This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

Downlink Power Control for VBR Video Streaming in Cellular Networks: A Majorization Approach

Yingsong Huang, Shiwen Mao, and Yihan Li
Department of Electrical and Computer Engineering, Auburn University, Auburn, AL

***Abstract*—In this paper, we investigate the problem of optimal power control for multiuser variable bit rate (VBR) video streaming in a cellular network with orthogonal channels. We adopt a deterministic model for VBR video traffic that incorporates video frame and playout buffer characteristics, and formulate a *constrained stochastic optimization* problem. We then develop a majorization-based solution approach. For the case of a single VBR video session with relaxed peak power constraint, we develop a power optimal algorithm with low complexity. We prove the power optimality of the proposed algorithm and the uniqueness of the global optimum, and demonstrate that the proposed algorithm is also smoothness optimal. For the case of multiuser VBR video streaming, we develop a heuristic algorithm that selectively suspends some video sessions when the peak power constraint is violated. The proposed algorithms are evaluated with trace-driven simulations, and are shown to achieve considerable power savings and improved video quality over a conventional "lazy" scheme.**

I. Introduction

With the wide availability of intelligent handheld devices and smart phones, there is a compelling need to support video in traditional cellular networks. Such applications not only are bandwidth intensive, but also involve user quality of experiences. They will greatly stress the capacity of traditional cellular networks and call for joint design and optimization of networking and control mechanisms across multiple layers.

Among various forms of compressed videos, *variable bit rate* (VBR) videos can offer constant and better quality over constant bit rate (CBR) videos with the same bit budget; a highly desirable advantage for video applications [1]. However, VBR videos also pose great challenges to network scheduling and control, due to the high variability and complex autocorrelation structure in VBR video traffic [2].

In our prior work [1], we investigate the problem of supporting VBR videos in a multi-cell network, where capacity is limited by inter-cell interference. We develop a reformulation-linearization technique (RLT) approach as well as a distributed algorithm based on dual decomposition. In this paper, we focus on the downlink of a cell with orthogonal channels, where the base station (BS) streams multiple VBR videos to mobile users in the cell. We consider buffer underflows (causing stalled display) and overflows (causing missing frames and error propagations in the following received and decoded frames) as user viewing performance measure, and aim to minimize the total power consumption at the BS.

We adopt a deterministic model for VBR video traffic that incorporates video frame and playout buffer characteristics, and formulate a *constrained stochastic optimization* problem. We show that the problem fits well with *majorization theory*, which concerns with partial ordering of real vectors and order-preserving functions, and develop a majorization-based solution approach. First, we prove that the objective function of the formulated problem is *Schur-convex* with the order-preserving property. Secondly, we investigate the case of a single VBR video session with relaxed peak power constraint. A majorization-based power optimal algorithm with low complexity is developed. We prove the power optimality of the proposed algorithm and the uniqueness of the global optimum, and demonstrate that the proposed algorithm is also smoothness optimal. Thirdly, we investigate the case of multiuser VBR streaming, where power allocations for the users are coupled with the peak power constraint. A heuristic algorithm is developed to selectively suspend some video sessions, if they will not incur underflow in the next time slot when the peak power constraint is violated. The proposed algorithms are evaluated with trace-driven simulations, and are shown to achieve considerable power savings over a conventional "lazy" scheme. This finding is somewhat *counter-intuitive*, since the lazy scheme seems to be energy efficient by only transmitting the minimal amount of video data in each time slot.

The rest of this paper is organized as follows. The system model and problem statement are presented in Section II. We transform the problem into a majorization problem in Section III. The proposed algorithms are described in Section IV and simulation results are presented in Section V. We review related work in Section VI. Section VII concludes the paper.

II. System Model and Problem Statement

A. Network and Video System Model

We consider the downlink of a cellular network. There are N mobile users in the cell. Let $\mathcal{U} = \{1, 2, \cdots, N\}$ denote the set of users. A BS streams multiple VBR videos to the mobile users. Each user occupies a downlink channel, which is a spectral/time resource slot, the nature of which depends on the specific multiple access technique adopted. We assume that the downlink channels within a cell are orthogonal, due to perfect synchronization of the spreading codes or the use of guard times or frequencies.

