TD(05)045

4th COST 290 MCM, Würzburg, October 13-14, 2005

Multiuser capacity and fairness evaluation of channel/QoS aware multiplexing algorithms

J.T. Entrambasaguas, M.C. Aguayo-Torres, G. Gómez, J.F. Paris
Departamento de Ingeniería de Comunicaciones, Universidad de Málaga
Campus Universitario de Teatinos s/n, Málaga, Spain E-29070
{jtem,aguayo,ggomez,paris}@ic.uma.es

Abstract

In this work multiplexing of variable-rate sources over generic multiple shared fading channels is addressed. A number of user multiplexing algorithms based on Channel State Information (CSI) and/or Quality of Service (QoS) indicators have been proposed in the literature. It has been shown that, as wireless channel behaviour changes in time and is distinct for each user, capacity gain due to (transmission) time and multiuser diversities can be obtained. If channel multiplexing is present, multichannel diversity gain can also be added up. Moreover, since many current traffic sources are variable rate, additional statistical gain is achieved because of (source) multiuser and time diversities.

Despite of the copious work devoted to user multiplexing, algorithms' performance comparison (if presented) is often carried out over very specific environments, what makes difficult to extract general conclusions about the multiuser capacity and the treatment given to each user by the proposed schemes.

In this work a generic variable-rate multiuser and multichannel system using adaptive modulation is addressed. Multiuser capacity and fairness is evaluated for different common CSI and/or QoS-aware user multiplexing algorithms. Results show the joint diversity gain and the QoS in terms of throughput and delay experienced by different flows.

Keywords

QoS, user multiplexing, adaptive resource allocation, diversity, capacity, delay, fairness

Working Group 1

Multiuser capacity and fairness evaluation of channel/QoS aware multiplexing algorithms

J.T. Entrambasaguas, M.C. Aguayo-Torres, G. Gómez, J.F. Paris
Departamento de Ingeniería de Comunicaciones, Universidad de Málaga
Campus Universitario de Teatinos s/n, Málaga, Spain E-29070
{jtem,aguayo,ggomez,paris}@ic.uma.es

1. INTRODUCTION

A fundamental characteristic of the wireless channel is the fading of the channel strength due to constructive and destructive interference between multipahts and, as a consequence, the random change of the channel capacity along the time [Goldsmith, 1997]. Commonly, service rate is constant during a session and hence, the bit error rate suffered by the transmitted data is variant. However, a variety of adaptation techniques such as adaptive modulation [Chung, 2001] and coding [Ling, 2003]or variable-rate spreading [Yang, 2004] has been introduced in modern wireless systems to exploit this (*transmission) time diversity*. These mechanisms provide the system the ability to match the effective throughput to the channel conditions of a specific user in order to keep Bit Error Rate (BER) constant. As a consequence, service rate is a function of time as those examples shown in Figure 1 and physical layer is no longer well modelled as a fixed-rate bit pipe. Note that as fading is time-correlated, the service rate is also so.

In a multiuser environment, another important characteristic is that channel quality varies asynchronously for different users. Such variations in the channel conditions can be exploited to increase the system throughput. An "opportunistic scheduler" assigns the channel to a user having the best channel condition at a given time. This

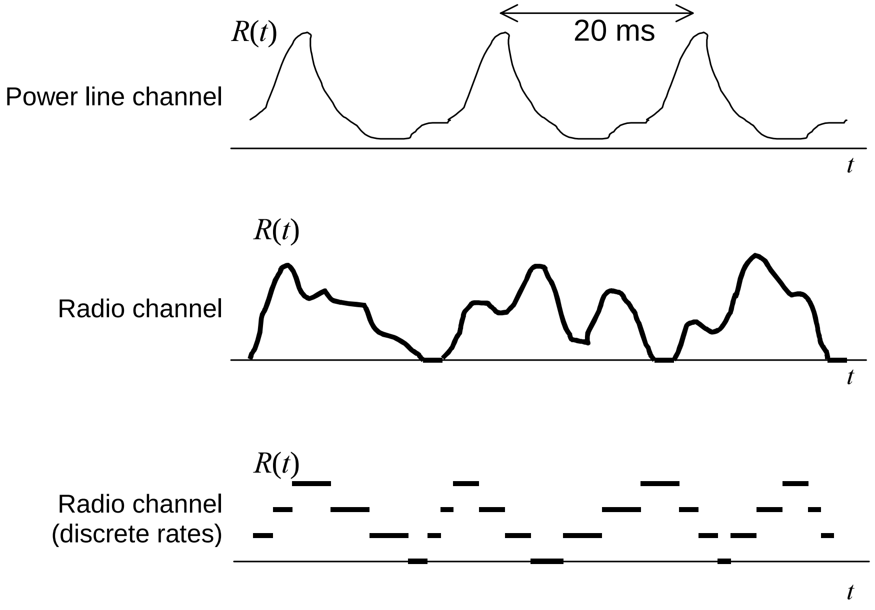

Figure 1. Several examples of variant service rate. The modelling function can be continuous or discrete

phenomenon is known as *multiuser diversity* [Knopp, 1995]. Exploiting this diversity results in a higher total system capacity as the number of users increases.

In a shared channel like the wireless link, resources are organized in physical channels [Glisic, 2005]. Each physical channel is assigned to a particular user for a certain period of time, which is usually called “frame” (but hereinafter referred to as “symbol”). Since the digital transmission appeared, the most common way to obtain the physical channels has been Time Division Multiplexing (TDM). In the last decade other techniques as Code Division Multiplexing (CDM) or Orthogonal Frequency Division Multiplexing Access (OFDMA) has began to be used. Actually, all those techniques are based in the same concept: orthogonal pulses multiplexing[1]. Physical channels can also be obtained by antenna multiplexing.

When TDM or CDM is the channel multiplexing mechanism employed, all the channels are equal for a specific user during one symbol period (frame). However, if other channel multiplexing mechanisms as Frequency Division Multiplexing (FDM) are used, physical channels (subcarriers) may experience different fading levels and have different behaviour during the symbol period. As on single channel links, adaptive modulation can be employed to track the assigned channel quality, associating the set of channels with different transmission rates. Mechanisms to exploit this *multichannel diversity* are well known in contexts as bit loading for OFDM transmissions [Chow, 1995] [Keller, 2000].

Copious works consider this transmission diversity (time, multiuser, multichannel) to multiplex users over channel in OFDMA environments. This is usually referred to as time and slot allocation [Song, 2005] [Wong, 1999]. Several aims can be considered: throughput maximization, constant rate transmission, minimum power allocation, etc. It has been shown that the assignment of system resources to the user experiencing the current Best Channel (BC) quality maximises the aggregated bit rate under different assumptions (about users’ targets BER) [Jang, 2003].

The aforementioned 3D transmission diversity (time, multiuser, multichannel) is present whatever the source behaviour. Many current traffic sources are variable rate, although time correlation is frequently present in the generated information rate. Figure 2 presents several examples of variant information rate. Even sources commonly described as constant bit rate, like voice, can be considered as variant rate. This can be envisioned as a kind of *source time diversity*. Because of traffic burstiness, resources requirements of a user fluctuate in a time scale similar to the information rate coherence time.

Moreover, the information rate varies with time asynchronously for different users. Exploiting the *source multiuser diversity* will allow more users to be accommodated on the system, what in wireline systems (with no transmission diversity) is often termed as statistical gain [Jha, 2002].

---

[1] Although in CDMA the orthogonality principle is not always met.

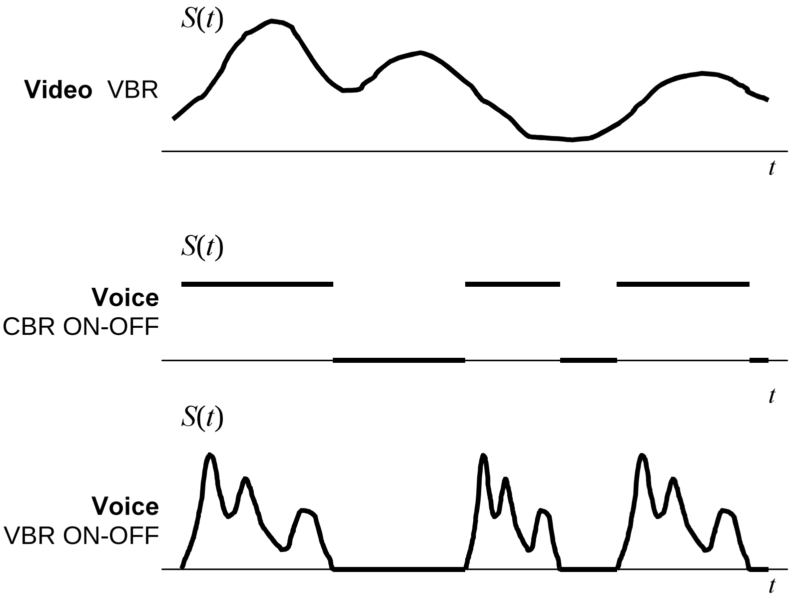

Figure 2. Several examples of variant information rate.

Delay tolerance of data traffic makes feasible to defer the transmission when their channel conditions are unfavourable and, as a consequence, to maximize the overall throughput. However, many current applications Quality of Service (QoS) requirements are not only related to the average throughput [3GPP, 2001] [Jha, 2002]. For multimedia networking many applications such as music and video streams are delay-sensitive. Even for Web browsing, response time is still an important performance criterion. Depending of the specific application, QoS features should provide the required performance of mean delay, maximum delay or delay variance (jitter). At the physical layer, fundamental communication limits established by information theory are, in many cases, well understood. However, when delay is taken into account, much less is known about fundamental performance tradeoffs.

