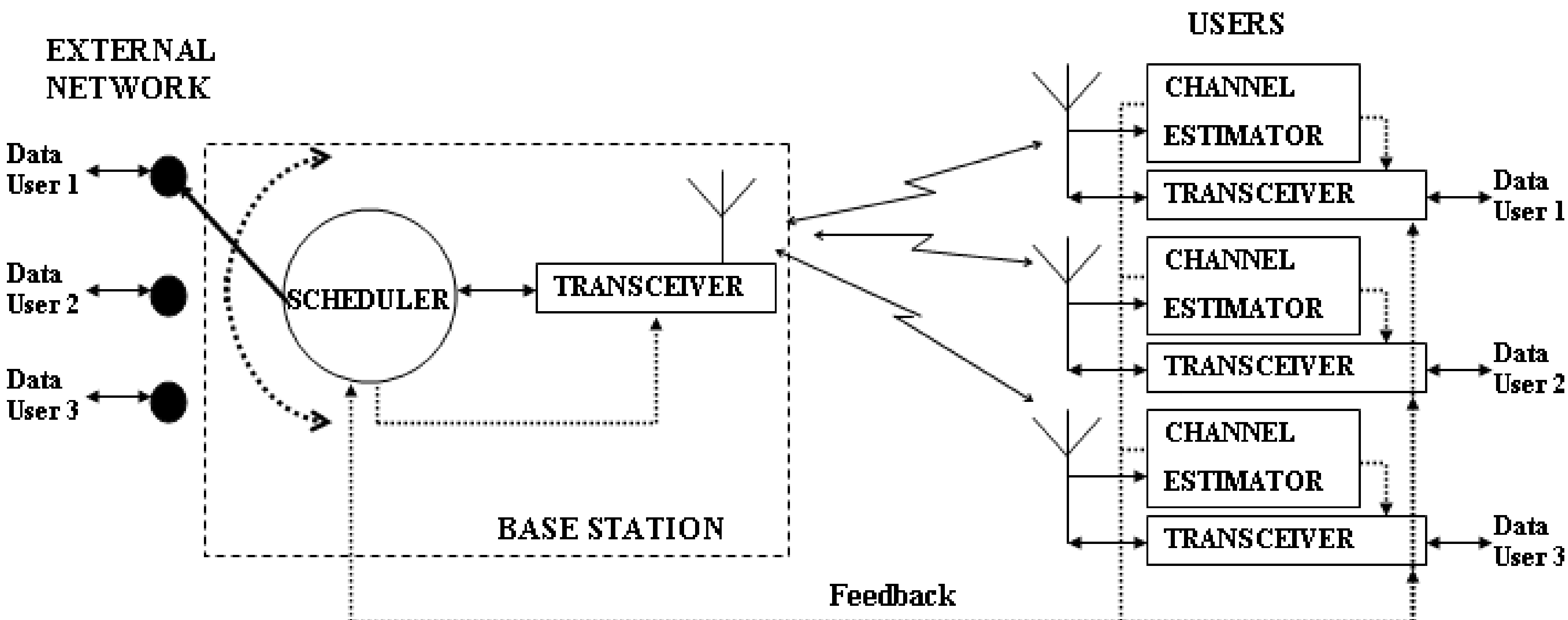
- What is Opportunistic Scheduling?
- Multiuser Diversity, MUD
- Calculating the MASSE
- Max CNR Scheduling, MCS
- Selective Multiuser Diversity, SMUD
- Optimal Rate, Reduced Feedback, ORRF
- Nested ORRF, NORRF
- Nested SMUD, NSMUD

What is Opportunistic Scheduling?

- The traditional scheduling algorithm has been *Round robin* (GSM)
  - The time-slots are assigned to the users in a sequential manner, independently of the channel conditions
  - Resource fair, but not necessary performance fair
- Opportunistic scheduling exploits the varying quality of the fading channel to increase the *Maximum Average System Spectral Efficiency* (MASSE)
- The future challenge will be to design algorithms that are both opportunistic and operate according to the QoS-demands from the applications
- Efficient adaptive modulation and coding is necessary to implement the scheduling algorithms

System Model

- Time Division Multiplexed (TDM) system with N mobile users
- Uplink and downlink at different frequencies
- Assumes that the carrier-to-noise-ratios (CNRs) of the users' channels are i.i.d.



Multiuser Diversity (MUD)

- *Diversity* in wireless systems arises because of independently fading channels
- Traditional forms of diversity: space, time, and frequency
- Viswanath and Tse: *Multiuser diversity*
- Background: With many users in a cell, there is high probability of finding a user with a good channel at any time
- MUD concept: To obtain the highest rate, the user with the best channel has to be chosen at all times
- Observe: While traditional forms of diversity gives better link spectral efficiency, multiuser diversity increases the *system* spectral efficiency

[1]

Calculating the MASSE

Calculating the Maximum Average System Spectral Efficiency (MASSE)

Optimum rate adaptation (no power adaptation):

$$\frac{C_{ora}}{W} = \int_0^\infty \log_2(1+\gamma) p_{\gamma^*}(\gamma)\, d\gamma$$

Optimum power and rate adaptation:

$$\frac{C_{opra}}{W} = \int_{\gamma_0}^\infty \log_2 \left(\frac{\gamma}{\gamma_0} \right) p_{\gamma^*}(\gamma)\, d\gamma,$$

