RATE-OPTIMAL MULTIUSER SCHEDULING WITH REDUCED FEEDBACK LOAD AND ANALYSIS OF DELAY EFFECTS

Vegard Hassel[*], *Mohamed-Slim Alouini*[†], *Geir E. Øien*[*], *David Gesbert*[‡]

[*]Norwegian University of Science and Technology, Dept. of Electronics and Telecommunications, Trondheim, Norway, emails: {vegard, oien@iet.ntnu.no}
[†]University of Minnesota, Dept. of Electrical and Computer Eng., Minneapolis, USA, email: {alouini@ece.umn.edu}
[‡]Eurecom Institute, Sophia-Antipolis, France, email: {gesbert@eurecom.fr}

ABSTRACT

In this paper we propose a multiuser scheduling algorithm that has the maximum average system spectral efficiency, but obtains a significant reduction in feedback load compared to full feedback by using a feedback threshold. An expression for the threshold value that minimizes the feedback load is found. Novel closed-form expressions are also found for the system spectral efficiency when using M-ary quadrature amplitude modulation. Finally, we analyze the impact of scheduling delay and outdated channel estimates.

1. INTRODUCTION

The ever-increasing demand for new applications in wireless communication systems makes efficient transmission scheduling between users a priority. The scheduling algorithm that maximizes the average system spectral efficiency among all time division multiplexing (TDM) based algorithms, is the one where the user with the highest carrier-to-noise ratio (CNR) is served at all times [6]. Here, we refer to this algorithm as *Max CNR scheduling* (MCS). One drawback of this rate-optimal policy is that the scheduler has to have full feedback from all users for every time-slot. To reduce the feedback load, the *Selective multiuser diversity* (SMUD) algorithm was introduced [3]. In that scenario, only the users that have a CNR above a CNR threshold should send feedback to the scheduler. If the scheduler does not receive feedback, a random user is chosen. Consequently, the SMUD algorithm introduces a reduction in capacity and it is not possible to set an optimal threshold value without deciding an outage probability. The algorithm proposed in this paper also employs a feedback threshold. However, if none of the users succeed to exceed the CNR threshold, the scheduler requests full feedback, and selects the user with the highest CNR. Consequently, the best user is always selected, but the feedback load is significantly reduced compared to the MCS algorithm. We will refer to this scheduling technique as the *Optimal rate, reduced feedback* (ORRF) algorithm.

2. SYSTEM MODEL

We consider a single base station that serves N users using TDM. Before performing scheduling, the base station is assumed to receive perfect information about the users' CNRs. It is assumed that the channels of all users are i. i. d. slowly-varying, flat Rayleigh fading channels with average received CNR $\overline{\gamma}$.

3. ANALYSIS OF THE FEEDBACK LOAD

For the ORRF scheme, the probability of full feedback is given by inserting $\gamma = \gamma_{th}$ into:

$$P_{\gamma^*}(\gamma) = P_\gamma^N(\gamma), \tag{1}$$

where $P_\gamma(\gamma)$ is the cumulative distribution function (CDF) of the CNR for a single user. Differentiating (1), inserting expressions for the Rayleigh distribution and using binomial expansion we obtain the probability density function (PDF) for the user with the highest CNR:

$$p_{\gamma^*}(\gamma) = \frac{N}{\overline{\gamma}} \sum_{n=0}^{N-1} \binom{N-1}{n} (-1)^n e^{-(1+n)\gamma/\overline{\gamma}}. \tag{2}$$

By using (1), the feedback load can be expressed as a weighted sum of full feedback load and feedback for the SMUD algorithm. If the load of full feedback is set to unity, it can be shown that the normalized average feedback load is given by:

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

A plot of the feedback load as a function of γ_{th} is shown in Fig. 1 for $\overline{\gamma}$= 15 dB.