Fairness is another problem to be confronted as the idea of multiuser diversity is implemented. In wireline networks, fairness is usually guaranteed by dedicating a certain amount of service share to a flow. However, the fairness issue in wireless networks is more complicated [Cao, 2001]. Each user can have a distinct average link capacity and then a percentage of time could not guarantee real fairness among them. Moreover, traffic characteristics are not necessarily equal for all users and, even more, QoS requirements (in terms of throughput and delay) are determined by the specific service. Considering users' differences is necessary in order to give appropriate treatment to all users.

User multiplexing from different points of view is a popular research topic in wireless systems, although under very distinct names as (dynamic) resource allocation, scheduling or subcarrier and slot allocation. Copious work focuses on a single channel [Chen, 2001] [Lee, 2004] [Rhee, 2001] [Guo, 2002] although multiple shared channels are also addressed in OFDMA environments [Song, 2005] [Damji, 2004] [Song, 2002]. In a set of the proposals, delay is considered as an issue [Berry, 2004] [Kwok, 2003] [Xiao, 2003] [Zhang, 2004]. Treatment given to each user by the proposed schemes has been analysed for equal [Kulkarni, 2003] [Chen, 2002] or distinct user requirements [Svedman, 2004] [Johnsson, 2005]. However, not many studies evaluating a wide set of proposed algorithms can be found [Shakkottai, 2001] [Hanssen, 2004]. On the other hand, performance comparison is generally carried out over very specific environments, what makes difficult to extract general conclusions about their multiuser capacity and fairness.

In this document, the performance of a generic multichannel and multiuser system using Adaptive Quadrature Amplitude Modulation (AQAM) is addressed. A simplified system model is employed in order to draw conclusions more easily. Physical channel conditions, random traffic arrival, QoS requirements and user fairness are taken into consideration in the system model.

Different resource allocation algorithms proposed in the literature based on Channel State Information (CSI) and Quality of Service (QoS) indicators are analysed over fast fading shared channels. Our focus is given to the characterization of multiuser and multichannel diversity gain for different CSI and QoS-aware user multiplexing schemes as well as the QoS experienced by the different flows.

The remainder of this document is organized as follows. Section 2 briefly describes the system model, including user multiplexing, transmission and information source models. In Section 3 several commonly proposed user multiplexing algorithms are described in detail and their performance is evaluated in Section 4. Finally, some concluding remarks and further work proposals are given in Section 5.

2. SYSTEM MODEL

User multiplexing over a multichannel wireless system in a single cell scenario is considered. N_u users share the transmission resources, which are organized in N_c physical channels. Physical time is divided into units that along this work will be referred to as symbols as it represents the discrete time unity.

Figure 3 uses fluid model to represent the process suffered by user information. Bits produced by the i^{th} source at rate $S_i[n]$ are loaded in a bucket (queue) and extracted at the rate $R_i[n]$, which is assigned by the user multiplexing algorithm. As CSI, the channel state matrix $\gamma[n]$ is assumed to be available at the multiplexor without error or delay. If QoS measurements are taken into account, the delay already suffered by the early bit in the queue $w_i[n]$ is employed as indicator, being $V_i[n]$ the amount of information still stored in the queue.

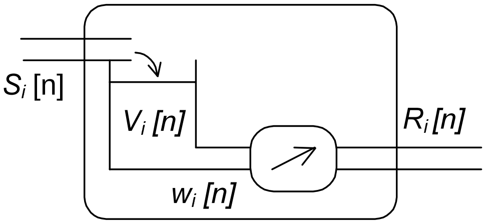

Figure 3. Process suffered by user information

A. MULTICHANNEL MODEL

The matrix $\gamma[n]=\left\{\gamma_{i,k}[n],\ i=1\ldots N_u,\ k=1\ldots N_c\right\}$ represents the received instantaneous Signal to Noise Ratio at symbol *n* for each pair {user *i,* channel *k*}*.* For a specific user and channel, flat Rayleigh fading is considered. Complex channel response of the kth channel for the ith user $h_{i,k}[n]$ follows a Gaussian distribution, and hence, for the received signal-to-noise ratio (SNR), defined as:

$$\gamma_{\mathrm{i,k}}[n]=\frac{\left|h_{i,k}[n]\right|^2}{N_o B_T} \qquad (1)$$

the following exponential distribution is assumed:

$$p(\gamma_{i,k})=\frac{1}{\overline{\gamma_i}}\exp\left(-\frac{\gamma_{i,k}}{\overline{\gamma_i}}\right),\gamma_{i,k}\geq 0 \qquad (2)$$

where $\overline{\gamma_i}$ is the average SNR for the ith user, which is equal for all channels. If perfect power control is assumed, losses between users and base station are compensated, being $\overline{\gamma_i}=\overline{\gamma_j},\ \nabla i,j$. Path loss is considered to be always compensated by power control though large scale shadowing can be included in the model as a lognormal distribution of $\overline{\gamma_i}$.

In the short term, transmitted power is constant and, thus, the received SNR is time-variant although slow enough to consider that there is no change during a symbol (frame). An auto-regression model governs channel correlation between symbols. The channel response at nth symbol is computed as

$$h_{i,k}^{ind}[n]=(1-\rho_G)h_{i,k}^{unc}[n]+\rho_G h_{i,k}^{ind}[n-1] \qquad (3)$$

where $h_{i,k}^{unc}[n]$ is the time-uncorrelated complex channel response. The parameter ρ_G determines the rate at which channel changes: the lower ρ_G, the more uncorrelated channel samples. $\rho_G=0$ models uncorrelated channel responses for each symbol while $\rho_G=1$ results in a constant channel. It has been considered that all channels and users exhibit the same normalized Doppler frequency and hence ρ_G is the same for any user and channel.

In wireless environments it is commonly assumed that at the nth symbol, the kth channel for the ith user $h_{i,k}[n]$ is independent of all of the others users *j≠i*. On the other hand, distinct channel processes $h_{i,k}[n]$ and $h_{i,k'}[n]$ are not independent for a specific user *i*. It has been considered that correlation between channel decreases with the distance between them and the channel response for the ith user has been evaluated as follows:

$$\overrightarrow{h_i}[n] = \begin{pmatrix} 1 & \rho_K & \rho_K^2 & \dots & \rho_K^{N_c-1} \\ \rho_K & 1 & \rho_K & \dots & \rho_K^{N_c-2} \\ \rho_K^2 & \rho_K & 1 & \dots & \rho_K^{N_c-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_K^{N_c-1} & \rho_K^{N_c-2} & \rho_K^{N_c-3} & \dots & 1 \end{pmatrix} \cdot \overrightarrow{h_i^{ind}}[n] \qquad (4)$$

$\overrightarrow{h_i^{ind}}[n]$ is a column vector $h_{i,k}^{ind}[n],\ k = 0 \dots N_c$ of channel-independent complex channel responses while $\overrightarrow{h_i}[n]$ elements ($h_{i,k}[n],\ k = 0 \dots N_c$) include the channel dependence. ρ_K zero value models independent channel responses (e.g. FDM with bands far away enough) while for $\rho_K = 1$ all channels have the same simultaneous response for the ith user (e.g. TDM). The OFDMA scheme can be modelled through an intermediate value determined by the coherence bandwidth.

B. USERS AND SOURCES

Variant rate information sources are considered in this work. Information rate generated by ith user along the symbol n $S_i[n]$ is modelled by a time-correlated stochastic process[2]. The number of bits generated follows a lognormal distribution of average $\overline{S_i}$. An auto-regression model governs information rate correlation between symbols. The parameter ρ_S determines the rate at which channel changes

$$S_i[n] = (1-\rho_S)S_i^{unc}[n] + \rho_S S_i[n-1] \qquad (5)$$

where $S_i^{unc}[n]$ is the time-uncorrelated information rate. The parameter ρ_S determines how fast information rate changes. For usual wireless channels and sources this parameter is lower than ρ_G: channel changes faster than information source. In this document it has been considered that source time-correlation ρ_S is equal for all users.

The offered load to the system $\overline{S}$ is the average of the total information rate generated by users. Due to source independence, it can be calculated as the sum of the average information rate, $\overline{S} = \sum_{i=1}^{N_u} \overline{S_i}$.

Information produced by sources is buffered in separated queues per user that are supposed to be long enough to accommodate all the bits waiting to be transmitted, i.e. no bit is ever discarded. Within one queue, bits are served following an FIFO discipline.

---

[2] This model is worse than that of discrete events for bursty sources with random interval between packets. However, it can be more useful for continous rate as voice or image, and it is more suitable to the model used for the physical channels: a matrix of stochastics processes, $\gamma_{i,k}$[n].

C. MULTIPLEXING AND TRANSMISSION

User multiplexing algorithm allocates channels to users in a symbol per symbol basis: i.e. every new symbol, the system assigns each physical channel to a single user. Let $R_{i,k}[n]\neq 0$ be the assigned rate (in bits) to the user i over the channel k at the symbol n. Then, $R_{j,k}[n]=0$, $j\neq i$.