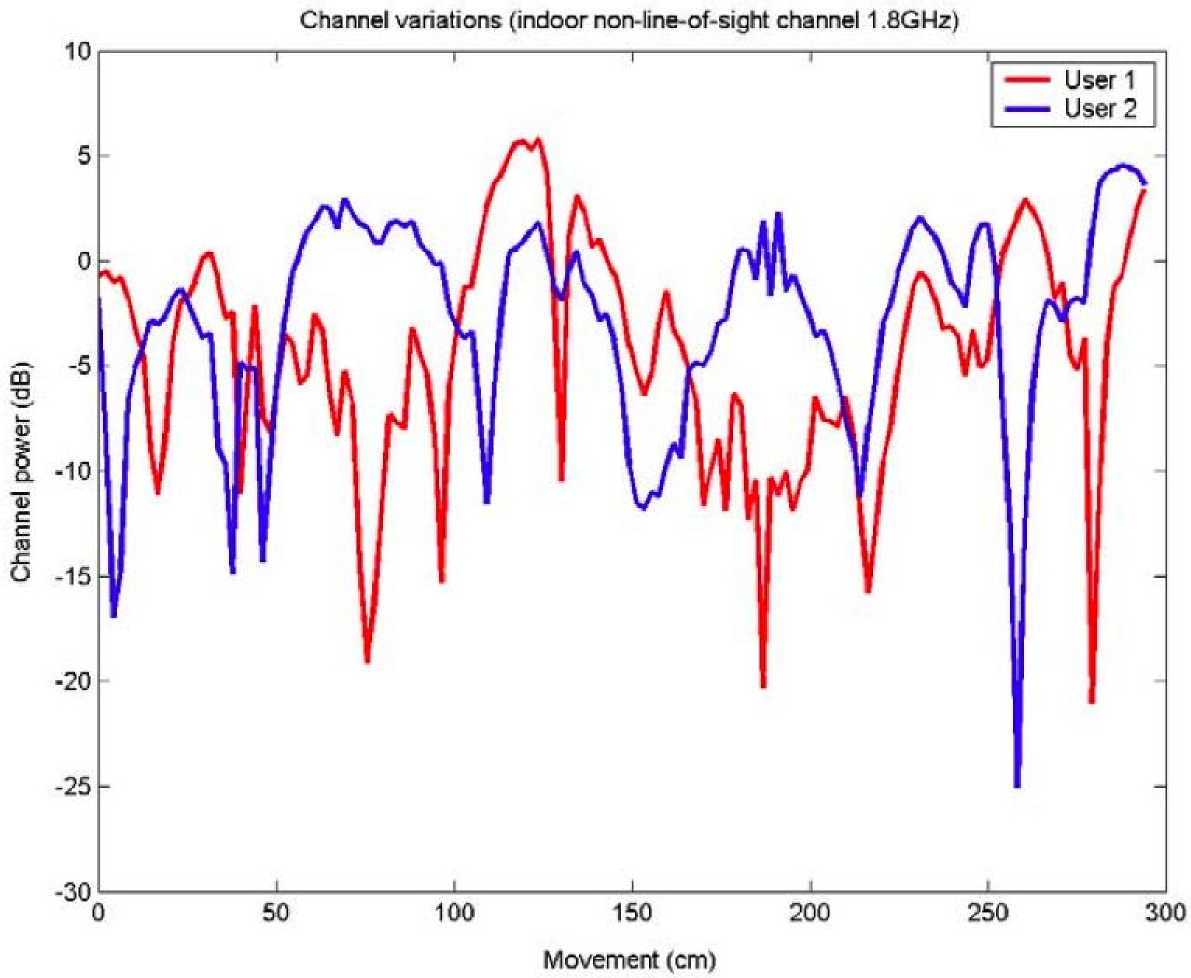
where γ_0 is found from the power constraint:

$$\int_{\gamma_0}^\infty \left(\frac{1}{\gamma_0} - \frac{1}{\gamma} \right) p_{\gamma^*}(\gamma)\, d\gamma \; = \; 1$$

[2]

Rate-Optimal TDM Scheduling - I

Ride the peaks!



[3, 4]

Rate-Optimal TDM Scheduling - II

The Max CNR scheduling (MCS) policy can be stated as:

$$i^*(t) = \operatorname*{argmax}_{1 \leq i \leq N} \gamma_i(t)$$

Equivalently:

$$i^*(t) = \operatorname*{argmax}_{1 \leq i \leq N} R_i(t)$$

- Also called the *greedy algorithm*
- Only fair if the users' CNRs are i.i.d. (Coherence time)
- Only optimal if the capacity degradation due to feedback is ignored

[3]

Rate-Optimal TDM Scheduling - III

The cumulative distribution of the best user is found by using *order statistics*:

$$P_{\gamma^*}(\gamma) = P_\gamma^N(\gamma),$$

The PDF of the best user is found by differentiating the CDF with respect to γ:

$$p_{\gamma^*}(\gamma) = N \cdot P_\gamma^{N-1}(\gamma) \cdot p_\gamma(\gamma),$$

For a Rayleigh channel we get the following PDF:

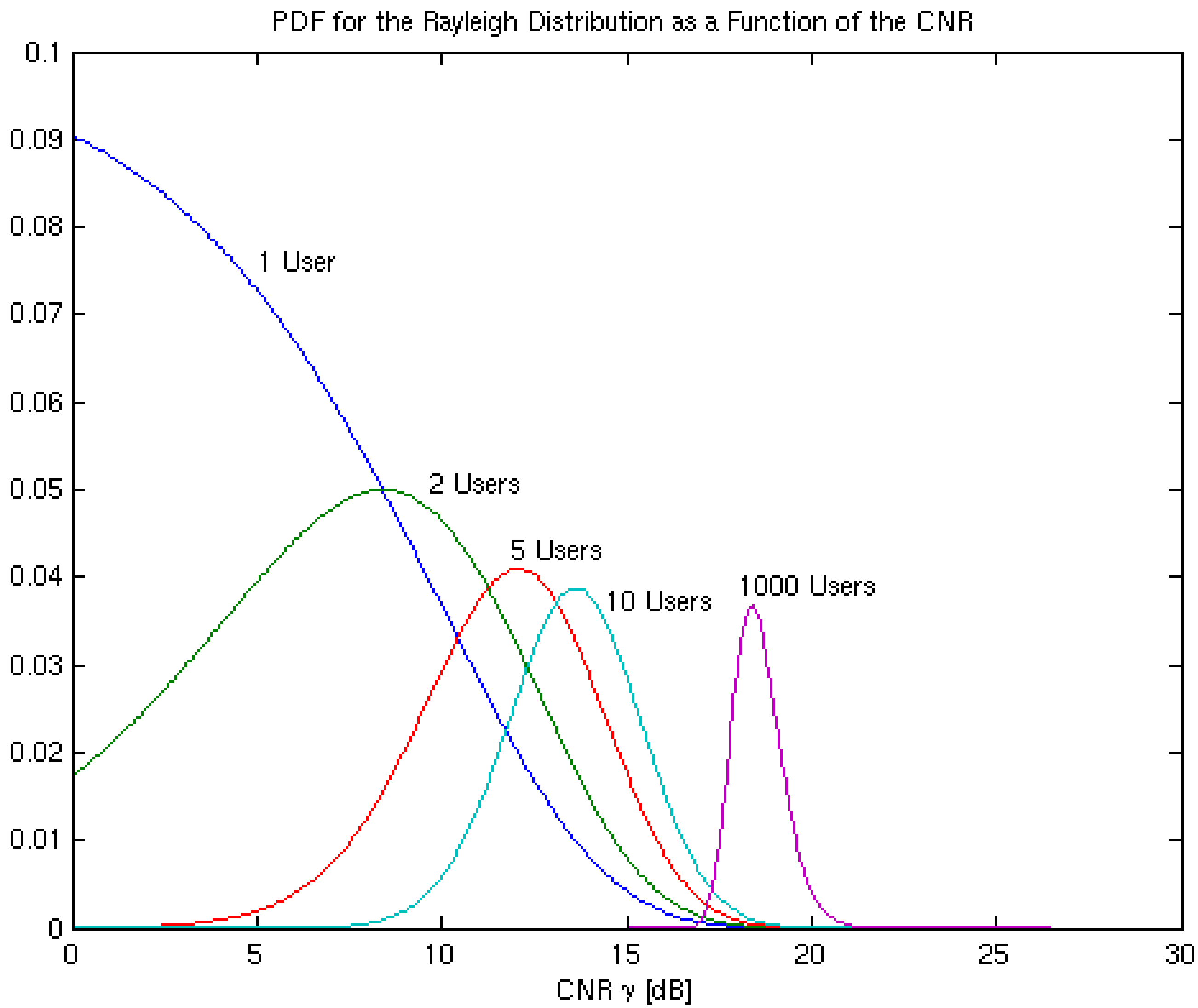
$$p_{\gamma^*}(\gamma) = N \cdot (1 - e^{-\gamma/\overline{\gamma}})^{N-1} \cdot \frac{e^{-\gamma/\overline{\gamma}}}{\overline{\gamma}}$$

