

A Delay Analysis for Opportunistic Transmission in Fading Broadcast Channels

Masoud Sharif and Babak Hassibi

Department of Electrical Engineering

California Institute of Technology, Pasadena, CA 91125

Email: {masoud,hassibi}@systems.caltech.edu

Abstract— We consider a single-antenna broadcast block fading channel (downlink scheduling) with n users where the transmission is packet-based and all users are backlogged. We define the delay as the minimum number of channel uses that guarantees all n users successfully receive m packets. This is a more stringent notion of delay than average delay and is the worst case delay among the users. A delay optimal scheduling scheme, such as round-robin, achieves the delay of mn . In a heterogeneous network and for the optimal throughput strategy where the transmitter sends the packet to the user with the best channel conditions, we derive the moment generating function of the delay for any m and n . For large n and in a homogeneous network, the expected delay in receiving one packet by all the receivers scales as $n \log n$, as opposed to n for the round-robin scheduling. We also show that when m grows faster than $(\log n)^r$, for some $r > 1$, then the expected value of delay scales like mn . This roughly determines the time-scale required for the system to behave fairly in a homogeneous network. We then propose a scheme to significantly reduce the delay at the expense of a small throughput hit.

We further look into two generalizations of our work: *i*) the effect of temporal channel correlation and *ii*) the advantage of multiple transmit antennas on the delay. For a channel with memory of two, we prove that the delay scales again like $n \log n$ no matter how severe the correlation is. For a system with M transmit antennas, we prove that the expected delay in receiving one packet by all the users scales like $\frac{n \log n}{M + O(M^2/n)}$ for large n and when M is not growing faster than $\log n$. Thus, when the temporal channel correlation is zero, multiple transmit antenna systems do not reduce the delay significantly. However, when channel correlation is present, they can lead to significant gains by “decorrelating” the effective channel through means such as random beamforming.

Keywords: Cellular networks, Information Theory, Queuing Theory.

I. INTRODUCTION

Resource allocation in wireless systems aims for two conflicting goals, firstly providing quality of service such as delay and fairness to users, and secondly maximizing the throughput of the system. A fundamental property of wireless channels is their time variation due to multi-path effects and the mobility of the users. This implies that at each channel use some users have favorable channel conditions and other users incur deep fades. In fact, assuming a block fading model for the channel and having full CSI in the transmitter, it can be shown that sending to the user with the best channel conditions maximizes the sum rate of the single antenna broadcast channel. In fact, this opportunistic way of transmission has been proposed in

Qualcomm’s High Data Rate (HDR) system (1xEV-DO). Other variations of this scheduling that do not require full CSI in the transmitter are studied in [1], [2].

However, there is a price to pay for maximizing the throughput which is fairness among users and delay in sending packets. Assuming users have different signal to noise ratios, the throughput optimal scheduling will provide much less service to the user with the lowest signal to noise ratio (SNR) compared to that of the user with the highest SNR. Even in a homogeneous network where users have equal SNRs and so the system is long-term fair, there is no delay guarantee for transmitting a packet to a specific user as the transmission is probabilistic, i.e. at each channel use each user will be chosen with some probability. The other extreme would be to use a round robin type scheduling that fairly gives service to all users and can guarantee a fixed delay for transmitting a packet to each user. In applications with delay constraints, one may wonder how bad the worst case delay (or the delay for the most unfortunate user) for the throughput optimal strategy is.

In this paper, we consider a broadcast channel with n backlogged users. The transmission is packet based and the channel is assumed to be block Rayleigh fading and changes independently from one block to the other. We define the delay as the minimum number of transmissions that guarantees all the users will receive m packets successfully. This notion of delay is clearly stronger than the average delay in the sense that it guarantees the reception of m packets by *all* users. Disregarding the throughput, the minimum delay of mn can be achieved by round-robin scheduling. However, the throughput optimal strategy has to contend with delay hits. The overriding question in this paper is to characterize the delay for the throughput optimal strategy, e.g. to determine its mean and other moments. We further look into the effect of temporal correlation of the channels on the delay as well as the effect of employing multiple antennas in the transmitter. Finally, we propose an algorithm to reduce the delay at the expense of a little hit in the throughput of the system.

Scheduling in broadcast channels has been considered by several authors [3]–[6]. In [4], stabilizing parallel queues in the transmitter is considered, where the connectivity of queues are random to capture deep fades in the wireless channel. In [6], the authors incorporate the channel state information in their scheduling while providing delay constraints for packets. Analyzing the average delay (over the users) can be also done

using the results for G/G/1 queues and it can be shown that the average delay is of the order of the number of users [7], [8]. However, in order to provide delay guarantee for all users, we have to study the delay for the most unfortunate user in the system. Clearly the worst case delay is a function of the number of users and their SNRs (or the probability of being chosen as the best user at each channel use). While these works give many insights and algorithms, they leave open the question of how large the worst case delay is using the throughput optimal strategy. This is the main goal of this paper. We also propose a practical scheme to reduce the delay with a little throughput hit.

This paper is organized as follows. Section II introduces our channel model and our notation. Section III deals with characterizing the delay for single antenna broadcast fading channels. Then in Section IV, we take a first look into the effect of channel correlation on the expected delay. Section V generalizes the results of Section III to multi-antenna broadcast channels. Finally Section VI proposes an algorithm to reduce the delay at the expense of a little reduction in the throughput.

II. SYSTEM MODEL AND ASSUMPTIONS

In this paper we consider a single antenna broadcast channel with n receivers. We assume a block fading channel with a coherence interval of T , and where the channel changes independently after T seconds. The transmission is assumed to be packet based and the length of each packet is T . However, if the length of the packet is smaller than T , the results in this paper can be easily generalized. Later on in this paper, we relax the independence assumption on the channel and look into channels with temporal correlation as well.

In the transmitter we assume there are n queues corresponding to each receiver and that there is always a packet available to be transmitted to any user (i.e., backlogged users). Fig. 1 illustrates the arrangement of queues in the transmitter. In fact, the main challenges for the scheduler are to first balance the service among all the users and to second exploit the multiuser diversity in the channel in order to maximize the throughput of the system. Any scheduling strategy implies a probability for choosing each user at each channel use that may depend on the signal to noise ratio (SNR) of the user, the length of the queue of users, and the statistics of the channel. For the throughput optimal strategy, this probability only depends on the SNR of the user and the channel statistics. For i.i.d channels, it is clear that these probabilities are only functions of users' SNRs. However, in the case of temporal correlation, the probabilities at each channel use will depend on the previous states of the channel and form a Markov process.

For each block of T channel uses, the received signal at the i 'th user at time t can be written as,

$$y_i(t) = \sqrt{\rho_i} h_i(t) S(t) + n_i(t), \quad i = 1, \dots, n, \quad (1)$$

where $h_i(t)$ is the effect of channel and $n_i(t)$ is additive white noise and that both are i.i.d. circularly symmetric complex Gaussian distributed with zero mean and variance of one

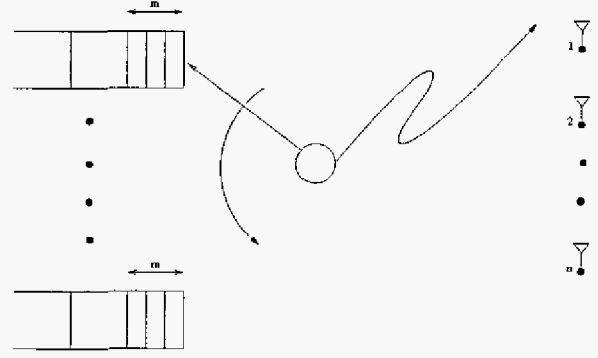


Fig. 1. n parallel queues in the transmitter corresponding to n user

modeling. Here ρ_i is the SNR of the i 'th user and $S(t)$ is the transmitted symbol at time t . We further assume independent memoryless channel which implies that the channel changes independently to another value after the coherence interval of T .

