Intelligent Wireless Video Communication: Source Adaptation and Multi–User Collaboration

Zhu Li[1], Jianwei Huang[2], Aggelos K. Katsaggelos[3], Mung Chiang[2]
[1]Multimedia Research Lab (MRL), Motorola Labs, Schaumburg, IL 60196, USA
Email: zhu.li@motorola.com
[2]Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA
Email: {jianweih, chiangm}@princeton.edu
[3]Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, IL 60260, USA
Email: aggk@ece.northwestern.edu

ABSTRACT

A key challenge to the vision of ubiquitous mobile multimedia access is the communication resource limitation in multiple access wireless channels. In this paper we review several source adaptation techniques and wireless collaboration schemes. We then formulate the multi-user video communication problem as a network utility maximization (NUM) problem., and solve it using a pricing scheme. The pricing scheme co-ordinates resource allocation, source adaptation in a distributed manner, and leads to efficient packet scheduling to achieve better utilization of communication resources and end-to-end video quality.

Key words: video adaptation, collaborative video streaming, optimization decomposition, pricing control, rate-distortion modeling

I. INTRODUCTION

The rapid advances in computing and communication technology have created a plethora of multimedia capable mobile devices, connected by a variety of wireless and wired networks. Serving and consuming multimedia content among mobile devices over heterogeneous networks present challenges to video coding and adaptation, as well as communication and networking technologies. A critical problem is how to efficiently serve video streaming sessions over a multiple access wireless channel with shared communication resources. The demand for better video quality needs to be reconciled with the limitation in communication resources, especially for the currently deployed wireless cellular systems (e.g., [IS95], [CDMA2000]) which are designed typically for applications like voice and lower bit rate data. Higher data rate video in the wired network need to be adapted through a variety of adaptation schemes, like transcoding [Xin05], [Vetro03], and scalable video stream extraction [Ohm05], before it can be accommodated by the wireless channel. The time varying rate-distortion characteristics of video streams need to be exploited in resource allocation and source adaptation to achieve better efficiency.

The problem of source adaptation has been well explored in the video coding community. Basically, it deals with the problem of re-shaping the video stream (after coding), to suit user preferences and network constraints. A good review of the state of art can be found in [Chang05]. Two leading solutions for adaptation are transcoding [Vetro03][Xin05], and scalable video stream extraction, including Fine Granularity Scalability for spatial quality adaptation [Li01][Wu00], and more recently, Motion Compensated Temporal Filtering (MCTF) [Ohm05] for temporal scalability. Moreover, for wireless channels at very low bit rate (VLBR), adaptation schemes need to be fully content-aware and carefully tradeoff spatial and temporal distortions, [Li05], [Liu05], in order to deliver useful video quality to the end user.

In order to coordinate resource allocation among users, and making appropriate adaptation decisions, it is natural to employ cross layer optimization approaches, i.e, jointly considering application layer source coding/adaptation options and lower layer communication resource allocations. There exists a rich body of literature on this

topic, as shown in some recent works in [Zhang05], [Schaar03], [Zheng03], [Yoo04] and [Zhao02], and the references therein. In a wireless multiple access scenario, not all information are available at a single central node, and therefore a distributed optimization solution is desirable.

To address these challenges, we take an optimization approach by formulating the problem as a Network Utility Maximization (NUM) problem, with the objective of delivering the best possible total utility (video quality) to the users, under the communication resource constraints. Solving the problem directly in its primal form is difficult due to the coupling constraints. We consider using the dual decomposition technique (e.g., [Liu04], [Saraydar02]) to solve the problem, which leads to a resource pricing control sub-problem at the base station, and a surplus maximization sub-problem at each mobile end user. These sub-problems can be solved in a distributed fashion with minimum information exchange. This is based on the view that communication layering is actually optimization decomposition [Chiang06] in both horizontal (among users) and vertical (cross layer) directions.