To make integration easier, it is often preferable to use *binomial expansion*:

$$p_{\gamma^*}(\gamma) = \frac{N}{\overline{\gamma}} \sum_{n=0}^{N-1} \binom{N-1}{n} (-1)^n e^{-(1+n)\gamma/\overline{\gamma}}$$

[3]

Rate-Optimal TDM Scheduling - IV



[3]

Analyzing Feedback

- The MCS algorithm assumes that each user feeds back its CNR for every time-slot
- We will look at how to reduce the feedback load of the MCS algorithm
- *Normalized Feedback load* (NFL) is defined as the average ratio of users that give feedback for every time-slot
- Reduced feedback from the mobile users will reduce their power consumption and will contribute to a higher system spectral efficiency

SMUD: Selective Multiuser Diversity - I

- Algorithm to reduce the feedback load
- The scheduler asks for the instantaneous CNR level only from the users that have CNR above a threshold, γ_{th}
- If none of the users feed back their CNR, a random user is chosen
- The algorithm is not rate-optimal, but gives a significant reduction in the feedback load

$$i^*(t) = \begin{cases} \text{rand}(i), & \text{if all } \gamma_i(t) \leq \gamma_{th} \\ \underset{1 \leq i \leq N}{\text{argmax}}\, \gamma_i(t), & \text{if it exists a } \gamma_i(t) > \gamma_{th} \end{cases}$$

[5]

SMUD: Selective Multiuser Diversity - II

CDF:

$$P_{\gamma^*}(\gamma) = \begin{cases} P_\gamma^{N-1}(\gamma_{th}) \cdot P_\gamma(\gamma), & \gamma \leq \gamma_{th} \\ P_\gamma^N(\gamma), & \gamma > \gamma_{th} \end{cases}$$

PDF can be found by differentiating the CDF with respect to γ:

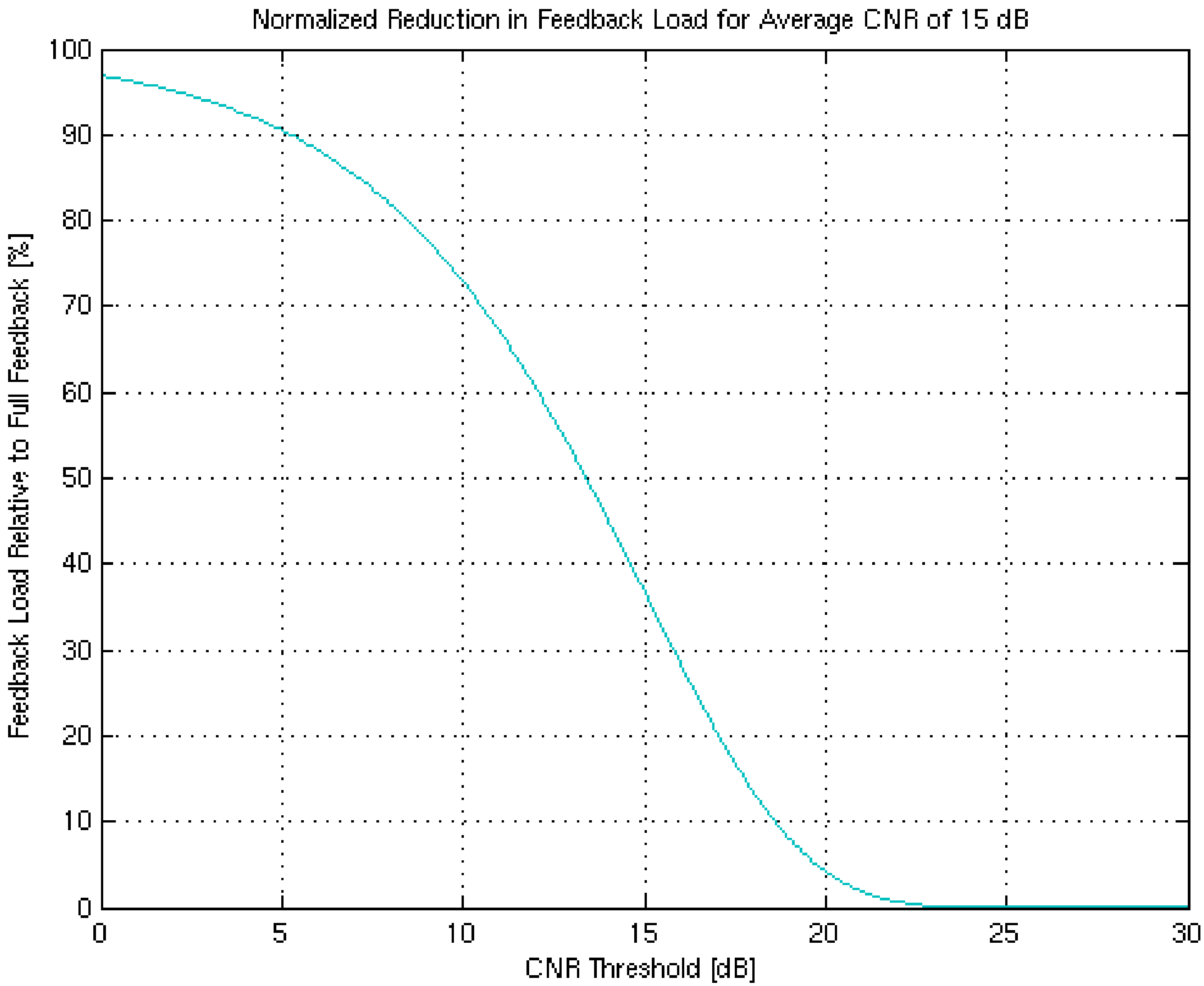
$$p_{\gamma^*}(\gamma) = \begin{cases} P_\gamma^{N-1}(\gamma_{th}) \cdot p_\gamma(\gamma), & \gamma \leq \gamma_{th} \\ N \cdot P_\gamma^{N-1}(\gamma) \cdot p_\gamma(\gamma), & \gamma > \gamma_{th} \end{cases}$$