None, one or several channels can be assigned to a user every symbol period. In total, user *i* is allowed to transmit $R_i[n]=\sum_{k=1}^{N_c} R_{i,k}[n]$ bits at symbol n.

Most multiplexing policies take into consideration physical channels' state. Let $r_{i,k}[n]$ be the achievable throughput (in bits/symbol) of user *i* at channel *k* under a given BER target, which is assumed to be a continuous number. For constant transmitted power, $r_{i,k}[n]$ is a function of the Signal to Noise Ratio $\gamma_{\mathrm{i,k}}[n]$ expressed as

$$r_{i,k}[n]=\log_2\left(1+\beta\gamma_{i,k}[n]\right) \qquad (6)$$

where β is a constant related to BER with value $\beta=\frac{1.5}{-\ln(5BER_T)}$ for uncoded QAM. Channel coding reduces the value in a quantity related to the coding gain.

If BER is fixed, the matrix $\underset{\sim}{r}[n]=\left\{r_{i,k}[n],\ i=1\ldots N_u,\ k=1\ldots N_c\right\}$ represents the physical Channel State Information (CSI) for the set of channels and users and it can be used instead of $\underset{\sim}{\gamma}[n]$. Note that $r_{i,k}[n]$ represents the potential rate that the user *i* might get from the kth channel, while $R_{i,k}[n]$ is the rate actually assigned, which is obtained only if the pair (k,i) is selected, that is:

$$R_{i,k}[n]=\begin{cases} r_{i,k}[n] & \text{if } k \text{ channel is assigned to user } i \\ 0 & \text{other case} \end{cases} \qquad (7)$$

If a single user were present in the system, the maximum attainable rate is the joint ergodic capacity of the channels, calculated as:

$$\overline{C_{su}}(\overline{\gamma}_i)=\sum_{k=1}^{N_c}\int r_{i,k}(\gamma_{i,k})p(\gamma_{i,k})d\gamma_{i,k} \qquad (8)$$

and the throughput is never greater than this value. This is not the case of a multiuser scenario as with the proper user multiplexing into channels the throughput per user can overcome the single user capacity as a result of the (transmission) multiuser diversity.

3. MULTIPLEXING ALGORITHMS

Efficient user multiplexing is the key issue to make use of traffic variation and exploit the transmission diversity. Fixed multiplexing algorithms do not adequate resource

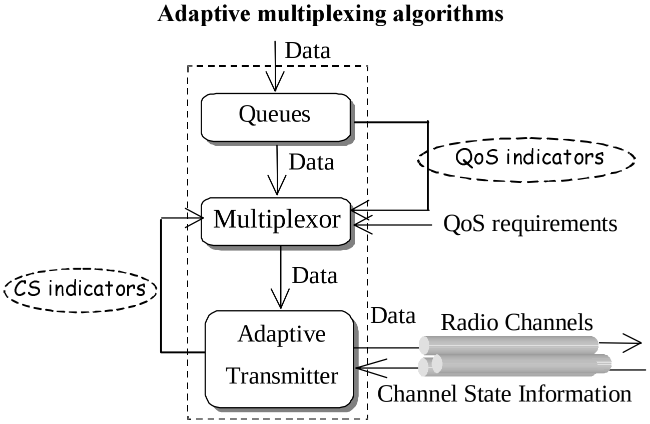

Figure 4. Adaptive multiplexing algorithms

allocation to the channel and source conditions, wasting the attainable diversity gain. Adaptive strategies try to track dynamically the varying conditions of both source and channel. In order to do this, QoS and/or CS indicators must be available to the multiplexor, as depicted in Figure 4. As CS indicators, the instantaneous SNR or the potential rates can be used. The throughput assigned to each user or the maximum delay suffered by their loaded data are the usual QoS measurements.

Eight different multiplexing algorithms, classified in Table 1, have been evaluated in this work. Only one fixed algorithm, Round Robin (RR), has been analysed. Largest Delay First (LDF) considers only QoS indicators. The rest correspond to strategies that adapts their behaviour to channel quality. Best Channel (BC) and Proportional Fair (PF) does not take into consideration QoS measurements. The other four strategies are cross-layer policies, able to simultaneously consider channel and QoS states. While Fair Throughput (FT) uses directly throughput measurements, delay already suffered by the loaded data is the QoS indicator employed by the other three algorithms. Note, however, that as data is never discarded, delay is an indirect throughput measurement that also includes source behaviour in it.

In Table 1, a comparison among algorithms' key features is provided. The different fields in the table indicate whether the different algorithms take into account the channel or queue state, QoS requirements or fairness among users (i.e. whether it avoids service starvation for users due to other users) [Ericsson, 2004].

Fixed Algorithms	**Adaptive Algorithms**		
	Throughput Adaptive	**Channel Adaptive**	**Delay Adaptive**
Round Robin (RR)		Best Channel (BC)	Largest Delay First (LDF)
		Proportional Fair (PF)	
	Fair Throughput (FT)		
		Largest Delay First to Best Channel (LDF-BC)	
		Modified Largest Weighted Delay First (M-LWDF)	
		Exponential Rule (ER)	

Table 1. Fixed vs. Adaptive algorithms

Algorithm	Channel State	Queue state	QoS requirements
RR	No	No	No
BC	Yes	No	No
PF	Yes	No	No
FT	Yes	No	No
LDF	No	Yes	No
LDF-BC	Yes	Yes	No
M-LWDF	Yes	Yes	Yes
ER	Yes	Yes	Yes

Table 1. Summary of the presented multiplexing algorithms

A. ROUND ROBIN (RR)

Round Robin (RR) is a fixed cyclic algorithm without priorities, which dispenses the bandwidth equally among the different flows independently of their priorities or radio channel conditions, i.e. transmission on channel k, $u(k)$, is assigned to the following user in a cyclic order.

$$u(k) = u(k-1) + 1, k = 1 : N_c \qquad (9)$$

Although this is the simplest multiplexing approach, this algorithm is only fair when service rate is constant and equal for all flows. Furthermore, this strategy does not work well over varying channels. Hence, a low efficiency in terms of system throughput and QoS differentiation is expected.

D. BEST CHANNEL (BC)

Best Channel (BC) strategy is adaptive to the channel state, giving priority to those users with higher potential transmission rate $r_{i,k}[n]$, i.e. each channel is assigned to the user that may transmit with the highest number of bits/symbol.

$$u(k) = max_i(r_{i,k}[n]),\ k = 1 : N_c \qquad (10)$$

This algorithm maximizes the total system efficiency. However, under this strategy good average SNR users get more average throughput than low SNR users. As a result, maximum system capacity is attained at the expense of fairness. Moreover, as users in deep fades might experience a bad channel for prolonged periods of time, long delays are expected.

E. PROPORTIONAL FAIR (PF)

This scheme is similar to Best Channel multiplexing; however, instead of using instantaneous potential transmission rate $r_{i,k}[n]$, PF uses the ratio:

$$u(k) = max_i\left(\frac{r_{i,k}[n]}{\bar{r}_i}\right),\ k = 1 : N_c \qquad (11)$$

That is, a user whose SNR is relatively better compared with his average SNR, will be given high priority. In such a way, users will be scheduled when their link is at its best, while in BC multiplexing, the link of the scheduled user is the one that is the best of all the users. The possibility of a user with a very bad link being neglected all the time is thus reduced tremendously. On the other hand, starvation periods are not solved by this policy.

F. FAIR THROUGHPUT (FT)

FT is a QoS-aware strategy, which gives priority to the instantaneous potential throughput vs. mean source throughput ratio:

$$u(k) = max_i \left(\frac{r_{i,k}[n]}{\overline{S_i}} \right), \quad k = 1 : N_c \qquad (12)$$

Note that this strategy is also considering QoS targets of the source in terms of throughput, but not delay.

G. LARGEST DELAY FIRST (LDF)

LDF is adaptive to the delay, providing the transmission turn to the flow that has been suffering the largest delay in its queue. In this strategy, channels are assigned on an arbitrary order.

$$u(k) = max_i \left(w_i[n] \right), \quad k = 1 : N_c \qquad (13)$$

Once a channel is assigned, delays are re-calculated for all data flows. Note that several channels might be assigned to the same flow until it has not the largest delay.

The multiplexor only needs to time stamp arriving data packets of all users. For constant bit rate sources keeping track of the current queue length is enough to implement this multiplexing algorithm.

H. LARGEST DELAY FIRST TO BEST CHANNEL (LDF-BC)

LDF-BC is an enhanced LDF strategy, where its key feature is that multiplexing decision depends on both the states of the queues and current channel conditions. As in LDF case, transmission turn is given to the flow with largest delay, but now the best channel for selected flow is chosen.

$$u(max_k(r_{j,k})) = j, \text{ with } j = max_i \left(w_i[n] \right) \qquad (14)$$

This enhancement provides LDF-BC better system efficiency than LDF, as a bad channel cannot be wasted on a delayed user if it has not enough quality.

I. MODIFIED LARGEST WEIGHTED DELAY FIRST (M-LWDF)

The basic idea behind this strategy is to consider both the waiting time in the queues, the channel capacity for each user as well as the delay tolerance of the service. In particular, M-LWDF serves the following queue

$$u(k) = max_i(g_i \cdot w_i[n] \cdot r_{i,k}[n]),\ \ k = 1{:}N_c \qquad (15)$$

In the previous expression, $w_i[n]$ is the delay already suffered by the loaded data in the queues, $r_{i,k}[n]$ is the potential rate of user i, and g_i introduces a QoS factor suggested to be $g_i = a_i / \overline{r_i}$ [Andrews, 2000], being

$$a_i = -\frac{\log(\delta_i)}{T_i}$$

where δ_i is the desired probability to fulfil the delay requirement T_i.