We investigate the challenging problem of streaming multiple VBR videos. We assume the wireline segment of a video session path is reliable with sufficient bandwidth, while the last-hop wireless link is the bottleneck. Thus the corresponding video data is always available at the BS before the scheduled transmission time. It is non-trivial to accurately model

978-1-4244-9268-8/11/$26.00 ©2011 IEEE

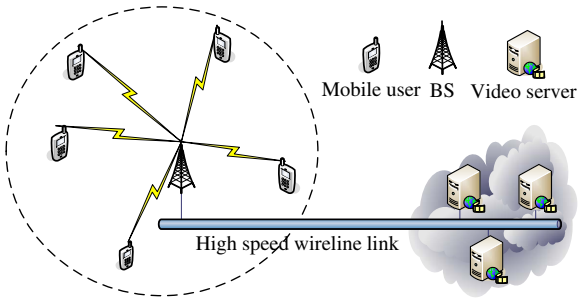


Fig. 1. Cellular network and video streaming system model.

VBR video traffic, which exhibits both strong asymptotic self-similarity and short-range correlation [2]. To this end, we adopt a *deterministic model* that considers frame sizes, frame intervals, and playout buffers [3]. Let $D_n(t)$ denote the *cumulative consumption curve* of the n-th user, representing the cumulative amount of bits consumed by the decoder at time t. The cumulative consumption curve is determined by video characteristics such as frame sizes and frame rates. User n has a playout buffer of b_n bits. Its video has T_n frames, which translates to Ω_n bits in total. We can derive a *cumulative overflow curve* for user n as

$$B_n(t) = \min\{D_n(t-1) + b_n, D_n(T_n)\},\ 0 \le t \le T_n. \quad (1)$$

$B_n(t)$ is the maximum number of received bits at time t without overflowing user n's playout buffer. Finally we define *cumulative transmission curve* $X_n(t)$ as the total amount of bits transmitted to user n at time t.

The three curves are illustrated in Fig. 2. We assume time is divided into time slots, and power control is performed on time slot units. Note that the frame intervals of the VBR videos do not have to be aligned, and time slots do not have to be aligned with the frame intervals. To simplify notation, we assume the frames are aligned and the duration of a time slot is τ in the rest of this paper. A feasible transmission schedule will produce a cumulative transmission curve $X_n(t)$ that lies in between $D_n(t)$ and $B_n(t)$, i.e., causing neither underflow nor overflow at the playout buffer. In practice, $D_n(t)$'s are known for stored videos and can be delivered to the BS as metadata during the session setup phase, and $B_n(t)$'s are then derived as in (1) based on playout buffer size information.

The BS allocates a transmit power to each user in each time slot. Let $\vec{P}(t) = [P_1(t), \cdots, P_N(t)]$ be the power allocation in time slot t. The Signal to Interference-plus-Noise Ratio (SINR) at user n can be written as $\gamma_n(t) = G_n(t)P_n(t)/\eta_n(t)$, where $G_n(t)$ is the path gain from BS to user n and $\eta_n(t)$ is the the noise power at user n in time slot t. We assume block fading channels, where the $G_n(t)$'s are i.i.d. random variables with a certain distribution, for $t = 1, \cdots, T_n$ [4]. The downlink data rate can be written as $c_n(t) = B_w \log(1 + \kappa\gamma_n(P_n(t)))$, where B_w is the channel bandwidth and κ depends on the transceiver design, such as modulation and channel coding. Without loss of generality, we use the Shannon capacity as an upper bounding approximation:

$$c_n(t) = B_w \log(1 + \gamma_n(P_n(t))). \quad (2)$$

Once the link capacity is determined, $c_n(t)\tau$ bits of video data will be delivered to user n in that time slot. The cumulative transmission curve $X_n(t)$ can be written as

$$X_n(0) = 0;\ \ X_n(t) = X_n(t-1) + c_n(t)\tau. \quad (3)$$

A feasible transmission schedule should cause neither playout buffer underflow nor overflow, i.e., satisfying $D_n(t) \le X_n(t) \le B_n(t)$, for all t and for all n.

B. Problem Statement

The problem is to find the optimal feasible transmission schedules $\{X_n(t),\ 0 < t \le T_n\}$, for all user $n \in \mathcal{U}$, such that the total transmit power consumption can be minimized. From (2), the required transmit power for user n is

$$P_n(t) = (2^{c_n(t)/B_w} - 1)\eta_n(t)/G_n(t). \quad (4)$$

A peak power constraint may be applied at the BS, i.e., $\sum_{n\in\mathcal{U}} P_n(t) \le \bar{P}$, for all t. We then formulate the following *constrained stochastic optimization problem*, to minimize the expectation of the total transmit power.

$$\text{minimize} \ \ \textstyle\sum_{n\in\mathcal{U}} \sum_{t=1}^{T_n} \mathbb{E}[P_n(t)] \quad (5)$$

$$\text{subject to:} \ \ D_n(t) \le X_n(t) \le B_n(t), \text{for all } n, t \quad (6)$$

$$\textstyle\sum_{n\in\mathcal{U}} P_n(t) \le \bar{P}, \text{for all } t. \quad (7)$$