The structure of this paper is summarized as follows. In Section II, we review the video source adaptation techniques, and a detailed discussion on video summarization for very low bit rate wireless communication. In Section III, we discuss various schemes for collaboration in wireless video communication, such as pricing and auction schemes. In Section IV, we propose a dual-based pricing scheme for *average resource allocation* and *source adaptation coordination* in wireless video communications. Most previous work on pricing has focused on *instantaneous resource allocation* in *elastic data* communications. Here, we propose using a video summarization technique to adapt video source to operate at VLBR range, and to determine average resource consumption levels locally based on its own utility function and the price announced by the base station. We further develop collaborative joint packet scheduling scheme to meet the unique delay constraints of video frames. Solutions for CDMA uplink and downlink cases are developed. Numerical results are presented and discussed in Section V, which demonstrate that our algorithms achieve much better performance than several existing heuristics. We finally conclude in Section VI.

II. VIDEO SOURCE ADAP–TATIONS

Adaptation is a key functionality component in modern multimedia communication system. In general, video adaptation serves two purposes: 1) to meet resource limitations in communication, storage, computation, and terminal capability, 2) to satisfy user preferences in video consumption.

Resource limitations can take the forms of bandwidth, energy in communication, disk size in storage devices, display size in hand-held devices, battery energy and computational power in mobile devices. User preference can be expressed as video signal quality (frame size, PSNR and frame rate). It can also be expressed as needs at object / structural / syntactical / semantic levels, for which in addition to signal level adaptation, visual analysis techniques need to be incorporated to help adaptation decisions.

There are two basic techniques for video adaptation, *transcoding* and *scalable coding*, each has its pros and cons for different applications.

A. Transcoding

In a unicast or small group multicast scenario, a transcoder [Vetro03][Xin05] can be employed either at the video source or some network edge devices to convert the video contents to an appropriate format, with desired quality/form and rate. The benefit of transcoding is the great flexibility in serving individual adaptation need. However, it also puts extra computational burden on edge devices in the network, which may not be practical in certain applications.

Video transcoding can be viewed as a two-pass video coding technique that can be used to adapt to the rate constraints, and/or user preferences in scale and spatial-temporal distortion. There are many scenarios for transcoding. It may be simply a format conversion, for example, from MPEG-2 to H.264. It may also be a bit rate reduction for wireless delivery, e.g, from 1.5Mbps to 64kpbs. It can also be a frame resizing, to fit a SDTV program onto a QCIF sized cell-phone display, or any combination of these goals.

One straight-forward solution is known as Open-Loop Pixel Domain (OLPD) transcoding. It simply concatenates a decoder with an encoder, with pixel domain scaling, selective coefficients transmission or re-quantization to achieve format conversion and quality/bit rate change.

For pure bit rate reduction at the original frame size and frame rate, OLPD can be implemented without the explicit Discrete Cosine Transform (DCT) / Inverse Discrete Cosine Transform (IDCT), by simply dropping certain coefficients from the bit stream. This is shown in Fig. 1.
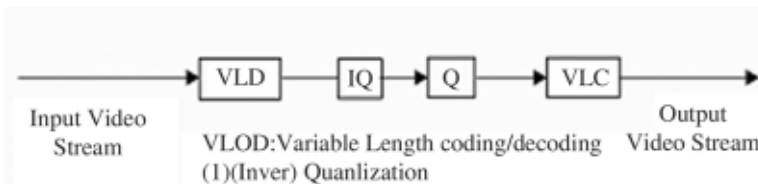

Fig.1 Open-loop transcoding[Sun96]

The main problem with OLPD transcoding is the quality degradation caused by "drifting", i.e., decoding motion compensation mismatch due to re-quantization. Frequent re-inser-

tion of intra frames can partly solve the problem, but also leads to rate increases. "Drifting" can be resolved by Closed-Loop Pixel Domain (CLPD) transcoding [Sun96], where the prediction is done against transcoded frame. A closed loop Motion Compensation (MC) is added to predict from re-quantized frames, through reusing the motion vectors (MV) [Shanableh00]. Since MC accounts for about 60~70% of computational cost, this is a big saving. To save cost in DCT, the end-of-block is predicted for the transcoded MBs, and only partial DCT is performed to further save time [Youn99].

