DOI : 10.1109/COMNET.2009.5373549

Account for fading in the dynamic performance evaluation of OFDMA cellular networks

Bartłomiej Błaszczyszyn
INRIA & ENS
45 rue d'Ulm, 75005 Paris France
E-mail: Bartek.Blaszczyszyn@ens.fr

Mohamed Kadhem Karray
France Telecom, R& D Division
38/40 rue Général Leclerc, 92794 Issy-Moulineaux France
E-mail: mohamed.karray@orange-ftgroup.com

***Abstract*—The objective of the present paper is to build a model which permits to capture and analyze the principal impacts of *fading* and *multiuser diversity gain* on the dynamic performance of an OFDMA cellular network.**

To this end, assuming Markovian arrivals and departures of customers that transmit some given data-volumes, as well as some temporal channel variability (fading), we study the mean *throughput* (and *delay*) that the network offers to users in the long-term evolution of the system. Explicit formulas are obtained in the case of *allocation policies*, which may or may-not take advantage of the fading, called respectively *opportunistic* and *non-opportunistic*.

The main practical results of the present work are the following. Firstly we evaluate for the non-opportunistic allocation the *degradation due to fading* compared to AWGN (that is, a decrease of at least 13% of the throughput). Secondly, we evaluate the *gain induced by the opportunistic* allocation. In particular, when the traffic demand per cell exceeds some value (about 2.5 Mbps in our example), the gain induced by opportunism compensates the degradation induced by fading compared to AWGN.

***Index Terms*—Radio communication, Communication system performance, Dynamics, Information theory, Markov processes.**

I. INTRODUCTION

Cellular networks provide streaming services (i.e. real-time such as voice calls, video streaming, etc.), which require predefined transmission rates, and carry elastic traffic (data), which accepts fluctuations of the rates. Users arrive to the network and require some service. Once they are admitted, they are allocated some *resource* (power, bandwidth, code) for the duration of their service, then they depart from the network. Besides the *dynamics* due to user arrivals and departures, the channel conditions vary due to *fading*.

Growing traffic requires better planning and/or dimensioning of the networks. This task can be substantially simplified by analytical study of the dynamic network performance.

The objective of the present paper is to *account for the effect of fading in the dynamic performance evaluation* of the downlink of an OFDMA cellular network.

To this aim, the following two groups of elements are crucial. On one hand, the network geometry and the *resource allocation*, which can be chosen by the network designer. On the other hand, *channel conditions* and the user traffic (intensity of arrivals and requested service) can only be predicted by the designer. In order to optimize the network performance, the resource allocation should be adapted to the channel and traffic conditions. Besides better coding (up to the *information theory*'s limit) one is looking for allocations that are opportunistic; i.e., take advantage of the actual channel state and user position, preserving some fairness of the mean service rates offered to users. More precisely, the *opportunistic* allocation assumes that at each time, each bandwidth portion is allocated to the user with the most favorable fading state. Once the resource allocation is given, the QoS (blocking probability for streaming traffic and delay for elastic traffic) perceived by the users may be evaluated using *queueing theory*.

Individual elements of the above puzzle (i.e. information theory, resource allocation and queueing theory) are often studied and optimized separately. The main contribution of this paper is a global approach that combines these elements to deduce the performance of the non-opportunistic and the opportunistic resource allocations. In doing so we systematically use a separation of the times scales of different elements of the network dynamics, described in Section II-A.

The main practical results of the present work are the following. Firstly we evaluate for the non-opportunistic allocation the *degradation due to fading* compared to AWGN. Secondly, we evaluate the *gain induced by the opportunistic* allocation.

The remaining part of this paper is organized as follows. In Section I-A we list these existing results in the literature that are used in our global approach. Our model of the OFDMA network is presented in Section II. In Section II-A we describe the network dynamics and its decomposition into three time scales. They correspond respectively to information theory, resource allocation and queueing theory which are studied in Sections III, IV and V respectively. The numerical results are described in Section VI. Finally Section VIII describes some interesting perspectives.

A. Related Work

A growing interest in OFDMA has resulted in many publications on performance evaluation of such networks. It is not in the scope of this paper to make a thorough review of this literature. For some references on this field see for example those cited in [4]. Here we remark only those results that are directly related to our approach.

Our starting point is our work in [4] assuming AWGN channels between the users and the base stations. As regard to the performance of an opportunistic allocation policy over fading, we extend the approach of [5] and [14] for HSDPA to OFDMA networks. Moreover we use the known results for the stability and stationary distribution of the multi-class extension of the processor sharing queue; see e.g. [8], [16].

The above results are used as building blocks to evaluate the dynamic performance of the non-opportunistic and the opportunistic resource allocations in OFDMA multi-cell networks.

II. MODEL ASSUMPTIONS

We will consider a wireless network composed of several base stations (BS). Each BS is equipped with a single antenna (no MIMO) and its total power is limited to some given maximal value. The same frequency spectrum is available to all BS (frequency reuse factor equal to one). Each BS allocates disjoint sub-carriers to its users without macrodiversity (each user is served by exactly one BS). Thus, any given user receives only other-BS interference that is the sum of powers emitted by other BS on the sub-carriers allocated to him by his BS.