Due to orthogonal channels, transmission in one channel does not interfere with those in other channels. We first relax the peak power constraint (7) (i.e., the case when $\bar{P}$ is large). Then, Problem (5) can be decomposed into N sub-problems, each minimizing the transmit power of a video session.

$$\text{minimize} \ \ \textstyle\sum_{t=1}^{T_n} \mathbb{E}[P_n(t)], \text{for all } n \in \mathcal{U} \quad (8)$$

$$\text{subject to:} \ \ D_n(t) \le X_n(t) \le B_n(t), \ \text{for all } n, t. \quad (9)$$

For given $B_n(t)$ and $D_n(t)$, the feasible transmission schedule satisfying (9) is not unique. The i-th feasible transmission schedule is a piece-wise linear curve that can be represented as a vector $\vec{C}_n^i = [c_n^i(1), \cdots, c_n^i(T_n)]$, where $c_n^i(t) \ge 0$ is the data rate in time slot t, for all t. Let $\vec{C}_n^* = [c_n^*(1), \cdots, c_n^*(T_n)]$ be the optimal solution to (8). For a given VBR video, all the feasible transmission schedules transmit the same amount of video data, i.e., $\sum_{t=1}^{T_n} c_n^i(t) = \sum_{t=1}^{T_n} c_n^*(t) = \Omega_n$, for all i. Furthermore, the total transmit power for a feasible schedule can be viewed as a function $\Phi : \mathcal{R}^{T_n} \to \mathcal{R}$ with

$$\Phi(\vec{C}_n^i) = \textstyle\sum_{t=1}^{T_n} (2^{c_n^i(t)/B_w} - 1)\eta_n(t)/G_n(t). \quad (10)$$

Given such an interpretation of Problem (8), the objective is to find an optimal feasible vector $\vec{C}_n^*$, such that its total power P_n^*, obtained through the mapping $\Phi(\cdot)$, is the minimum among all feasible vectors $\vec{C}_n^i$. This interpretation fits well with the *majorization* theory, which provides useful order preserving results for inequality problems [5]. Applying these results, we design an optimal algorithm for solving the decomposed sub-problem (8) in Section IV-A. We will examine then the case of multiuser VBR video streaming coupled with the peak power constraint in Section IV-B.

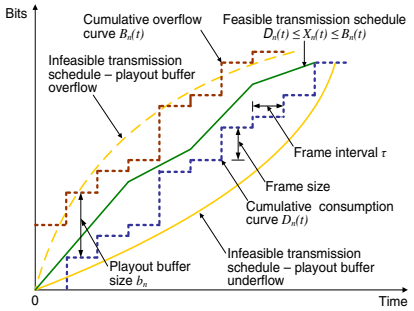


Fig. 2. Transmission schedules for video n.

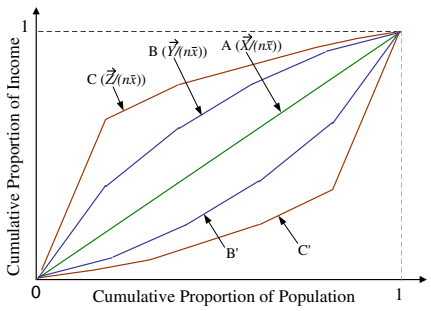


Fig. 3. Lorenz curves.

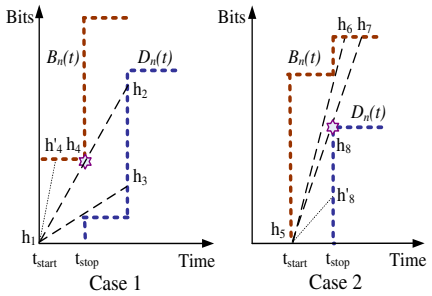


Fig. 4. Determine the next transmission rate.

III. Inequality and Theory of Majorization

A. Majorization Preliminaries

Majorization theory formalizes the intuitive notion of inequality and describes when the components of a vector are "less spread out" or "more nearly equal" than the components of another vector [5]. It concerns with how to order vectors with nonnegative real components and order-preserving functions. Majorization theory has been used to solve communications and networking problems [3], [6], [7]. In this paper, we apply majorization to the problem of optimal power control for streaming VBR videos. For simplicity, all the vectors in this section are row vectors.

Definition 1: Consider two n-dimensional vectors $\vec{X}$ and $\vec{Y}$ with nonnegative real elements. Let the elements be ordered non-increasingly and re-indexed as $x_1 \ge \cdots \ge x_n \ge 0$ for $\vec{X}$ and $y_1 \ge \cdots \ge y_n \ge 0$ for $\vec{Y}$. $\vec{X}$ is said to be majorized by $\vec{Y}$, denoted as $\vec{X} \prec \vec{Y}$, if $\sum_{i=1}^{t} x_i \le \sum_{i=1}^{t} y_i$, $t = 1, \cdots, n-1$ and $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ [5].

