Cross-Layer Resource Allocation via Geometric Programming in Fading Broadcast Channels

Kibeom Seong
Dept. of Electrical Engineering
Stanford University
Stanford, CA 94305 USA
Email: kseong@stanford.edu

Ravi Narasimhan
Dept. of Electrical Engineering
University of California
Santa Cruz, CA 95064 USA
Email: ravi@soe.ucsc.edu

John M. Cioffi
Dept. of Electrical Engineering
Stanford University
Stanford, CA 94305 USA
Email: cioffi@stanford.edu

***Abstract*— In a fading broadcast channel (BC), *Queue Proportional Scheduling* (QPS) is presented via geometric programming (GP). Given the current queue state, QPS allocates a data rate vector such that the expected rate vector averaged over all fading states is proportional to the current queue state vector as well as on the boundary of the ergodic BC capacity region. One well known throughput optimal policy for a fading BC is *Maximum Weight Matching Scheduling* (MWMS), which maximizes the inner product of the queue state vector and achievable rate vector. Simulation results for Poisson packet arrivals and exponentially distributed packet lengths demonstrate that QPS provides a significant decrease in average queuing delay compared to MWMS. In addition, QPS is shown to guarantee fairness among users in terms of average queuing delay.**

I. Introduction

In multiuser communication systems, the transmit power and rate of each user are often determined based on the channel capacity region. This information theoretic approach to resource allocation, which ignores the randomness in packet arrivals and queuing, cannot guarantee stability of queuing systems. The network capacity region is defined as a set of all packet arrival rate vectors for which queue lengths can remain finite [1]. Though an arrival rate vector is within the network capacity region, resource allocation based only on channel state information may cause a certain user's backlog to become unacceptably large, resulting in long queuing delay as well as packet loss.

To account for queuing parameters, a cross-layer approach to resource allocation has been recently proposed in [2], [3], [4] and the references therein. These works show that consideration of both channel and queue states allows the entire network capacity region to be achieved in fading broadcast channels (BC) and multiple-access channels (MAC). A scheduling policy that achieves the network capacity region is called throughput optimal. One well-known throughput optimal scheduling algorithm for the fading BC and MAC is *Maximum Weight Matching Scheduling* [2]. This policy achieves throughput optimality by allocating power and data rate that maximize the inner product of the queue state vector and achievable rate vector. Recent applications of MWMS can be also found in OFDM downlink systems [5] and MIMO downlink systems [4]. For the fading MAC, [6] shows that MWMS minimizes the sum of each user's average queuing delay if symmetric channels and equal packet arrival rates are assumed. This property is a consequence of the polymatroidal structure of the MAC capacity region [7]. However, there are no such structural properties in the fading BC capacity region so that even with symmetry assumptions, MWMS cannot guarantee the minimum average queuing delay.

In this paper, we propose another throughput optimal scheduling policy for a fading BC, *Queue Proportional Scheduling* (QPS), which has more desirable delay and fairness properties than MWMS. For the current queue state, QPS allocates a data rate vector such that the expected data rate vector averaged over all fading states is a boundary point of the ergodic BC capacity region and is proportional to the current queue state vector. Utilizing degradedness of BC for each fading state as well as convexity of the BC capacity region, QPS is formulated as a geometric program, which is a special form of convex optimization problem with very efficient algorithms [8]. Simulations with Poisson packet arrivals and exponentially distributed packet lengths show that QPS provides significantly smaller average queuing delay than MWMS. Moreover, with the QPS policy, fairness among users can be always guaranteed in terms of average queuing delay.

The organization of this paper is as follows: Section II provides the model of fading broadcast channels and queuing systems. Together with a description of MWMS, the QPS policy is proposed and formulated as a geometric program in Section III. Section IV presents numerical results and discussion, and concluding remarks are given in Section V.

Notation: Vectors are bold-faced. $\mathbb{R}^n$ denotes the set of real n-vectors and $\mathbb{R}^n_+$ denotes the set of nonnegative real n-vectors. The symbol $\succeq$ (and its strict form $\succ$) is used to denote the componentwise inequality between vectors: $\mathbf{x} \succeq \mathbf{y}$ means $x_i \geq y_i$, $i = 1, 2, \cdots, n$.