It has been proved analytically in [Andrews, 2000] that M-LWDF multiplexing algorithm is throughput optimal. This means that with M-LWDF multiplexing, the maximal possible number of users can be supported.

J. EXPONENTIAL RULE (ER)

Exponential Rule algorithm is based on the following two-factors' equation:

$$u(k) = max_i\left(g_i \cdot r_{i,k}[n] \cdot exp\left(\frac{a_i w_i(t) - \overline{aw}}{1 + \sqrt{\overline{aw}}}\right)\right),\ \ k = 1 : N_c \qquad (16)$$

being the notation similar to M-LWDF, and $\overline{aw} = \frac{1}{N_u}\sum_{i=1}^{N_u} a_i w_i[n]$. Note that the first term $g_i \cdot r_{i,k}[n]$ is dominant as long as delays do not grow too large from the average. Otherwise, the exponential factor will give priority to the user with largest delay.

4. PERFORMACE RESULTS

Simulations have been carried out in order to analyse the previously described multiplexing algorithms. In all cases, source and channel time correlation ρ_S and ρ_G are fixed to 0.95 a 0.975, i.e. correlation between symbols is high for both information rate and channel response. β = 4dB has been employed, corresponding to a target BER of about $5 \cdot 10^{-3}$ for encoded QAM.

Results are always presented as a function of the total offered load relative to the single user average capacity, $\overline{S} / \overline{C_{su}}$. The number of users is kept fixed and information rate is increased in short steps.

Delay in symbols is shown in the following figures. Maximum, average, 90th percentile or deviation (jitter) are presented, as their behaviour is quite similar. Note that this is an indirect throughput measure. If a user is underprovisioned, queues get unstable and delay grows towards infinite. On the other hand, if throughput is kept over offered load, maximum delay is bounded.

A set of tests have been conducted and analysed on the different scheduling strategies:

1. No QoS differentiation and independent channels
2. Delay differentiation among data flows
3. Different throughput requirements of the data sources
4. Shadowing effect
5. Channels dependence for the same user
6. Multiuser and Multichannel diversity gain

Main simulation parameters on the different test environments are summarized in Table 2.

Test Environment	**N_u**	**$\frac{N_u}{N_c}$**	$\bar{S}/C_{su}$	$\bar{\gamma}_i$ **(dB)**	ρ_K	$\frac{\mathrm{Max}(\bar{S}_i)}{\mathrm{Min}(\bar{S}_i)}$	**Max. Delay (symbols)**
1	40	4	[0.01:2]	20	0	1	1
2	40	4	[0.01:2]	20	0	1	**1,2,3,4**
3	40	4	[0.01:2]	20	0	**50**	1
4	**[8:100]**	4	[0.01:2]	20	0	1	1
5	[20:100]	**[2:10]**	[0.01:2]	20	0	1	1
6	12	**[1:4]**	[0.01:2]	20	0	1	1
7	40	4	[0.01:2]	20	**[0:1]**	1	1
8	40	4	[0.01:2]	**lognormal**	0	1	1

Table 2. Simulation parameters for the different test environments

A. EFFICIENCY COMPARISON

The scope of this test is to provide a performance comparison in terms of system efficiency and QoS provision for all the multiplexing algorithms described in section 3.

As summarized in Table 2 Enviroment 1, a set of 40 users is sharing the resources from 10 uncorrelated Rayleigh fading channels with same SNR average (i.e. shadowing effect is neglected). Furthermore, sources and QoS requirements (delay and mean information rate) are equal for all the data flows.

Figure 5 show the average delay suffered by all data flows for the eight evaluated multiplexing mechanisms vs. total offered load relative to the single user capacity, $\bar{S}/\bar{C}_{su}$. Delay results are used as an indicator of the “fairness” of the algorithm as well as to determine the maximum offered load that keeps the system stable.

Starting from the RR strategy, it does not provide any multiuser or multichannel diversity gain (i.e. offered load is below 1 to keep queues stable) since RR is not channel-aware, offering the worst performance in terms of system efficiency.

On the contrary, BC and PF provides the best efficiency for some users (close to 2) at the expense of increasing the delay and being unfair among users in terms of delay. In this particular case, PF offers similar behaviour that BC since SNR average is similar for all users. FT is able to reduce users differences but lower system efficiency is achieved.

LDF tries to solve the QoS unfairness, achieving very similar and low delays for all users at the expense of limiting the system efficiency. However, introducing the enhancement of LDF-BC, efficiency is increased up to 40% more.

Finally, M-LWDF and ER provides the best joint results for QoS and system efficiency, achieving M-LWDF a slightly higher efficiency (allowing to reach around 1.5 offered load) than ER.

B. DELAY REQUIREMENTS DIFFERENTIATION

In this section (Environment 2) the impact of having data sources with different delay requirements is evaluated on the two algorithms that are delay-aware: M-LWDF and ER. Delay requirements have been distributed among users into four groups (T_i={1,2,3,4} symbols) while the desired probability to fulfil the delay requirement is δ_i=0.1

Performance results for M-LWDF and ER are illustrated in Figure 6. Delay differentiation is clearly observed in four sets of curves in the average delay for both cases. The delay which is accomplished the 90% of time is also shown. Although system efficiency is similar for both algorithms, ER presents a fairer delay distribution among users than M-LWDF thanks to its exponential term.

C. SOURCE DIFFERENTIATION

When data sources are generated with different mean rates $\bar{S}_i$ (Environment 3), some of the strategies (RR, BC, PF and FT) are not able to deliver all flows a similar delay. On the contrary, the higher $\bar{S}_i$ the sooner it causes infinite delay, as illustrated in Figure 7.

Despite of the rest of algorithms only consider instantaneous delay as a decision indicator (and not throughput), they handle quite efficiently that situation, being again M-LWDF and ER the ones performing better.

D. MULTIUSER AND MULTICHANNEL DIVERSITY GAIN

In this section, we quantitatively discuss the performance gain obtained by multiuser and multichannel diversities, i.e. quantify the increment of efficiency as the number of users and channels increases. The following tests have been performed when increasing the number of users:

a) Keep constant the users per channel ratio (Environment 4). As expected, the higher number of users and channels the higher system efficiency due to multiuser diversity (see Figure 8). Multiuser diversity gain grows faster for low number of users. When the number of user is high, no more diversity gain can be reached as finding a potential transmitter is very likely.

b) Keep constant he number of channels (Environment 5). Figure 9 depicts average delay vs. total offered load relative to the single user capacity. In this case, system efficiency is not improved as the number of users increases in the system. This situation causes the mean delays to be longer as more users share the existing channels (even for the same system occupation). It can be observed that if bounded delay is desired, allowable throughput cannot be increased by multiuser diversity .

c) Keep constant the number of users (Environment 6). In Figure 10 the number of channels is increased. It can be observed that as the number of channels increase, delay diminishes and maximum capacity grows.

E. CHANNELS DEPENDENCE

In this section (Environment 7), the impact of channels correlation on the performance of the different multiplexing algorithms is studied.

Figure 11 shows the delay results for different channel dependence factor (ρ_K). First observe that the system efficiency decreases as the channel correlation increases. That is as expected, since the more independent the channels are for the same user, the higher multichannel diversity gain. Note that when ρ_K=1 (equivalent to a single channel scheme), although there is no multichannel diversity at all, still multiuser diversity might lead to a system efficiency higher than 1.

F. SHADOWING EFFECT

In order to model the shadowing effect (Environment 8), the received SNR average for the different users has been computed as a lognormal distribution with standard deviation 6dB.

The impact of shadowing in the algorithms' performance is illustrated in Figure 12.

In the RR case, as the multiplexing algorithm simply picks the following user in a cyclic order, those users with lower SNR will use lower modulation levels in average, leading to higher delays, although service time is equal for all of them.

BC results show that users with a better link are selected more often at the expense of lower SNR users. This effect is eased by PF since it also takes into account the average quality of the channel via the $1/\overline{r}_i$ term, although its general performance is still quite poor and unfair.

It is also interesting to highlight the difference between M-LWDF and ER algorithms in a shadowing environment. Although the maximum capacity looks similar in both cases, ER presents a better fairness among users in terms of delay.

5. CONCLUSIONS

The variable-rate wireless channel capacity allows the possibility of introducing new strategies for an efficient resource management while maximizing the QoS.

The performance of a set of CSI and QoS-aware multiplexing algorithms over a multiuser and multichannel environment has been presented. Specifically, this work has studied the delay distribution provided by the different multiplexing strategies. The offered load and the maximum number of admissible users treated in a fair way have been compared for distinct user multiplexing mechanisms.

Channel-adaptive schedulers avoid sending data over a channel in bad conditions. It has been proved that general performance is better for those techniques taking into account both the channel conditions and the service state of the data sources.

Simulation results have shown that both the M-LWDF and ER algorithms tend to outperform the rest of strategies in terms of system efficiency ($\overline{S}/C_{su} \approx 1.5$ in the simulated conditions). When introducing QoS differentiation or shadowing effect in the system, ER gives a slightly fairer treatment than M-LWDF.

