Source: **Bernd Friedrichs, Bosch Telecom, Germany**

Title: **On the Statistical Multiplex Gain of HA Systems**

Agenda item: **HA PHY / DLC**

Document for:

Decision	
Discussion	X
Information	X

Decision/action requested

The contribution should assist the bandwidth design of HA systems and may serve as background information for the statistical effects of traffic multiplexing.

Abstract

The statistical multiplex gain is one of the major performance measures of wireless point-to-multipoint systems and an important criterion for the design of the bandwidth of Hiperaccess systems. The relations between the multiplex gain, the bandwidth, the burstiness of the data sources, the number of terminals and the quality requirements is described in detail, both in theory as well as with numerical results. In addition, the effects of clustering are investigated, if a broadband system is replaced by several narrowband systems. This contribution is restricted to real-time traffic, the more complicated effects due to buffering and traffic shaping for specific DLC architectures in case of non-real-time services are not considered.

1. The Traffic Model and Fundamental Definitions

We consider a PMP architecture with TDM/TDMA using FDD. For the transmission in downlink from the base station to the terminals as well as for the uplink a single broadband carrier is supposed. As an alternative for the uplink, the separation of the broadband carrier into several narrowband carriers is also taken into consideration.

For a sector of the PMP system the allocated spectral ressource is shared among many users, where the term user could represent a terminal or a connection. Only a part of the ressource is allocated to a specific user, especially by time slots in case of TDM/TDMA, ie. a single user occupies the complete spectrum, but only for certain time slots. The sharing processes may be similar but completely independent for downlink and uplink. The channel bandwidth B determines the **total sector rate**

$$r_s = B/(1+\beta),$$

where β denotes the roll-off factor. To ease the notation, we assume QPSK modulation and error-control coding with a code rate of 1/2, so a symbol corresponds exactly to an information bit.

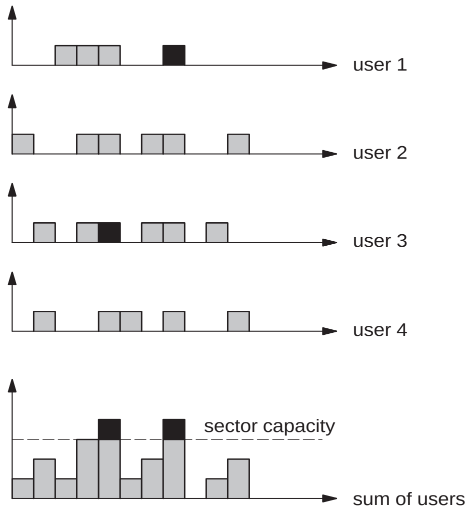

Figure 1: Simple on/off model for real-time services

This contribution is restricted to the simple on/off model for the user data without any kind of buffering as shown in **Figure 1**. We assume that all users have the same source statistics and the cells of different users are statistically independent and are transmitted in time slots. In the literature, this model is also addressed as rate envelope multiplexing (REM) or bufferless call admission control (CAC). The **cell loss rate** (CLR) is the probability that the sum of the traffic rates exceeds the total sector rate, related to the area above the sector capacity which is marked with black boxes in Figure 1. These black cells are declared as lost cells. The model describes real-time traffic only. In contrast, for non-real-time traffic, the peaks of the total load can be smeared over time by buffers (i.e. traffic shaping) to increase the sector capacity at the expense of delay or delay variation.

A TDM/TDMA time frame is divided into N_{cf} time slots. For simplification we assume that a time slot corresponds to an ATM cell. A user can occupy a maximum of one time slot per frame, ie. one cell per frame is transmitted. Therefore N_{cf} is the **guaranteed maximum number of collision-free users per sector.** For an active user, r_s is the data rate within the allocated slot and $r_p = r_s / N_{cf}$ is the data rate with respect to averaging over the frame. Usually, r_p is called the **peak data rate per user (**or peak cell rate in case of ATM). Let

$$P(\text{user is active with a data rate of } r_p) = p = \frac{1}{b} \ ,$$
$$P(\text{user is passive with a data rate of } 0) = 1\text{-}p = \frac{b\text{-}1}{b},$$

where P denotes probability and b is the so-called **burstiness** of the user data source. Thus $r_a = r_p / b$ is the **average user data rate** (or sustainable cell rate in case of ATM). Hence, in summary,

$$b = \frac{r_p}{r_a} = \frac{\text{peak data rate}}{\text{average data rate}} \quad \text{and} \quad N_{cf} = \frac{r_s}{b \cdot r_a} = \frac{r_s}{r_p}.$$