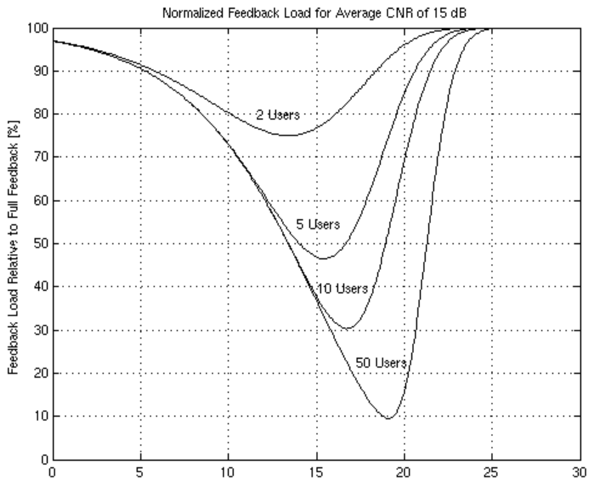

Figure 1: Normalized feedback load as a function of γ_{th} with $\overline{\gamma}$=15 dB.

The expression for the threshold value that minimizes the average feedback load can be found by differentiating (3) with respect to γ_{th} and setting the result equal to zero. For a Rayleigh fading channel, with CDF $P_\gamma(\gamma) = 1 - e^{-\gamma/\overline{\gamma}}$, the optimum threshold is found to be:

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

This work was in part supported by the Center for Transportation Studies (CTS), Minneapolis, USA.

4. CONSTANT-POWER, VARIABLE-RATE M-QAM SPECTRAL EFFICIENCIES

From [1] we know that the link spectral efficiency for continuous-rate M-QAM can be approximated by

$$\log_2(M) \approx \log_2\left(1 + \frac{3\,\gamma}{2\,K_0}\right), \qquad (5)$$

where M is the constellation size and $K_0 = -\ln(5\,\mathrm{BER}_0)$. Taking the expectation of the expression in (5), using integration by parts, L'Hôspital's rule, and [5, (3.352.2)], we obtain the following *maximum average system spectral efficiency* (MASSE):

$$\frac{< R >_{acr}}{W} = \frac{N}{\ln 2}\sum_{n=0}^{N-1}\binom{N-1}{n}\frac{(-1)^n}{1+n}e^{\frac{2K_0(1+n)}{3\overline{\gamma}}}\,E_1\left(\frac{2K_0(1+n)}{3\overline{\gamma}}\right), \qquad (6)$$

where *acr* denotes *adaptive continuous rate* M-QAM. The MASSE [Bit/Sec/Hz] is defined as the maximum average sum of spectral efficiency within a cell, shared between all users' up-links and down-links.

For physical systems we use *adaptive discrete rate* (ADR) M-QAM where the CNR range is divided into $K+1$ *fading regions*, with constellation size $M_k = 2^k$ assigned to the kth fading region. The MASSE is now given by [1]:

$$\frac{< R >_{adr}}{W} = \sum_{k=1}^{K} k\,p_k, \qquad (7)$$

where p_k is the probability that the CNR falls into region k. For a Rayleigh channel, this probability is given by

$$p_k = \left(1 - e^{-\gamma_{k+1}/\overline{\gamma}}\right)^N - \left(1 - e^{-\gamma_k/\overline{\gamma}}\right)^N. \qquad (8)$$

5. CONTINUOUS-POWER, VARIABLE-RATE M-QAM SPECTRAL EFFICIENCIES

Inserting (2) into [4, (25)], the following closed-form expression for the MASSE using adaptive continuous power and rate M-QAM is obtained:

$$\frac{< R >_{acr,pa}}{W} = \frac{N}{\ln 2}\sum_{n=0}^{N-1}\binom{N-1}{n}\frac{(-1)^n}{1+n}E_1\left(\frac{(1+n)\gamma_K}{\overline{\gamma}}\right), \qquad (9)$$

where $\gamma_K = \gamma_0/K_1 = -\gamma_0(2/3)\ln(5\mathrm{BER}_0)$ is the optimal cut-off CNR level below which data transmission is suspended. Correspondingly, inserting (2) into [4, (22)], the closed-form expression for the power constraint yields:

$$\sum_{n=0}^{N-1}\binom{N-1}{n}(-1)^n\left[\frac{e^{-(1+n)\gamma_K/\overline{\gamma}}}{(1+n)\gamma_0/\overline{\gamma}} - \frac{1}{K_1}E_1\left(\frac{(1+n)\gamma_K}{\overline{\gamma}}\right)\right] = \frac{\overline{\gamma}}{N}. \qquad (10)$$