Assuming that all packets have C_0 information bits, we consider a packet to be dropped if outage occurs, i.e., if the instantaneous capacity C goes below C_0 at the time of the transmission [9]. For the round-robin scheduling, the instantaneous capacity is $C = \log(1 + \rho|h_i|^2)$ which does not depend on n . For the throughput optimal strategy, C however is the maximum of $\log(1 + \rho|h_i|^2)$ over $1 \leq i \leq n$, i.e., $C = \max_{1 \leq i \leq n} \log(1 + \rho|h_i|^2)$.

If we assume that the error probability is simply the outage probability (a reasonable assumption for long packets [10]), we have $P_e = \Pr(C < C_0)$. The throughput is therefore $R = C_0(1 - P_e) = C_0 \Pr(C \geq C_0)$. Note that for any value of C_0 , the throughput optimal strategy is to send to the best user as this would minimize P_e . Conversely, for any fixed value of P_e , sending to the strongest user maximizes the throughput as this would allow for the largest possible C_0 . It is also worth mentioning that the maximum of n i.i.d. exponential random variables (the $|h_i|^2$) behaves almost surely as $\log n$. Therefore, we do not need to use power control to compensate for the channel variation as the maximization automatically prevents having deep fades for large number of users with high probability. Thus, for the throughput optimal scheduling, it is quite reasonable to assume that all the packets have the same amount of information, i.e., C_0 , independent of the time and channel condition.

We also define the delay in the broadcast channel as the number of channel uses (denoted by $D_{m,n}$) required to guarantee that all the users will receive m packets successfully. It is clear from the definition of $D_{m,n}$ that this notion of delay refers to the worst case delay among users (or the delay for the most unfortunate user). Of course, the delay is a random variable and depends on the number of users n , the number of packets m and also the scheduling algorithm. A delay-optimal strategy is round-robin scheduling which clearly achieves the optimal delay of mn . However, round-robin does

not maximize the throughput. Throughput optimal strategies, on the other hand, will have to contend with delay hits. The next section deals with the delay for the throughput optimal scheduling.

III. DELAY ANALYSIS FOR SINGLE-ANTENNA BROADCAST CHANNELS

Throughput optimal transmission is a probabilistic scheduling which implies that each user will be given service with some given probability. Analyzing the average delay (over all the users) can be done using Kingman's result as the queue of each user is a G/G/1 queue [7]. In fact, it is quite easy to see that the set up of the broadcast problem is similar to the set up for a multiple access channel using slotted ALOHA with the only difference that P_e in a broadcast channel should be replaced by the probability of collision in slotted ALOHA systems. The average delay for slotted ALOHA scheduling has been studied and it is proved that the average delay behaves like $O(n)$ using Kingman's result [8]. However analyzing the worst case delay (or the delay for the most unfortunate user in the system) requires considering all the n parallel queues of n users together [11].

In this section, we assume that at each channel use the transmitter chooses the i 'th user as the best user with the probability p_i (which depends on its SNR), and drops the packet with probability P_e . We then obtain the moment generating function of the random variable $D_{m,n}$. This is done by first considering the simple case in which the network is homogeneous and $P_e = 0$. Then we generalize the result to the case where we have a heterogeneous network and a non-zero P_e . We further look into the asymptotic behavior of $D_{m,n}$ for different regions of m and n at the end of this section.

A. Homogeneous Network with No Dropping Probability

When users are homogeneous, and assuming throughput optimal scheduling, the transmitter chooses the user with the best reception. Clearly, the i 'th user has is the best out of n users with probability $\frac{1}{n}$ since the users are statistically identical. The random variable $D_{m,n}$ is basically the minimum number of channel uses to guarantee all n users have been chosen at least m times.

This problem can be restated as the coupon collector problem [12] which is studied by several authors in the mathematics literature (see also chapter 6 of [13]). To be more precise, users can be seen as people carrying coupons and the transmitter is the collector that chooses randomly and uniformly from the n people and collects his/her coupon. The question is how many times should the collector choose to guarantee that everybody has given at least m coupons. In fact we can state the mean value of $D_{m,n}$ based on a result found in [14].

Theorem 1: (Newman and Shepp [14]) Consider a homogeneous broadcast system with n users. We assume that at each channel use, the transmitter sends to the user with the

best channel condition. Then, we have,

$$E(D_{m,n}) = n \int_0^\infty (1 - (1 - S_m(t)e^{-t})^n) dt, \quad (2)$$

for any m and n where $S_m(t) = \sum_{k=0}^{m-1} \frac{t^k}{k!}$.

Proof: See [14] for the proof. ■

Inspired by the proof of Theorem 1, we can derive the moment generating function of $D_{m,n}$ defined as

$$F(z) = \sum_{i=0}^{\infty} z^i \Pr\{D_{m,n} > i\} = \sum_{i=0}^{\infty} z^i b_i. \quad (3)$$

where $b_i = \Pr\{D_{m,n} > i\}$. Using the generating function $F(z)$ in (3), we can obtain all the moments of $D_{m,n}$ with a little effort and by taking higher derivatives of $F(z)$ at $z = 1$ as shown in (A.3) (refer to Appendix A for more details). For example, in Appendix A, it is shown that using the definition of $F(z)$ in (3),

$$\begin{aligned} E(D_{m,n}) &= F(1) \\ \sigma_{D_{m,n}}^2 &= 2F'(1) + F(1) - (F(1))^2. \end{aligned} \quad (4)$$

Next Theorem obtains $F(z)$ and generalizes the result of Theorem 1.

Theorem 2: Considering the setting of Theorem 1, we can write the moment generating function of $D_{m,n}$ defined in (3) as,

$$F(z) = \frac{n}{z} \int_0^\infty e^{-\frac{n}{z}t} (e^{nt} - (e^t - S_m(t))^n) dt, \quad (5)$$

where $S_m(t)$ is defined in Theorem 1.