Due to the statistical multiplex, the traffic peaks on all user data streams will usually not coincide, so more than N_{cf} users can be served if $b > 1$. Let N_{eff} be the effective maximum number of users (which are multiplexed together) under the condition that the required capacity $N_{eff} r_p$ exceeds the true capacity r_s of the sector only with a small probability, which is related to the CLR (see below for an exact definition of CLR). Note that $N_{eff} r_p > r_s$ is equivalent to $N_{eff} > N_{cf}$. Apparently, N_{eff} is bounded as follows:

$$N_{cf} \le N_{eff} \le b \cdot N_{cf},$$

where the upper bound is satisfied with equality if every time slot is occupied. The **statistical multiplex gain** *G* with respect to a specific CLR is defined as [2,3]

$$\begin{aligned} G = \frac{N_{eff}}{N_{cf}} &= \frac{\text{max \# users with statistical multiplex}}{\text{max \# users with static collision} - \text{free multiplex}} \\ = \frac{N_{eff} \cdot r_p}{r_s} &= \frac{\text{required total sector rate with static collision} - \text{free multiplex}}{\text{required total sector rate with statistical multiplex}} \end{aligned}$$

Another important parameter is the **spectrum utilization** *U*, defined as

$$U = \frac{\text{average total data rate in sector with statistical multiplex}}{\text{total sector rate}} = \frac{G}{b}.$$

A further way to look at statistical multiplexing could be based on the so-called effective bandwidth $r_{eff} = r_s / N_{eff}$, thus $G = r_p / r_{eff}$. However, it is of minor importance whether *G* or *U* or N_{eff} or r_{eff} is used for the description of multiplex effects, since these parameters are linked together by very simple relations. Apparently, $1 \le G \le b$ and $0 \le U \le 1$, given that $r_s \ge r_p$ (otherwise one single user exceeds the sector capacity). The variables *G* and *U* are not continuous but fractions of integers.

A perfect DLC layer is assumed, ie. it is always guaranteed to transmit up to the maximum number of N_{cf} cells and only the exceeding cells are declared as lost. Now the CLR is defined as follows [3]:

$$CLR = \frac{\text{average number of lost cells}}{\text{average total number of cells to be transmitted}} = \frac{E(y)}{E(x_{binomial})},$$

where $x_{binomial}$ is the random variable describing the total number of cells to be transmitted and the variable $y = \max(0, x_{binomial} - N_{cf})$ describes the number of lost cells, where *y* is a function of

$x_{binomial}$. Obviously, $x_{binomial}$ is binomial distributed with the mean $\mu = N_{eff}\,p$ and the variance $\sigma^2 = N_{eff}\,p(1-p)$. Hence, the exact expression for the CLR is

$$CLR = \frac{\sum_{k>N_{cf}}^{N_{eff}} (k-N_{cf}) \cdot \binom{N_{eff}}{k} \cdot p^k (1-p)^{N_{eff}-k}}{N_{eff} \cdot p} \qquad \text{(A)}$$
$$= 1 - \frac{1}{r_a / r_s \cdot N_{eff}} \cdot \sum_{k<N_{cf}} (k-N_{cf}) \cdot \binom{N_{eff}}{k} \cdot p^k (1-p)^{N_{eff}-k}$$

For given parameters r_a / r_s and b and CLR, N_{eff} is the maximum integer such that the ratio above is less or equal to *CLR*.