When the radio signal propagates (from a BS to a given user) it is attenuated. We model the propagation-loss as the product of three factors called distance-loss, shadowing and fading. The first factor is due to the *distance* between the transmitter and the receiver. The *shadowing* is due attenuation by the obstacles between the transmitter and the receiver. The *fading* is due to reflection on the obstacles in the neighborhood of the receiver which generated multiple paths. We account only for distance and fading effects in the present study.

We assume that the bandwidth of each sub-carrier is smaller than the *coherence frequency* of the channel, so that we can consider that the *fading* in each sub-carrier is *flat*. That is, the output of the channel at a given time depends on the input only at the same instant of time. We don't make any assumption on the *correlation* of the fading processes corresponding to different subcarriers (for a given user and a given BS). However, the fading processes for different users or base stations are assumed independent.

Time is divided into time-slots of length smaller than the *coherence time* of the channel, so that, for a given sub-carrier, the fading remains *constant during each time-slot* and the fading process in different time-slots may be assumed *ergodic*. (Such model for fading generalizes the so-called *quasi-static* model where the fading process at different time-slots is assumed to independent and identically distributed.)

The fading in each time-slot and each sub-carrier is a *Rayleigh* distributed random variable.

The codeword duration equals the time-slot, which is assumed sufficiently large so that the *capacity* within each time-slot may be defined in the *asymptotic* sense of the information theory.

Users perform *single user detection*; thus the interference is considered as additive noise and added to the additive white Gaussian noise (AWGN). The statistical properties of the interference are not known a priory since they depend of the codings of the other users. Moreover the signals transmitted by different base stations are assumed independent.

For each sub-carrier, and each time-slot, efficient (e.g. turbo) codes are used to obtain bit-rates close to the capacity given by information theory (i.e. the maximal rate for which there exist coding schemes with error probability vanishing when the length of the code is sufficiently large). In particular, for the AWGN channel the capacity equals

$$C = w \log_2(1 + \mathrm{SNR}) \qquad (1)$$

where w is the bandwidth and SNR designates the signal to noise power ratio.

A. Separation of time scales

We assume that there are a few time scales each associated to the evolution of some stochastic process. Roughly speaking, we assume that the duration of each time scale is sufficiently large so that the corresponding process converges to its stationary state; and sufficiently small so that the processes of the larger time scales don't evolve within this duration.

We describe now the times scales from the fastest to the slowest one.

The fastest time scale corresponds to *information theory*. The user number and positions as well as powers and bandwidths allocated to them are assumed fixed. Then the bit-rate of each user can not exceed the capacity given by information theory (i.e. the maximal rate for which there exist coding schemes with error probability vanishing when the length of the code is sufficiently large). This *information theory constraint* (i.e. existence of a coding scheme corresponding to the bit-rate) together with the maximal power and bandwidth constraints constitute the *resource allocation constraints*.

The intermediate time scale corresponds to the *resource allocation* problem. The user number and positions are assumed fixed. The network attempts to allocate powers, bandwidths and bit-rates (coding schemes) to all users respecting the resource allocation constraints.

This problem may be reformulated in terms of some condition on the bit-rates. More precisely, by *feasibility condition* we mean a condition regarding bit-rates of different users which guarantees the existence of powers and bandwidths satisfying the resource allocation constraints.

It is the role of the *load control* to guarantee this condition. The load control comprises the *admission control* (for streaming traffic) which checks at the arrival instant of each new user the *feasibility condition* and decides whether it can be admitted or not; and the *congestion control* (for elastic traffic) which may modify the user bit-rates to satisfy the *feasibility condition*. Once this is done, the network allocates powers and bandwidths to all users supporting the (feasible) bit-rates.

The slowest time scale corresponds to *queueing theory*, where we consider the stochastic process of user positions which is driven by *call arrivals and durations or data volumes*. This process is subject at each time the feasibility condition described above. The stationary state (or distribution) of this process permits to calculate the quality of service perceived by the users.

B. Power allocation

The network may operate a power adaptation. Such power adaptation may be either a *power control* to transmit just what is necessary to compensate the interference as described in [4, §III.C] or a *water filling* attempting to maximize the cell capacity by allocating the powers among the users according to their fadings as described in [15, Theorem 2 p.119].

We shall assume that the signal transmitted by each BS has a *given power* constant over time (which equals to the maximal authorized value) and a *constant power spectral density* (i.e. the power is the same for all sub-carriers). The performance obtained with this assumption gives a *lower bound* of the performance of a network which operates a power adaptation. We shall carry our analysis consider further lower bounds. The dimensioning based on such lower bounds is conservative.

C. Resource allocation

We always assume that each *user (receiver) knows* its own fading state; a situation called CSIR (Channel State Information at Receiver) in literature.