Proof: We evaluate $F(z)$ defined in (3) by the same trick as in [14] to derive the mean of $D_{m,n}$. In fact, $F(z)$ can be evaluated by noting that b_i is the probability of failure in obtaining m packets at all the n users up to and including the i 'th trial. Therefore, b_i is simply the polynomial $(\frac{1}{n}x_1 + \dots + \frac{1}{n}x_n)^i$ evaluated at $x_1 = \dots = x_n = 1$ after excluding all terms which have all x_i 's with exponent larger than $m - 1$. Therefore, we may write

$$F(z) = \sum_{i=0}^{\infty} z^i \frac{\{ (x_1 + \dots + x_n)^i \}}{n^i} \quad (6)$$

where $\{\cdot\}$ denotes the operator that removes all the terms which have all x_i 's with exponent less than $m - 1$. Considering the following identities [14],

$$\frac{z^i i!}{n^i} = \frac{n}{z} \int_0^\infty e^{-\frac{n}{z}t} t^i dt, \quad (7)$$

$$\begin{aligned} \{e^{x_1 + \dots + x_n}\} &= \sum_{i=0}^{\infty} \frac{\{(x_1 + \dots + x_n)^i\}}{i!} \\ &= e^{x_1 + \dots + x_n} - \prod_{i=1}^n (e^{x_i} - S_m(x_i)), \end{aligned} \quad (8)$$

where the second equality in (8) follows by noting that the second term in the right hand side just subtract out the terms

with all x_i 's larger than m . We may then replace the integral form for $\frac{1}{n!}$ using (7) in (6) to get,

$$\begin{aligned} F(z) &= \sum_{i=0}^{\infty} \int_0^{\infty} \frac{n}{z} e^{-\frac{n}{z}t} t^i dt \times \frac{\{(x_1 + \dots + x_n)^i\}}{i!} \\ &= \frac{n}{z} \int_0^{\infty} e^{-\frac{n}{z}t} \sum_{i=0}^{\infty} \frac{\{(x_1 + \dots + x_n)^i\}}{i!} dt \\ &= \frac{n}{z} \int_0^{\infty} e^{-\frac{n}{z}t} \left(e^{tx_1 + \dots + tx_n} - \prod_{i=0}^n (e^{tx_i} - S_m(tx_i)) \right) dt \\ &= \frac{n}{z} \int_0^{\infty} e^{-\frac{n}{z}t} (e^{nt} - (e^t - S_m(t))^n) dt. \end{aligned} \quad (9)$$

where we used (8) to get the second equality. ■

B. Heterogeneous Network with Dropping Probability

For the special case of a homogeneous network, we derived the mean and variance of $D_{m,n}$ in Theorem 1 and 2. In what follows, we generalize the results to a more general setting in which users may have different SNRs and also a packet may be dropped with probability P_e (if outage capacity occurs). Having said that, and assuming a memoryless i.i.d. channel, the transmitter chooses the i 'th user with probability p_i that depends on the user's SNR and its channel conditions for the throughput optimal strategy.

The following Theorem states the moment generating function of $D_{m,n}$ for this general setting and for any m and n . The Theorem is a generalization of the result of Newman and Shepp [14] stated in Theorem 1.

Theorem 3: Suppose we have n users such that the probability of choosing the i 'th user is $p_i = \frac{\alpha_i}{n}$ and the probability of dropping a packet is P_e . Then the moment generating function for $D_{m,n}$ defined in (3) is,

$$F(z) = \frac{n}{z} \int_0^{\infty} e^{-\frac{n}{z}t} \left(e^{nt} - e^{nP_e t} \prod_{i=1}^n (e^{t\beta_i} - S_m(t\beta_i)) \right) dt, \quad (10)$$

where $\beta_i = (1 - P_e)\alpha_i$. In particular

$$E(D_{m,n}) = n \int_0^{\infty} \left(1 - \prod_{i=1}^n (1 - S_m(\beta_i t) e^{-\beta_i t}) \right) dt,$$

where $S_m(t)$ is as defined in Theorem 1.

Proof: Similar to the proof of Theorem 2, we derive the moment generating function of $D_{m,n}$ as defined in (3). Since we have a non-zero probability of dropping a packet, we may assume that there is a fictitious user ($n+1$ 'th one) corresponding to the case where the packet is lost; therefore whenever a packet is dropped, we may assume that $n+1$ 'th user has been chosen to be transmitted to. Assuming that P_e is the probability of dropping a packet, the probability of choosing the fictitious user is P_e and the probability of choosing the i 'th user and sending successfully is $\frac{\alpha_i}{n}(1 - P_e)$ for $i = 1, \dots, n$. Therefore, the delay $D_{m,n}$ is the number of channel uses that guarantees having m packets in all the n users (i.e. except for the fictitious user).

The moment generating function $F(z) = \sum_{i=0}^{\infty} z^i \Pr(D_{m,n} > i) = \sum_{i=0}^{\infty} z^i b_i$ where b_i is the probability of failure in sending m packets to n users up to and including the i channel uses and is equal to the polynomial $(\frac{\beta_1}{n}x_1 + \dots + \frac{\beta_n}{n}x_n + P_e x_{n+1})^i$ evaluated at $x_1 = \dots = x_{n+1} = 1$ after removing all the terms that have all x_1, \dots, x_n exponents larger than m . Therefore, we can write $F(z)$ as,

$$\begin{aligned} F(z) &= \sum_{i=0}^{\infty} \frac{z^i}{n^i} \{(\beta_1 x_1 + \dots + \beta_n x_n + nP_e x_{n+1})^i\} \\ &= \frac{n}{z} \int_0^{\infty} e^{-\frac{n}{z}t} \sum_{i=0}^{\infty} \frac{\{(\beta_1 x_1 + \dots + nP_e x_{n+1})^i\}}{i!} dt \\ &= \frac{n}{z} \int_0^{\infty} e^{-\frac{n}{z}t} \left\{ e^{t\beta_1 x_1 + \dots + t\beta_n x_n + nP_e x_{n+1}} - \right. \\ &\quad \left. e^{nP_e x_{n+1} t} \prod_{i=1}^n (e^{t\beta_i x_i} - S_m(t\beta_i x_i)) \right\} dt \\ &= \frac{n}{z} \int_0^{\infty} e^{-\frac{n}{z}t} \left(e^{nt} - e^{nP_e t} \prod_{i=1}^n (e^{t\beta_i} - S_m(t\beta_i)) \right) dt. \end{aligned} \quad (11)$$

where we used the identity in (7) to deduce the second equality. We also used the following identity which is analogous to (8)

$$\{e^{x_1 + \dots + x_{n+1}}\} = e^{x_1 + \dots + x_{n+1}} - e^{x_{n+1}} \prod_{i=1}^n (e^{x_i} - S_m(x_i)) \quad (12)$$

to obtain the third equality. In (12), $S_m(t)$ is as defined in Theorem 1 and the operator $\{\cdot\}$ removes the terms that have the exponents of x_1, \dots, x_n larger than m . Eq. (12) can be easily proved by noting that the polynomials in the left removes all the term from the exponential function that have all x_i 's for $i = 1, \dots, n$ with exponents larger than $m-1$.

Using the relationship between $F(z)$ and its moments shown in Appendix A (see also (4)) and having $F(z)$ derived in (11), we can obtain the mean of $D_{m,n}$ as stated in the Theorem. ■

We can further generalize the setting of Theorem 3 in two different directions. First we can consider the delay in sending m_i packets to the i 'th user for $i = 1, \dots, n$. We can further assume that we are only interested in receiving the packets for a subset of the set of all users. The following Corollary states the moment generating function for the aforementioned set up.