2. Numerical Evaluation

Since the evaluation of the weighted sum of binomial coefficients in equation (A) is not a simple matter if N_{eff} attains large values (e.g. more than 1000...10000), several approximations can be found in the references. In [2], the CLR is firstly replaced by the probability of collisions and $x_{binomial}$ is secondly approximated by a Gaussian distribution:

$$CLR \approx P(x_{binomial} > N_{cf})$$
$$\approx P(x_{Gaussian} > N_{cf}) = Q\left(\frac{N_{cf} - N_{eff}\,p}{\sqrt{N_{eff}\,p(1-p)}} \right). \qquad \text{(B)}$$

This equation can be solved for *G* analytically by some simple manipulations:

$$G = \frac{b \cdot r_a / r_s}{4} \cdot \left(\sqrt{\frac{4}{r_a / r_s} + \alpha^2 (b-1)} - \alpha \sqrt{b-1} \right)^2, \quad \text{where} \quad \alpha = Q^{-1}(CLR).$$

This result was used in [1] and is also represented by the dotted lines in the Figures 2 to 6 to demonstrate the errors caused by the two approximations in (B). Another approximation for the weighted binomial sum is given by Lindberger`s equations [3] (let $l = -2\log_{10}(CLR)$ and $a = 1 + l/100$):

$$\frac{1}{U} = \frac{b}{G} = \left\{ \begin{array}{ll} a(1+3z(1-1/b) & \text{for } z \le \min(1, b/3) \\ a(1+3z^2(1-1/b)) & \text{for } 1 \le z \le \sqrt{b/3} \\ a \cdot b & \text{otherwise} \end{array} \right\}, \quad \text{where} \quad z = l \cdot b \cdot r_a / r_s. \qquad \text{(C)}$$

The results of Lindberger`s approximation are shown in Figure 7. However, the optimum approach is the exact evaluation of the CLR. The weighted binomial sum in equation (A) can be exactly evaluated at least for $N_{eff} \le 1000$, whereas for larger values of N_{eff} an exact evaluation of a Gaussian approximation of the binomial sum should be performed as follows:

$$CLR \approx \frac{E(\max(0, x_{Gaussian} - N_{cf}))}{E(x)} \qquad \text{(approximation for large } N_{eff} \text{ only)}$$

$$= \frac{(\mu - N_{cf})Q\left(\frac{N_{cf}-\mu}{\sigma}\right) + \frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{(N_{cf}-\mu)^2}{2\sigma^2}\right)}{\mu} \qquad \text{(exact)} \qquad \text{(D)}$$

3. Results for Statistical Multiplex Gain and Spectrum Utilization

The solid lines in the Figures 2 to 6, 8, 9 display the exact results based on equations (A) and (D). The dotted lines in Figures 2 to 6 refer to the simple Gaussian approximation of the collision probability as stated in equation (B). Figure 7 displays Lindberger's approximation according to equation (C). In most Figures the results are displayed over the average user rate to total sector rate, but for comparison in Figure 7 the multiplex gain is also displayed over the peak user rate to total sector rate.

The **Figures 2** and **4** illustrate *G* for CLR = 10^{-6} and **Figure 3** for CLR = 10^{-12}. Increasing *b* implies increasing *G* in case of large r_s, and $G \to b$ for $r_a / r_s \to 0$. However, for small r_s contrary relations are observed, i.e. increasing *b* reduces *G*, and especially for $r_a / r_s > 1/b$, i.e. $r_p > r_s$, the sector capacity is smaller than the data rate of a single user, hence *G* = 0. Such detailed observations are not possible if the analysis is restricted to the simple approximation according to equation (B). For lower CLR (ie. less collisions or higher requirements), drastic losses in *G* are observed, e.g. for r_a / r_s =0.0005 and *b*=256, i.e. N_{cf} =7.8, *G* reduces from 23 for CLR = 10^{-6} to 2 for CLR = 10^{-12}.

Figure 5 displays *U*. We observe $U \to 1$ for $r_a / r_s \to 0$, *U=1/b* for large r_a / r_s and *U=0* for $r_a / r_s > 1/b$.