II. System Model

In this paper, a block fading channel is assumed where the fading state is constant over one scheduling period and each scheduling period undergoes independent and identically distributed (i.i.d.) fading. Also, both transmitter and receivers are assumed to have perfect knowledge of channel state information (CSI) so that the transmitter can perform superposition coding and each receiver can use successive

0-7803-9392-9/06/$20.00 (c) 2006 IEEE

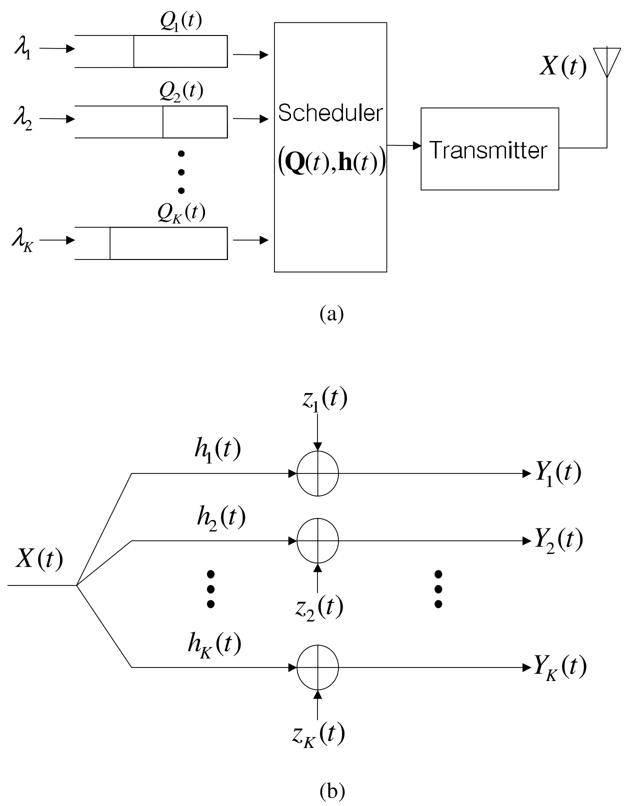

Fig. 1. (a) Block diagram of the queuing system and scheduler. (b) Fading broadcast channel models.

decoding [9]. In addition, the transmitter is assumed to have a peak power constraint of P on each transmission. Consider a Gaussian broadcast channel with a single transmitter sending independent messages to K users over two-sided bandwidth $2W$. At time t, the received signal of user i is expressed as

$$Y_i(t) = h_i(t)X(t) + z_i(t), \quad i = 1, \cdots, K \tag{1}$$

where the transmitted signal $X(t)$ is composed of K independent messages, the complex channel gain of user i is denoted by $h_i(t)$, and $z_i(t)$'s are i.i.d. zero-mean Gaussian noise with power N_0W. As in [10], the channel gain can be combined with the noise component by defining an effective noise $\tilde{z}_i(t) = z_i(t)/h_i(t)$. Then, the equivalent received signal is given by

$$Y_i(t) = X(t) + \tilde{z}_i(t), \quad i = 1, \cdots, K \tag{2}$$

where the power of $\tilde{z}_i(t)$ conditioned on the channel gain is defined as $n_i(t) = N_0W/|h_i(t)|^2$. Without loss of generality, $W = 1$ is assumed throughout this paper for simplicity. The effective noise power $\mathbf{n} = [n_1 \ n_2 \ \cdots \ n_K]^T$ is considered as the denotation of a fading state. We define the ergodic BC capacity region as the set of all long-term average rate vectors achievable in the fading BC with arbitrarily small probability of error. For each fading state, the channel is equivalent to a degraded Gaussian BC described above. A power control policy $\mathcal{P}$ over all possible fading states is defined as a function that maps from any fading state $\mathbf{n}$ to each user's transmit power $P_i(\mathbf{n})$. Let Ω denote the set of all power policies satisfying the sum power constraint P which is given by

$$\Omega = \left\{ \mathcal{P} : \sum_{i=1}^{K} P_i(\mathbf{n}) \le P, \text{ for any } \mathbf{n} \right\}. \tag{3}$$