The NFL for the SMUD algorithm can be shown to be given by:

$$\bar{F} = 1 - P_\gamma(\gamma_{th})$$

[5]

SMUD: Selective Multiuser Diversity - III



[5]

ORRF: Optimal-Rate, Reduced Feedback - I

- Based on the SMUD algorithm

- If all users fail to meet the CNR threshold value, the base station requests full feedback

- Obtains optimal rate, but has a significant reduction in feedback compared to the MCS algorithm

[5, 6]

ORRF: Optimal-Rate, Reduced Feedback - II

For the ORRF algorithm, it can be shown that the *normalized average feedback load* can be expressed as:

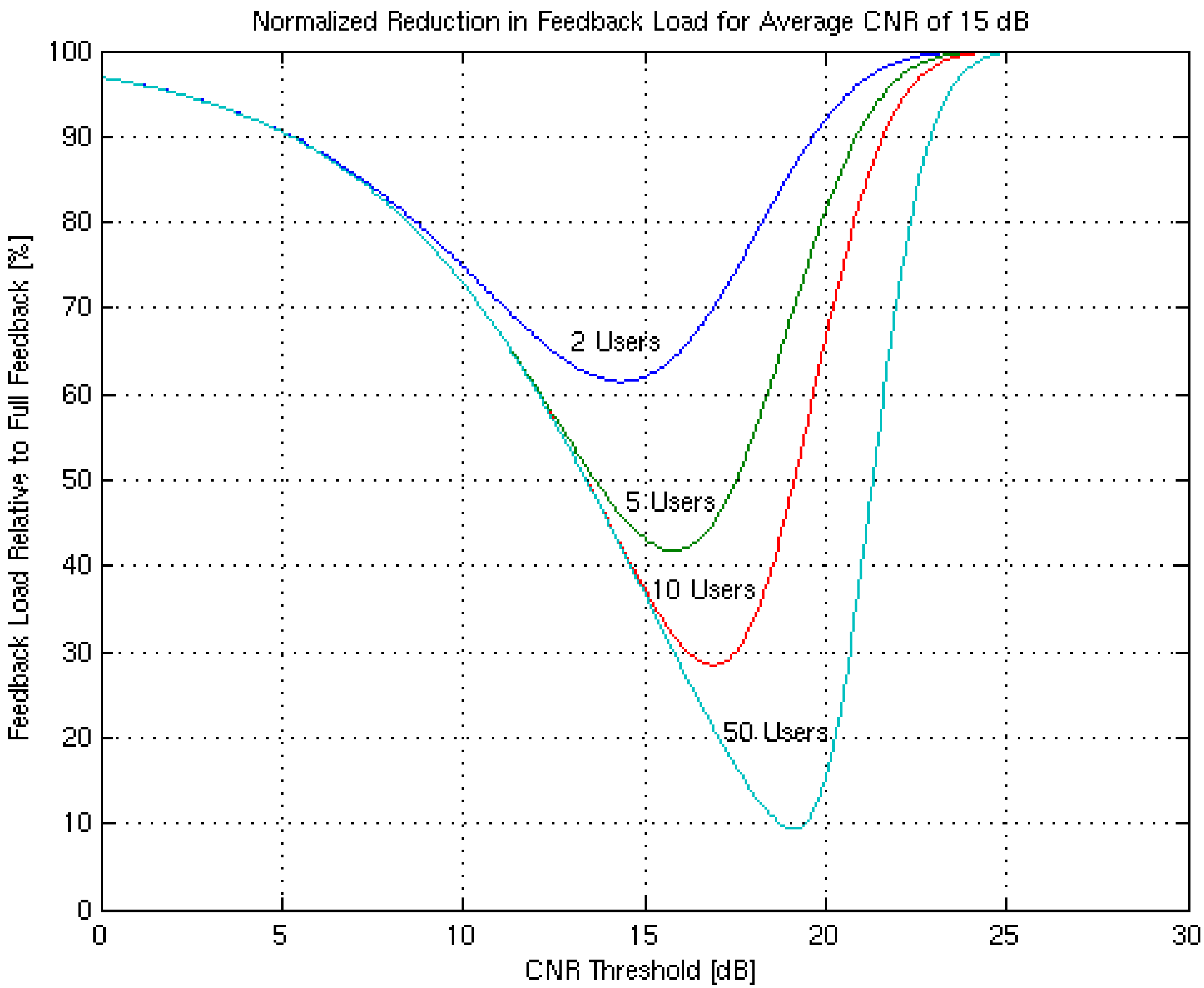
$$\bar{F} = 1 - P_\gamma(\gamma_{th}) + P_\gamma^N(\gamma_{th}), \;\; N = 2, 3, 4, \cdots$$

Differentiating this expression with regard to γ_{th} and setting the expression equal to zero gives the optimal threshold value:

$$\gamma_{th}^* = -\overline{\gamma} \ln(1 - (1/N)^{\frac{1}{N-1}}), \;\; N = 2, 3, 4, \cdots$$