Figure 6 shows the effective number of users N_{eff}. Apparently, high *G* and *U* do not only require huge spectrum, but also a huge number of users. Example: For 1000 effective users, the total sector rate must be 2000 times higher than the average rate (in detail slightly depending on the *b*), but *U* depends strongly on the *b* and ranges from 1 (low *b*) to 0 (high *b*). Hence, good spectral utilization requires many users and a lot of spectrum and also low burstiness, however, we have to keep in mind, that the spectral utilization is only *1/b* for traditional systems without statistical multiplexing.

Figure 7 shows *G* by using Lindberger's approximation. The numerical evalation of equation (C) is very simple compared to (A), but this has to be paid with a considerable error in the results as the comparison with Figure 2 indicates.

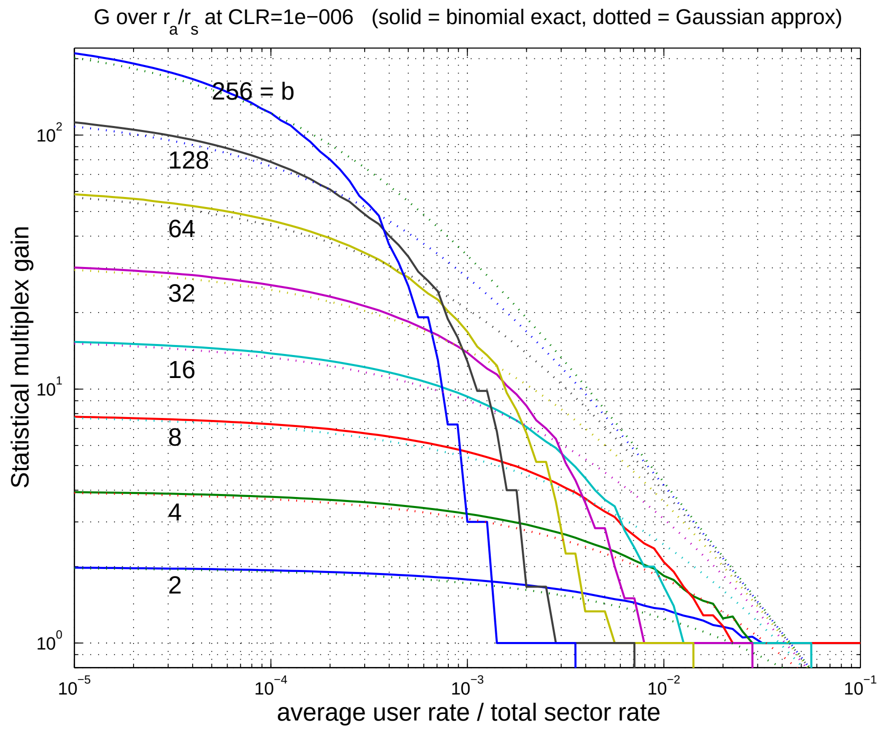

Figure 2: Statistical multiplex gain for $CLR = 10^{-6}$

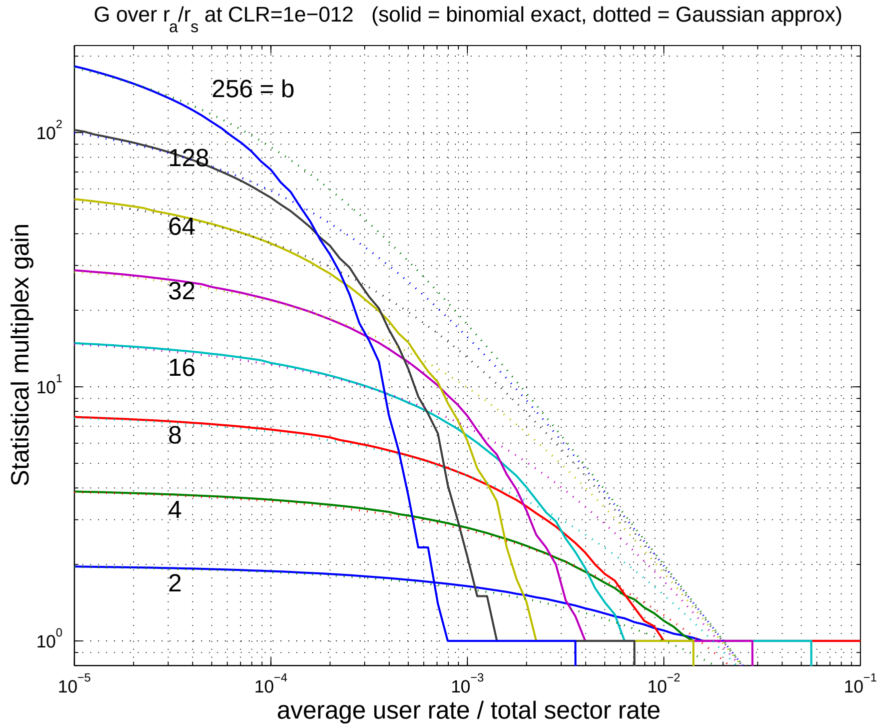

Figure 3: Statistical multiplex gain for $CLR = 10^{-12}$

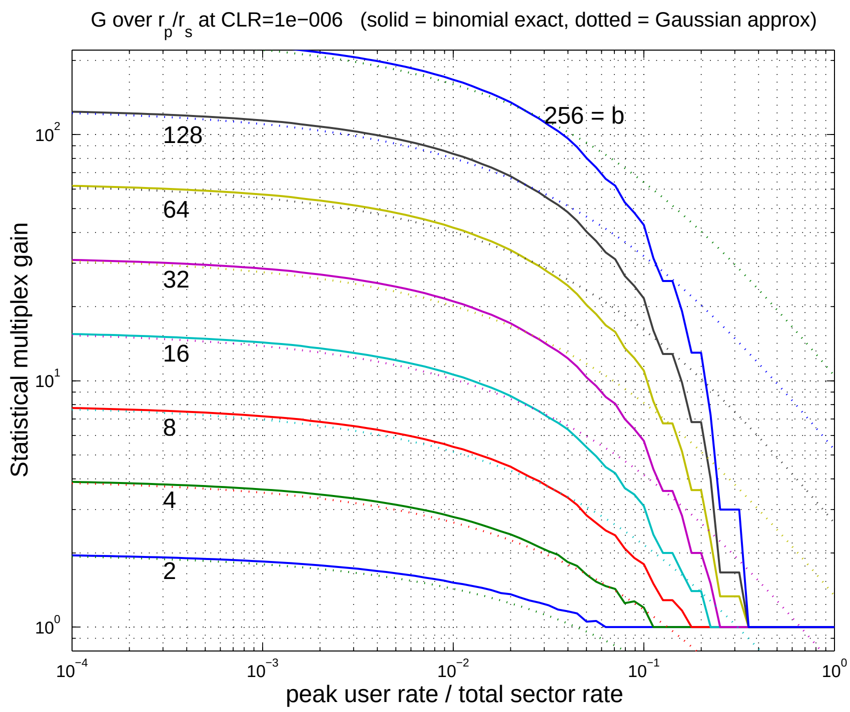

Figure 4: Statistical multiplex gain for $CLR = 10^{-6}$ **displayed over** r_p / r_s