For example, we have $[1/n, \cdots, 1/n] \prec [1/(n-1), \cdots, 1/(n-1), 0] \prec [1/2, 1/2, 0, \cdots, 0] \prec [1, 0, \cdots, 0]$. It can be seen that the total weight of 1 becomes more and more concentrated into fewer and fewer elements in the above majorized vector sequence. When such partial ordering is obtained, an *order-preserving function* can be applied to the vectors and the outcomes remain monotonic with respect to the partial order of the vectors.

Definition 2: A real-valued function $\phi(\cdot)$ defined on a set $\mathcal{S} \subset \mathcal{R}^n$ is said to be *Schur-convex* on $\mathcal{S}$, if $\vec{X} \prec \vec{Y}$ implies $\phi(\vec{X}) \le \phi(\vec{Y})$, for all $\vec{X}, \vec{Y} \in \mathcal{S}$ [5].

Schur-convex functions have the order-preserving property, which makes majorization useful for solving optimization problems. The following fact can be used to verify whether a function is Schur-convex.

Fact 1: If $\phi(\cdot)$ is symmetric and convex, then ϕ is Schur-convex. Consequently, $\vec{X} \prec \vec{Y}$ implies $\phi(\vec{X}) \le \phi(\vec{Y})$ [5].

If we choose $\phi = \sum_i f(x_i)$ and $f(\cdot)$ is continuous and convex, then we have the following strong fact.

Fact 2: $\sum_i f(x_i) \le \sum_i f(y_i) \Leftrightarrow \vec{X} \prec \vec{Y}$ holds true for all continuous convex functions $f : \mathcal{R} \to \mathcal{R}$ [5].

We can extend the concept of majorization to the case of a set of vectors or multiple segments of a long vector. The proof is omitted due to lack of space.

Lemma 1: Consider $\vec{X} = [\vec{X}_1, \cdots, \vec{X}_K]$ and $\vec{Y} = [\vec{Y}_1, \cdots, \vec{Y}_K]$, where each element vector $\vec{X}_i$ and $\vec{Y}_i$ has dimension J_i and satisfying $\vec{X}_i \prec \vec{Y}_i$, $i = 1, \cdots, K$. Then we have $\vec{X} \prec \vec{Y}$.

Consider three real vectors $\vec{X} = [\bar{x}, \cdots, \bar{x}]$, $\vec{Y} = [y_1, \cdots, y_n]$, and $\vec{Z} = [z_1, \cdots, z_n]$, where $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} z_i = n\bar{x}$. If the elements in each vector are non-increasing, we can plot the cumulative curves for the normalized vectors $\vec{X}/(n\bar{x})$, $\vec{Y}/(n\bar{x})$ and $\vec{Z}/(n\bar{x})$, which are piece-wise linear curves A, B, and C in Fig. 3. The i-th point on the cumulative curve is the sum of the first i elements of the corresponding vector. If each element of the vectors is the income of an individual, these are known as *Lorenz Curves* that evaluate social income inequality [8]; each point $(a\%, b\%)$ on the curve indicates the top a% of the population earn b% of the income.

In Fig. 3, the straight line A is for vector $\vec{X}/(n\bar{x})$, i.e., equal distribution. Unequal distributions $\vec{Y}/(n\bar{x})$ and $\vec{Z}/(n\bar{x})$ produce curves B and C, respectively, which are bent in the middle. As the bow-shaped curves are bent, concentration increases. We call B and C *concave bow curves*. The bow curve closer to A represents more even distribution [5], which leads to $\vec{X} \prec \vec{Y} \prec \vec{Z}$. Furthermore, we can permute $\vec{Y}$ and $\vec{Z}$ by reordering their elements, to obtain new vectors $\vec{Y}'$ and $\vec{Z}'$ as shown in Fig. 3. We call the resulting curves B' and C' *convex bow curves*. Since the order of elements in a vector plays no role in majorization (see Definition 1), $\vec{X} \prec \vec{Y}' \prec \vec{Z}'$ still holds true in this convex bow curve case.

B. Schur-convexity of Problem (8)

As discussed, Problem (8) fits well with majorization theory with a mapping function (10). To solve the problem, we need to find the optimal rate vector $\vec{C}_n^*$ that is majorized by all other feasible transmission rate vectors $\vec{C}_n^i$, as $\vec{C}_n^* \prec \vec{C}_n^i$, for all i. If the mapping (10) is Schur-convex, then the total transmit power to achieve $\vec{C}_n^*$ will also be dominated by those of other feasible transmission rate vectors. That is, the minimum power is found for Problem (8). Due to random path gains and noise powers, stochastic majorization (rather than ordinary majorization) should be used, which investigates the inequality properties related to random variables [5]. We have the following theorem for the mapping (10) in Problem (8). The proof is omitted due to lack of space.