Under superposition coding and successive decoding with the optimal ordering, the capacity of user i for a fading state $\mathbf{n}$ is

$$R_i\left(\mathbf{P}(\mathbf{n})\right) = \log\left(1 + \frac{P_i(\mathbf{n})}{n_i + \sum_{k=1}^{K} P_k(\mathbf{n})\mathbf{1}\left[n_i > n_k\right]}\right) \tag{4}$$

where $\mathbf{1}\left[\cdot\right]$ is the indicator function which is equal to 1 if its argument is true and 0 otherwise. Then, the Gaussian BC capacity region for the fading state $\mathbf{n}$ and transmit power P is

$$C(\mathbf{n}, P) = \{R_i : R_i \le R_i\left(\mathbf{P}(\mathbf{n})\right), \ i = 1, 2, \cdots, K, \text{ where } \sum_i P_i(\mathbf{n}) = P\}. \tag{5}$$

Let $C_{BC}(\mathcal{P})$ denote the set of achievable rates averaged over all fading states for a power policy $\mathcal{P}$

$$C_{BC}(\mathcal{P}) = \{R_i : R_i \le \mathbf{E}_{\mathbf{n}}\left[R_i(\mathbf{P}(\mathbf{n}))\right], \ i = 1, 2, \cdots, K\}. \tag{6}$$

By the theorem in [10], with the sum power constraint P as well as perfect CSI at the transmitter and receivers, the ergodic capacity region of a fading broadcast channel is given by

$$C_{erg}(P) = \bigcup_{\mathcal{P} \in \Omega} C_{BC}(\mathcal{P}) \tag{7}$$

where the region $C_{erg}(P)$ is convex.

The queuing system and scheduler are modeled as the following. K data sources generate packets according to independent Poisson arrival processes $\{A_i(t), i = 1, \cdots, K\}$, which are stationary counting processes with $\lim_{t\to\infty} A_i(t)/t = a_i < \infty$, and $\mathrm{var}(A_i(t+T) - A_i(t)) < \infty$ for $T < \infty$. Packet lengths in bits $\{X_i\}$ are i.i.d. exponentially distributed and satisfy $\mathbf{E}[X_i] = \gamma_i < \infty$, and $\mathbf{E}[{X_i}^2] < \infty$. We assume packet lengths are independent of packet arrival processes; thus, user i's arrival rate in bits is given by $\lambda_i = a_i\gamma_i$.

The transmitter has K output queues assumed to have infinite capacity. Packets from source i enter queue i and wait until they are served to receiver i. The scheduling period is denoted by T_s, and without loss of generality, we assume $T_s = 1$. At time t, the achievable data rate vector is within the capacity region $C\left(\mathbf{n}(t), P\right)$ defined in (5). $Q_i(t)$ denotes the number of bits waiting to be sent to user i at time t. A time interval $[t, t+1)$, with $t = 0, 1, 2, \cdots$, is denoted by the *time slot* t, and $Z_i(t)$ is defined as the number of arrived bits at user i's queue during the time slot t. Then, after a scheduling period, user i's queue state vector is equal to $Q_i(t+1) = \max\{Q_i(t) - R_i(\mathbf{n}(t), \mathbf{Q}(t)), 0\} + Z_i(t)$. The allocated rate vector at time slot t, $\mathbf{R}(\mathbf{n}(t), \mathbf{Q}(t))$ is determined by the scheduler based on both queue states and channel conditions. Fig. 1 summarizes the overall system.

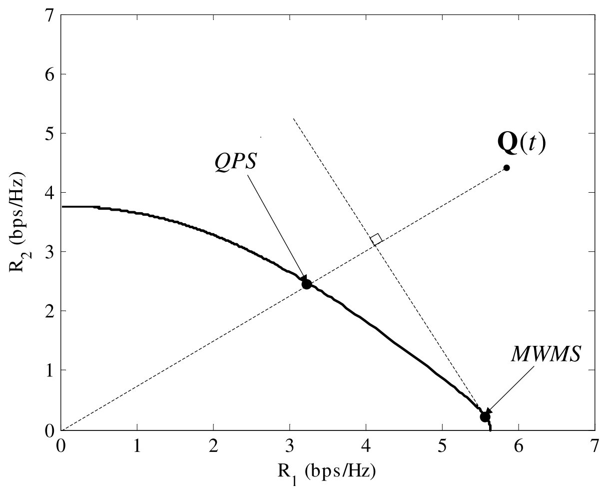

Fig. 2. Ergodic capacity region of two user Rayleigh fading BC, and expected rate vectors of QPS and MWMS when the queue state vector is $\mathbf{Q}(t)$ (User 1's average SNR=13dB and user 2's average SNR=7dB).