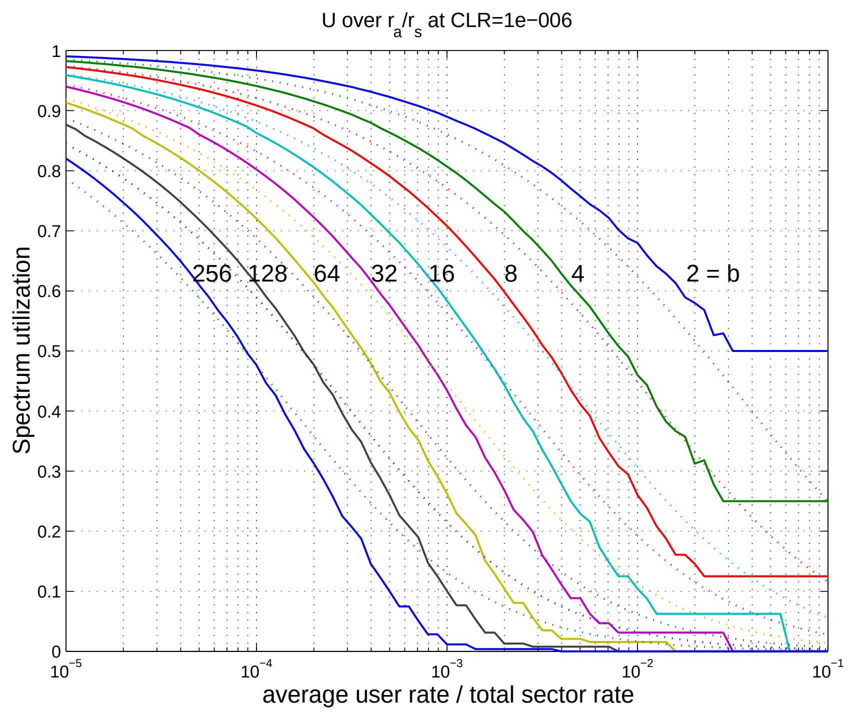

Figure 5: Spectrum utilization for $CLR = 10^{-6}$

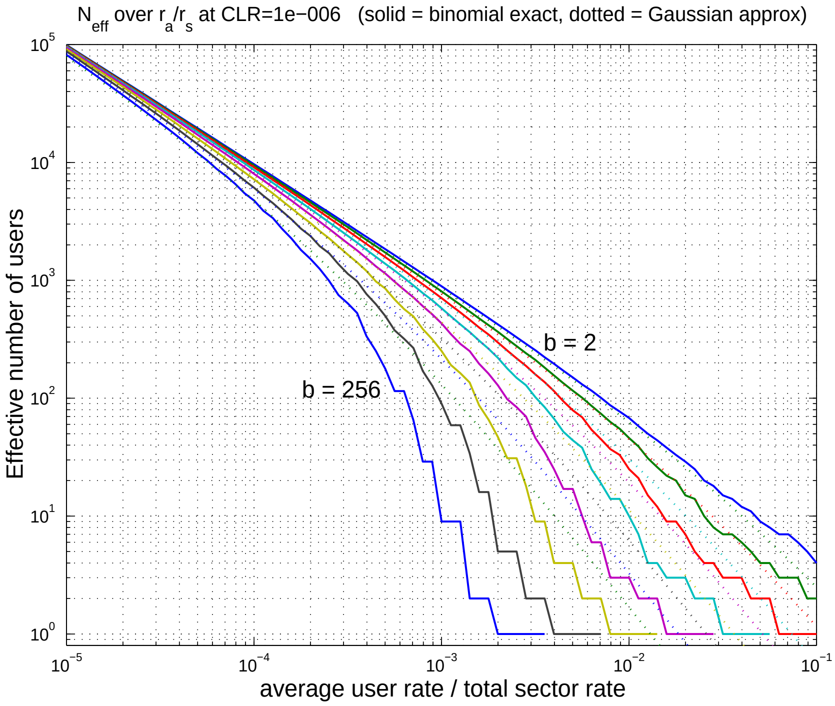

Figure 6: Efficient number of users N_{eff} **for** $CLR = 10^{-6}$ **displayed over** r_p / r_s