Theorem 1: The objective function of (8) is an increasing Schur-convex function.

With Theorem 1, solving Problem (8) is equivalent to finding the optimal rate vector $\vec{C}_n^*$, such that $\vec{C}_n^* \prec \vec{C}_n^i$, for all i. Then the total power associated with $\vec{C}_n^*$ is the minimum since the mapping (10) is order-conserving. As discussed in Section III-A, the feasible rate vector that is closest to equal distribution (i.e., curve A in Fig. 3) will be majorized by all other feasible rate vectors. Therefore, we transform Problem (8) to finding a transmission schedule with the most evenly distributed rates for all the time slots.

IV. Power Optimal Algorithm

A. Optimal Algorithm for Problem (8)

From Section III-A, an evenly distributed rate vector $\vec{C}_n^{opt} = [\Omega_n/(T_n\tau), \cdots, \Omega_n/(T_n\tau)]$ is majorized by all feasible schedules, i.e., $\vec{C}_n^{opt} \prec \vec{C}_n^i$, for all i. However, due to the high variability of VBR video frames, limited playout buffer size, random path gains and noise powers, $\vec{C}_n^{opt}$ may not always be feasible. In general, each feasible schedule is piece-wise linear with a set of rate changing points, where the rate is increased or decreased to prevent buffer underflow or overflow. $\vec{C}_n^{opt}$ is a special case with no such rate change points.

The algorithm in Table I, termed PMA, can generate a piece-wise linear schedule, while keeping each piece as long as possible and rate variation as small as possible. The operation of the algorithm is illustrated in Fig. 4. Starting from t_{start} (e.g., h_1 in Fig. 4), PMA first computes two probe lines:

- One through the starting point and one of the future corner points of $B_n(t)$, which can go the furthest into the future without causing buffer underflow or overflow (e.g., lines h_1h_2 in Case 1 and h_5h_6 in Case 2 of Fig. 4). The rate of this probe line is $C_{max}(t) = \frac{B_n(t)-X_n(t_{start})}{t-t_{start}}$.
- The other through the starting point and one of the future corner points of $D_n(t)$, which can go the furthest into the future without causing buffer underflow or overflow (e.g., lines h_1h_3 in Case 1 and h_5h_7 in Case 2 of Fig. 4). The rate of this probe line is $C_{min}(t) = \frac{D_n(t)-X_n(t_{start})}{t-t_{start}}$.

All feasible transmission curves should lie in between these two probe lines in order to go that far. Furthermore, when the probe lines end, they hit *both* on either $B_n(t)$ or $D_n(t)$. Otherwise, we can always adjust one of the probe lines to make them go even further into the future. For example, see lines h_1h_3 and h_1h_4' in Case 1 of Fig. 4. We can use line h_1h_2, which goes further into the future, to replace line h_1h_4', and both probe lines hit $D_n(t)$ eventually (also see lines h_5h_6 and h_5h_8' in Case 2).

If both probe lines hit $D_n(t)$ (i.e., Case 1 in Fig. 4), any feasible schedule for this interval will also hit $D_n(t)$, since they must lie in between the two probe lines. We then trace back the upper probe line (i.e., line h_1h_2) to find the latest time when the buffer is full (i.e., point h_4 at time t_{stop}). Then segment h_1h_4 will be chosen as the transmission schedule for this interval, with rate $\frac{B_n(t_{stop})-X_n(t_{start})}{t_{stop}-t_{start}}$.

If both probe lines hit $B_n(t)$ (i.e., Case 2 in Fig. 4), any feasible schedule for this interval will also hit $B_n(t)$. We then trace back the lower probe line (i.e., line h_5h_7) to find the latest time when the buffer is empty (i.e., point h_8 at time t_{stop}). Then segment h_5h_8 will be chosen as the transmission schedule for this interval, with rate $\frac{D_n(t_{stop})-X_n(t_{start})}{t_{stop}-t_{start}}$.

After the transmission schedule for $[t_{start}, t_{stop})$ is determined, we set $t_{start} = t_{stop}$ and repeat the above procedure to find the schedule for the next time interval. In Table I, the algorithm probes for the longest feasible rate starting from t_{start} in Steps 4–10. In Steps 11–14, the transmission rate for the interval $[t_{start}, t_{stop})$ is determined depending on which of the two cases it is as illustrated in Fig. 4. Steps 16–17 are for the case that the rate does not change in the time slot. Step 19 resets the variables to start the computation for the next segment of $X_n(t)$. Finally, Steps 21–23 transmit the frames following the computed schedule.