We shall consider two resource allocation schemes:

1) *Non-opportunistic*: The BS doesn't take into account the fading when allocating the sub-carriers to the users. This may be due to the fact that the BS *(transmitter) doesn't know* the fading states of its users.
2) *Opportunistic*: The BS allocates takes into account the fading when allocating the sub-carriers to the users. This requires that the BS *(transmitter) knows* the fading states

of its users; a situation called CSIT (Channel State Information at Transmitter) in literature. The gain obtained by the opportunistic allocation is called the *multiuser diversity* gain.

D. Possible model extensions

Even though our model seems to have many simplifying assumptions, we will show that it permits to capture and analyze the principal impacts of fading and multiuser diversity gain in cellular OFDMA networks. Moreover the following remarks show how to extend the model.

In [12] it is observed that for OFDMA systems implementing a family of M-QAM modulations (as those described in [10]) with some BER target, the AWGN capacity formula (1) should be replaced by

$$C = w\log_2\left(1+\frac{\mathrm{SNR}}{\Gamma}\right)$$

where $\Gamma = -\ln(5\times\mathrm{BER})/1.5$ (which is larger than 1 for $\mathrm{BER} < e^{-1.5}/5 \simeq 0.0446$). Thus accounting for real coding schemes may be tentatively taken into account in our approach by an appropriate modification of the AWGN capacity formula. (For a similar question for HSDPA networks, it is observed in [13, Figure 11.1 p.175] that HSDPA coding offers a capacity which is approximately the third of the AWGN capacity.)

The assumption that the fading is constant over a time slot may be replaced by the less restrictive assumption that the fading is ergodic and perfectly known by the receiver (see [6]).

In what follows we analyze the three time scales of our model described in Section II-A.

III. Information theory

We assume that the users don't move at the considered time scale. Consider a given user served by a given BS u. For each sub-carrier, and each time-slot, say $[0,T]$, the channel output $Y(t)$ is related to the transmitted signal $X(t)$ by

$$Y(t) = S_u\times X(t)+Z(t)+I(t),\quad t\in[0,T]$$

where S_u represent the fading assumed constant, $Z(t)$ designate the noise assumed AWGN with power spectral density N_0 and $I(t)$ is the interference. (We skip for the moment the propagation loss induced by the distance.)

Lets denote by p be the power in a given sub-carrier, that is

$$E\left[|X(t)|^2\right]=p$$

For each interfering BS v, let $X_v(t)$ be its transmitted signal and S_v be the fading assumed constant. Then the interference equals to

$$I(t)=\sum_{v\neq u}S_vX_v(t)$$

Since the signals transmitted by different base stations are assumed independent (and centred),

$$E\left[|I(t)|^2\right]=\sum_{v\neq u}|S_v|^2E\left[|X_v(t)|^2\right]=p\sum_{v\neq u}|S_v|^2$$

Using [17, Theorem 18], we may show that the worst noise process distribution (not necessarily white nor Gaussian) for capacity with given second moment, is the AWGN. We deduce that the capacity C within the considered time-slot is lower bounded by

$$C\geq w\log_2\left(1+\frac{p|S_u|^2}{wN_0+p\sum_{v\neq u}|S_v|^2}\right)$$

where w is the bandwidth of a sub-carrier.

Remark 1: An alternative way to show the above inequality is to use [1, Lemma 1 p.30] giving the capacity of the interference channel. It is frequent to see in Literature the equality sign in place of the inequality in the above display, as for example [15, p.118]. We don't see why the equality may be justified in this generality (see [2, §4 p.811] for a similar question).

We will now introduce propagation loss L_v induced by the distance between BS v and the given user. In order to account for these losses, the above formula should be replaced by

$$C\geq w\log_2\left(1+\frac{p|S_u|^2/L_u}{wN_0+p\sum_{v\neq u}|S_v|^2/L_v}\right)\tag{2}$$

Since the number of interfering BS is large and since the fading processes for different BS are assumed independent, it is reasonable (by the law of large numbers) to assume that $\sum_{v\neq u}|S_v|^2/L_v$ (where the $|S_v|^2$ are realizations of i.i.d. random variables with expectation 1) is approximately $\sum_{v\neq u}1/L_v$. Thus we get

$$C\geq w\log_2\left(1+\beta|S_u|^2\right)\tag{3}$$

where

$$\beta=\frac{p/L_u}{wN_0+p\sum_{v\neq u}1/L_v}\tag{4}$$

which may be viewed as the *Signal to Interference and Noise Ratio* (SINR) in our model.

IV. Resource allocation

A. Non-opportunistic allocation

We consider a given sub-carrier and multiple time-slots. Recall that the fading for different time-slots are i.i.d. random variables. We assume that each receiver knows its own fading state. Then using (3) and the ergodicity of the fading process, we deduce that the capacity averaged over a large number of time-slots is lower bounded by

$$\bar{C}\geq wE\left[\log_2\left(1+\beta|S_u|^2\right)\right]\tag{5}$$

where the expectation is over the fading random variable S_u and β is given by (4). As observed in [7, §I], the above formula, called *ergodic capacity*, is suitable for the analysis of the performance of *elastic bit-rate* traffic. Averaging over a large number of time-slots corresponds to exploiting the so-called *time-diversity*.