For the discrete rate scenario, the CNR range is also here divided into $K+1$ bins. However now the system transmits the constellation size M_k when $\gamma_0^* M_{k+1} \geq \gamma \geq \gamma_0^* M_k$, where γ_0^* is found by optimizing the MASSE with regard to the power constraint in [4, (32)]. The expression for the MASSE in the rate-discrete case, is given by (7), and the following expression for p_k is obtained:

$$p_k = \left(1 - e^{-\gamma_0^* M_{k+1}/\overline{\gamma}}\right)^N - \left(1 - e^{-\gamma_0^* M_k/\overline{\gamma}}\right)^N. \qquad (11)$$

Inserting (2) into [4, (32)], the following closed-form expression for the power constraint for ADR M-QAM is obtained:

$$\sum_{k=1}^{K}(M_k - 1)\sum_{n=0}^{N-1}\binom{N-1}{n}(-1)^n \times\left[E_1\left(\frac{(1+n)\gamma_0^* M_k}{\overline{\gamma}}\right) - E_1\left(\frac{(1+n)\gamma_0^* M_{k+1}}{\overline{\gamma}}\right)\right] = \frac{\overline{\gamma}K_1}{N}. \qquad (12)$$

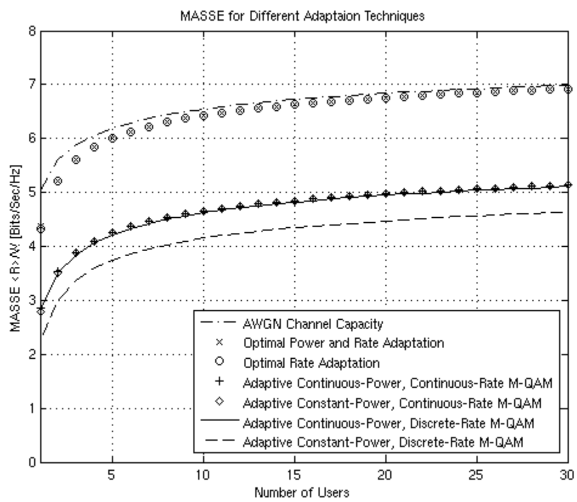

Figure 2: Maximum average system spectral efficiency for ORRF scheduling with $\overline{\gamma}$=15 dB, 7 fading regions and $\mathrm{BER}_0 = 10^{-3}$.

Fig. 2 shows how the MASSE varies with the number of users for different power and rate adaptation policies. We see that for discrete-rate M-QAM, the gain due to power adaptation of about 0.5 Bit/Sec/Hz is independent of the number of users.

6. M-QAM BIT-ERROR-RATES

The BER of coherent M-QAM with two-dimensional Gray coding over an additive white Gaussian noise (AWGN) channel can be approximated by [4]:

$$\mathrm{BER}(M, \gamma) \approx 0.2\exp\left(-\frac{3\gamma}{2(M-1)}\right). \qquad (13)$$

This equation is the inverse of (5). Consequently, it can be easily shown that the constant-power ACR M-QAM scheme always operates at the target BER. Because the discrete assignment of constellation sizes in ADR M-QAM, this scheme has to operate at a BER lower than the target. The average BER for ADR M-QAM using constant power can be calculated as [1]:

$$< \mathrm{BER} >_{adr} = \frac{\sum_{k=1}^{K} k\,\overline{\mathrm{BER}}_k}{\sum_{k=1}^{K} k\,p_k}, \qquad (14)$$

where

$$\overline{\mathrm{BER}}_k = \int_{\gamma_k}^{\gamma_{k+1}} \mathrm{BER}(M_k, \gamma)p_{\gamma^*}(\gamma)d\gamma. \qquad (15)$$

Inserting (13) into (15) we obtain the following expression for the average BER within a fading region:

$$\overline{\mathrm{BER}}_k = \frac{0.2N}{\overline{\gamma}} \sum_{n=0}^{N-1} \binom{N-1}{n} (-1)^n \frac{e^{-\gamma_k a_{k,n}} - e^{-\gamma_{k+1} a_{k,n}}}{a_{k,n}}, \tag{16}$$

where $a_{k,n}$ is given by:

$$a_{k,n} = \frac{1+n}{\overline{\gamma}} + \frac{3}{2(M_k - 1)}. \tag{17}$$

By applying the power adaptation policy given by [4, (22)] it can be shown that the M-QAM schemes using continuous power adaptation always operates at the target BER.