TABLE I
Power Minimization Algorithm (PMA-1)

1	BS obtains b_n, D_n, B_n, and B_w for all user n;
2	$t = 1, t_{start} = 0, t_{stop} = t_{c_1} = t_{c_2} = 1, C_{min} = 0,$ $C_{max} = \infty$;
3	DO:
4	Calculate $C_{max}(t)$ and $C_{min}(t)$ over interval $[t_{start}, t]$;
5	IF ($C_{min} \le C_{min}(t)$ & $C_{min}(t) \le \min\{C_{max}, C_{max}(t)\}$)
6	$C_{min} = C_{min}(t)$ and $t_{c_1} = t$;
7	END IF
8	IF ($C_{max} \ge C_{max}(t)$ & $C_{max}(t) \ge \max\{C_{min}, C_{min}(t)\}$)
9	$C_{max} = C_{max}(t)$ and $t_{c_2} = t$;
10	END IF
11	IF ($C_{min} > \min\{C_{max}, C_{max}(t)\}$)
12	Select C_{min} from t_{start} to $t_{stop} = t_{c_1}$;
13	ELSEIF ($C_{max} < \max\{C_{min}, C_{min}(t)\}$)
14	Select C_{max} from t_{start} to $t_{stop} = t_{c_2}$;
15	ELSE
16	$t++$;
17	CONTINUE;
18	END IF
19	$t_{start} = t_{stop}, t_{stop} = t_{c_1} = t_{c_2} = t_{start} + 1,$ $t = t_{start} + 1, C_{min} = 0, C_{max} = \infty$;
20	WHILE (some time slots are not assigned a rate)
21	DO:
22	Measure the channel gain of the time slot, calculate power using (4), and transmit the video data;
23	WHILE (more video frames to transmit)

We next show that the algorithm given in Table I computes the optimal solution to Problem (8). The proofs are omitted due to lack of space.

Theorem 2: The power minimization algorithm PMA is optimal to Problem (8).

Corollary 2.1: The power optimal transmission scheme $\vec{C}_n^*$ is unique for given $B_n(t)$ and $D_n(t)$.

Corollary 2.2: The computational complexity of Algorithm PMA is $\mathcal{O}(T_n^2)$.

Note that Algorithm PMA is executed during the session setup time. It only incurs a small initialization delay. In our simulations with VBR video traces, we find the execution time is usually negligible. When the channel statistics are changed (i.e., due to handoff), the schedule will be recomputed for the remaining video frames.

Corollary 2.3: The power optimal transmission schedule $\vec{C}_n^*$ is also the smoothest one among all feasible schedules.

TABLE II
POWER MINIMIZATION ALGORITHM FOR MULTIUSER VIDEOS (PMA-m)

1	Execute power minimization algorithm PMA to compute transmission schedules for all active users;
2	DO:
3	Measure channel gains of the current time slot and calculate the transmit powers using (4);
4	IF (peak power constraint is violated)
5	Select the users who won't have underflow even without transmission in this time slot;
6	Sort the selected users in decreasing order of powers;
7	DO:
8	Decrease the power of the selected users by the order;
9	WHILE (peak power constraint is not satisfied)
10	END IF
11	Transmit the videos and recalculate the optimal transmission scheme for the paused mobile users for the next time interval;
12	WHILE (there are more video frames to transmit)

B. Multiuser Video Transmissions

We now consider Problem (5) to compute transmission schedules for N VBR video sessions, which are coupled by the peak power constraint (7). Due to the peak power constraint and random channel gains, the individually calculated transmit powers may violate (7) in some time slots. The problem is further complicated by the random channel gains, which is not available a priori (except for the statistics of the channels).

To solve Problem (5), we develop a heuristic algorithm, termed PMA-m, as presented in Table II. The PMA-m algorithm uses PMA to compute transmission schedules for all active users. Then based on current channel state information, it computes the power needed to achieved the rate for each user, and checks the peak power constraint $\sum_{n\in\mathcal{U}_n} P_n(t) \leq \bar{P}$. If the constraint is not violated, each user's video data will be transmitted at the computed power. Otherwise, as in Steps 4–10, PMA-m select those users who will not have buffer underflow if their transmissions are suspended in the following time slot, and sort them in the deceasing order of their required powers. Starting with first user, PMA-m decreases the powers of the users in the list; if the first user's power reaches 0 W but the peak power constraint is till not satisfied, PMA-m starts to reduce the power of the second user in the list; and so forth until the peak power constraint is satisfied. In some extremely severe channel conditions, the total power $\bar{P}$ cannot even support the minimum required bit rate for all the users. Some users have to be paused and the current frames be discarded. The corresponding playout of such a user will be frozen until the next time slot. Finally, the transmission schedules for the suspended users will be recomputed using PMA as in Step 11 and the above procedure is repeated.