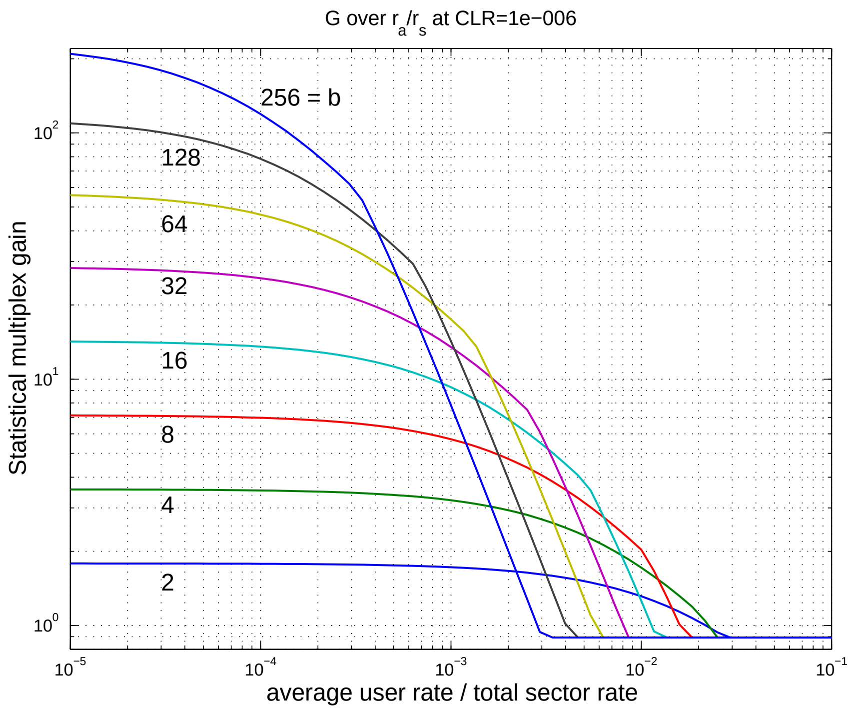

Figure 7: Lindberger`s approximation (compare with Figure 2)

4. Effects due to User Clustering

We consider now the effects caused by a clustering of the users (or terminals). Such a clustering occurs, if a broadband carrier for a sector (eg. 28 MHz) is replaced by several narrowband carriers (eg. 4 carriers each with 7 MHz). If the terminals have a fixed allocation to a specific narrowband carrier and can not change dynamically between the narrowband carriers, then a cluster size of 4 (for this example) results. Such a scenario is supposed for the **Figures 8** (with b = 8) and **9** (with b = 256) displaying the spectral utilization over the ratio of average user rate to total sector rate for cluster sizes of 1 (implying identical results as in Figure 2), 2, 3 and 4. The maximum data rate per narrow carrier equals the total sector rate divided by the number of clusters.

The curves for cluster size of 2, 3 and 4 are approximately obtained by left-shifts with these factors from the curve for cluster size 1, only some small unimportant differences occur due to quantization effects. Obviously, clustering can imply severe reductions in the efficiency, and a numerical example is given in Table 1.

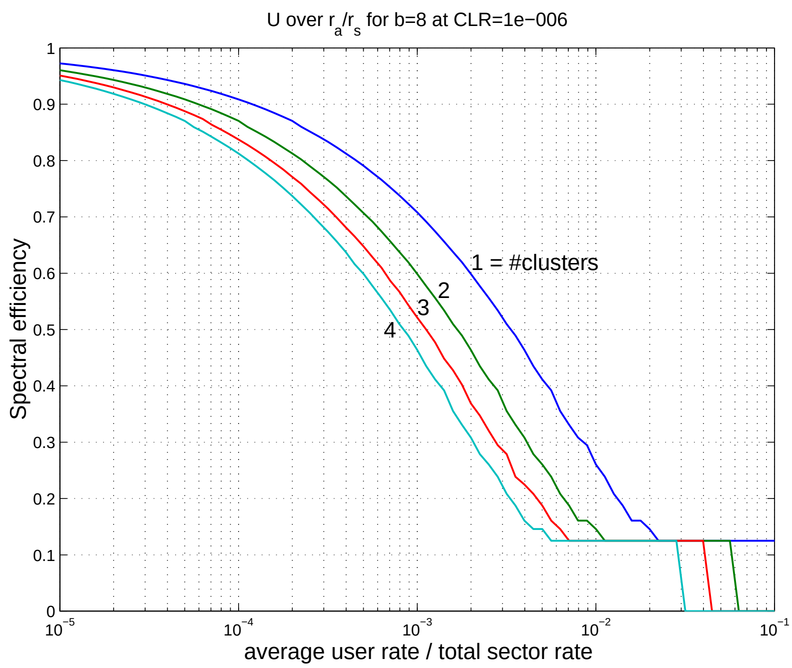

Figure 8: Spectral efficiency for b=8 with user clustering

	no clustering	2 clusters	4 clusters
b = 8, implies N_{cf}= 256	U = 0.87 N_{eff} = 4350	U = 0.81 N_{eff} = 4050	U = 0.73 N_{eff} = 3650
b = 256, implies N_{cf} = 19	U = 0.31 N_{eff} = 1550	U = 0.07 N_{eff} = 350	U = 0.02 N_{eff} = 100