One obvious temptation is to do transcoding in the DCT domain, which saves the time in projection back and forth between the pixel and the DCT domains. Early implementations are reported in [Chang95], [Assuncao98]. A diagram of Open Loop Transform Domain (OLTD) transcoding is shown in Fig. 2.
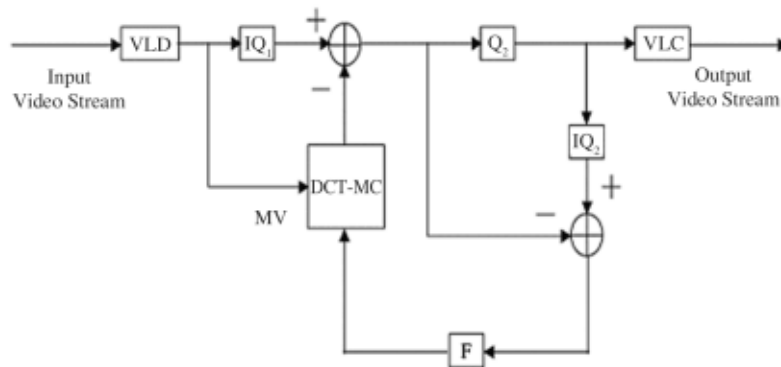

Fig.2 Open-loop DCT domain transcoding [Sun96]

The computationally expensive part of the OLTD scheme is DCT-MC, which is derived based on the assumption that DCT, IDCT and MC operations are linear. In [Chang95], the MBs in DCT domain are reconstructed from its 4 motion predicted neighbouring MBs in the previous frame. Several fast algorithms [Merhav97, Song00] also exist to improve the performance of DCT-MC, but there is a fundamental flaw in linearity assumption of DCT-MC, which causes "drifting" problem when the linearity assumption fails. The OLTD scheme can not handle frame size and frame rate changes either, this leads to a closed loop solution similar to the pixel domain case. An additional DCT-MC is introduced to predicate from the frame difference between the transcoded frames. This gives the CLTD scheme flexibility in handling frame resizing and frame drops, in addition to the quantization changes.

B. Scalable bit stream extraction

In a broadcast or large group multicast scenario, since there are large variations in adaptation need among receivers, a scalable coding solution [Ohm05] may be more appropriate, e.g. where the source is coded once, and the bit stream can be extracted and re-packetized to suit various adaptation needs without transcoding. The computational costs involved are typically much smaller than the transcoding case, and is suitable for source stream adaptation at routers and wireless access points.

There are three types of scalability in video, namely, frame size, frame rate and frame quality. A scalable video coding diagram is shown in Fig. 3. There are 3 layers of coding, each corresponds to a frame size that is $\frac{1}{4}$ of its previous layer.

Within each coding layer, frame rate scalability can be achieved through Motion Compensated Temporal Filtering [Ohm94](MCTF), or similar Hierarchical B frame schemes. An example with dyadic prediction structure is shown in Fig. 4. For the GoP size of 8 frames, this structure gives us 4 different frame rates of 8, 4, 2 and 1 frames per GoP, e.g., if we drop all B frames in green, the frame rate is halved.

For each frame, the quality scalability can be achieved by applying bit plane coding techniques like Fine Granularity Scalability (FGS) [Li01], or Progressive FGS [Wu00]. The advantage of FGS approach is that bit stream can be truncated at almost any point, and still can be decoded with reconstruction quality corresponding to number of bits recovered.

Armed with these scalable coding features, the adaptation task can be achieved by dropping certain bits/ packets in the bit stream to meet target bit rate, re-sizing frames for display size constraint, and adjusting spatio-temporal qualities in a graceful manner. The extraction process is illustrated in Fig. 5. A global bit stream for a video sequence is a collection of (droppable) packets that have dependency along the prediction direction in frame size, frame rate (in frame per sec, or fps), and frame quality. The adaptation process is therefore a graph pruning process with certain quality change/ degradation associated with each valid packet drop combinations.