Low channels dependence provides the multiplexor a higher flexibility to assign best channels. When channel dependence grows the number of equivalent independent channels is reduced, leading to a decrease in the maximum capacity of the system.

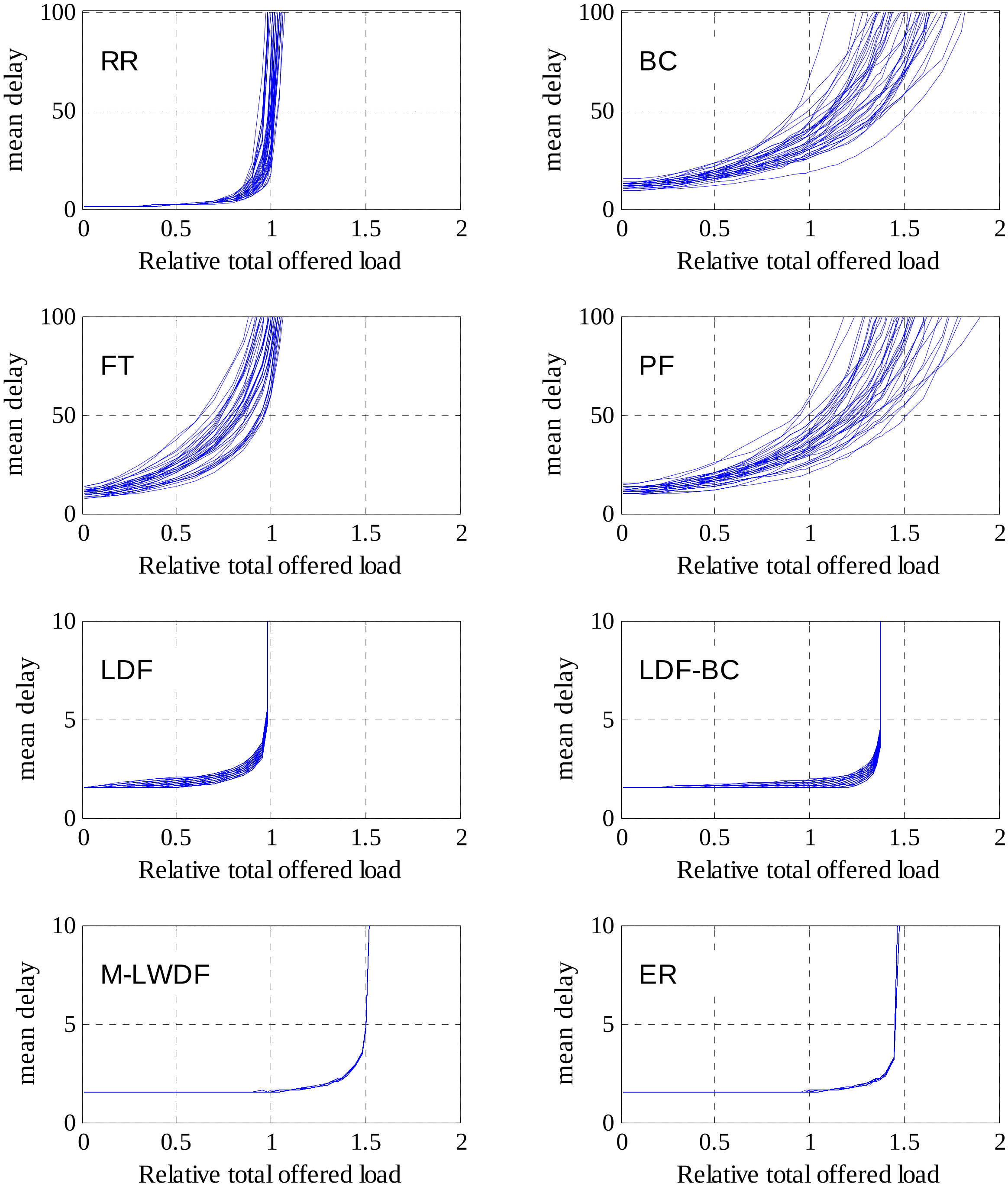

Figure 5. Mean delay vs. total offered load relative to the single user capacity, $\bar{S}/\bar{C}_{su}$. All users have equal average quality, offered load and QoS requirements.

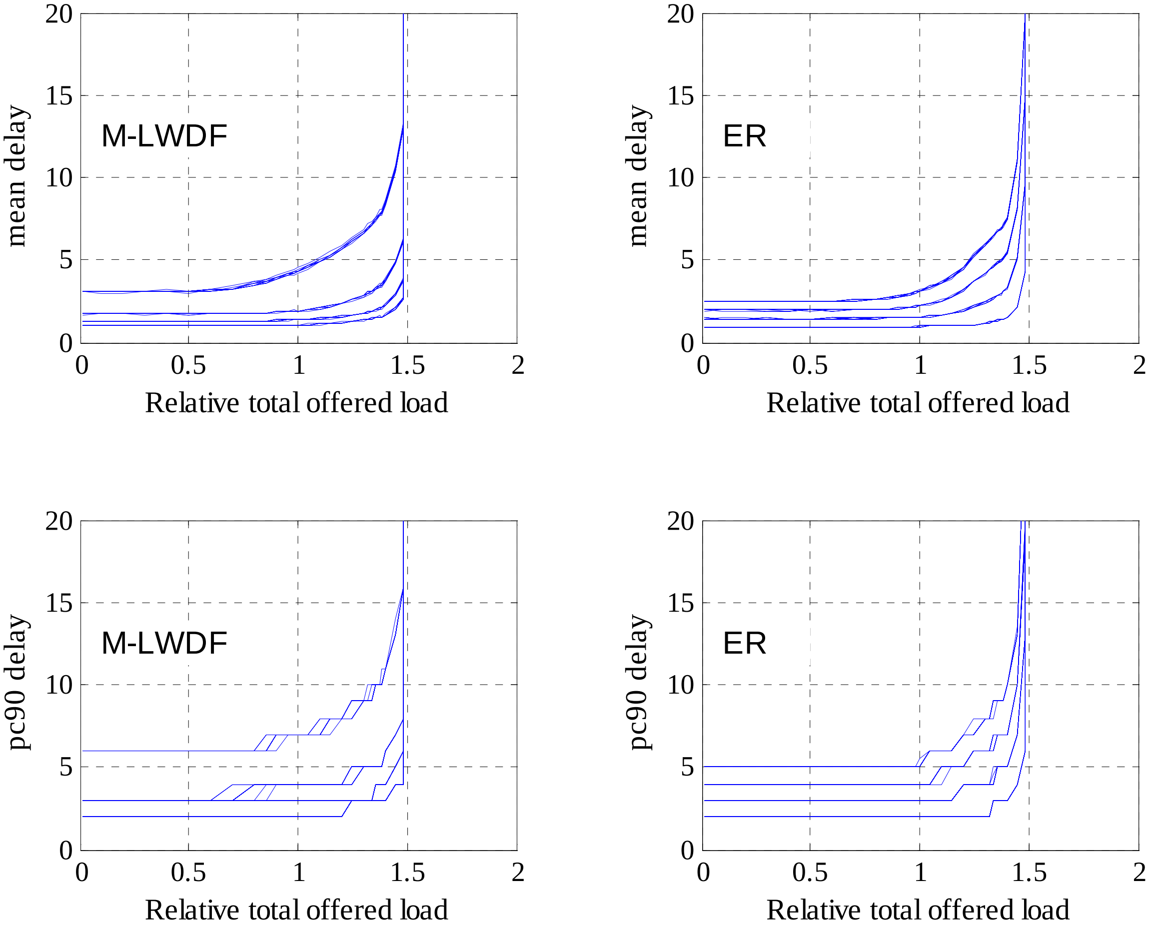

Figure 6. Average delay and delay which is kept 90% of time vs. total offered load relative to the single user capacity, $\overline{S}/\overline{C}_{su}$. All users have equal average quality and offered load. Four groups of QoS requirements are defined T_s={1,2,3,4}.