Remark 2: The above bound may be derived directly from (2). Indeed, the function $x\to\log(1+1/x)$ is convex (its second derivative is $\frac{2x+1}{x^2(x+1)^2}$), and by Jensen's inequality

$$E\left[\log_2\left(1+\frac{p|S_u|^2/L_u}{wN_0+p\sum_{v\neq u}|S_v|^2/L_v}\right)\right]$$
$$\geq wE\left[\log_2\left(1+\beta|S_u|^2\right)\right]$$

Remark 3: Note that the formula

$$E\left[\log_2\left(1+\beta|S_u|^2\right)\right]$$

is similar to the capacity of a channel with ergodic fading in [6] (or i.i.d. fading in [11, Equation (2)]) and CSIR (Channel State Information known at Receiver; i.e. the receiver knows the fading). Nevertheless, the fading varies during a codeword there, whereas in the model considered in the present paper the fading is constant during a codeword.

Since $|S_u|$ is Rayleigh distributed, i.e. it has the probability density function

$$f_{|S_u|}(s)=2se^{-s^2}1_{s\geq 0}$$

then $H = |S_u|^2$ has exponential distribution

$$f_H(h) = e^{-h} 1_{h\geq 0}$$

In this case,

$$E\left[\log_2\left(1+\beta|S_u|^2\right)\right] = \int_0^\infty \log_2(1+\beta h) e^{-h} dh = \frac{1}{\ln 2} e^{\beta^{-1}} \operatorname{Ei}\left(1, \beta^{-1}\right)$$

where

$$\operatorname{Ei}(1,x) = \int_1^\infty \frac{e^{-xt}}{t} dt$$

is the *exponential-integral function*.

Remark 4: FREQUENCY DIVERSITY. We consider now a given time-slot and large number n of sub-carriers. Assume that the fading variables for different sub-carriers are i.i.d. (or more generally the fading process is ergodic with respect to the number n of subcarries). Then again, by the law of large number, the capacity of a large number n of sub-carriers is lower bounded by

$$nwE\left[\log_2\left(1+\beta|S_u|^2\right)\right]$$

Thus the ergodic capacity is also appropriate for *streaming traffic* (where we can't always count on time-diversity) when the number of sub-carriers allocated to each user is large enough.

Remark 5: OUTAGE CAPACITY. Consider now the case of *streaming traffic* when the number of sub-carriers allocated to each user is not large. This is typically the case for voice traffic. Let n be the number of sub-carriers allocated to a given user. We deduce from (3) that the capacity in a given time-slot is bounded by

$$C \geq w\sum_{k=1}^{n}\log_2\left(1+\beta\left|S_u^k\right|^2\right)$$

where S_u^k designate the fading for the k-th sub-carrier ($k \in \{1,\ldots,n\}$).

For streaming traffic, we aim to obtain a bit-rate at each time-slot larger that some fixed value, say r. As observed in [7, §I], the relevant performance indicator in this case, is the so-called *outage probability*, defined as the probability that the capacity in a given time-slot is less than the desired bit-rate r. The objective is then to assure that the outage probability doesn't exceed some maximal value, say P_o. We deduce from the above inequality that it is enough to assure that

$$P\left(w\sum_{k=1}^{n}\log_2\left(1+\beta\left|S_u^k\right|^2\right) < r\right) \leq P_o \qquad (6)$$

For methods to approximate the outage probability see [15, §6.2.4 p.155-159].

1) Feasible bit-rates: Consider a cell u serving some users denoted by $m \in u$. In what follows we will establish a feasibility condition for the bit rates $(r_m)_{m\in u}$.

Let $(w_m)_{m\in u}$ satisfying

$$\sum_{m\in u} w_m \leq W \qquad (7)$$

be given. By (5) we deduce that the following bit-rates are feasible

$$r_m = w_m E\left[\log_2\left(1+\beta_m|S_u|^2\right)\right], \quad m\in u \qquad (8)$$

where w_m is the bandwidth allocated to user m and

$$\beta_m = \frac{p/L_{u,m}}{wN_0 + p\sum_{v\neq u} 1/L_{v,m}} \qquad (9)$$

is the SINR for user m.

With this definition of r_m, the total-bandwidth constraint (7) is equivalent to

$$\sum_{m\in u}\frac{r_m}{E\left[\log_2\left(1+\beta_m|S_u|^2\right)\right]} \leq W$$

or equivalently,

$$\sum_{m\in u} r_m\gamma_m \leq 1 \qquad (10)$$

where

$$\gamma_m = \left[WE\left[\log_2\left(1+\beta_m|S_u|^2\right)\right]\right]^{-1} \qquad (11)$$

Each bit-rate allocation satisfying (10) will be called *non-opportunistic allocation*. Note that the value of the bit-rate do not depend on the current value of the fading states but only on their statistics.

B. Opportunistic allocation

We assume that the BS knows the fading states of all the users in its cell and that each user knows its own fading state.