C. Video summarization

On top of these basic coding/adaptation schemes, more intelligent solutions that serve users' preference in viewing time compression and browsing can be developed through analysis, structuring, and summarizing video content. Good reviews can be found in [Wang00], [Pitas06]

In general, the first task of structuring the video sequence is to segment it into video shots (i.e, visually consistent segment shot by a single camera at the same scene). Various shot boundary effects, from a simple cut, to fade in/outs, dissolves, and wipes introduce different characteristics and challenges to the shot segmentation task. A good solution depends on appropriate features and metrics for visual consistence, as well as a good temporal behavior modeling and detection algorithm.

Reviews of various shot segmentation techniques can be found in [Hanjalic02], [Lienhartt01] and [Pitas06].

Video summarization is to use a subset to best represent the original sequence. This is an intuitive and well-defined problem. If the resulting video summaries consist of segments of frames continuous in time, they are called *video skims*, otherwise they are known as *video highlights*.

Video highlights are typically generated based on image features, and/or object level information. It is usually formulated as a clustering problem, [Hanjalic99], or a rate-distortion problem [Li05], or object level heuristic constraint [Kim02].

Video skims [Gong01] require more video program structure and domain knowledge, in addition to the image features and object level

Fig.3 Scalable coding architecture[JVT05]

information. Typically it operates at a larger time scale than video highlights, and select video shots that are absolutely crucial according to certain criterion.

Hierarchical video highlights/skims are useful for large video collection/program browsing. Obviously it is more difficult than simple video summarization. Semantic level information is needed to achieve good results.

In this paper, we will show examples of source adaptation for video streaming over VLBR wireless channels. We employ a video summarization and transcoding solution to address the spatio-temporal quality tradeoffs at VLBR. We define the adaptation utility as the video summarization quality, and the problem can be solved by a smart frame selection scheme [Li05] as described next.

Let us consider a segment of n original frames, denoted by $V=\{f_0, f_1, ..., f_{n-1}\}$. Also denote the corresponding video summary of m frames as $S=\{f_{l_0}, f_{l_1}, \Lambda, f_{l_{m-1}}\}$, where $m \leq n$. In other words, we select only a subset of m frames out of the original sequence, and code them at higher quality than the case if all frames are sent, to
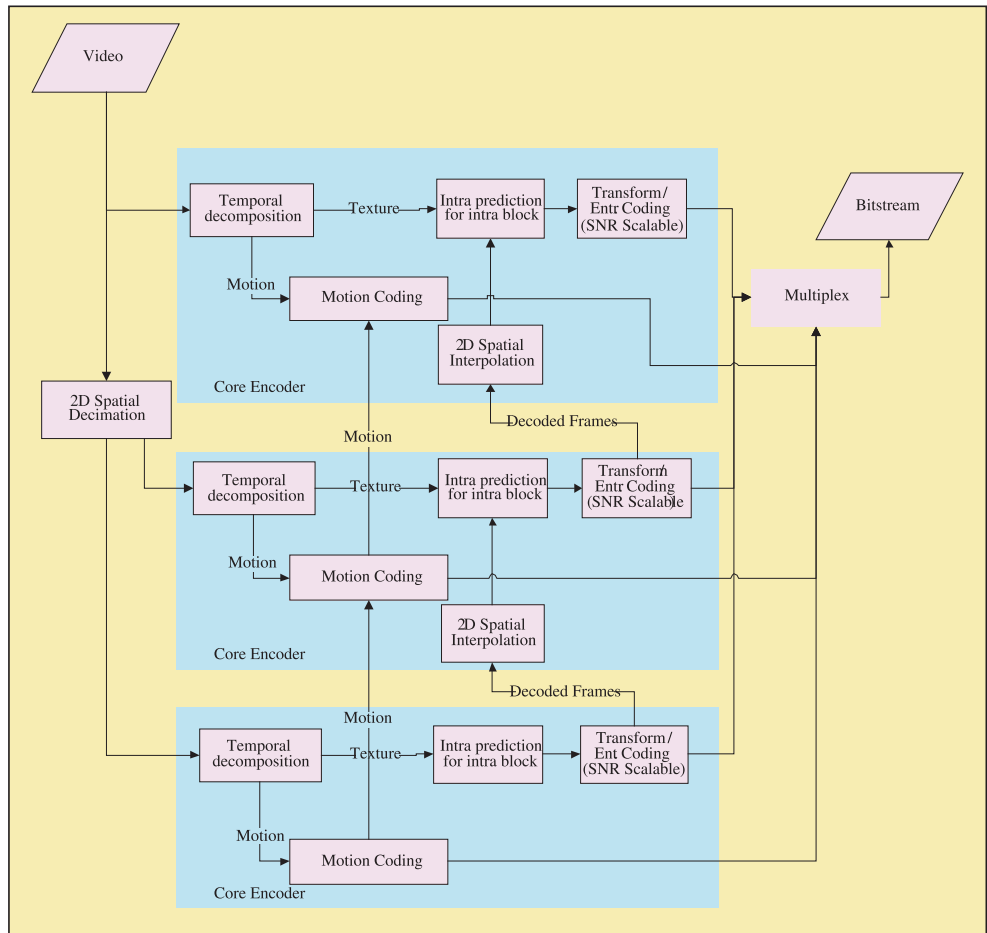
achieve a more satisfactory received video quality.