Corollary 1: Consider the setting of Theorem 3. Defining $\mathbf{m} = (m_1, \dots, m_i)$ and $D_{\mathbf{m}}$ as the minimum number of channel uses that guarantees receiving of m_j packets at the j 'th user for $j = 1, \dots, i$, we can write the moment generating function for $D_{\mathbf{m}}$ as,

$$\begin{aligned} F(z) &= \sum_{i=0}^{\infty} z^i \Pr(D_{\mathbf{m}} > i) \\ &= \frac{n}{z} \int_0^{\infty} e^{-\frac{n}{z}t} \left(e^{nt} - e^{nP_e t + \sum_{k=i+1}^n \beta_k t} \prod_{p=1}^i (e^{t\beta_p} - S_{m_p}(t\beta_p)) \right) dt \end{aligned}$$

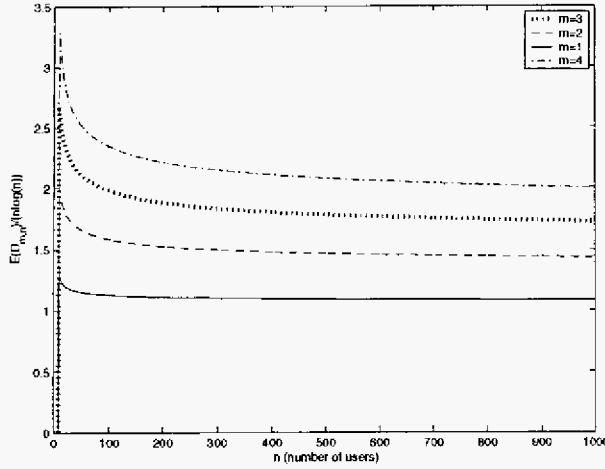


Fig. 2. $\frac{E(D_{m,n})}{n \log n}$ for different values of m and n

Proof: Proof follows along the same line as the proof for Theorem 3. We omit the proof for the lack of space (see [15]).

As a simple consequence of Corollary 1, the expected delay in receiving m_j packets at the j 'th user for $j = 1, \dots, i$ and denoted by $D_{\mathbf{m}}$ can be written as,

$$E(D_{\mathbf{m}}) = \int_0^\infty \left(1 - \prod_{k=1}^i (1 - S_{m_k}(\beta_k t) e^{-\beta_k t}) \right) dt \quad (13)$$

where $\mathbf{m} = (m_1, \dots, m_i)$. Clearly Eq. (13) reduces to the mean derived in Theorem 3 when m_j 's equals m and i equals n . Assuming that users are divided in different subgroups where each subgroup has its own delay constraint, Corollary 1 will allow us to obtain the delay guarantee for any subset of the set of all users.

To get more insight on the behavior of the expected delay, we can numerically evaluate the expected delay. Considering n equally likely users and that the probability of dropping a packet is P_e , we can state that,

$$E(D_{m,n}) = \frac{n+1}{1-P_e} \int_0^\infty (1 - (1 - S_m(x) e^{-x})^n) dx, \quad (14)$$

by a simple change of variable in the integral for the expected delay in (13).

Fig. 2 shows the expected delay for $m = 1, 2, 3, 4$ and for different number of users for a homogeneous network. It is clear that when n is large and $m = 1$, the growth in the expected delay is like $n \log n$. Also Fig. 2 implies that the expected delay does not grow linearly with m (for small values of m). In fact it converges to $n \log n$ although the convergence seems to be quite slow. It is worth mentioning that using round robin we can achieve a minimum delay of mn for any m and n . The next subsection makes the above observation more precise.

C. Asymptotic Analysis of the Moments of $D_{m,n}$

In the previous subsection, we obtained the moments of $D_{m,n}$ for a general setting and for any m and n in closed form. However, it is hard to speculate how the mean and variance of the delay behave as functions of m and n . In order to get a better insight into the behavior of the delay, we derive some asymptotic results for the moments of $D_{m,n}$ and for different regions of m and n .

Theorem 4: Assuming a homogeneous network and that a packet will be dropped with probability P_e ,

1) For m fixed and $n \rightarrow \infty$, we have

$$E(D_{m,n}) = \frac{n \log n}{1-P_e} + n(m-1) \log \log n + o(n),$$

$$\sigma_{D_{m,n}}^2 = O(n^2).$$

2) For $m = \log n$ and $n \rightarrow \infty$, we have

$$E(D_{m,n}) = \alpha \frac{1}{1-P_e} n \log n + O(n \log \log n). \quad (15)$$

where $\alpha = 3.146$ is the solution to the equation $\alpha - \log \alpha = 2$.

3) For $m = (\log n)^r$ where $r > 1$ is fixed and $n \rightarrow \infty$, then

$$E(D_{m,n}) = \frac{1}{1-P_e} n (\log n)^r + o(n (\log n)^r)$$

$$= \frac{1}{1-P_e} mn + o(mn).$$

4) For n fixed and $m \rightarrow \infty$,

$$E(D_{m,n}) = \frac{1}{1-P_e} nm + o(m). \quad (16)$$

Proof: See [15] for the proof. ■

Assuming $m = 1$ and using the result of Theorem 4, we can state that the delay converges to the mean almost surely using Chebychev's inequality as,

$$\Pr \left\{ |D_{1,n} - n \log n + o(n \log \log n)| \leq n \sqrt{\log \log n} \right\}$$

$$\geq 1 - \frac{1}{\log n}, \quad (17)$$

for large n . This implies that the delay hit for sending the first packet successfully to all the users is increased from the minimum of n for the round robin scheduling to $n \log n$ for the opportunistic transmission for large n . So the delay degradation due to exploiting the channel variation and maximizing the throughput of the system is a multiplicative factor of $\log n$. It would be also interesting to investigate the scaling law of the variance of $D_{m,n}$ when m also grows to infinity; this would imply the type of convergence to the mean for different regions of m and n .

Remark 1: For a homogeneous network, as opportunistic transmission (or throughput optimal scheduling) is long term fair (i.e. the probability of choosing all the users is the same), we know that for sufficiently large m , the expected delay should behave like mn . This is confirmed by the fourth part of Theorem 4. Interestingly, Theorem 4 further implies that if m

grows faster than $(\log n)^r$ where r is fixed and greater than one the expected delay behaves like mn . This has implications for the time scale after which the system behaves fairly. Moreover, if m grows logarithmically with n , the expected delay is only off by a constant factor of $\alpha = 3.14$, compared to the minimum delay mn . Therefore, our result can be seen as the short term behavior of the delay for different regions of m and n .

As mentioned, the largest delay hit is when we focus on sending a few packets, i.e. $m = 1$ or m is small, to all users. The delay hit gets less when we focus on sending more and more packets (i.e., when m gets larger). Therefore, in the rest of the paper, we mainly focus on the delay for sending the first packet, i.e. $D_{1,n}$. We will try to characterize the effect of channel temporal correlation, as well as the use of multiple transmit antennas on the delay for sending the first packet. We also propose a scheduling to reduce the $D_{1,n}$ without sacrificing too much on the throughput.

IV. A STUDY OF THE EFFECT OF CORRELATION ON THE DELAY

Previously, we considered a block fading i.i.d. channel in which the channel is changing independently from one block to the other. In this section, we make our model for the channel more realistic by considering correlation between consecutive blocks of the channel. Clearly, if there is a large correlation between two consecutive blocks, the likelihood of choosing the user that has the best channel in the first block as the user with the best channel conditions again in the next block is higher, and therefore, the delay deteriorates as the correlation increases. Temporal correlation basically implies that at each channel use, the probabilities of choosing users depends on the states of channel in the previous time slots.

To make the observation more clear, we consider a channel with memory L , meaning that blocks separated by L blocks are independent. We can prove an upper bound for the delay by considering the following simplistic scheduling and noting that the channel is independently changing from its samples L blocks back: we send to the user that is maximum in the i 'th block for L consecutive blocks and switch to another user by independently choosing the user which has the best channel after L channel uses. Using this simplistic scheduling, we can state an upper bound for the delay as $D_{m,n}^L \leq LD_{m,n}$.