1) Analysis for a given subcarrier: We account now for the fading variations at the resource allocation time scale. In what follows, by *opportunistic allocation we mean that at a given time-slot, each subcarrier is allocated to the user with the most favorable fading state*. We will calculate the ergodic throughput of each user per time-slot under this policy.

If we allocate a given subcarrier to a user $m \in u$, then we deduce from (3) that, for a given time-slot, by (3), we can find a coding scheme to support the following bit-rate

$$R_m = w\log_2\left(1+\beta_m|S_{u,m}|^2\right), \quad m\in u \qquad (12)$$

where β_m is the SINR given by (9) and $S_{u,m}$ is the fading for the considered user, subcarrier and time-slot. We may write the above equation as

$$R_m = w\phi\left(\beta_m|S_{u,m}|^2\right)$$

where the function ϕ is defined by

$$\phi(x) = \log_2(1+x)$$

Let us denote by R'_m the bit-rate effectively allocated by our opportunistic policy to a given user m averaged over a large number of time-slots. From the temporal ergodicity, we deduce that

$$R'_m = E\left[R_m \times 1\left\{|S_{u,m}| = \max_{n\in u}|S_{u,n}|\right\}\right]$$

Proposition 1: For the opportunistic allocation, the bit-rate averaged over a large number of time-slots is

$$R'_m = w\int_0^{+\infty}\phi(\beta_m x)e^{-x}\left(1-e^{-x}\right)^{M-1}dx \qquad (13)$$

$$= \frac{w}{\beta_m}\sum_{k=0}^{M-1}\binom{M-1}{k}(-1)^k\mathcal{L}_\phi\left(\frac{k+1}{\beta_m}\right) \qquad (14)$$

where M is the number of the users in the cell and $\mathcal{L}_\phi$ is the Laplace transform of ϕ, that is

$$\mathcal{L}_\phi(u) = \int_0^{+\infty}\phi(x)e^{-ux}dx$$

Proof: From (12) and the ergodicity of the fading processes over the time-slots, we deduce that

$$\begin{aligned}
&R'_m\\
&= wE\left[\phi\left(\beta_m |S_{u,m}|^2\right) 1\left\{|S_{u,m}| = \max_{n\in u}|S_{u,n}|\right\}\right]\\
&= w\int_{\mathbb{R}_+^M}\phi(\beta_m x_m)\, 1\left\{x_m = \max_{n\in u} x_n\right\}\prod\nolimits_{n\in u} e^{-x_n}dx_n\\
&= w\int_{\mathbb{R}_+}\phi(\beta_m x)\, e^{-x}\left(\int_{\mathbb{R}_+^{M-1}}\prod\nolimits_{n\neq m} 1\{x_n\le x\}\, e^{-x_n}dx_n\right)dx\\
&= w\int_{\mathbb{R}_+}\phi(\beta_m x)\, e^{-x}\left(\prod\nolimits_{n\neq m}\int_{\mathbb{R}_+} 1\{x_n\le x\}\, e^{-x_n}dx_n\right)dx\\
&= w\int_{\mathbb{R}_+}\phi(\beta_m x)\, e^{-x}\left(1-e^{-x}\right)^{M-1}dx\\
&= w\int_0^{+\infty}\phi(\beta_m x)\, e^{-x}\left(1-e^{-x}\right)^{M-1}dx
\end{aligned}$$

Using the binomial formula

$$\left(1-e^{-x}\right)^{M-1} = \sum_{k=0}^{M-1}\binom{M-1}{k}(-1)^k e^{-kx}$$

we get

$$\begin{aligned}
R'_m &= w\sum_{k=0}^{M-1}\binom{M-1}{k}(-1)^k\int_0^{+\infty}\phi(\beta_m x)\, e^{-(k+1)x}dx\\
&= w\sum_{k=0}^{M-1}\binom{M-1}{k}(-1)^k\int_0^{+\infty}\phi(y)\, e^{-\frac{k+1}{\beta_m}y}\frac{dy}{\beta_m}\\
&= \frac{w}{\beta_m}\sum_{k=0}^{M-1}\binom{M-1}{k}(-1)^k\mathcal{L}_\phi\left(\frac{k+1}{\beta_m}\right)
\end{aligned}$$

■

Corollary 1: For $\phi(x) = \log_2(1+x)$ we have

$$\mathcal{L}_\phi(u) = \frac{1}{\ln 2}\frac{e^u}{u}\operatorname{Ei}(1,u) \tag{15}$$

and for $\phi(x) = x^n$ we have

$$\mathcal{L}_\phi(u) = \frac{n!}{u^{n+1}}$$

Example 1: Assume that β_m is constant in m, that is $\beta_m = \beta$, for all $m\in u$. We deduce from (14) that

$$R'_m = \frac{w}{\beta}\sum_{k=0}^{M-1}\binom{M-1}{k}(-1)^k\mathcal{L}_\phi\left(\frac{k+1}{\beta}\right)$$

In particular when $\phi(x) = \log_2(1+x)$, $\mathcal{L}_\phi$ is given by (15).