Assuming all m frames can be received at the receiver side, we can then reconstruct the original n frame sequence as $V_S' = \{f_0', f_1', \Lambda f_{n-1}'\}$ by substituting the missing frames with the most recent frame that is in the summary S. The video summary average distortion caused by the missing frames, is given as,

$$D(S) = \frac{1}{n} \sum_{k=0}^{n-1} d(f_k, f_k'),$$

Where $d(f_k, f_k')$ is the distortion between the original frame f_k and the reconstructed frame f_k'. (If frame f_k is chosen as one of the m frames, then $f_k = f_k'$ and $d(f_k, f_k') = 0$.) Therefore, the
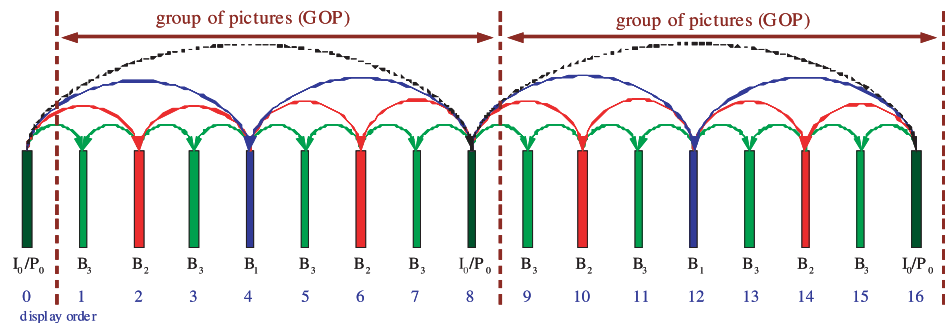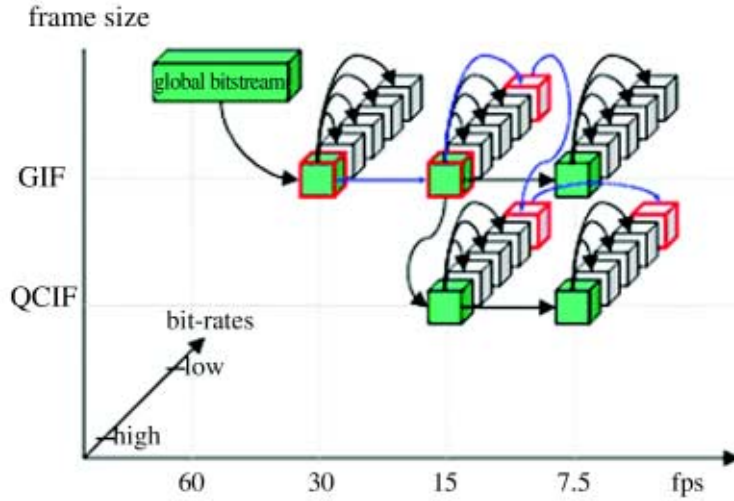