III. Queue Proportional Scheduling via Geometric Programming

In this section, QPS is introduced and formulated via geometric programming (GP), which is a convex optimization problem with efficient algorithms to obtain the globally optimal solution. GP uses monomial and posynomial functions. A monomial function has the form of $h(\mathbf{x}) = cx_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$, where $\mathbf{x} \succ 0$, $c \geq 0$ and $a_i \in \mathbb{R}$. A posynomial is a sum of monomials $f(\mathbf{x}) = \sum_k c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}$. Then, GP takes the following form,

$$\begin{array}{lrcl} \text{minimize} & f_0(\mathbf{x}) & & \\ \text{subject to} & f_i(\mathbf{x}) & \leq & 1 \\ & h_j(\mathbf{x}) & = & 1 \end{array} \quad (8)$$

where f_0 and f_i are posynomials and h_j are monomials. Although this is not a convex optimization problem, with a change of variables: $y_i = \log x_i$ and $b_{ik} = \log c_{ik}$, we can convert it into a convex form as the following:

$$\begin{array}{lrcl} \text{minimize} & p_0(\mathbf{y}) & = & \log \Sigma_k \exp(\mathbf{a}_{0k}^T \mathbf{y} + b_{0k}) \\ \text{subject to} & p_i(\mathbf{y}) & = & \log \Sigma_k \exp(\mathbf{a}_{ik}^T \mathbf{y} + b_{ik}) \leq 0 \\ & q_j(\mathbf{y}) & = & \mathbf{a}_j^T \mathbf{y} + b_j = 0. \end{array} \quad (9)$$

This problem can be easily solved by using efficient interior point methods.

As shown in [11], when an average rate vector $\mathbf{R}^* = \mathbf{E_n}[\mathbf{R}(\mathbf{n})]$ lies on the boundary surface of $C_{erg}(P)$ in the fading BC, $\mathbf{R}^*$ is a solution to the optimization problem $\max_{\mathbf{r}} \boldsymbol{\mu}^T \mathbf{r}$ such that $\mathbf{r} \in C_{erg}(P)$ for some $\boldsymbol{\mu} \in \mathbb{R}_+^K$. Also, for a given $\boldsymbol{\mu}$, $\mathbf{R}^*$ solves the above problem if and only if there exists the rate allocation $\mathbf{R}(\mathbf{n}) \in \mathbb{R}_+^K$ such that for any fading state $\mathbf{n}$, $\mathbf{R}(\mathbf{n})$ is a solution to the optimization problem $\max_{\mathbf{r}} \boldsymbol{\mu}^T \mathbf{r}$ where $\mathbf{r} \in C(\mathbf{n}, P)$. At time slot t, MWMS assigns the data rate vector $\mathbf{R}_{MWMS}(\mathbf{n}(t), \mathbf{Q}(t)) \in C(\mathbf{n}(t), P)$ which satisfies

$$\mathbf{E}_{\mathbf{n}(t)}\left[\mathbf{R}_{MWMS}\left(\mathbf{n}(t), \mathbf{Q}(t)\right)\right] = \arg\max_{\mathbf{r}} \sum_{i=1}^{K} \alpha_i Q_i(t) r_i \quad \text{such that } \mathbf{r} \in C_{erg}(P) \quad (10)$$

where $\mathbf{r} = [r_1 \ r_2 \ \cdots \ r_K]^T$ and α_i is the user i's priority weight which is set to 1 for all users if everyone has the same priority. Let $\mathbf{Q}'(t) = [\alpha_1 Q_1(t) \ \cdots \ \alpha_K Q_K(t)]^T$. Then, by the theorem in [11],

$$\mathbf{R}_{MWMS}\left(\mathbf{n}(t), \mathbf{Q}(t)\right) = \arg\max_{\mathbf{r}} \mathbf{Q}'(t)^T \mathbf{r} \quad \text{such that } \mathbf{r} \in C\left(\mathbf{n}(t), P\right). \quad (11)$$

This algorithm tends to allocate higher data rate to the user with longer queue or better channel conditions.