Example 2: Assume that

$$\phi(x) = \frac{x}{\ln 2}$$

In this case,

$$\mathcal{L}_\phi(u) = \frac{1}{u^2\ln 2}$$

We deduce from (14) that

$$R'_m = \left(\sum_{k=0}^{M-1}\binom{M-1}{k}(-1)^k\frac{1}{(k+1)^2}\right)\frac{w\beta_m}{\ln 2}$$

It is interesting to note that in some particular cases, we may decompose the expression (14) of R'_m as follows

$$R'_m = \frac{1}{h(M)}\frac{w}{W}\frac{1}{\gamma'_m} \tag{16}$$

for some function $h(M)$ of the number of users and some function γ'_m of W and β_m (which depends on the positions of users). This is the case for Examples 1 and 2 above. The particular form (16) allows to make explicit calculus of the performance at the queueing theory time-scale as will be shown in Proposition 3 below.

Remark 6: Note that, in the above two examples, the random variables $\left\{\frac{R_m}{E[R_m]}\right\}_{m\in u}$ are identically distributed, where the expectation is taken with respect to $S_{u,m}$. This case is called *symmetric* in [5, §II] where it is shown that

$$R'_m = \frac{G(M)}{M}E[R_m] \tag{17}$$

where $G(M) = E\left[\max_{m=1,\dots,M}\frac{R_m}{E[R_m]}\right]$ (which equals $\sum_{k=1}^M\frac{1}{k}$ for Example 2). Comparing Equations (16) and (17), we deduce the following relations

$$h(M) = \frac{M}{G(M)}$$

and

$$\gamma'_m = \frac{w}{WE[R_m]}$$

Our Proposition 1 may be seen as an extension of the result in [5, §II] to the non-symmetric case (i.e. when the random variables $\left\{\frac{R_m}{E[R_m]}\right\}_{m\in u}$ are not necessarily identically distributed).

2) Feasible bit-rates: The bit-rate for user m over all the bandwidth, say r'_m, is related to the bit-rate R'_m per sub-carrier by

$$r'_m = \frac{W}{w}R'_m \tag{18}$$

where R'_m is given by (13).

Example 3: In the particular case where r'_m is given by (16), then

$$r'_m = \frac{1}{h(M)}\frac{1}{\gamma'_m} \tag{19}$$

V. Queueing theory

A. Traffic dynamics

Denote the geographic region covered by the a cell by $\mathbb{D}$ that is assumed to be a bounded subset of the plane $\mathbb{R}^2$. Consider only elastic bit-rate calls whose inter-arrival times to $\mathbb{D}$ are independent and identically distributed (i.i.d.) exponential random variables with rate λ (mean $1/\lambda$). The position of each arrival is picked at random in $\mathbb{D}$ according to the uniform distribution. We assume that users don't move during their calls. Each call requires to transmit a given volume of data (amount of bits that has to be sent or received), which is modeled by an independent of everything else, exponential random variable with parameter μ. The quantity $\rho = \lambda/\mu$ is called the *traffic demand* (expressed in Mbps[1]) per cell. Users are served by the BS according to some bit-rate allocation policy.

The set of positions of all users served at a given time is called *configuration of users*. Let $\mathbb{M}$ be the set of all possible configurations (this can be formalized e.g. on the basis of the theory of point processes). We denote by $\{N_t\}_{t\ge 0}$ the process describing the evolution in time of the user configurations in $\mathbb{D}$ (due to arrivals and departures). It takes its values in $\mathbb{M}$. If the process $\{N_t\}_{t\ge 0}$ isn't ergodic, then the mean number of users in the system grows unboundedly in the long run of the system. This situation has to be avoided; in which case we say that the system is *stable* (or equivalently ergodic).

One distinguish two milestones of the analytical *evaluation of the network performance*: identification of its stability region, and the evaluation of the steady state characteristics (e.g. the mean throughput and delay).

[1]The abbreviation Mbps designates "Kilo-bit per second".

B. Non-opportunistic case

Define the *critical traffic demand* ρ_c defined by

$$\rho_c = \frac{1}{\bar{\gamma}}$$

where $\bar{\gamma}$ is the expectation of γ_m with respect to the position of the user m uniformly distributed in the cell.

Proposition 2: Assume any non-opportunistic work-conserving bit-rate allocation satisfying Condition (10). If the traffic demand per cell

$$\rho < \rho_c$$

then the system is stable.