V. SIMULATION STUDY

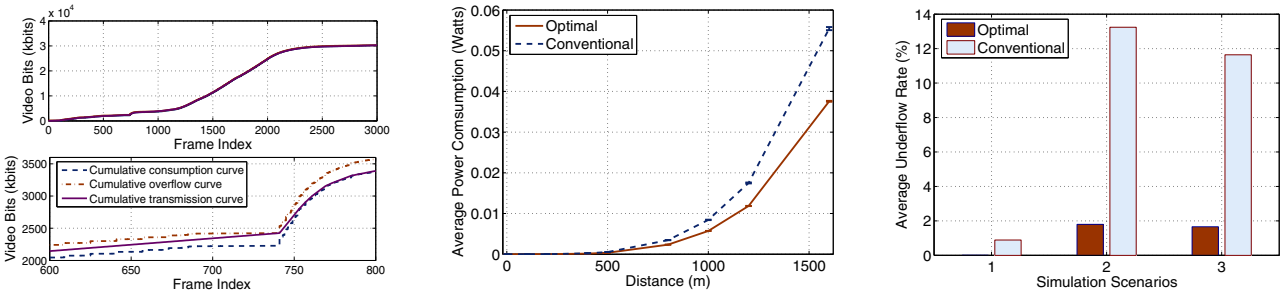
We demonstrate the performance of the proposed optimal power control algorithm through trace-driven simulations. We simulate the downlink of a cell with 1 mile radius. The channels are assumed to be orthogonal, each with $B_w = 1$ MHz bandwidth. We assume that bit errors can be corrected by error correction codes. The path gain averages are $\bar{G}_n = d_n^{-4}$, where d_n is the physical distance from the BS to user n. We assume log-normal fading with zero mean and 8 dB standard deviation. The device temperature is 290 Kelvin and the equivalent noise bandwidth is $B_w = 1$ MHz. The BS streams movies *Star Wars*, *NBC News*, and *Tokyo Olympics* to active users. The video traces are obtained from the Video Trace Lib at Arizona State University.

We first investigate the performance of the power optimal algorithm. In the simulation, the BS streams $3,000$ frames to a mobile user located at different distances to the BS. The cumulative consumption, overflow and transmission curves of the *Star Wars* video session are plotted in Fig. 5(a). It can be seen that the transmission schedule always lie in between the cumulative consumption and overflow curves, indicating that there is no playout buffer underflow or overflow events.

We next compare the optimal power algorithm with a conventional transmission scheme with respect to the average power consumption at the BS. In each time slot, the conventional scheme only transmits the video data that is needed by the decoder at the end of the time slot. It achieves a cumulative transmission curve that connects all the corner points of $D_n(t)$ (also called the "*lazy*" scheme). Intuitively, such lazy approach should be energy efficient since it always transmits the minimal amount of data as needed. However, we will see that the proposed algorithm outperform this lazy approach in the simulations. In Fig. 5(b), we plot the average power consumption achieved by the two schemes for increased distance to the BS. Each point in the figure is the average of $10,000$ simulation runs. The 95% confidence intervals are plotted as vertical bars in the figure, which are all very small.

It can be seen that the proposed algorithm outperforms the conventional scheme for the entire range of distances examined. When the distance to BS is small, both schemes use small transmit powers and the power savings are not very big. However, when the distance is increased, channel fading has a bigger impact on interference and channel capacity. The proposed algorithm achieves considerable power savings than the conventional scheme. When the distance is 1,600 m, the total power of the proposed scheme is only 46.62% of that of the conventional approach, corresponding to a 54.34% normalized improvement. We also obtain the average execution time of the proposed algorithm, under the same setting but for $20,000$ *Star Wars* frames. We find that the average execution time is about 0.06 s on an IBM Laptop with 1.83 GHz processor and 2 GB RAM.

Finally, we examine the buffer underflow events. The following scenarios are simulated: i) $\bar{P} = 1$ W; the movies are *Star Wars*, *NBC News*, and *Tokyo Olympics*; 50 mobile users; $B_w = 1$ MHz; ii) the same setting as i) except that $B_w = 125$ KHz; iii) $\bar{P} = 10$ W; the HD movies are *Terminator 2*, *From Mars to China*, and *Sony Demo*; 20 mobile users; $B_w = 1$ MHz. The HD movies have larger frame sizes and higher variability in frame sizes. The buffer underflow rates are plotted in Fig. 5(c), each being the ratio of the number of underflow frames over the total number of frames. PMA-m achieves considerably lower underflow rates for all the three scenarios. The PMA-m underflow rates are 0.056%, 13.60%,



(a) Transmission curves of *Star Wars*. (b) Average power consumption. (c) Underflow rates.