On the other hand, QPS allocates the following data rate vector at time slot t.

$$\mathbf{R}_{QPS}\left(\mathbf{n}(t), \mathbf{Q}(t)\right) \in C\left(\mathbf{n}(t), P\right) \quad \text{such that}$$
$$\mathbf{E}_{\mathbf{n}(t)}\left[\mathbf{R}_{QPS}\left(\mathbf{n}(t), \mathbf{Q}(t)\right)\right] = \mathbf{Q}'(t)\left(\max_{\mathbf{Q}'(t)x \in C_{erg}(P)} x\right) \quad (12)$$

where x is a scalar. Assuming equal priority on each user, $\mathbf{Q}'(t) = \mathbf{Q}(t)$. Then, the average rate vector under the QPS policy, $\mathbf{E}_{\mathbf{n}(t)}[\mathbf{R}_{QPS}(\mathbf{n}(t), \mathbf{Q}(t))]$ is proportional to the queue state vector, and also lies on the boundary surface of the ergodic capacity region. Therefore, for some $\boldsymbol{\mu} \in \mathbb{R}_+^K$, the data rate vector assigned by QPS at time slot t can be expressed as

$$\mathbf{R}_{QPS}\left(\mathbf{n}(t), \mathbf{Q}(t)\right) = \arg\max_{\mathbf{r}} \boldsymbol{\mu}^T \mathbf{r} \quad \text{such that } \mathbf{r} \in C\left(\mathbf{n}(t), P\right). \quad (13)$$

Under the QPS policy, $\boldsymbol{\mu}$ is determined based on the current queue state vector as well as the ergodic BC capacity region. On the other hand, MWMS only considers the queue state vector in the derivation of $\boldsymbol{\mu}$.

For the queue state vector $\mathbf{Q}(t)$, Fig. 2 illustrates two distinct expected rate vectors supported by MWMS and QPS. Two user Rayleigh fading BC is considered where user 1's average signal-to-noise ratio (SNR) is 13dB and user 2's average SNR is 7dB. From Fig. 2, we can expect that as the queue state changes, MWMS exhibits more fluctuations in the average rate vector compared to QPS. According to queuing theory, lower variance in service rate or arrival rate provides smaller queuing delay [12]. Therefore, QPS is expected to have smaller average queuing delay than MWMS as demonstrated in the next section.

By utilizing degradedness of BC for each fading state as well as convexity of the ergodic BC capacity region, the rate allocation of QPS can be formulated as GP. Assume that recent M fading states are sampled, which are denoted by $\left\{\mathbf{n}^{(1)}, \cdots, \mathbf{n}^{(M)}\right\}$. To reduce the correlation among samples, sampling period needs to be determined in consideration of fading coherence time. In this paper, the sampling period is simply assumed equal to one scheduling period due to i.i.d. block fading over each scheduling time. Without loss of generality, $\mathbf{n}^{(M)}$ is assumed to denote the current fading state $\mathbf{n}(t)$. Then, consider a family of M parallel broadcast channels, such that in the mth component channel, user i has noise variance $n_i^{(m)}$, rate and power denoted by $R_i^{(m)}$ and $P_i^{(m)}$. Note that each BC channel has a power constraint of P. At time slot t, QPS allocates the data rate vector $\mathbf{R}_{QPS}(\mathbf{n}^{(M)}, \mathbf{Q}(t))$ that is a solution of the following optimization problem.

$$\frac{1}{M}\sum_{m=1}^{M}\mathbf{R}_{QPS}\left(\mathbf{n}^{(m)}, \mathbf{Q}(t)\right) = \mathbf{Q}(t)\left(\max_{\mathbf{Q}(t)x\in C_{erg}(P)} x\right)$$
$$\mathbf{R}_{QPS}\left(\mathbf{n}^{(m)}, \mathbf{Q}(t)\right) \in C\left(\mathbf{n}^{(m)}, P\right) \quad \text{for all } m \quad (14)$$