Consider the particular allocation

$$r_m = \frac{1}{M\gamma_m} \tag{20}$$

Then, at the steady state, the throughput for the users, denoted $\bar{r}$, equals

$$\bar{r} = \rho_c - \rho$$

and the delay for a typical transmission is

$$\bar{T} = \frac{1}{\mu\bar{r}} = \frac{1}{\mu\left(\rho_c - \rho\right)}$$

Moreover, the number of users at the steady state equals

$$\frac{\rho}{\rho_c - \rho}$$

Proof: See [9] and [13, §10.2]. ■

C. Opportunistic case

Proposition 3: Consider the opportunistic bit-rate allocation (19). If the traffic demand per cell

$$\rho < \lim_{n\to\infty} 1/h\left(n\right)$$

then the system is stable. Moreover, at the steady state, the throughput for the users, denoted $\bar{r}'$, equals

$$\bar{r}' = \frac{1}{\bar{\gamma}'}\frac{1}{\mathcal{H}(\bar{\gamma}'\rho)}$$

where the function $\mathcal{H}(s)$ is defined for $s > 0$ by

$$\mathcal{H}(s) = \frac{\mathbf{E}[H(X+1)]}{\mathbf{E}[H(X)]}, \quad H\left(M\right) = \begin{cases} \prod_{k=1}^{M} h(k) & \text{if } M \geq 1 \\ 1 & \text{if } M = 0 \end{cases}$$

where X is a Poisson random variable with parameter s.

Moreover, the delay for a typical transmission is

$$\bar{T} = \frac{1}{\mu\bar{r}'} = \frac{\bar{\gamma}'\mathcal{H}(\bar{\gamma}'\rho)}{\mu}$$

and the number of users at the steady state equals

$$\bar{\gamma}'\rho\mathcal{H}(\bar{\gamma}'\rho)$$

Proof: See [5, Proposition 3.1] and [13, §11.1]. ■

1) Approximation: Note that Proposition 3 assumes a particular case when the decomposition (19) holds true as in Examples 1 and 2.

Unfortunately, for the more general allocation (18) we do not know a closed form for the performance; especially because the geometric parameters β_m depend of the location of the user m. The idea is then to replace these parameters by a constant β_0 calculated as follows.

β_0 is determined in such a way that the average over the cell of the factors γ_m given by Equation (11) should remain unchanged when the β_m are replaced by β_0. (By Proposition 2, this is equivalent to say that the performance of the non-opportunistic allocation (20) remains unchanged when the β_m are replaced by β_0.)

We will approximate the performance of our opportunistic scheme (18) by replacing β_m with β_0 and applying Proposition 3.

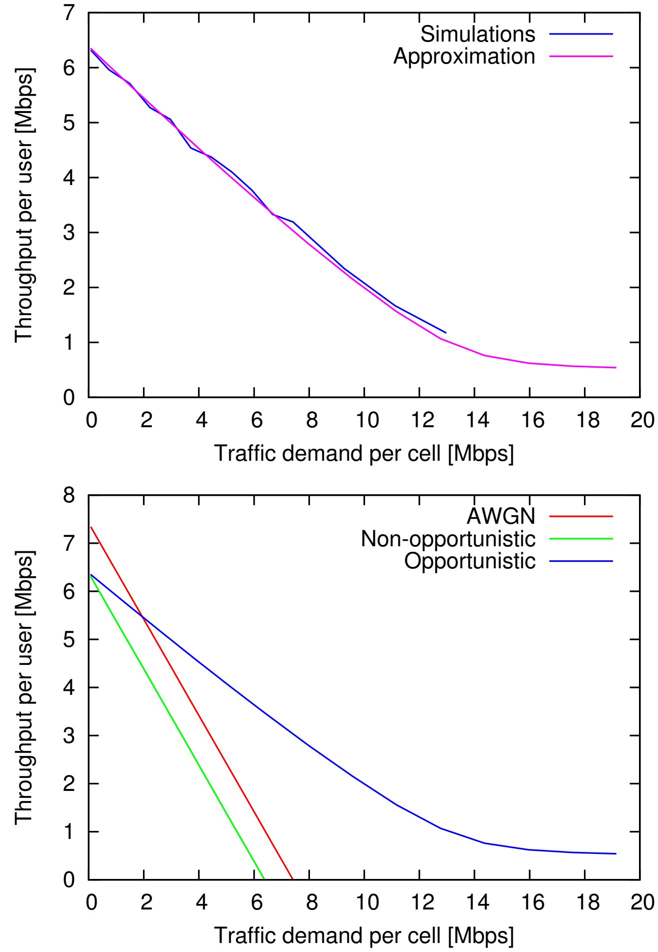

Fig. 1. Top: Validation of the approximation for the opportunistic allocation. Bottom: Comparison of the perfomance of the different allocations.

VI. Numerical results

A. Model specification

In order to obtain numerical values, we consider the most popular *hexagonal network model*, where the base stations are placed on a regular infinite hexagonal grid. Let R be the radius of the disc whose area is equal to that of the hexagonal cell served by each base station, and call R the *cell radius*. We assume a distance-loss $L(r) = (Kr)^\eta$, with $\eta = 3.38$, $K = 8667$. This means that the distance-loss between BS u and user m is equal to $L(|x_u, x_m|)$ where x_u, x_m denote, respectively, the geometric location of u and m and $|\cdot|$ is the Euclidean distance.

We consider elastic traffic with traffic demand assumed (spatially) uniform over the cell. We consider a *traffic demand* per cell ρ varying from 0 to 20Mbps.