Fig. 5. Simulation results for the PMA and PMA-m algorithms.

and 14.26% of that of the conventional scheme. Therefore, PMA-m achieves not only considerable energy savings, but also much better video quality at the mobile users.

VI. Related work

VBR video over wired networks is well-studied under two types of models: *statistical* or *deterministic* models. The former models VBR traffic with stochastic process and observed long-range-dependence in VBR trace [2], [9], which provides valuable knowledge for the nature of VBR traffic. The latter method uses piecewise-constant-rate transmission and transport (PCRTT) method to optimize one or several objects while preserving the continuous video playout [3], [10], [11]. In an interesting work [3], Salehi, et al. applied majorization to VBR video smoothing with a smoothness optimal algorithm. The proof of Theorem 2 follows a similar approach as in [3]. These prior work are based on the assumptions of a single video session and constant rate channels, which may not be directly applied to the case of wireless networks.

There are two pieces of closely related work on VBR video over wireless channels [12], [13]. Both papers are focused on one VBR stream over a given time-varying wireless channel. In [12], it was shown that the separation between a delay jitter buffer and a decoder buffer is in general suboptimal, and several critical parameters are derived for the system. In [13], the authors studied the frequency of jitters under both network and video system constraint and provided a framework for quantifying the trade-offs among several system parameters.

In this paper, we consider multiuser VBR video streaming within a cellular network with orthogonal channels. We jointly consider power control and video traffic/palyout information for power minimization. Our stochastic majorization theory based approach is quite different from the prior works [12], [13], which allow us to develop effective algorithms with low complexity and proven optimality.

VII. Conclusion

In this paper, we studied the problem of downlink multiuser VBR video streaming in cellular networks. Our formulation took into account the interactions among power control, fading channels, and VBR video traffic/playout characteristics. We formulated a constrained stochastic optimization problem aiming to minimize the BS power consumption and to avoid playout buffer overflow or underflow. Majorization-based algorithms were developed to solve the formulated problem with highly competitive solutions. The performance of the proposed algorithms were validated with trace-based simulations.

Acknowledgment

This work is supported in part by the US National Science Foundation (NSF) under Grants CNS-0953513, ECCS-0802113, and IIP-1127952, and through the NSF Wireless Internet Center for Advanced Technology at Auburn University.

References

[1] Y. Huang and S. Mao, "Downlink power control for variable bit rate video over multicell wireless networks," in *Proc. IEEE INFOCOM'11*, Shanghai, China, Apr. 2011, pp. 2561–2569.

[2] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 24, no. 4, pp. 269–280, 1994.

[3] J. Salehi, Z.-L. Zhang, J. Kurose, and D. Towsley, "Supporting stored video: reducing rate variability and end-to-end resource requirements through optimal smoothing," *IEEE/ACM Trans. Networking.*, vol. 6, no. 4, pp. 397–410, Aug. 1998.

[4] J. Lee, R. Mazumdar, and N. Shroff, "Downlink power allocation for multi-class wireless systems," *IEEE/ACM Trans. Networking*, vol. 13, no. 4, pp. 854–867, Aug. 2005.

[5] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York, NY: Academic Press, 1979.

[6] S. Ulukus and A. Yener, "Iterative transmitter and receiver optimization for CDMA networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1879–1884, Nov. 2004.

[7] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[8] B. C. Arnold, *Majorization and the Lorenz Order: A Brief Introduction*. New York, NY: Springer-Verlag, 1987.

[9] J. Beran, R. Sherman, M. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1566–1579, Feb./Mar./Apr. 1995.

[10] S. Liew and H. Chan, "Lossless aggregation: a scheme for transmitting multiple stored VBR video streams over a shared communications channel without loss of image quality," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 6, pp. 1181–1189, Aug. 1997.

[11] W.-C. Feng and M. Liu, "Critical bandwidth allocation techniques for stored video delivery across best-effort networks," in *Proc. IEEE ICDCS'00*, Taipei, Taiwan, Apr. 2000, pp. 56–63.

[12] T. Stockhammer, H. Jenkac, and G. Kuhn, "Streaming video over variable bit-rate wireless channels," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 268–277, Apr. 2004.

[13] G. Liang and B. Liang, "Balancing interruption frequency and buffering penalties in VBR video streaming," in *Proc. IEEE INFOCOM'07*, Anchorage, AK, May 2007, pp. 1406–1414.