7. CONSEQUENCES OF DELAY

In the previous sections, it has been assumed that there is no delay from the instant where the channel estimates are obtained and fed back to the scheduler, to the time when the optimal user is transmitting. For real-life systems, we have to take delay into consideration. We have analyzed two delay scenarios. In the first scenario a *scheduling delay* arises because the scheduler receives channel estimates, takes a scheduling decision and notifies the selected user, who does not necessarily have to be the best user anymore. The second scenario deals with *outdated channel estimates*, which leads to both a scheduling delay and suboptimal modulation constellations with increased BERs.

7.1 Impact of Scheduling Delay

In this subsection we will assume that the scheduling decision is based on a channel estimate at time t, whereas the data are sent over the channel at time $t+\tau$. We will assume that the transmitter uses a perfect channel estimate available at time $t+\tau$, to determine the transmission rate.

To investigate the influence of scheduling delay, we want to develop a PDF for the CNR at time $t+\tau$. Let α and α_τ be the channel gains at time t and $t+\tau$, respectively. Assuming that the average power gain remains constant over the time delay τ for a slowly-varying Rayleigh channel, (i.e. $\Omega = E[\alpha^2] = E[\alpha_\tau^2]$), and using the same approach as in [1] it can be shown that the conditional PDF $p_{\alpha_\tau|\alpha}(\alpha_\tau|\alpha)$ is given by:

$$p_{\alpha_\tau|\alpha}(\alpha_\tau|\alpha) = \frac{2\alpha_\tau}{(1-\rho)\,\Omega} I_0\left(\frac{2\sqrt{\rho}\,\alpha\alpha_\tau}{(1-\rho)\,\Omega}\right) e^{-\frac{(\alpha_\tau^2+\rho\alpha^2)}{(1-\rho)\,\Omega}}. \tag{18}$$

where ρ is the correlation factor between α and α_τ and $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind [5]. Assuming Jakes Doppler spectrum, the correlation coefficient can be expressed as $\rho = J_0^2(2\pi f_D \tau)$, where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind and f_D [Hz] is the maximum Doppler frequency shift [1]. Recognizing that (18) is similar to [2, Eq. (A-4)], gives the following PDF at time $t+\tau$ for the ORRF algorithm expressed in terms of γ_τ and $\overline{\gamma}$ [2, Eq. (5)]:

$$p_{\gamma_\tau^*}(\gamma_\tau) = \sum_{n=0}^{N-1} \binom{N}{n+1} (-1)^n \frac{\exp\left(-\frac{\gamma_\tau}{\overline{\gamma}(1-\rho\frac{n}{n+1})}\right)}{\overline{\gamma}(1-\rho\frac{n}{n+1})}. \tag{19}$$

Note that for $\tau = 0$ ($\rho = 1$) this expression reduces to (2), as expected. When τ approaches infinity ($\rho = 0$) (19) reduces to the Rayleigh PDF for one user. This is logical since for large τs, the scheduler will not have useful feedback information, and will select users independent of their CNRs.

Inserting (19) into [4, Eq. (8)], using binomial expansion, integration by parts, L'Hôspital's rule and [5, Eq. (3.352.2)], it can be shown that we get the following expression for the MASSE after a delay τ:

$$\frac{<C>_{ora}}{W} = \frac{1}{\ln 2} \sum_{n=0}^{N-1} \binom{N}{n+1} (-1)^n e^{\frac{1}{\overline{\gamma}(1-\rho\frac{n}{n+1})}} \times E_1\left(\frac{1}{\overline{\gamma}\left(1-\rho\frac{n}{n+1}\right)}\right). \tag{20}$$