Fig.4 Dyadic MCTF structure[JVT05]

Fig.5 Bit stream extraction [JVT05]

summarization problem for wireless transmission can be formulated as a constrained optimization problem,

$$\min D(S), \, s.t. X(S) \leq X_{max}. \qquad (1)$$

where $X(S)$ is the total communication resources needed to transmit video summary S. For the single user case, the solution to Eq. (1) can typically be found via Lagrangian relaxation and a fast search on Lagrangian multiplier that traces out optimal solutions with a convex hull approximation. The relaxed problem is solved with a *Dynamic Programming (DP)* approach, that traverses a trellis of all possible video summary frame transitions.

For multi-user wireless communication case, the shared resource constraint has certain structure in dual decomposition, which we will explore in Section III.

III. WIRELESS COLLABORATIONS IN COMMUNICATION

In a multi-user environment, the adaptations of various users need to coordinate with each other to achieve the optimal overall system performance. In many wireless networks, there is not a single centralized entity that has all system information (e.g., channel states and users' utility functions), thus it is desirable to design distributed user collaboration schemes. In this section, we will discuss multi-user collaborations through two mechanisms: pricing and auction.

A. User collaborations through pricing

Pricing has been used extensively in economics to coordinate the selfish behaviors of various entities (e.g., consumers and producers). Various pricing schemes have also been used for resource allocation in wireless networks (e.g., [Huang06b], [Marbach03] and the references therein). Here we consider a scheme based on dual decomposition of a network utility

maximization problem (e.g., [Palomar06]). We focus on average resource allocation (i.e., average transmission time in uplink and average transmission power in downlink) instead of instantaneous resource allocation, since the latter will be further determined by the user adaptation described in Section II.

As an example, we consider a typical multiple access wireless cellular network with a base station serving mixed voice and video traffics. We only consider the Quality of Service (QoS) of the video users and regard the voice users as background traffic. We characterize a video user i's end-to-end QoS by a utility function $U_i(x_i)$, which is an increasing and strictly concave function of the communication resource allocated to user i, x_i. This can model various commonly used video quality measurements such as the video rate-PSNR function, and video rate-summarization distortions [Li05]. It is well known from information theory [Cover91] that the rate-distortion functions for a variety of sources are convex, and in practice, they are usually convex as well. Thus the utility functions (defined as negative distortion) are concave.

We want to solve the following Network Utility Maximization (NUM) problem, which aims at achieving the maximum total network QoS of N video users, subject to limited total communication resource X_{max}, i.e.,

$$\max_{\{x_i \geq 0\}, 1 \leq i \leq N} \sum_i U_i(x_i), \, s.t. \sum_i x_i \leq X_{max}. \qquad (2)$$

where each user's utility function can be represented by certain received video quality function, and in this paper, we will use the inverse video summarization distortion as utility.

Solving (2) directly requires a centralized approach due to the coupling resource constraint. However, a distributed solution is often more desirable, since the base station typically does not know the utility functions of individual mobile users.