In the following Theorem, we present a tighter bound for the case when the channel has a memory of two, i.e. $L = 2$. Here we assume a Gaussian fading channel as in (1) such that $h_i(t)$ and $h_i(t+1)$ are jointly Gaussian with correlation factor β where $h_i(t)$ is the channel for the i 'th user at the t 'th block. Here β has to be less than or equal to $\frac{1}{2}$ to make the covariance matrix positive definite. Since $L = 2$, $h_i(t)$ and $h_i(t+m)$ for $m \geq 2$ are independent.

Theorem 5: Consider a homogeneous network and a Gaussian channel with memory of two and that $P_e = 0$. Then the expected delay in sending one packet to all users (denoted by $D_{1,n}$) when sending to the user with the best channel at each

channel use is,

$$E(D_{1,n}) = (n-1) \frac{1}{1-P_{max}} \sum_{i=1}^n \frac{1}{i}, \quad (18)$$

where P_{max} is the probability that a user is chosen in two consecutive blocks and can be written as,

$$P_{max} = \int_0^\infty \int_0^\infty \left(\int_0^{a_1} \int_0^{a_2} f(r_1, r_2) dr_1 dr_2 \right)^{n-1} f(a_1, a_2) da_1 da_2, \quad (19)$$

$$f(r_1, r_2) = (1 - \beta^2) r_1 r_2 e^{-r_1^2/2} e^{-r_2^2/2} I_0(\beta r_1 r_2), \quad (20)$$

where β is the correlation between the consecutive blocks of the channel and $I_0(\cdot)$ is modified Bessel function.

Proof: Suppose the probability of choosing a user that is the best user at time i as the best user at time $i+1$ is P_{max} and the probability of choosing any other user at time i as the best user is P_0 . Since memory of the channel is two and the users who are not chosen in the last channel use are equally likely to be chosen for the current transmission, we have $P_{max} + (n-1)P_0 = 1$.

To obtain the delay in sending one packet to all users, we write the delay $D_{1,n}$ as

$$D_{1,n} = \sum_{i=1}^n r_i \quad (21)$$

where r_i denotes the number of channel uses after sending at least one packet to $i-1$ users and before completing the transmission of at least one packet to i users. It is clear that r_i has a Geometric distribution (i.e. $\Pr(r_i = k) = (1-p_i)^k p_i$ for $k = 1, \dots$) with the parameter $1-p_i$ denoting the probability of choosing a user among the $i-1$ users that have already been chosen. Therefore, $p_1 = 0$ and

$$p_i = 1 - P_{max} - (i-2)P_0 = (1 - P_{max}) \frac{n-i+1}{n-1}, \quad (22)$$

for $2 \leq i \leq n$, where the first equality follows by noting that p_i is equal to the probability of not choosing from a set of i users including the user that has the best channel for the last channel use. The second equality follows from the fact that $P_{max} + (n-1)P_0 = 1$. It is quite easy to show that the first moment of r_i is $\frac{1}{p_i}$. Therefore, taking the expectation from both sides of (21), we get

$$E(D_{1,n}) = \sum_{i=1}^n \frac{1}{p_i} = (n-1) \frac{1}{1-P_{max}} \sum_{i=1}^n \frac{1}{i} \quad (23)$$

In what follows we compute P_{max} which is a function of n , β and is independent of the user index i as the network is homogeneous. Clearly, channels corresponding to different users are independent. However, for any i , $h_i(t)$ and $h_i(t-1)$ have a jointly complex Gaussian distribution with correlation factor of β . Noting that P_{max} denotes the probability that a user has the best channel if it had the best channel condition

in the last channel use, we obtain

$$P_{\max} = \Pr \left\{ |h_1(t)| \geq |h_i(t)|, i = 2, \dots, n \right. \\ \left. |h_1(t-1)| \geq |h_i(t-1)|, i = 2, \dots, n \right\} \\ = \int_0^\infty \int_0^\infty (F(a_1, a_2))^{n-1} f(a_1, a_2) da_1 da_2 \quad (24)$$

where $h_i(t)$ denotes the channel for the i 'th user at time instance t as in (1) and where $F(\cdot, \cdot)$ and $f(\cdot, \cdot)$ denote the joint CDF and PDF of $|h_i(t)|$ and $|h_i(t-1)|$ for $1 \leq i \leq n$. For any i , $h_i(t)$ and $h_i(t-1)$ have a jointly complex Gaussian distribution with correlation factor of β . It is then straightforward to derive the joint distribution of $|h_i(t)|$ and $|h_i(t-1)|$ as in (20). Replacing (20) in (24) leads to (19). ■

Theorem 5 can be generalized to the case with a memory of L , however, the expected delay is going to be related to the probabilities that a user that has the best channel condition at time i again gets chosen as the best user at time $i+j$ for $j \leq L$. We omit the derivation for the case where $L > 2$ and focus on analyzing the effect of β (the correlation factor) on the expected delay for a channel with memory of two and for large n . To do so, we have to obtain the behavior of P_{\max} for large n which is done in the next Lemma.

Lemma 1: For large n , P_{\max} defined in (19) can be bounded by,

$$P_{\max} \leq B\left(n, \frac{1-2\beta}{1-\beta^2}\right) + \frac{n}{\sqrt{1-4\beta^2}} B\left(n, \frac{1}{1-\beta^2}\right)$$

where $B(m, n)$ is the Beta function and β is the correlation factor which has to be less than $\frac{1}{2}$.

Proof: Refer to [15] for the proof. ■

Remark 2: It is worth noting that $B(n, \gamma)$ behaves like $O(\frac{1}{n^\gamma})$ for large n and fix γ [16]. Therefore, for fixed $\beta < \frac{1}{2}$, P_{\max} tends to zero. Using the result of Lemma 1 and Theorem 5, we can write the expected delay as

$$E(D_{1,n}) = \frac{1}{1 - O\left(\frac{1}{n^{\frac{\min(\beta^2, 1-2\beta)}{1-\beta^2}}}\right)} n \log n + O(n \log \log n).$$

Clearly, the expected delay still scales like $n \log n$ which implies that correlation for a channel with memory of two does not really affect the delay for the throughput optimal scheduling. In fact, Eq. (25) shows that, for $L = 2$, the upper bound for the delay, i.e., $LD_{1,n}$, is loose by a factor of roughly 2.

V. DELAY IN MULTI ANTENNA BROADCAST CHANNELS

Multiple transmit antennas have been shown to significantly improve the throughput of a broadcast channel. It is shown that dirty-paper coding achieves the sum rate capacity of a Gaussian broadcast channel [17]–[19]. However, beamforming has long been proposed as a heuristic method to mitigate the interference in the transmitter and to send multiple beams to different users. Although, beamforming is not optimal in

achieving the sum rate capacity, its throughput does scale the same as that of dirty paper coding for a system with many users and has much less complexity than that of dirty paper coding [20].

In this paper, for a system with M transmit antennas, we assume a simple model in which the base station transmits to M different receivers at each channel use. This is certainly a valid model for beamforming or channel inversion, though it does not fit the dirty paper scheduling in which the transmitter sends information to all the users at each time. However, as far as the scaling law of the sum rate capacity is concerned, when M is either fixed or growing logarithmically with n , it can be shown that beamforming, channel inversion, and random beamforming all give the optimal scaling law for the sum rate throughput [20], [21].