From (4) and (5), if $\mathbf{R}^{(m)}$ is on the boundary of the capacity region for the mth component channel, solving power in terms of rate yields

$$\sum_{i=1}^{l} P_{\pi_m(i)}^{(m)} = \sum_{i=1}^{l}\left(n_{\pi_m(i)}^{(m)} - n_{\pi_m(i-1)}^{(m)}\right) \times \exp\left(\ln 2\sum_{j=i}^{K} R_{\pi_m(j)}^{(m)}\right) - n_{\pi_m(l)}^{(m)}, \quad l = 1, \cdots, K \quad (15)$$

where $\pi_m(\cdot)$ is the permutation such that $n_{\pi_m(1)}^{(m)} < n_{\pi_m(2)}^{(m)} < \cdots < n_{\pi_m(K)}^{(m)}$ and $n_{\pi_m(0)}^{(m)} \equiv 0$. Then, as shown in [10], the capacity region of the mth Gaussian BC can be expressed as

$$C(\mathbf{n}^{(m)}, P) = \{R_{\pi_m(i)}^{(m)} : \sum_{i=1}^{K}\left(n_{\pi_m(i)}^{(m)} - n_{\pi_m(i-1)}^{(m)}\right) \times \exp\left(\ln 2\sum_{j=i}^{K} R_{\pi_m(j)}^{(m)}\right) - n_{\pi_m(K)}^{(m)} \le P \text{ and } R_{\pi_m(i)}^{(m)} \ge 0,\ i = 1, 2, \cdots, K\}. \quad (16)$$

Using this relation, (14) can be converted into

$$\begin{aligned}
\text{minimize} \quad & \log(\exp(-x)) \\
\text{subject to} \quad & \log\left(\exp\left(-R_i^{(m)}\right)\right) \le 0, \quad \forall\ i \text{ and } m \\
& \log\left(\exp\left(-Q_i^{(M)}\right)\exp\left(R_i^{(M)}\right)\right) \le 0, \quad \forall\ i \\
& \log\sum_{i=1}^{K}\left(\frac{n_{\pi_m(i)}^{(m)} - n_{\pi_m(i-1)}^{(m)}}{P + n_{\pi_m(K)}^{(m)}}\right) \times \exp\left(\ln 2\sum_{j=i}^{K} R_{\pi_m(j)}^{(m)}\right) \le 0, \quad \forall\ m \\
& \mathbf{Q}(t)x - \frac{1}{M}\sum_{m=1}^{M}\mathbf{R}^{(m)} = 0 \qquad (17)
\end{aligned}$$

where the second constraint is added to avoid allocating redundant power to some users with short queue lengths. If the optimization variable is defined as $\mathbf{y} = [x\ (\mathbf{R}^{(1)})^T\ \cdots\ (\mathbf{R}^{(M)})^T]^T \in \mathbb{R}^{(KM+1)\times 1}$, (17) is the standard geometric program with the globally optimal solution $\mathbf{y}^* = [x^*\ (\mathbf{R}^{*(1)})^T\ \cdots\ (\mathbf{R}^{*(M)})^T]^T$. Thus, the data rate vector supported under the QPS policy is $\mathbf{R}_{QPS}(\mathbf{n}^{(M)}, \mathbf{Q}(t)) = \mathbf{R}^{*(M)}$, and the corresponding power allocation can be obtained by solving (15) for $m = M$.

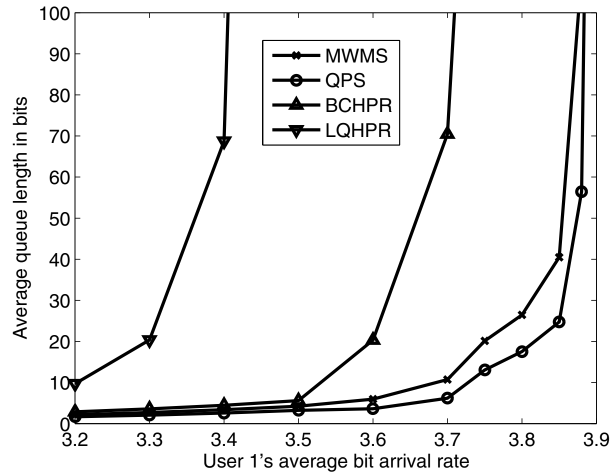

Fig. 3. Average queue length vs user 1's bit arrival rate under four scheduling policies (2 users, user 1's average SNR=13dB and user 2's average SNR=7dB, $\lambda_2 = 0.5\lambda_1$).