Table 1: A simple example demonstrating the effects of clustering for r_a/r_s=0.0002
(b=burstiness, r_a=average user rate, r_s=total sector rate,
N_{cf}=maximum #users without stat.multiplex, N_{eff}=effective #users with stat.multiplex)

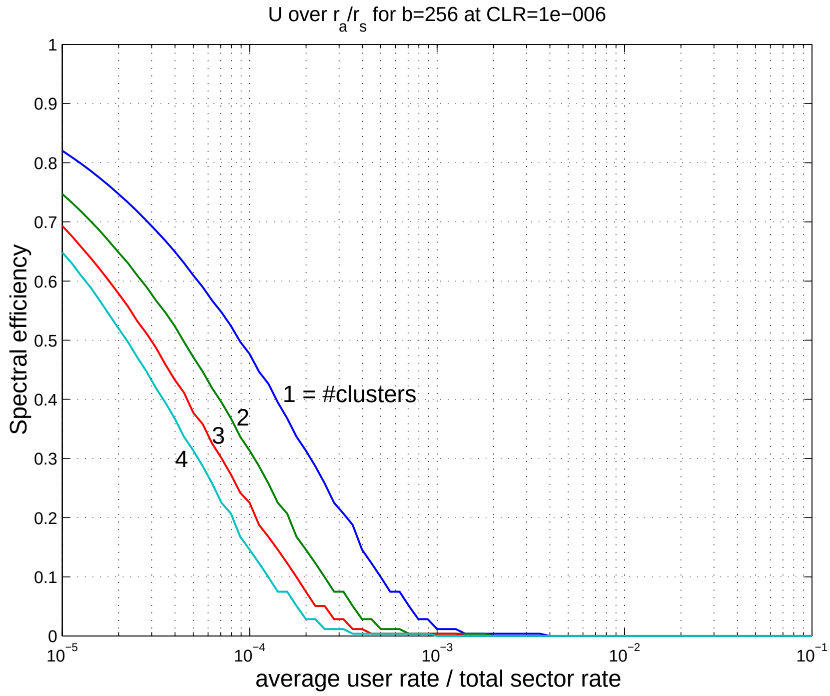

Figure 9: Spectral efficiency for b=256 with user clustering

5. Summary

The statistical multiplex gain as well as the spectrum utilization does not provide us with a tight rule for the design of the HA bandwidth. Higher bandwidth improves in general, but the range between poor und excellent performance is larger than a factor of 10...100 in the user data rates. Further details depend on the burstiness of the user data which is not known and depends strongly on the type of service, certainly the burstiness is between 1 (eg. FTP) and more than 1000 (eg. web surfing). A broadband system of bandwidth *B* has clearly a better performance than *n* narrowband systems with a bandwidth of *B/n*.

Non-real-time traffic allows buffering and traffic shaping and will improve the spectral efficiency of statistical multiplexing, depending on the tolerable delays and the details of the DLC architecture. This is for further investigation.

References

[1] B. Friedrichs: On the Channel Bandwidth of Hiperaccess Systems. ETSI BRAN document HA15bot2a.pdf, October 1999.

[2] D.McDysan, D.Spohn: ATM Theory and Applications. New York: McGraw-Hill 1998.

[3] J.Roberts et al (Eds.): Broadband network Traffic: Performance Evaluation and Design of Broadband Multiservice Networks, Final Report of Action COST 242. Berlin: Springer-Verlag 1996, LNCS Vol. 1155.