Using a similar derivation as for the expression above it can be shown that we get the following expression for the delayed MASSE using optimal power and rate adaptation:

$$\frac{<C>_{opra}}{W} = \frac{1}{\ln 2} \sum_{n=0}^{N-1} \binom{N}{n+1} (-1)^n E_1\left(\frac{\gamma_0}{\overline{\gamma}\left(1-\rho\frac{n}{n+1}\right)}\right), \tag{21}$$

with the following power constraint:

$$\sum_{n=0}^{N-1} \binom{N}{n+1} (-1)^n \left(\frac{e^{-\frac{1}{\overline{\gamma}(1-\rho\frac{n}{n+1})}}}{\gamma_0} - \frac{E_1\left(\frac{1}{\overline{\gamma}(1-\rho\frac{n}{n+1})}\right)}{\overline{\gamma}(1-\rho\frac{n}{n+1})} \right) = 1. \tag{22}$$

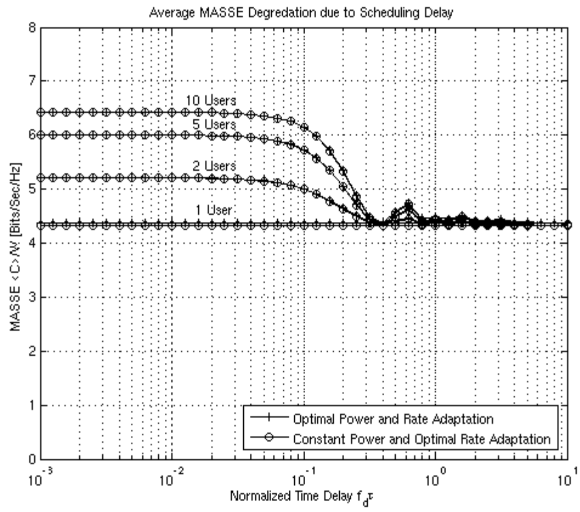

Figure 3: Average degradation in MASSE due to scheduling delay for ORRF using optimal power and rate adaptation and optimal rate adaptation.

From Fig. 3 we see that independent of the number of users and adaptation policy, the system will be able to operate satisfactory if the normalized delay is below the critical value of $2\cdot10^{-2}$. For normalized time delays above this value, we see that the MASSE converges towards the MASSE for one user.

7.2 Impact of Outdated Channel Estimates

We will now assume that the transmitter does not have a perfect channel estimate available at time $t+\tau$. Consequently, both the selection of a user and the decision of the constellation size is done at time t. This means that the channel estimates are outdated the same amount of time as the scheduling delay. The constellation size is thus not dependent on γ_τ, and the time delay in this case does not affect the MASSE. However, now the BER will suffer from degradation because of the delay and in [1] it is shown that the average BER, conditioned on γ is

$$\mathrm{BER}(\gamma) = \frac{0.2\gamma}{\gamma + \overline{\gamma}(1-\rho)K_0} \cdot e^{-\frac{\rho K_0 \gamma}{\gamma + \overline{\gamma}(1-\rho)K_0}}. \qquad (23)$$

The average BER can be found by using the following equation:

$$< \mathrm{BER} >_{acr} = \int_0^\infty \mathrm{BER}(\gamma) p_{\gamma^*}(\gamma)\, d\gamma. \qquad (24)$$

For discrete rate adaptation with constant power, the BER can be expressed by (14), replacing $\overline{\mathrm{BER}}_k$ with $\overline{\mathrm{BER}}'_k$, where:

$$\overline{\mathrm{BER}}'_k = \int_{\gamma_k}^{\gamma_{k+1}} \int_0^\infty \mathrm{BER}(M_k, \gamma_\tau) p_{\gamma_\tau|\gamma}(\gamma_\tau|\gamma) d\gamma_\tau p_{\gamma^*}(\gamma) d\gamma. \qquad (25)$$