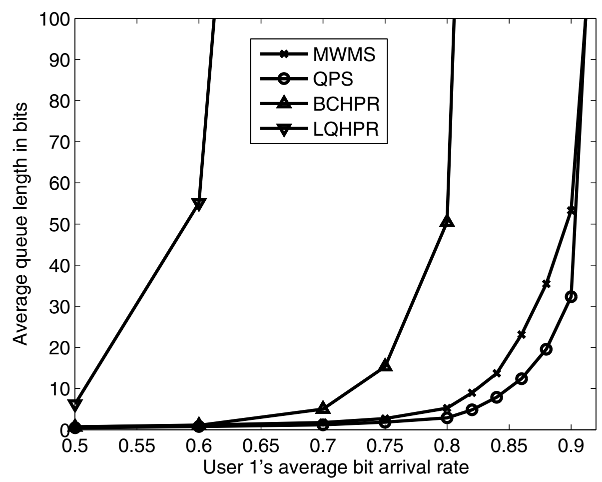

Fig. 4. Average queue length vs user 1's bit arrival rate under four scheduling policies (10 users, user i's average SNR (dB) $= 20 - (i-1)$ and $\lambda_i = \lambda_1$ for $i = 1, \cdots, 10$).

IV. Numerical results and discussion

This section presents simulation results with Poisson packet arrivals and exponentially distributed packet lengths to demonstrate stability, delay, and fairness properties of the QPS algorithm. In the simulation, average packet length for each user, scheduling period, and signal bandwidth are all equal to 1, and noise power is 0.1. In Fig. 3 and Fig. 4, average queue lengths over K users defined as $\lim_{t\to\infty}\mathbf{E}[\frac{1}{K}\sum_{i=1}^{K} Q_i(t)]$ are evaluated for different values of λ_1 when $K = 2$ and $K = 10$, respectively. Four scheduling algorithms are compared in both figures: QPS, MWMS, Longest Queue Highest Possible Rate

(LQHPR) and Best Channel Highest Possible Rate (BCHPR) [2]. LQHPR allocates full power to a user with the longest queue. Under the BCHPR policy, a user with better channel condition takes higher priority in rate allocation, and user i can be served only if user i has the best channel or some transmit power remains after clearing queue backlogs of users with higher priorities than user i.

For the two user case in Fig. 3, the Rayleigh fading BC channel presented in Fig. 2 is considered where the total power constraint $P = 2$, user 1's average SNR=13dB, and user 2's average SNR=7dB. Also, the bit arrival rate vector satisfies $\boldsymbol{\lambda} = \lambda_1[1\ 0.5]^T$. From Fig. 2, $\boldsymbol{\lambda} \in \text{int}\ C_{erg}(P)$ if and only if $\lambda_1 < 3.9$. Fig. 3 demonstrates that the average queue length of QPS is about 30% smaller than that of MWMS for any $\lambda_1 < 3.9$. Since MWMS is a throughput optimal policy, this observation corroborates throughput optimality of QPS, which is proved in [13]. Fig. 3 also shows that LQHPR and BCHPR, which are not throughput optimal, have much longer average queue lengths than MWMS. Simulation results with 10 users are presented in Fig. 4. The total transmit power is $P = 10$ and user i's average SNR is equal to $20 - (i - 1)$ (dB) for $i = 1, \cdots, 10$. Also, the bit arrival rate is identical for every user. With $\lambda_1 < 0.92$, QPS provides about 40-50% smaller average queue length than MWMS, which is a greater difference compared to the two user case.