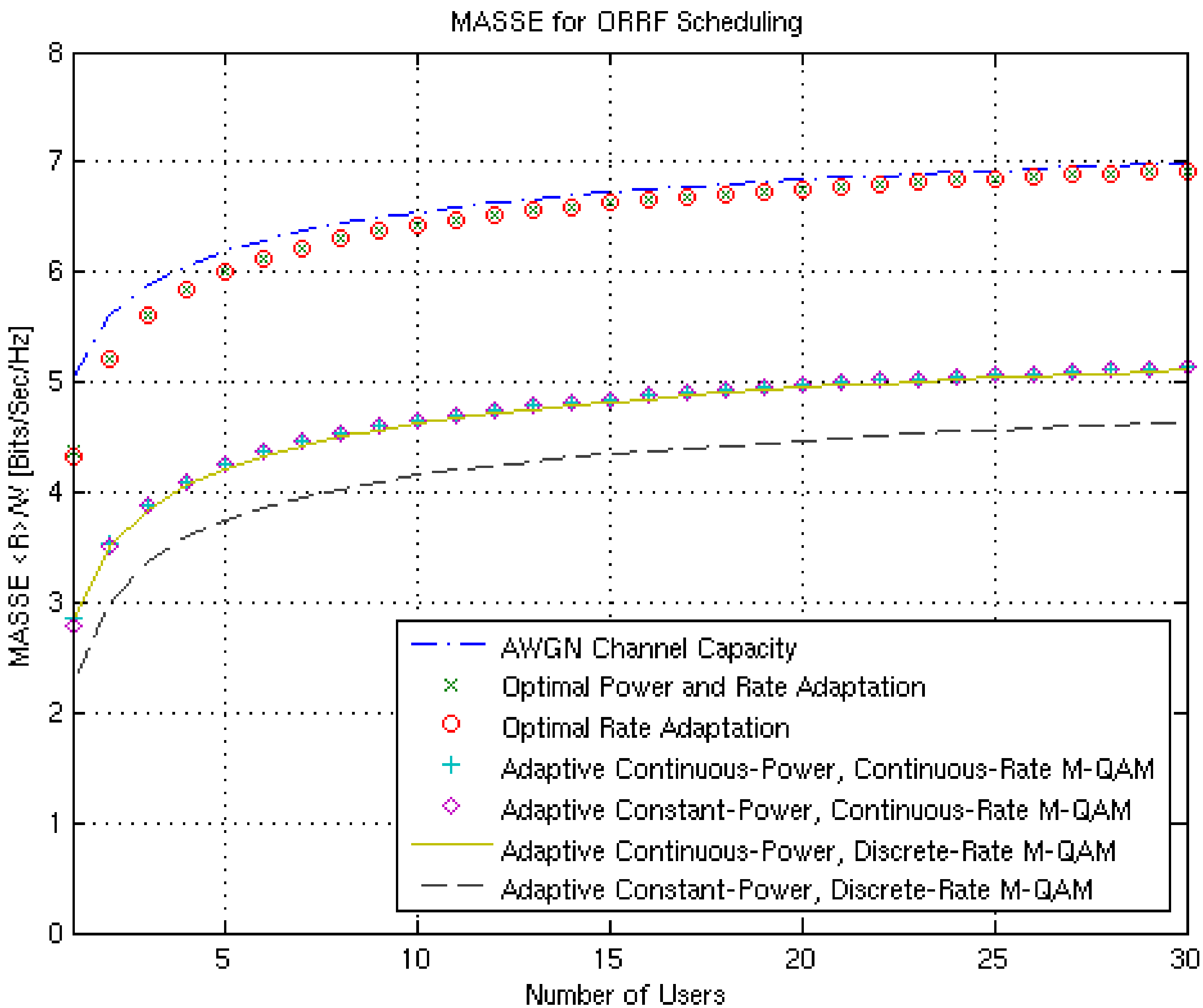
[6]

ORRF: Optimal-Rate, Reduced Feedback - III



[6]

ORRF: Optimal-Rate, Reduced Feedback - IV



[6]

ORRF: Optimal-Rate, Reduced Feedback - V

Delays can be analyzed using two different scenarios:

- Scenario 1: *Scheduling delay*:
  - Arises because the scheduler received channel estimates, takes a scheduling decision and notifies the selected user
  - The selected user is assumed to transmit using a constellation size based on a perfect channel estimate without delay
  - When the selected user is transmitting, it doesn't necessarily have to be the best anymore. Therefore, we will experience a degradation in MASSE compared to the optimal rate

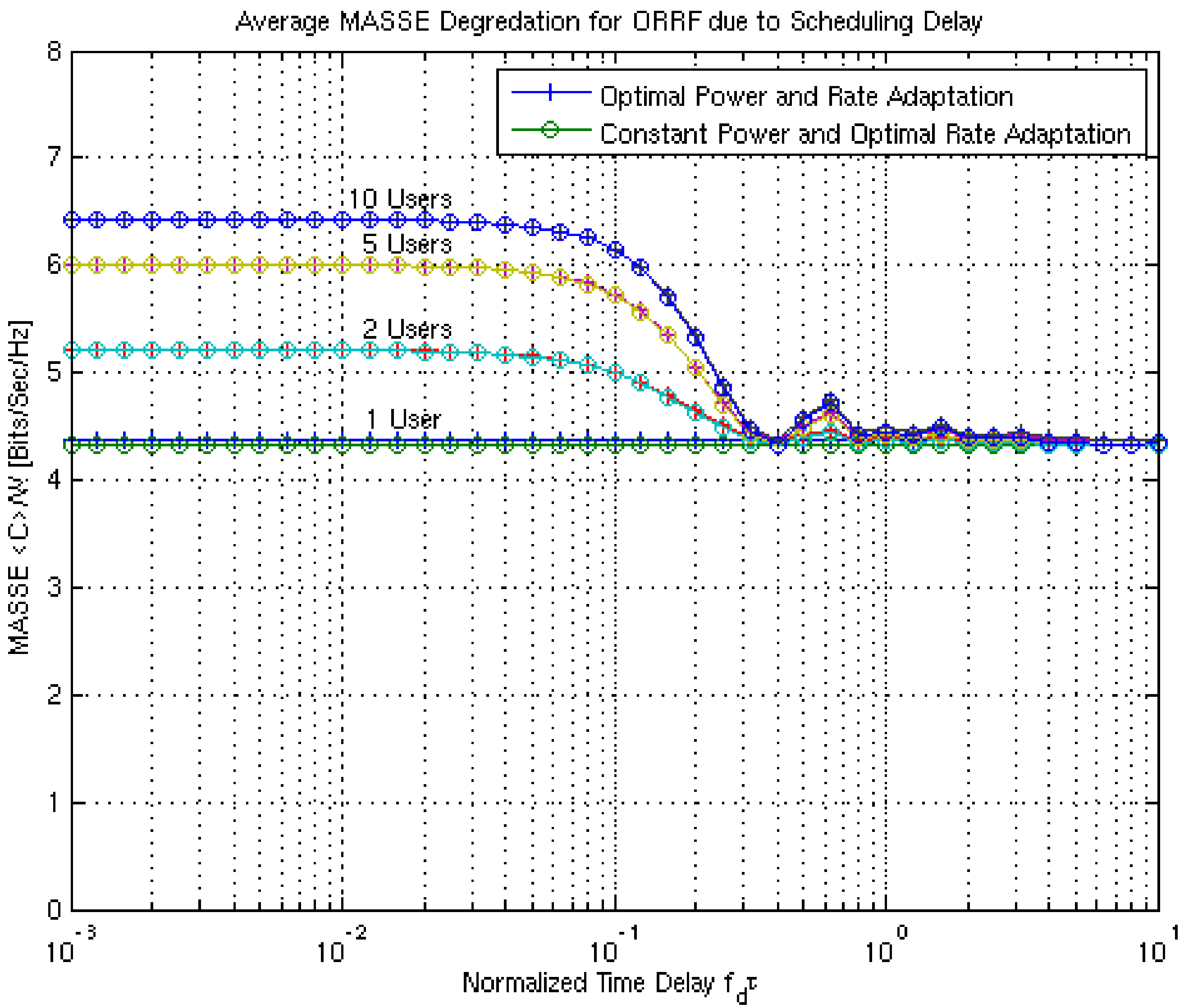
[6]