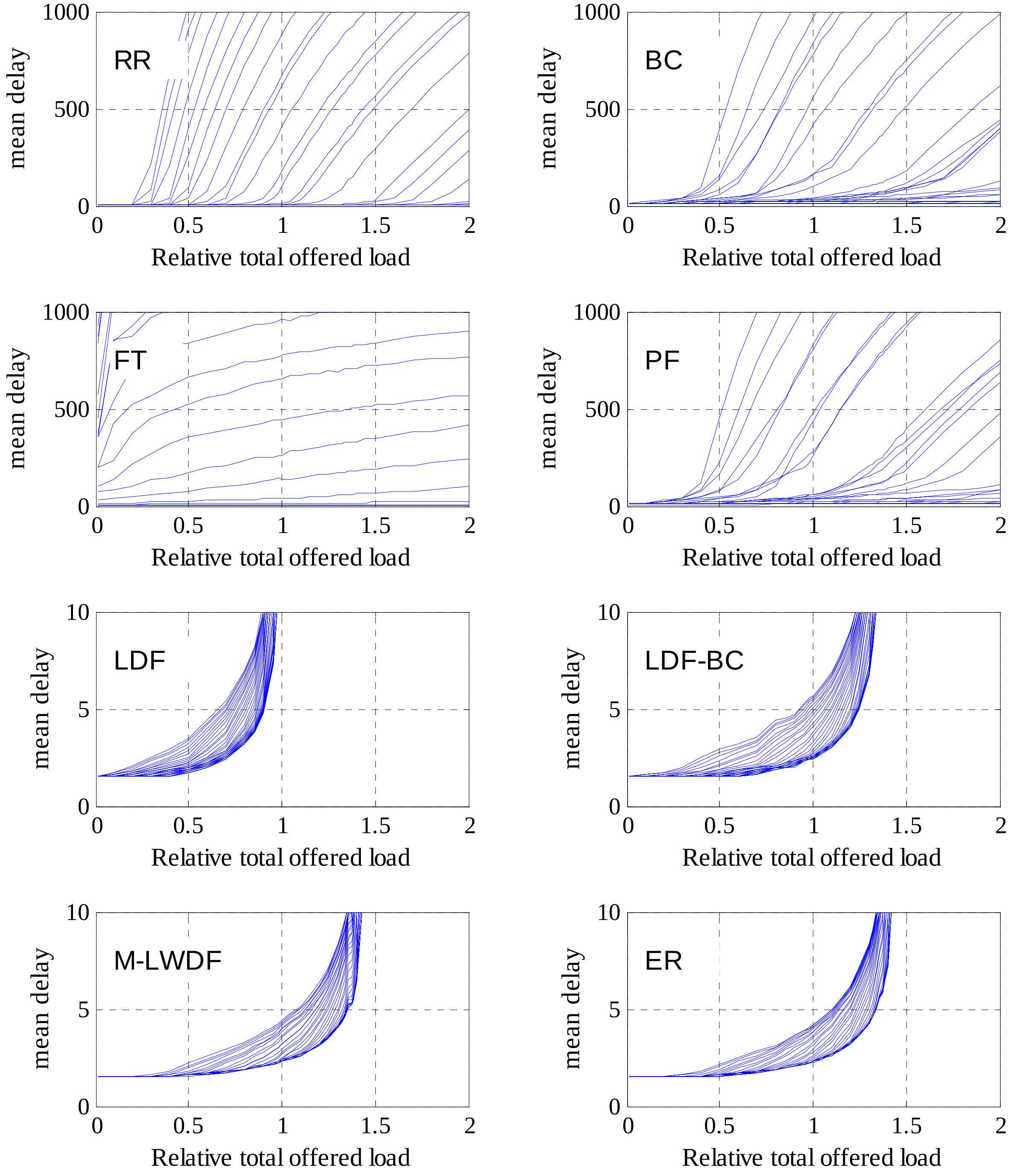

Figure 7. Average delay vs. total offered load relative to the single user capacity, $\overline{S}/\overline{C}_{su}$. Maximum offered load is 50 times minimum information rate.

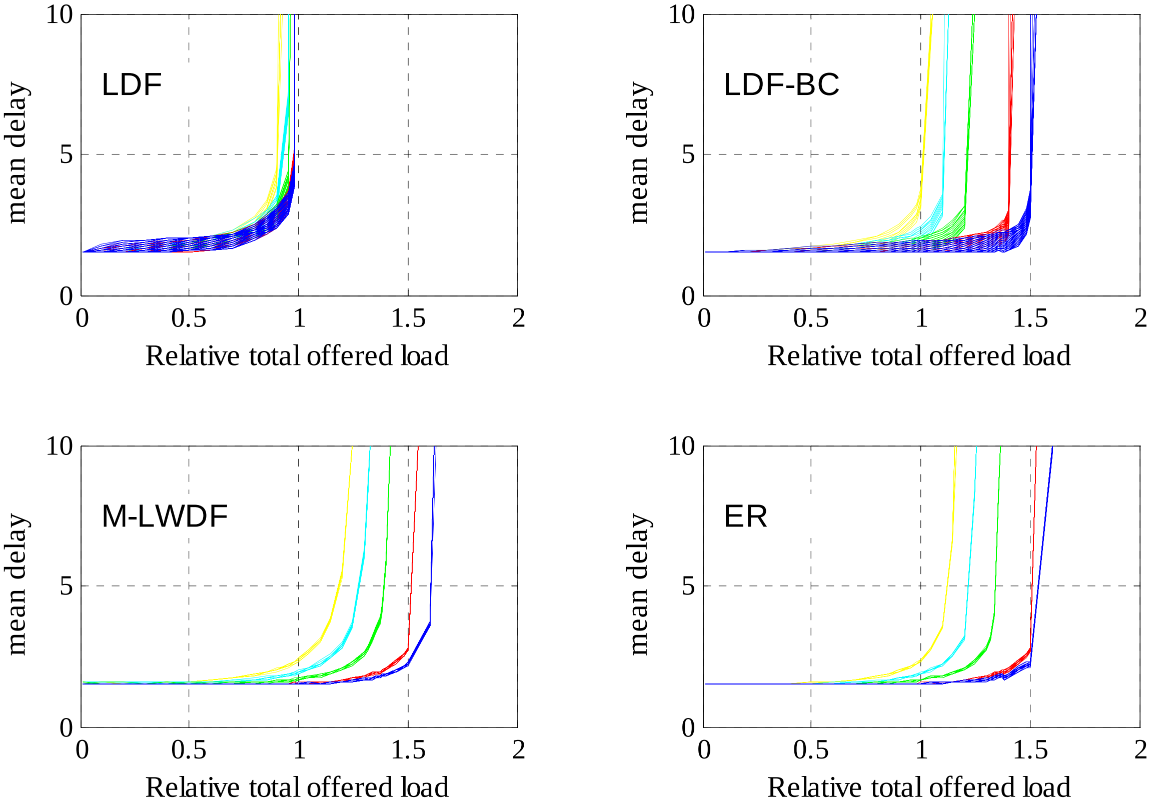

Figure 8. Average delay vs. total offered load relative to the single user capacity, $\overline{S}/\overline{C}_{su}$. Simulations have been carried out for 8 (yellow), 12 (cyan), 20 (green), 60 (red) and 100 (blue) users. The number of channels is a quarter of the number of users.

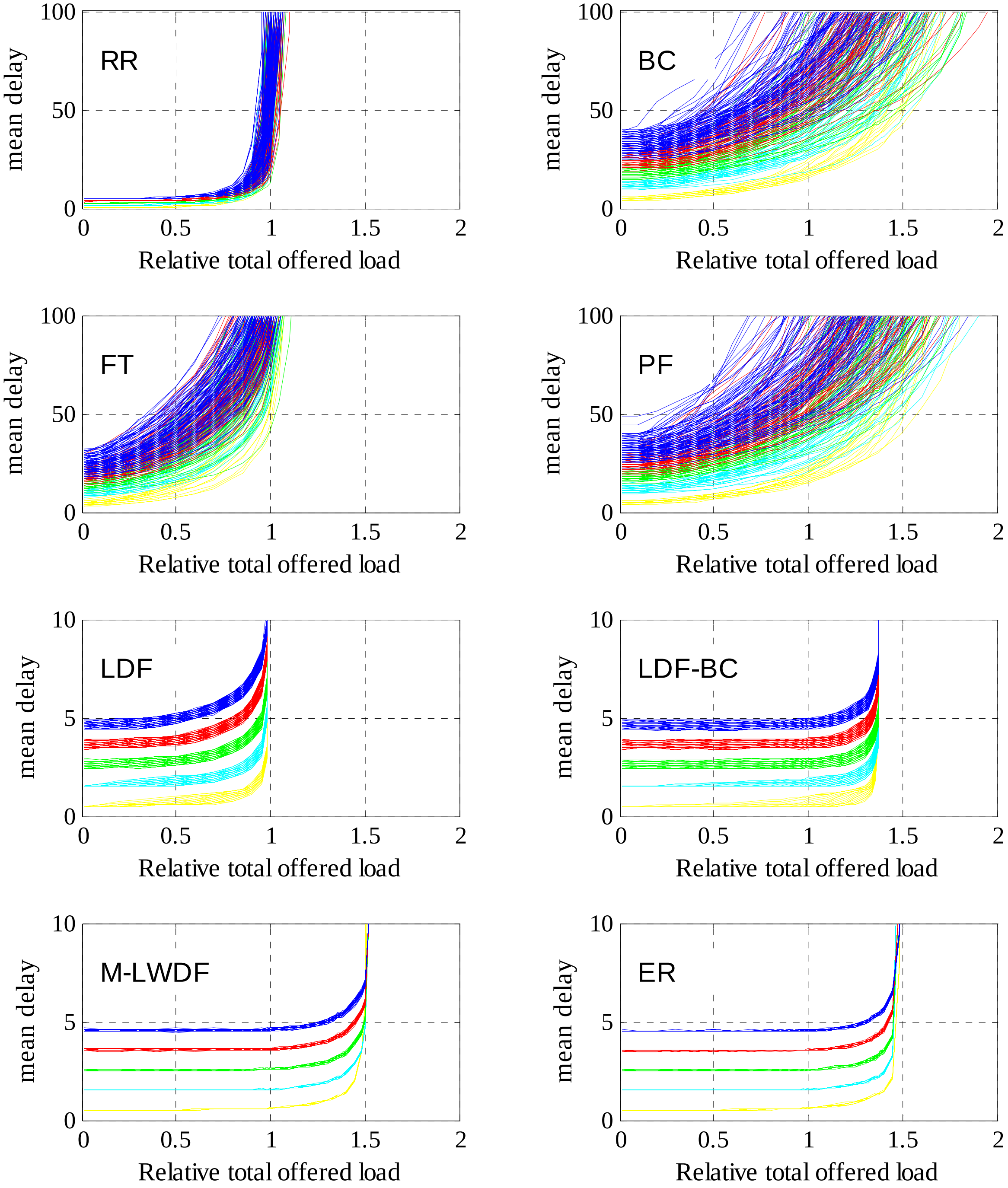

Figure 9. Average delay vs. total offered load relative to the single user capacity, $\overline{S}/\overline{C}_{su}$. Simulations have been carried out for 20 (yellow), 40 (cyan), 60 (green), 80 (red) and 100 (blue) users and 10 channels.