The fairness properties of QPS, MWMS and BCHPR with 10 users are illustrated in Fig. 5. User i's average SNR is equal to $20 - 0.5(i - 1)$ (dB) and $\lambda_i = 1.55(0.9)^{i-1}$ for $i = 1, \cdots, 10$. Fig. 5 presents each user's average queuing delay in slots for the above three scheduling policies. It is observed that fairness among users is not satisfied under the BCHPR, which provides intolerably long average queuing delay for users with worse channel conditions. MWMS tends to equalize each user's average queue length. Since each user has a different arrival rate, by Little's theorem [14], MWMS provides smaller average queuing delay for the user with higher arrival rate. On the other hand, the average queue length of QPS is shown to be proportional to the arrival rate vector so that each user's average queuing delay is equalized. Therefore, under the QPS policy, fairness among users can be guaranteed in terms of average queuing delay.

V. Conclusion

In fading broadcast channels, *Queue Proportional Scheduling* (QPS) is shown to have more desirable delay and fairness properties than *Maximum Weight Matching Scheduling* (MWMS). Utilizing degradedness of BC and convexity of the BC capacity region, QPS is formulated as a geometric program whose globally optimal solution can be obtained with very efficient algorithms. Numerical results demonstrate that QPS provides significantly smaller average queuing delay compared to MWMS for any arrival rate vector within the network capacity region. In addition, QPS is shown to satisfy fairness among users in terms of average queuing delay.

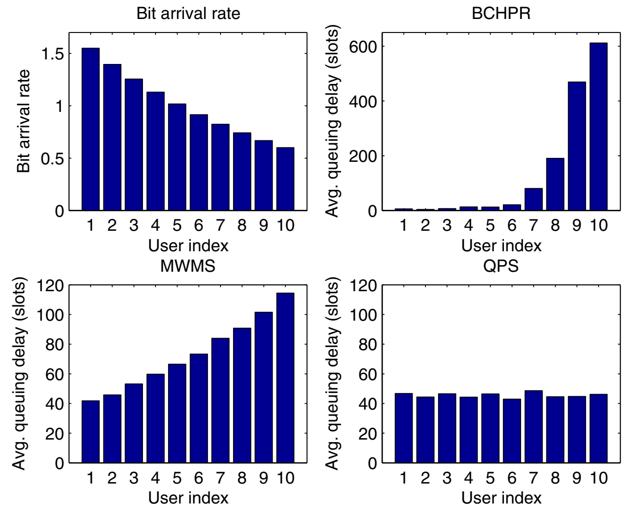

Fig. 5. Each user's average queuing delay under QPS, MWMS and BCHPR (10 users, user i's average SNR (dB) $= 20 - 0.5(i-1)$ and $\lambda_i = \lambda_1(0.9)^{i-1}$ for $i = 1, \cdots, 10$).

References

[1] M. Neely, E. Modiano and C. E. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Trans. Networking*, vol. 11, no. 1, pp 138-152, Feb 2003.

[2] R. Barry and E. Yeh, "Cross-layer wireless resource allocation," *IEEE Signal Processing Magazine*, pp 59 - 68, Sept. 2004.

[3] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, and P. Whiting, " Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, pp. 150-154, Feb. 2001.

[4] C. Swannack, E. Uysal-Biyikoglu, and G. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," *Proc. Allerton Conference*, 2004.

[5] G. Song, Y. Li, L. Cimini Jr, and H. Zheng, "Joint channel-aware and queue-aware data scheduling in multiple shared wireless channels," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Atlanta, Georgia, 2004.

[6] E. Yeh and A. Cohen, "Throughput and delay optimal resource allocation in multi-access fading channels," in *Proceedings of the International Symposium on Information Theory*, Yokohama, Japan, 2003, p. 245.

[7] D. Tse and S. Hanly, "Multi-access fading channels: Part I," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2796-2831, 1998.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2003.

[9] T. Cover and J. Thomas, *Elements of Information Theory*, New York: John Wiley and Sons Inc., 1997.

[10] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels-Part I: Ergodic capacity," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1083-1102, Mar. 2001.

[11] D. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," *Proc. of International Symposium on Information*, Ulm, Germany, June 1997, p.27.

[12] S. Asmussen, *Applied Probability and Queues*, New York: Springer, 2000.

[13] K. Seong, R. Narasimhan, and J. Cioffi, "Queue proportional scheduling in Gaussian broadcast channels," *Proc. IEEE International Conference on Communications*, Istanbul, Turkey, 2006.

[14] D. Bertsekas and R. Gallager, *Data Networks*. NJ: Prentice Hall, 1992.