ORRF: Optimal-Rate, Reduced Feedback - VI

- Scenario 2: *Outdated channel estimates*:
  - Leads to both a scheduling delay and suboptimal modulation constellations with increased BER
  - Both the scheduling decision and the decision of the modulation constellation is based on an outdated channel estimate
  - The selected user experiences a CNR degradation, but does not adjust its modulation constellation accordingly (as for the previous scenario)
  - The rate is not lowered, but the BER will increase with the degree of outdatedness

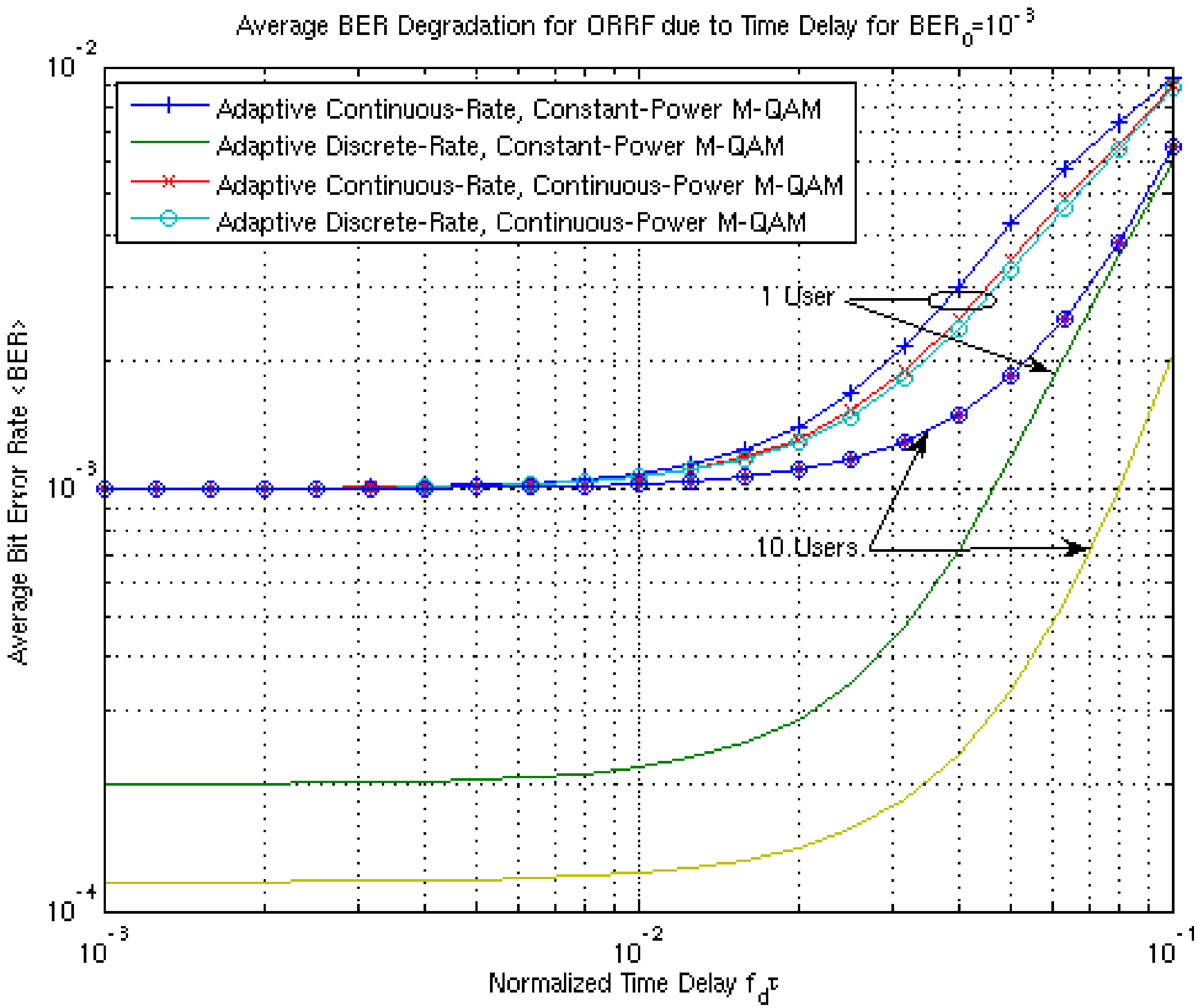
[6]

ORRF: Optimal-Rate, Reduced Feedback - VII



[6]

ORRF: Optimal-Rate, Reduced Feedback - VIII



[6]