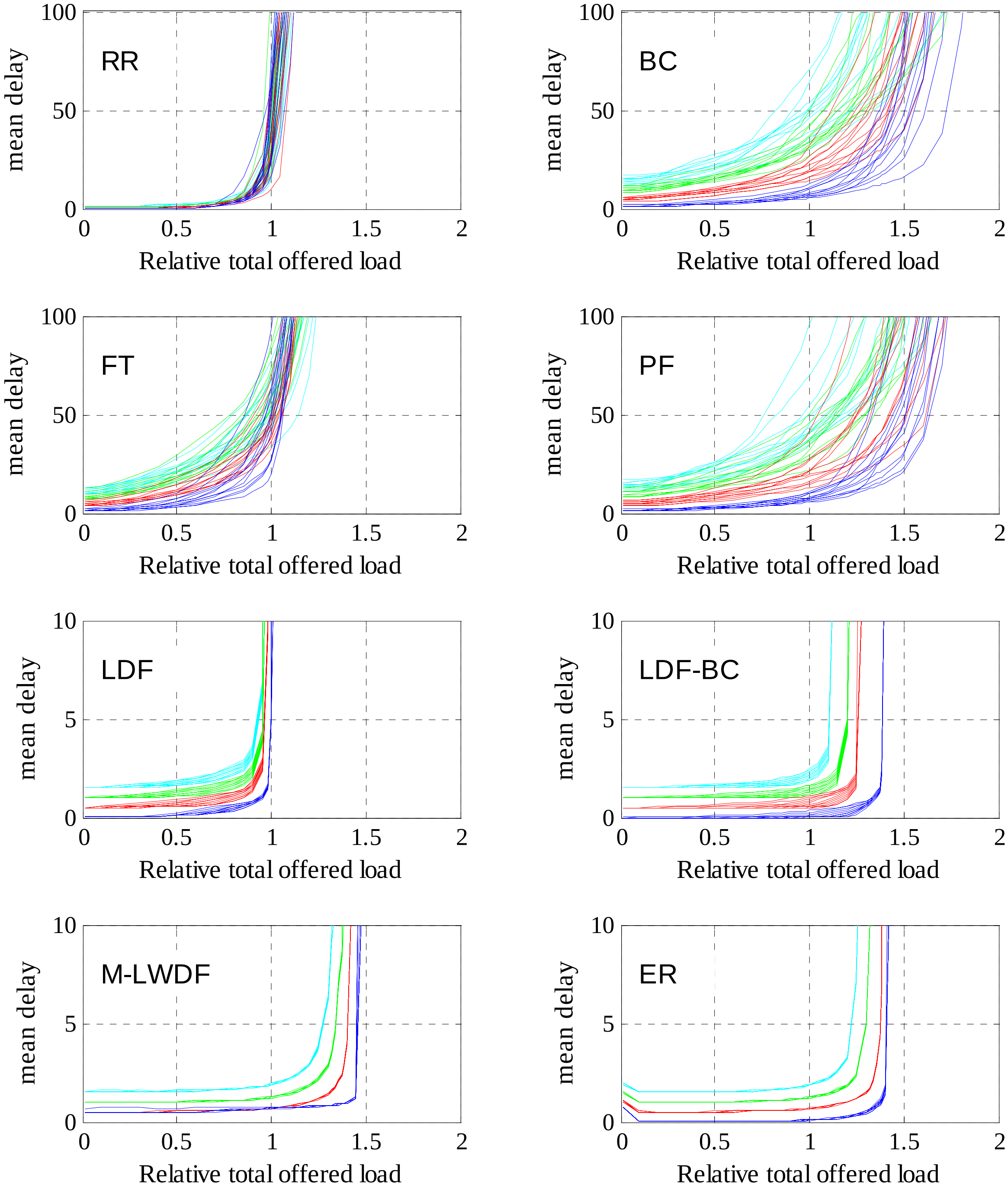

Figure 10. Average delay vs. total offered load relative to the single user capacity, $\overline{S}/\overline{C}_{su}$. Simulations have been carried out for 3 (cyan), 4 (green), 6 (red) and 12 (blue) channels and 12 users.

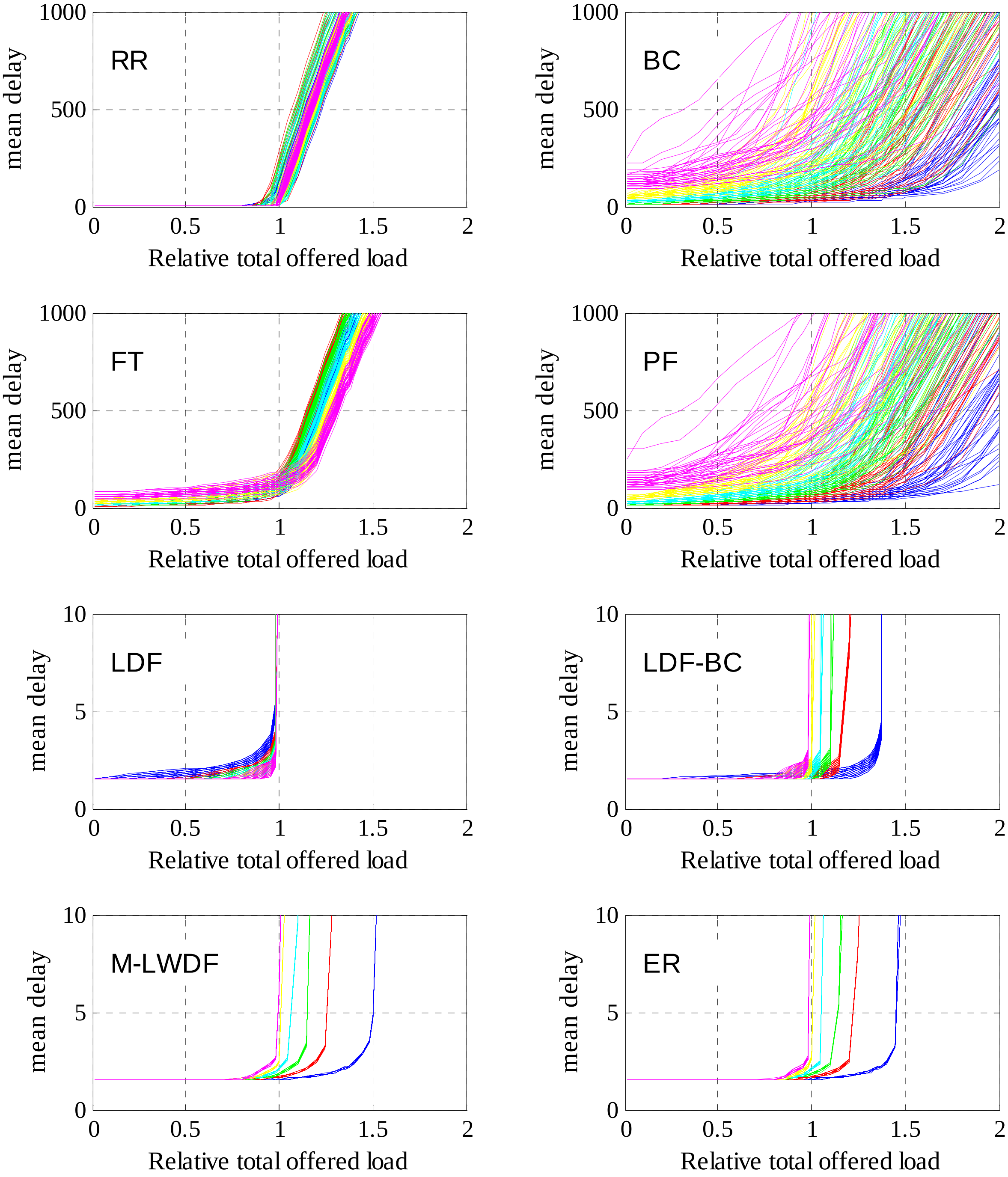

Figure 11. Average delay vs. total offered load relative to the single user capacity, $\overline{S}/\overline{C}_{su}$. Simulations have been carried out for ρ_K = 0 (blue), 0.2 (red), 0.4 (green), 0.6 (cyan), 0.8 (yellow) and 1 (magenta).