For a homogeneous network, our model for the multiple antenna transmitter implies that, at each channel use, the transmitter sends to M different users uniformly chosen from the pool of n users. The scheduling is certainly more balanced compared to the case where we have a single antenna system that works M times faster. This can be justified by noticing the fact that we exclude the possibility of sending to one user twice (or more) in each block of M transmissions and hence the scheduling is more balanced. Therefore a simple upper bound for the expected delay for any M is the one we derived in Theorem 3 divided by M . More specifically,

$$D_{m,n}(M) \leq \frac{1}{M} D_{m,n} \quad (25)$$

where $D_{m,n}(M)$ is the delay for sending m packets successfully to n users in an M -transmit antenna system and where $D_{m,n}$ is the delay for a single antenna broadcast system as in Theorem 1.

In fact we can compute exactly the expected delay in transmitting the first packet successfully, i.e. $E(D_{1,n}(M))$, for any n and M . Further generalization of the result to $m > 1$ is non trivial and we have not been able to do this; however, it is quite easy to show that $D_{m,n}(M) \leq m D_{1,n}(M)$. The next theorem presents the result for $m = 1$ and for any n and M .

Theorem 6: Consider a M -antenna broadcast channel such that at each channel use M different users are being chosen. Then the expected delay in sending one packet to all users can be written as,

$$E(D_{1,n}(M)) = \sum_{k=0}^{\infty} \sum_{r=1}^n \sum_{i=0}^{n-r} \frac{(-1)^{n-r-i} \binom{n}{r} \binom{n-r}{i}}{\binom{n}{M}^k} \binom{i}{M}^k$$

for any n and M .

Proof: Similar to the proof of Theorem 3, we first note that the mean of $D_{1,n}(M)$ can be written as,

$$E(D_{1,n}(M)) = \sum_{k=0}^{\infty} \Pr(D_{1,n}(M) > k) \quad (26)$$

In order to compute the probability of $D_{1,n} > k$, we define the auxiliary random variable $\mu_n^M(k)$ as the number of users that have received no packets after k channel uses in which

the transmitter sends to M different users. From the definition of μ_n^M , it is clear that $\mu_n^M \leq n$ and that $D_{1,n}(M) > k$ is equivalent to $\mu_n^M(k) > 0$. Therefore, Eq. (26) can be written as,

$$E(D_{1,n}(M)) = \sum_{k=0}^{\infty} \sum_{r=1}^n \Pr(\mu_n^M(k) = r) \quad (27)$$

The probability that $\mu_n^M(k) = r$ can be computed as follows. Assuming $\mu_n^M(k) = r$ implies that *only* $n - r$ users have received at least one packet in k channel uses. We then define the event S_i for $i = 0, 1, \dots, n - r$ as the event that at least $n - r - i$ users have not received any packets among $n - r$ users that are supposed to receive a packet. This implies that there are at most i users that the transmitter sends packets to. It is clear that for $1 \leq i \leq M$ probability of S_i is zero, since the transmitter certainly can transmit to M different users at each channel use. For $n - r \geq i \geq M$, however we can write the probability of S_i as

$$\Pr\{S_i\} = \binom{n}{r, i} \left(\frac{i}{M}\right)^k = \binom{n}{r} \binom{n-r}{i} \left(\frac{i}{M}\right)^k, \quad (28)$$

where we first chose two sets of users with cardinality r and i from the set of n users and then we distributed packets among i of them k times by choosing M different users at each time.

Considering the definition of $\mu_n^M(k) = r$ and the S_i 's, we can use the inclusion-exclusion principle (see chapter 4 of [12]) to obtain

$$\begin{aligned} \Pr(\mu_n^M(k) = r) &= \sum_{i=0}^{n-r} (-1)^{n-r-i} \Pr(S_i) \\ &= \frac{\binom{n}{r}}{\binom{n}{M}^k} \sum_{i=M}^{n-r} (-1)^{n-r-i} \binom{n-r}{i} \left(\frac{i}{M}\right)^k. \end{aligned}$$

Substituting the above identity in (27), we can write the expected delay as,

$$\begin{aligned} E(D_{1,n}(M)) &= \sum_{k=0}^{\infty} \Pr(\mu_n^M(k) > 0) \\ &= \sum_{k=0}^{\infty} \sum_{r=1}^n \frac{\binom{n}{r}}{\binom{n}{M}^k} \sum_{i=M}^{n-r} (-1)^{n-r-i} \binom{n-r}{i} \left(\frac{i}{M}\right)^k \end{aligned}$$

This completes the proof for the Theorem. ■

Remark 3: It is worth mentioning that we can also obtain the generating function $F(z) = \sum_{k=0}^{\infty} z^k \Pr(D_{1,n}(M) > k)$ that would lead to the moments of $D_{1,n}(M)$ for any M and n . In fact, $F(z)$ is equal to

$$F(z) = \sum_{k=0}^{\infty} \sum_{r=1}^n \sum_{i=0}^{n-r} (-1)^{n-r-i} \frac{z^k \binom{n}{r}}{\binom{n}{M}^k} \binom{n-r}{i} \left(\frac{i}{M}\right)^k \quad (29)$$

Using (4) and (29), we can easily obtain the variance (and other moments) of $D_{1,n}(M)$.

Although Theorem 6 gives us the exact value of the expected delay for any number of users, it does not make clear how

much improvement on the delay we can get in using multi-antenna transmitter over that of the single antenna system. We can in fact asymptotically analyze the expected delay derived in Theorem 6 for large number of users to get a better intuition about this result.

Theorem 7: Consider the setting of Theorem 6. Then the expected delay in sending at least one packet to all n users using an M -antenna transmitter derived in Theorem 6 behaves like

$$\begin{aligned} E(D_{1,n}(M)) &= \frac{\sum_{k=1}^n \frac{1}{k}}{\sum_{r=0}^{M-1} \frac{1}{n-r}} + O(1) \\ &= \frac{n \log n}{M + O(M^2/n)} + o(n). \end{aligned} \quad (30)$$

for large n and when M grows no faster than $\log n$.

Proof: Refer to [15] for the proof. ■

For the special case of $M = 1$, the problem reduces to the coupon collector problem when $m = 1$ (one packet). It can be easily shown that the expected delay is equal to $n \sum_{i=1}^n \frac{1}{i}$. Clearly the result of Theorem 6 confirms this result for one transmit antenna, i.e. $M = 1$.

Remark 4: As mentioned, using multiple transmit antennas in the transmitter should improve the delay. We may write the improvement (denoted by G) on the expected delay by using M transmit antennas over that of single antenna case as,

$$G = \frac{1}{\sum_{r=0}^{M-1} \frac{1}{n-r}} = M + O\left(\frac{M^2}{n}\right). \quad (31)$$

Eq. (31) implies that when M is not growing faster than $\log n$, the gain in delay is a factor of M which comes from the fact that we are transmitting packets M times faster. Therefore, multiple transmit antenna systems incur pretty much the same delay as that of a single antenna transmitter that operates M times faster when there is no channel correlation.