Here we use the dual decomposition technique [Boyd04], where the base station sets a price λ on the resource, and each mobile user determines its average resource request depending on the announced price and its own source utility characteristic by solving the following problem,

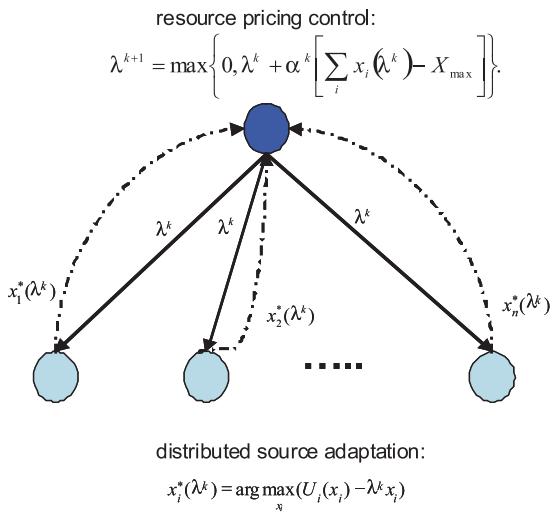$$\max_{x_i \geq 0} U_i(x_i) - \lambda x_i, \qquad (3)$$

which corresponds to maximizing the net utility (i.e., utility minus payment) based on the price λ announced by the base station. Denote the optimal solution of (3) as $x_i(\lambda)$, which is unique since the utility function is increasing and strictly concave. The mobile users will then feedback the values of $x_i(\lambda)$ to the base station. The base station adjusts λ using a sub-gradient search method,

$$\lambda^{k+1} = \max\left\{0, \lambda^k + \alpha^k\left[\sum_i x_i\left(\lambda^k\right) - X_{\max}\right]\right\}. \quad (4)$$

where α^k is a small step size at iteration k.

The two level optimizations in (3) and (4) together solve the *dual* problem of the original NUM problem (2) (which we call the *primal* problem). Base station controls the resource price according to (4), and each video source performs source adaptation to maximize its profit as in (3), in a distributed fashion. The only communication overheads incurred are price and resource requests from sources. This makes the solution scalable in a large network. The distributed solution is illustrated in Fig. 6.

resource pricing control:

$$\lambda^{k+1} = \max\left\{0, \lambda^k + \alpha^k\left[\sum_i x_i\left(\lambda^k\right) - X_{\max}\right]\right\}.$$

distributed source adaptation:

$$x_i^*(\lambda^k) = \arg\max_{x_i}(U_i(x_i) - \lambda^k x_i)$$

Fig.6 Illustration of dual decomposition technique

The connection between user collaboration and source adaptation is through (3), where each user chooses an adaptation scheme that gives a value of that x_i to solve (3).

B. User collaborations through auction

Another alternative for user collaborations is an auction scheme, where users submit bids and the resource is allocated based on the bids. Here the resource allocation to one user is dependant on the actions of the other users. We can model this user interaction as a game. Next we will introduce some background on game theory and auction. Detailed discussions can be found in [Huang05, Ras2001]. Then we will comment on how wireless video users can collaborate through an auction scheme.

Game theory [von47, Nas50] is a method to study interactive decision problems among intelligent rational decision makers. The essential elements of a game are the players, the actions, the payoffs and the information, know collectively as the rules of the game.

Players are the individuals who make decisions. A player's action set is the set of all the choices he can make. An action profile is the collection of e of all players' actions,

one from each player. For example, in an auction setting, players are the bidders and actions are the bids submitted by the bidders. A common action set for a bidder is the interval of $[0, \infty)$, i.e. he can submit any nonnegative bid.

A player's payoff is a function of the action profile, and describes how much the player gains from the game for each possible action profile.

A player's information can be characterized by an information set, which tells what kind of knowledge the player has at the decision instances. In order to maximize their payoffs, the players would design contingent plans known as strategies, which are mappings from one player's information sets to his actions. A strategy profile is a collection of the players' strategies, one from each player.

In simultaneous move games, or one-shot games, players choose their actions simultaneously and only once. Each player only has one information set, which is what he knows at the beginning of the game. This is the game of our focus. There is no iterative process involved in a one-shot game

A reasonable prediction of the outcome of a game is equilibrium, which is a strategy profile where each player chooses a best strategy to maximize his payoff. Among several available equilibrium concepts, we focus mainly on the Nash Equilibrium (NE). In a one-shot game, an NE is a strategy profile where no player can increase his payoff by deviating unilaterally. Note that there may be no NE or multiple NEs in a given game.