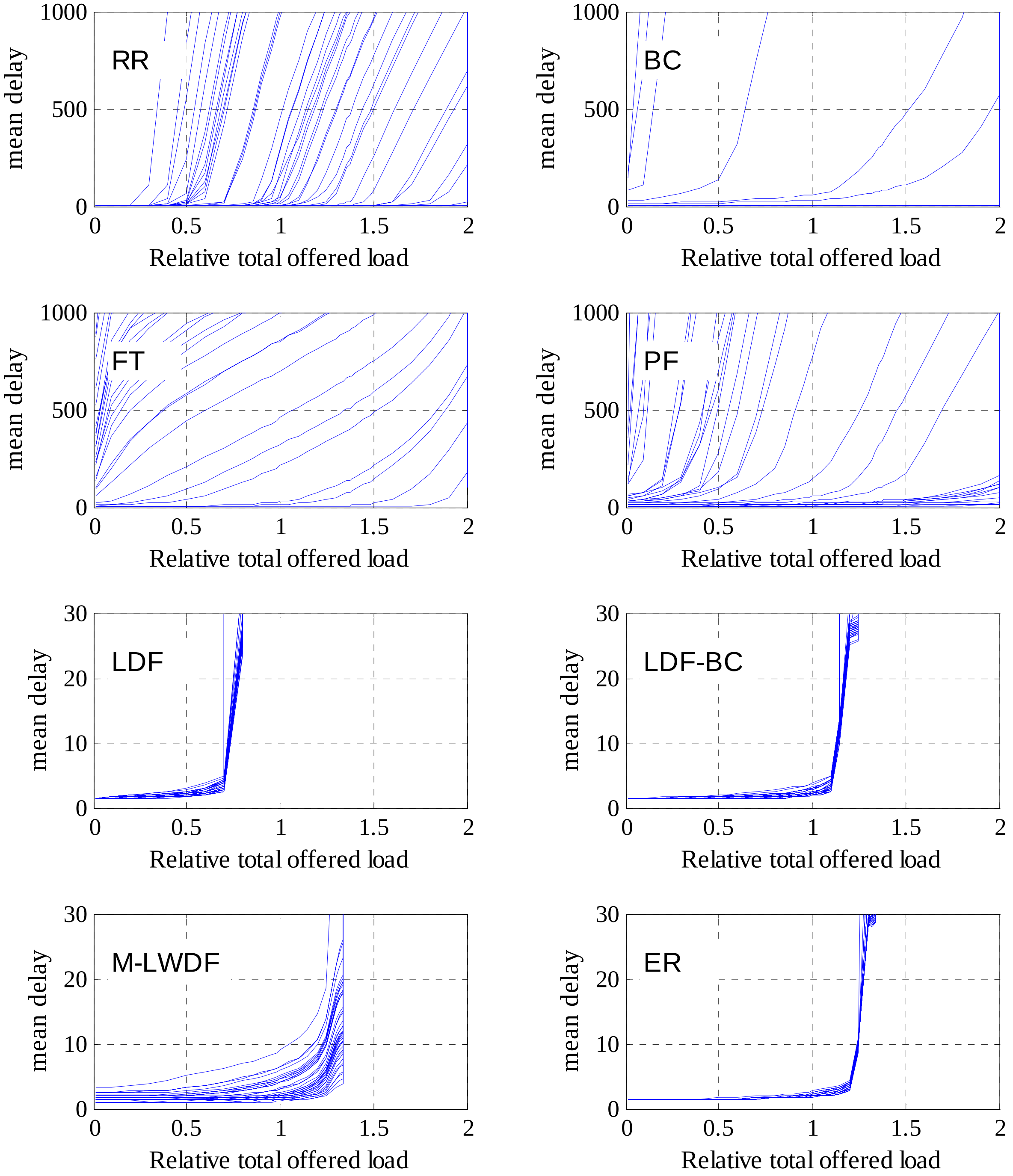

Figure 12. Average delay vs. total offered load relative to the single user capacity, $\overline{S}/\overline{C}_{su}$. Average SNR is a random variable with σ = 6dB

REFERENCES

[3GPP, 2001] "QoS concepts and architecture (Release 5)", 3GPP, 3rd generation partnership project, Technical specification group services and service aspects, TS 23.107, January 2001

[Andrews, 2002] "CDMA data QoS scheduling on the forward link with variable channel conditions", M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, P. Whiting, Bell Labs Technical Memorandum, April 2002

[Berry, 2004] "Cross-layer wireless resource allocation", R.A. Berry, E.M. Yeh, IEEE Signal Procesing Magazine, September 2004, pp. 59-68

[Cao, 2001] "Scheduling algorithms in broad-band wireless networks", Y. Cao, V.O.K. Li, Proceedings of IEEE, vol. 89, nº 1, January 2001, pp. 76-87

[Chen, 2002] "Scheduling policies for joint optimization of DLC and physical layer in mobile communication systems", B. Chen, R. Grünheid, H. Rohling, Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, vol. 2, September 2002, pp. 614-618

[Chen, 2001] "Framework for combined optimization of DLC and physical layer in mobile OFDM systems", B. Chen, F. Fitzek, J. Gross, R. Grünhei, H. Rohling, A. Wolisz, International OFDM workshop, September 2001

[Chow, 1995] "A practical multitone transceiver loading algorithm for data transmission over spectrally shapped channels", P.S. Chow, J.M. Cioffi, J.A.C. Bingham, IEEE Transactions on Communications, vol. 43, nº 2, February 1995, pp. 773-775

[Chung, 2001] "Degrees of freedom in adaptive modulation:-unified view", S.T. Chung, A.J. Goldsmith, IEEE Transactions on Communications, vol. 49, nº 9, September 2001, pp. 1561-1571

[Damji, 2004] "Adaptive downlink multi-carrier resource allocation for real-time multimedia traffic in cellular systems", N. Damji, T. Le-Ngoc, Proceedings of IEEE International Conference on Communications, vol. 7, June 2004, pp. 4258-4262

[Ericsson, 2004] "Revenue maximization in resource allocation: Applications in wireless communication networks", N.C. Ericsson, PhD Thesis, Uppsala University, September 2004

[Glisic, 2005] "Advanced wireless communications. 4G technologies", S.G. Glisic, Wiley & Sons, 2005

[Goldsmith, 1997] "The capacity of downlink fading channels with variable rate and power", A.J. Goldsmith, IEEE Transactions on Vehicular Technology, vol. 46, nº 3, August 1997, pp. 569-580

[Guo, 2002] "Class-based quality of service over air interface in 4G mobile networks", Y. Guo, H. Chaskar, IEEE Communications Magazine, March 2002, pp. 132-137

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

[Jha, 2002] "Engineering Internet QoS", S. Jha, M. Hassan, Artech House Publishers, , 2002

[Johnsson, 2005] "An adaptive cross-layer scheduler for improved QoS support of multiclass data services on wireless systems", K.B. Johnsson, D.C. Cox, IEEE Journal on Selected Areas in Communications, vol. 23, nº 2, February 2005, pp. 334-343

[Keller, 2000] "Adaptive multicarrier modulation:-convenient framework for time-frequency processing in wireless communications", T. Keller, L. Hanzo, Proceedings of IEEE, vol. 88, nº 5, May 2000, pp. 611-640

[Knopp, 1995] "Information capacity and power control in single-cell multiuser communications", R. Knopp, P.A. Humblet, Proceedings of IEEE International Conference on Communications, June, 1995

[Kulkarni, 2003] "Opportunistic scheduling policies for wireless systems with short term fairness constraints", S.S. Kulkarni, C. Rosenberg, Proceedings of IEEE Global Telecommunications Conference, vol. 1, 2003, pp. 533-537

[Kwok, 2003] "On channel adaptive multiple access control without contention queue for wireless multimedia services", Y.K. Kwok, V.K.N. Lau, Wireless Networks, July 2003, pp. 379

[Lee, 2004] "Optimal time slot allocation for multi-user wireless networks", N.H. Lee, K.C. Hwang, S. Bahk, K.B. Lee, Proceedings of IEEE Wireless Communications and Networking Conference, March 2004, pp. 2256-2261

[Ling, 2003] "Adaptive transmission with discrete code rates and power levels", L. Ling, R.D. Yates, P. Spasojevic, IEEE Transactions on Communications, vol. 51, nº 12, December 2003, pp. 2115-2125

[Rhee, 2001] "A wireless fair scheduling algorithm for 1xEV-DO systems", J.H. Rhee, T.H. Kim, D.K. Kim, Proceedings of IEEE Vehicular Technology Conference, vol. 2, October 2001

[Song, 2002] "Utility-based joint physical-MAC layer optimization in OFDM", G. Song, Y. Li, Proceedings of IEEE Global Telecommunications Conference, November 2002, pp. 671-675

[Song, 2005] "Cross-layer optimization for OFDM wireless networks – Part I: Theoretical framework", G. Song, Y. Li, IEEE Transactions on Wireless Communications, vol. 4, nº 2, March 2005, pp. 614-624

[Svedman, 2004] "A QoS aware proportional proportional fair scheduler for opportunistic OFDM", P. Svedman, S.K. Wilson, B. Ottersten, Proceedings of IEEE Vehicular Technology Conference, vol. 1, September 2004, pp. 558-562

[Wong, 1999] "Multiuser OFDM with adaptive subcarrier, bit and power allocation", C.Y. Wong, R.S. Cheng, K.B. Letaief, R.D. Murch, IEEE Journal on Selected Areas in Communications, vol. 17, nº 10, October 1999, pp. 1747-1758

[Yang, 2004] "Adaptive rate DS-CDMA systems using variable spreading factors", L.L. Yang, L. Hanzo, IEEE Transactions on Vehicular Technology, vol. 53, nº 1, January 2004, pp. 72-81

[Zhang, 2004] "Adaptive resource allocation and scheduling for multiuser packet-based OFDM networks", Y.J. Zhang, K.B. Letaief, Proceedings of IEEE International Conference on Communications, vol. 5, June 2004, pp. 2949-2953