The BS maximal total power is $\tilde{P} = 52$dBm, the common channel power P' is the fraction $\epsilon = 0.12$ of $\tilde{P}$, the total bandwidth is $W = 5$MHz and the ambient noise power $WN_0 = -103$dBm.

B. Results

Figure 1 (left) shows the throughput per user as function of the traffic demand per cell for the opportunistic allocation. We show the result of the simulations and of the approximation described in Section V-C1. The two curves are close which validates the considered approximation.

Figure 1 (right) shows the throughput per user as function of the traffic demand per cell for the AWGN, non-opportunistic and opportunistic allocation. Firstly, we observe that for the non-opportunistic allocation, the fading *decreases* the throughput per user of at least 13% compared to AWGN. Note moreover that the non-opportunistic allocation gives, like AWGN, the throughput that is a linear function of the demanded traffic in the bounded stability region. Secondly, the *gain* induced by opportunism increases with the traffic. When the traffic demand is very small, the gain induced by the opportunistic allocation is negligible. Nevertheless, when the traffic demand per cell exceeds some value (about 2.5 Mbps in our example) the gain induced by opportunism compensates the degradation induced by fading compared to AWGN.

VII. Conclusion

We have given analytic expressions for the mean throughput and delay in OFDMA cellular networks with elastic traffic. These formulas take into account in a simple but not simplistic way all important elements of the network performance: the network geometry, the traffic dynamics, fading and bit-rate allocation policy. They allow, in particular, to compare in a systematic way the effect on performance of fading for either non-opportunistic or opportunistic allocation policies.

VIII. Perspectives

Even though we assumed Rayleigh distribution for the fading, the approach may be extended to other distributions such as Rice, Nakagami, etc.

It may be useful to study the effect of grouping the sub-carriers into the so-called *chunks* (i.e. a group of sub-carriers and a number of time-slots); see [15, §6.2].

As observed in [15, §6.2], the non-opportunistic allocation scheme is suitable for fast varying fading, due to *high mobility* of users for example, which makes difficult to track of the rapidly varying channel. On the other hand, the opportunistic allocation scheme is suitable for slow varying fading as in the case of *fixed (or pedestrian)* users for example. Thus, from the point of view of the present paper, mobility decreases performance. It is interesting to couple this with the account of the speed effect at the queueing theory time scale in [3] which have shown that mobility ameliorates performance.

It is useful to study the effect of fading on the dimensioning of streaming traffic (see Remarks 4 and 5).

IX. Acknowledgments

We thank F. Baccelli at INRIA for motivating discussions on resource allocation and queueing theory. We thank also M. Debbah from Alcatel-Lucent for instructive exchanges about information theory.

References

[1] R Ahlswede. Multi-way communication channels. In *Proc. 2nd. Int. Symp. Information Theory (Tsahkadsor, Armenian S.S.R.)*, pages 23–52. Publishing House of the Hungarian Academy of Sciences, 1971.

[2] R. Ahlswede. The capacity region of a channel with two senders and two receivers. *Ann. Prob.*, 2:805–814, 1974.

[3] B. Błaszczyszyn and M.K. Karray. Performance Evaluation of Scalable Congestion Control Schemes for Elastic Traffic in Cellular Networks with Power Control. In *Proc. of IEEE INFOCOM*, 2007.

[4] B. Błaszczyszyn and M.K. Karray. Dimensioning of the downlink in OFDMA cellular networks via an Erlang's loss model. In *Proc. of European Wireless*, 2009.

[5] S. Borst. User-level performance of channel-aware scheduling algorithms in wireless data networks. In *Proc. of IEEE INFOCOM*, 2003.

[6] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Transactions on Information Theory*, 45:2007–2019, September 1999.

[7] G. Caire, G. Taricco, and E. Biglieri. Optimum power control over fading channels. *IEEE Transactions on Information Theory*, 45(5):1468–1489, July 1999.

[8] J.W. Cohen. *On Regenerative Processes in Queueing Theory*, volume 121 of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, 1976.

[9] J.W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica*, 12:245–284, 1979.

[10] A.J. Goldsmith and S.-G Chua. Variable-Rate Variable-Power MQAM for Fading Channels. *IEEE Trans. Commun.*, 45:1218–1230, 1997.

[11] A.J. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. *ieeeit*, 43:1986–1992, November 1997.

[12] J. Jang and K. B. Lee. Transmit power adaptation for multiuser OFDM systems. *IEEE J. Sel. Areas Commun.*, 21(2), February 2003.

[13] M.K. Karray. *Analytic evaluation of wireless cellular networks performance by a spatial Markov process accounting for their geometry, dynamics and control schemes*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 2007.

[14] H.J. Kushner and P.A. Whiting. Convergence of Proportional-Fair Sharing Algorithms Under General Conditions. *IEEE Transactions On Wireless Communications*, 3(4), July 2004.

[15] H. Liu and G. Li. *OFDM-Based Broadband Wireless Networks: Design and Optimization*. Wiley-Interscience, 2005.

[16] R. Serfozo. *Introduction to stochastic networks*. Springer, New York, 1999.

[17] C. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.