NORRF: Nested ORRF - I

- Employs multiple (nested) feedback thresholds
- Denoting the thresholds by $\gamma_{th,L} > \gamma_{th,L-1} > \cdots > \gamma_{th,0}$
(For convenience $\gamma_{th,L} = \infty$ and $\gamma_{th,0} = 0$)
- The base station initially requests feedback from those users whose CNR is above $\gamma_{th,L-1}$. If there are none, the threshold is successively lowered to $\gamma_{th,L-2}, \gamma_{th,L-3}, \cdots, \gamma_{th,0}$
- The best user is always selected, but the average feedback load is significantly reduced compared to the MCS algorithm
- A large number of thresholds will require a long guard interval and will therefore contribute to reducing the MASSE

[7]

NORRF: Nested ORRF - II

For the NORRF algorithm the normalized feedback load (NFL) can be expressed as:

$$\bar{F}_{\text{NORRF}} = \frac{1}{N} \sum_{l=0}^{L-1} \sum_{n=1}^{N} n \binom{N}{n} \left(P_\gamma(\gamma_{th,l+1}) - P_\gamma(\gamma_{th,l})\right)^n \cdot P_\gamma^{N-n}(\gamma_{th,l})$$

Using the binomial expansion formula, it can be shown that this expression can be written as:

$$\bar{F}_{\text{NORRF}} = \sum_{l=0}^{L-1} \left(P_\gamma(\gamma_{th,l+1}) - P_\gamma(\gamma_{th,l})\right) \cdot P_\gamma^{N-1}(\gamma_{th,l+1})$$

[7]

NORRF: Nested ORRF - III

Taking the gradient of this expression with regard to the thresholds $\gamma_{th,l}$ and setting the result equal to zero yields:

$$\gamma^*_{th,l} = P_\gamma^{-1}\left(S_l \cdot P_\gamma(\gamma_{th,l+1})\right), \;\; l = 1, 2, 3, \cdots, L-1$$

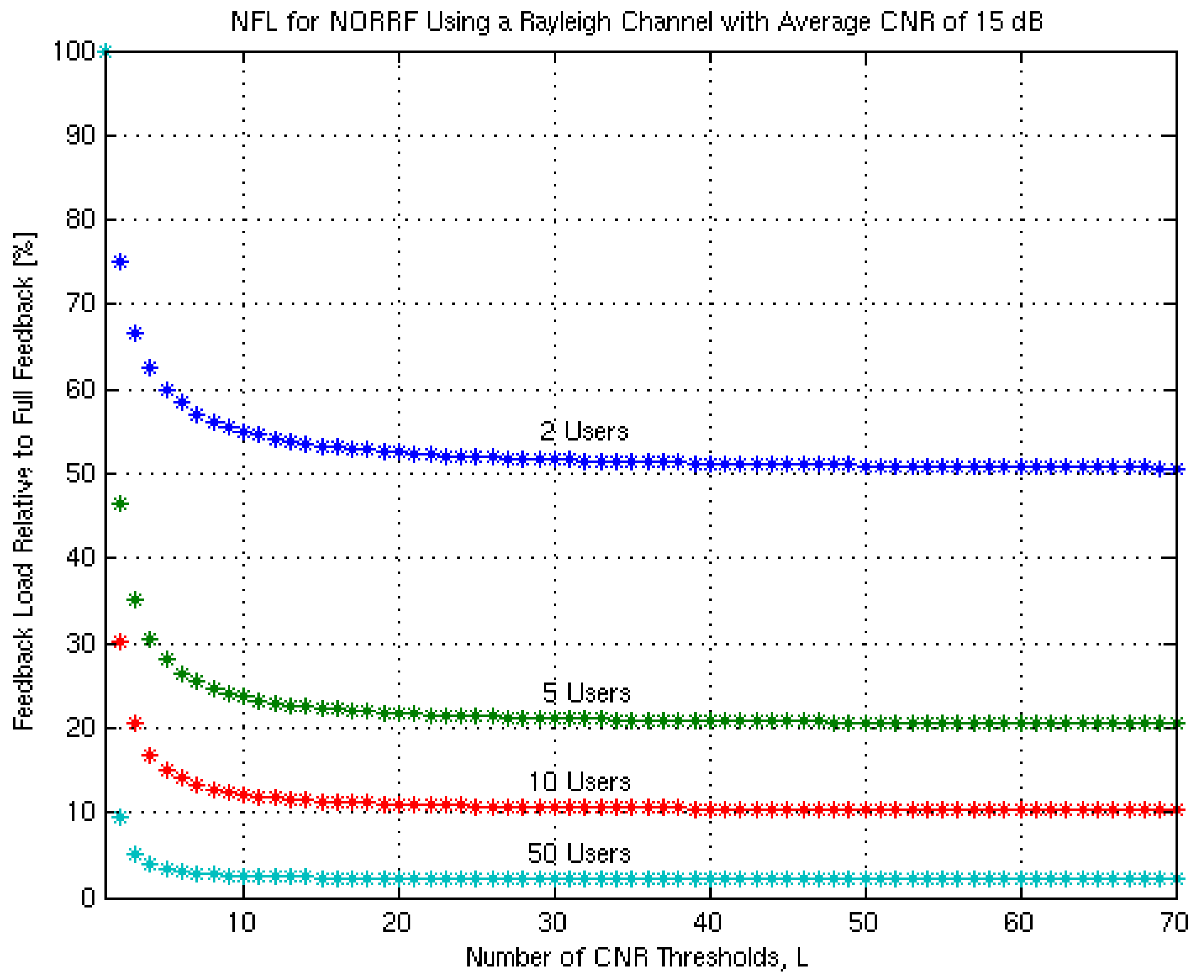
$P_\gamma^{-1}(\cdot)$ is the inverse CDF of the CNR for a single user
The expression for S_l is:

$$S_l = \begin{cases} N^{\frac{1}{1-N}}, & l = 1 \\ \left[N - (N-1)S_{l-1}\right]^{\frac{1}{1-N}}, & l = 2, 3, \cdots, L-1 \end{cases}$$

where $N \geq 2$.

[7]

NORRF: Nested ORRF - IV



[7]

NSMUD: Nested SMUD - I

For the NORRF algorithm the feedback load can be expressed as:

$$\bar{F}_{\text{NSMUD}} = \sum_{l=1}^{L-1} \left(P_\gamma(\gamma_{th,l+1}) - P_\gamma(\gamma_{th,l})\right) \cdot P_\gamma^{N-1}(\gamma_{th,l+1})$$

Taking the gradient of this expression and setting the result equal to zero gives the global solution $\gamma_{th,1} = \infty$. This however means that the NFL will be zero and no multiuser diversity gain will be experienced.