Although the gain on delay in using multiple transmit antennas is not that much, multiple transmit antennas can significantly improve the long term fairness in a heterogeneous network. More precisely, in [21], it is proved that if M grows logarithmically with the number of users, the probability of choosing each user become independent of its SNR and approaches to $\frac{1}{n}$. However, when there is channel correlation, multiple antenna systems can significantly reduce the delay by "decorrelating in time" the effective channel through means such as random beamforming [21], [22].

VI. TRADING DELAY WITH THE THROUGHPUT: d-ALGORITHM

Previously, we showed the delay hit in using the optimal throughput scheduling is a $\log n$ fold increase compared to the minimum achievable delay. In this section, we propose an algorithm that can reduce the expected delay for sending the first packet at the price of a little throughput degradation. The goal is to improve the $\log n$ fold degradation in the delay without too much reducing the throughput of the system.

In order to improve the delay, we have to introduce more options to the scheduler at each channel use. For single antenna

systems, this can be done by looking at the d best users in terms of capacity and transmit to the user among those d users that has received the least number of packets. We call this scheduling the d -algorithm. For a large number of users and fixed d , it is quite easy to show that the capacity of the best user and that of the d 'th best user is quite close almost surely. This in fact guarantees that the throughput degradation using our algorithm is not that much. The next Theorem quantifies the performance of the d algorithm precisely.

Theorem 8: Consider the setting of Theorem 1 and suppose the transmitter uses the d algorithm. We denote the expected delay in sending the first packet by $E(D_{1,n}^d)$. Then, for any d ,

$$E(D_{1,n}^d) = n \int_0^{1-\frac{d}{n}} \frac{1}{1-x^d} dx + O(1) \quad (32)$$

Asymptotically, we can further prove that if d is fixed,

$$\lim_{n \rightarrow \infty} \frac{E(D_{1,n}^d)}{E(D_{1,n})} = \lim_{n \rightarrow \infty} \frac{E(D_{1,n}^d)}{n \log n} = \frac{1}{d}. \quad (33)$$

Proof: In order to compute the expected delay, we again define the variable r_i as the number of channel uses after sending at least one packet to $i-1$ users and before completing the transmission of at least one packet to i users. Clearly r_i has a Geometric distribution as,

$$\Pr(r_i = k) = p_i^k (1 - p_i) \quad k = 0, 1, \dots \quad (34)$$

where p_i is the probability that both the best and the second best user have been chosen before, therefore

$$\begin{aligned} p_i &= 0 & 0 \leq i \leq d-1 \\ p_i &= 1 - \frac{\binom{i}{d}}{\binom{n}{d}}, & d \leq i \leq n-1 \end{aligned} \quad (35)$$

Noting that $D_{1,n} = \sum_{i=0}^{n-1} r_i$, and also using the fact that the mean value of r_i is $\frac{1}{p_i}$, we can obtain the expected value of $D_{1,n}$ as

$$E(D_{1,n}^d) = \sum_{i=d}^{n-1} \frac{1}{p_i} \leq \sum_{i=d}^{n-1} \frac{1}{1 - \left(\frac{i-d+1}{n}\right)^d} \quad (36)$$

where we used a simple upper bound for $\frac{\binom{i}{d}}{\binom{n}{d}}$. To evaluate the summation in the right hand side of (36), we may take integrals from $x = 1$ to $x = n-d+1$ from both sides of

$$\frac{1}{1 - (x/n)^d} \geq \frac{1}{1 - (\lfloor x \rfloor/n)^d} \geq \frac{1}{1 - ((x-1)/n)^d}, \quad (37)$$

to obtain

$$E(D_{1,n}^d) = n \int_0^{1-\frac{d}{n}} \frac{dx}{1-x^d} + O(1), \quad (38)$$

which complete the proof for the first part of the Theorem. To prove the second part, we define the integral in the right hand side of (38) as $G(n)$. Then it is quite easy to show that when d is fixed, we have

$$\lim_{n \rightarrow \infty} \frac{G(n)}{\log n} = \lim_{n \rightarrow \infty} \frac{d}{n(1 - (1-\frac{d}{n})^d)} = \frac{1}{d}. \quad (39)$$

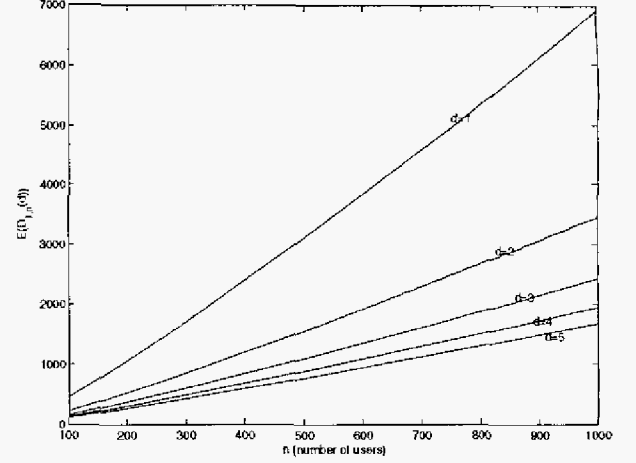


Fig. 3. Expected delay $E(D_{1,n}^d)$ for different values of d and n

where we used the L'Hopital's rule in (39). Considering that $E(D_{1,n})$ scales like $n \log n$ as proved in Theorem 4, the second part of the theorem immediately follows from (39). ■

Fig. 3 shows the delay improvement for different values of d and for different number of users. As d increases the delay improves though with less pace. Clearly, we can get most of the improvement by just checking the the best two users ($d = 2$) and further increasing d will not improve the expected delay as much as before.

Remark 5: It is clear that $D_{1,n}^d$ is always less than $D_{1,n}$ for any $d > 1$. Therefore, using the result of Theorem 4, the variance of $D_{1,n}^d$ is certainly less than or equal to $(n \log n)^2 + O(n^2)$. This upper bound, however, may not be tight. It would be interesting to obtain tighter bounds on the variance to see the type of convergence of $E(D_{1,n}^d)$ to $\frac{1}{d}n \log n$ (if it does at all).

There is of course a price to pay on the rate for the delay improvement. In order to see the throughput hit, we look into the ergodic throughput of the channel (denoted by $R(d)$) using the d algorithm defined as

$$E\{R(d)\} = E \log \left(1 + \rho \max_{1 \leq k \leq n}^k |h_k|^2 \right) \quad (40)$$

where \max^k denotes the k 'th maximum and k is a random variable uniformly distributed between 1 and d . The next Lemma proves that if d does not grow faster $\log n$, the throughput retains the same scaling law.

Lemma 2: Consider a broadcast channel with n users and the transmitter uses the d algorithm. Then, the throughput using this algorithm satisfies

$$\lim_{n \rightarrow \infty} E(R(d)) - E(R(1)) = 0, \quad (41)$$

when d is fixed.

Proof: See [15] for the proof. ■

Remark 6: It is worth mentioning that the transmitter may use a round-robin type scheduling and also exploits the

channel. This can be done by sending to the best user among n users at the first channel use, and then sending to the best user among $n - 1$ users that have not been chosen and so on. This method can make sure that the worst case delay is equal to n . The ergodic throughput of this scheme can be written as,

$$R_{RR} = E \left\{ \frac{1}{n} \sum_{k=1}^n \log \left(1 + \rho \max_{1 \leq i \leq k} |h_i|^2 \right) \right\} \quad (42)$$

Assuming that the channel is Rayleigh fading, we can show that in the limit of large n , the ratio of $E(R_{RR})$ over $E(R(1))$ is one. Of course, the convergence proved in Lemma 1 for d -algorithm holds in a stronger sense, and therefore when d is fixed, the throughput of the d algorithm is certainly much larger than $E(R_{RR})$.