Inserting (13) and (18) expressed in terms of γ_τ and γ into (25), we obtain the following expression for the average BER within a fading region:

$$\overline{\mathrm{BER}}'_k = \frac{0.2N}{\overline{\gamma}} \sum_{n=0}^{N-1} \binom{N-1}{n} (-1)^n \frac{e^{-\gamma_k c_{k,n}} - e^{-\gamma_{k+1} c_{k,n}}}{d_{k,n}}, \qquad (26)$$

where $c_{k,n}$ is given by

$$c_{k,n} = \frac{1+n}{\overline{\gamma}} + \frac{3\rho}{3\overline{\gamma}(1-\rho) + 2(M_k - 1)}, \qquad (27)$$

and $d_{k,n}$ by

$$d_{k,n} = \frac{1+n}{\overline{\gamma}} + \frac{3(1+n-\rho n)}{2(M_k - 1)}. \qquad (28)$$

Note that for zero delay ($\rho = 1$) $c_{k,n} = d_{k,n} = a_{k,n}$, and (26) reduces to (16), as expected.

Because we are interested in the average BER only for the CNRs for which we have transmission, the average BER for continuous-power, continuous-rate M-QAM is

$$< \mathrm{BER} >_{acr,pa} = \frac{\int_{\gamma_K}^\infty \mathrm{BER}(\gamma) p_{\gamma^*}(\gamma)\, d\gamma}{\int_{\gamma_K}^\infty p_{\gamma^*}(\gamma)\, d\gamma}. \qquad (29)$$

Correspondingly, the average BER for the continuous-power, discrete-rate M-QAM case is given by:

$$< \mathrm{BER} >_{adr,pa} = \frac{\int_{\gamma_0^* M_1}^\infty \mathrm{BER}(\gamma) p_{\gamma^*}(\gamma)\, d\gamma}{\int_{\gamma_0^* M_1}^\infty p_{\gamma^*}(\gamma)\, d\gamma}. \qquad (30)$$

Fig. 4 shows that the average system BER is satisfactory as long as the normalized time delay again is below the critical value 10^{-2} for the adaptation schemes using continuous power and/or continuous rate.

8. CONCLUSION

We have analyzed a scheduling algorithm that has optimal spectral efficiency, but reduced feedback compared with full feedback load. We obtained closed-form expressions for the optimal CNR threshold used in this algorithm. Novel closed-form expressions have also been found for the system spectral efficiency when using M-ary quadrature amplitude modulation. Both the impact of scheduling delay and outdated channel estimates have been analyzed.

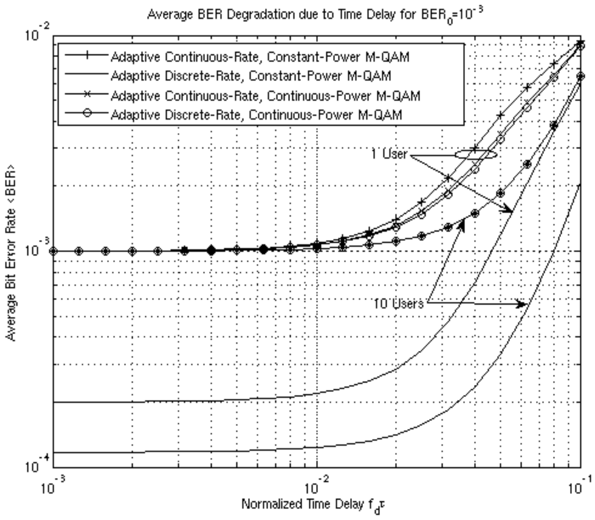

Figure 4: Average BER degradation due to time delay for ORRF using M-QAM rate adaptation with $\overline{\gamma}$=15 dB, 5 fading regions and $\mathrm{BER}_0 = 10^{-3}$.

REFERENCES

[1] M.-S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," in *Kluwer Journal on Wireless Communications*, vol. 13, pp. 119–143, May 2000.

[2] J. H. Barnard and C. K. Pauw, "Probability of Error for Selection Diversity as a Function of Dwell Time," in *IEEE Trans. Commun.*, vol. 37, no. 8, pp. 800–803, Aug. 1989.

[3] D. Gesbert and M.-S. Alouini, "How much feedback is multi-user diversity really worth?," in *Proc. ICC 2004*, Paris, France, June 2004, pp. 234–238.

[4] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power M-QAM for fading channels," in *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

[5] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products.* Address: Academic Press, San Diego, CA, USA, 2000.

[6] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. ICC 1995*, Seattle, WA, USA, June 1995, pp. 331–335.