[7]

NSMUD: Nested SMUD - II

- Setting $\gamma_{th,1} = 0$ always gives the base station enough feedback information to choose the best user

- With $\gamma_{th,1} = 0$ the minimization of the NFL gives the following solution:

$$\gamma^*_{th,l} = P_\gamma^{-1}\left(S_l \cdot P_\gamma(\gamma_{th,l+1})\right), \;\; l = 2, 3, \cdots, L{-}1$$

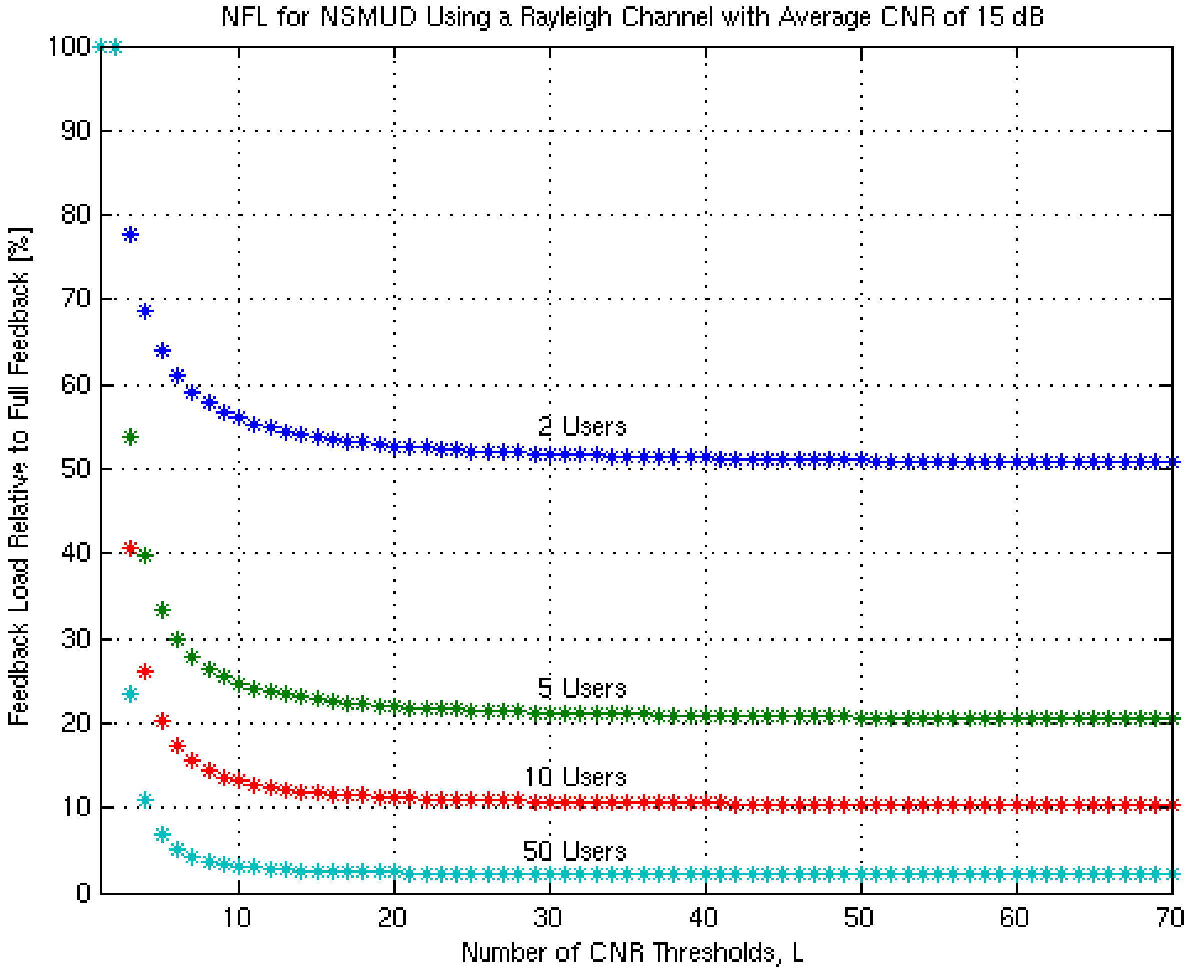
The expression for S_l is:

$$S_l = \left[N - (N-1)S_{l-1}\right]^{\frac{1}{1-N}}, \;\; l = 2, 3, \cdots, L{-}1$$

where $N \geq 2$.

[7]

NSMUD: Nested SMUD - III



[7]

Conclusion

- The ORRF algorithm obtains maximal spectral efficiency and a significant reduction in the feedback load compared to full feedback
- Analysis of delay effects shows that the performance degrades significantly when the delay exceeds certain values
- Using the ORRF and the SMUD algorithms in a nested fashion minimizes the feedback load when the number of thresholds grow large

References

[1] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.

[2] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. IT-43, pp. 1896–1992, Nov. 1997.

[3] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," in *IEEE Int. Conf. on Communications (ICC'95)*, (Seattle, WA), pp. 331–335, June 1995.

[4] M. Johansson, "Issues in multiuser diversity." `http://www.signal.uu.se/Research/PCCWIP/Visbyrefs/Johansson_Visby04.pdf`. Presentation at WIP/BEATS/CUBAN workshop Wisby, Sweden, Aug. 2004.

[5] D. Gesbert and M.-S. Alouini, "How much feedback is multi-user diversity really worth?," in *IEEE Int. Conf. on Communications (ICC'04)*, (Paris, France), pp. 234–238, June 2004.

[6] V. Hassel, M.-S. Alouini, G. E. Øien, and D. Gesbert, "Rate-optimal multiuser scheduling with reduced feedback load and analysis of delay effects." Submitted to *IEEE Int. Conf. on Comm. (ICC'05)*, (Seoul, South Korea), May 2005.

[7] V. Hassel, M.-S. Alouini, D. Gesbert, and G. E. Øien, "Minimizing feedback load for nested scheduling algorithms." Submitted to *IEEE Veh. Tech. Conf. (VTC'05-spring)*, (Stockholm, Sweden), May 2005.