VII. CONCLUSION

Assuming a block fading i.i.d. channel and a single antenna broadcast system with n backlogged users, we derive the moment generating function of the delay for any m and n and for a general heterogeneous network where a packet can be dropped with probability P_e . Asymptotically, for a homogeneous network where the throughput optimal scheduling is long-term fair (i.e., the probability of choosing users are equal), the result implies that the average delay in sending one packet to all users behaves like $n \log n$ as opposed to n for a round robin scheduling. We also prove that when m grows like $(\log n)^r$, for some $r > 1$, then to the first order the delay scales as $m n$. This roughly determines the time-scale required for the system to behave fairly.

We then consider the effect of correlation as well as the advantage of multiple transmit antenna on the delay. We derive the expected delay in sending one packet to all users $D_{1,n}$ when the channel has a memory of two. For a system with M transmit antennas, we obtain the moment generating function for $D_{1,n}$. For large n and when M does not grow faster than $\log n$, the expected delay behaves like $\frac{n \log n}{M + O(\log n/n)}$. Finally we propose an algorithm that without sacrificing too much on the throughput can significantly improve the delay. The algorithm always considers the first d user with the best channel conditions and transmits to the one that has received the least number of packets.

There are still questions remain to be answered. For example, in the model we considered, all the users always have a packet for transmission, it would be quite interesting to generalize the results to the case where each user have a random rate of arrival and analyze the behavior of the length of the longest queue in among n users. It is worth mentioning that analyzing the average length of the queues can be done using Kingman's result [7] as mentioned in Section 3. However deriving the expected length of the longest queue remains as an interesting open problem.

VIII. APPENDIX A

In this appendix we prove that using $F(z)$, we can generate all the moments of $D_{m,n}$ as in (4). Defining a_i as the

probability of success in sending at least m packet to all users in i channel uses, we may write the mean of $D_{m,n}$ as,

$$E(D_{m,n}) = \sum_{i=0}^{\infty} i a_i \quad (A.1)$$

As b_i was the probability of failure in obtaining m packets in all the receivers in up to and including the i 'th channel uses, it is clear that $a_i = b_{i-1} - b_i$. Therefore, we may write the mean value of $D_{m,n}$ as

$$E(D_{m,n}) = \sum_{i=1}^{\infty} i(b_{i-1} - b_i) = \sum_{i=1}^{\infty} b_i = F(1) \quad (A.2)$$

which leads to the first identity in (4). We can also write the second moment of $D_{m,n}$ as,

$$\begin{aligned} E(D_{m,n}^2) &= \sum_{i=1}^{\infty} i^2 a_i = \sum_{i=1}^{\infty} i^2 (b_{i-1} - b_i) \\ &= \sum_{i=0}^{\infty} (i+1)^2 b_i - \sum_{i=1}^{\infty} i^2 b_i \\ &= \sum_{i=0}^{\infty} (2i+1) b_i = 2F'(1) + F(1) \end{aligned}$$

which completes the proof of (4). We can similarly prove that the i 'th moment of $D_{m,n}$ can be written as

$$E(D_{m,n}^k) = \sum_{i=0}^{\infty} (i+1)^k b_i - \sum_{i=0}^{\infty} i^k b_i = \sum_{j=0}^{k-1} \binom{k}{j} \sum_{i=0}^{\infty} i^j b_i \quad (A.3)$$

It is worth noting that the inner summation can be found by taking the derivatives of $F(z)$. For example,

$$\sum_{i=0}^{\infty} i^2 b_i = \frac{d}{dz} \left(z \frac{dF(z)}{dz} \right) \Big|_{z=1} \quad (A.4)$$

Therefore, Eq. (A.3) and (A.4) imply that all the moments of $D_{m,n}$ can be obtained in terms of derivatives of $F(z)$ at $z = 1$.

REFERENCES

- [1] X. Qin and R. Berry, "Exploiting multiuser diversity for medium access control in wireless networks," in *Proc. of INFOCOM*, 2003, pp. 1084–1094.
- [2] S. Shamai and E. Telatar, "Some information theoretic aspects of decentralized power control in multiple access fading channels," in *Proc. Information Theory and Networking Workshop*, 1999.
- [3] J. I. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Info.*, vol. 25, no. 9, pp. 505–515, Sep. 1979.
- [4] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Trans. Inform.*, vol. 39, no. 2, Mar. 1993.
- [5] A. Eryilmaz, R. Srikant, and J. Perkins, "Stable scheduling policies for broadcast channels," in *Proc. IEEE Inter. Symp. Info.*, July 2002, p. 382.
- [6] M. Andrew, K. Kumaran, K. Ramanan, A. Stoytar, P. Whiting, and R. Vijaykumar, "Providing quality of service over a shared wireless link," *IEEE Communications magazine*, vol. 39, no. 2, pp. 246–251, Feb. 2001.
- [7] J. F. Kingman, "Inequalities in the theory of queues," *Journal of the Royal Statistical Society: Series B*, vol. 32, no. 1, pp. 102–110, Jan. 1970.

- [8] M. J. Ferguson, "On the control, stability, and waiting time in a slotted ALOHA random access system," *IEEE Trans. Comm.*, vol. 23, no. 10, Oct. 1975.
- [9] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Vehic. Tech.*, vol. 43, no. 2, pp. 359–378, May 1994.
- [10] G. Caire, G. Taricco, and E. Biglieri, "Optimum power allocation over fading channels," *IEEE Trans. Info.*, vol. 45, no. 5, pp. 1468–1489, July 1999.
- [11] A. Ephremides and R. Zhu, "Delay analysis of interacting queues with an approximate model," *IEEE Trans. Comm.*, vol. 35, no. 2, Feb. 1987.
- [12] W. Feller, *An introduction to probability theory and its applications*, John Wiley and Sons, Inc., 1967.
- [13] N. L. Johnson and S. Kotz, *Urn models and their application*, John Wiley and Sons, Inc., 1977.
- [14] D. J. Newman and L. Shepp, "The double dixie cup problem," *Amer. Math. Monthly*, vol. 67, no. 1, pp. 58–61, Jan. 1960.
- [15] M. Sharif and B. Hassibi, "A delay analysis for opportunistic transmission in fading broadcast channel," *to be submitted to IEEE Trans. Inform.* (download available at www.its.caltech.edu/~masoud), 2004.
- [16] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press Inc., London, 1965.
- [17] P. Viswanath and D. N. Tse, "Sum capacity of the vector Gaussian broadcast channel and downlink-uplink duality," *IEEE Trans. Inform.*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.
- [18] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform.*, vol. 49, no. 7, pp. 1691–1706, July 2003.
- [19] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum rate capacity of Gaussian MIMO broadcast channel," *submitted to IEEE Trans. Inform.*, 2002.
- [20] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *submitted to IEEE Trans. Comm.* (download available at www.its.caltech.edu/~masoud), 2004.
- [21] M. Sharif and B. Hassibi, "On the capacity of MIMO BC channel with partial side information," *submitted to IEEE Trans. Inform.* (download available at www.its.caltech.edu/~masoud), 2003.
- [22] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform.*, vol. 48, no. 6, pp. 1277–1294, June 2002.