An auction is a game that is used to determine how a good is allocated by the seller (auctioneer) among the buyers (bidders). An auction is specified by a set of rules that determine: (i) how bidders submit bids to express their evaluations of the good (or willingness to pay), (ii) how the good is allocated among the bidders as a function of the bids, and (iii) what are the payments from the bidders to the auctioneers as a function of the allocation and bids. In the downlink transmission case, the base station acts as the auctioneer and the mobile end users act as the bidders. The good is the total transmission power at the base station, which is divisible and allocated to bidders depending on their bids.

Auction-based resource allocation schemes have been considered in both wireline and wireless networks (e.g., [Sun06,Huang06b]). Auction mechanism for multimedia communications in wireline networks is reported in [Rei03], but not much work has been done in the wireless case.

In the wireless video communication case, we can design auction mechanism to allocate the resource to video users in the average sense similar as using pricing mechanism. To be specific, each user submits a bid (e.g., a nonnegative number) to the base station, representing his evaluation of the resource. The base station then allocates the resource to the users based on the bids. If the constrained resource is the transmission power, the base station can allocate the transmission power proportionally to users' bids. Users' payments are equal to their bids, and users' payoffs equal to the difference between the

achieved utilities and the payments. It is clear that users participate in a game, since one user's power allocation not only depends on its own bid but also depends on the other users' bids as well. Once a N.E. is found, each user can perform source adaptation to meet the average resource allocation.

Examples of source adaptation and mutli-user collaboration in wireless video communication are shown in Section IV.

IV. JOINT RESOURCE ALLOCATION AND SCHEDULING IN CDMA NETWORKS

CDMA technology has been widely deployed in the wireless multiple access systems, as in [IS95] and [CDMA2K]. There exist rich literatures in resource allocation for CDMA networks, as in [Kam01], [Lee02], [Wang05], [Xu04] and [Huang06]. However, most of previous work only considers elastic data traffic or multimedia traffic without explicating considering the delay deadlines. In this sections, we briefly review the two-stage resource allocation and scheduling algorithms for both CDMA uplink and downlink communications proposed in [Huang06c], [Li05b], and [Li06]. The first stage involves average resource allocation based on dual decomposition (i. e. solve a variation of problem (1) using Algorithm 1), and the second stage involves a greedy scheduling algorithm that tries to meet the video frames' individual deadlines.

A. Video streaming over CDMA uplink

The uplink communications refer to the transmissions from the mobile users to the base station, where users interfere with each other due to asynchronous transmissions.. Without considering complicated multi-user detection techniques, the receiver at the base station decodes each user's signal independently, regarding the received power from other users as interference.

We consider the case of mixed voice and streaming video transmissions in the uplink channel. The objective is to provide the best possible QoS to the video users (measured in their total utility), without interrupting the transmissions of voice users. In other words, we need to guarantee that the aggregated interference generated from the video users is small than a threshold, such that the voice users can reach their target signal-to-interference plus noise ratios (SINRs). This can be translated into an optimization problem subject to a total received power constraint from the video users at the base station, denoted as P_{max}. We want to optimize the received power function of each user video user i, $P_i(t)$, during sliding window $[0,T]$, such that the total utility of all video users during this time window is maximized.

$$\max_{\{P_i(t) \geq 0\}, 1 \leq i \leq N} \sum_{i=1}^{N} U_i\left(S_i\left(\mathbf{P}(t)\right)\right), s.t. \sum_{i=1}^{N} P_i(t) \leq P_{max}, \forall t \in [0,T] \quad (5)$$

Notice that $P_i(t)$ will be a time variant function due to the VBR nature of the video traffic. We use video summarization as the adaptation technique, and the corresponding summarization decision of user i is denoted by S_i, which depends on the total bits it can transmit during time $[0,T]$, and in turn depends on everyone's received power function $\mathbf{P}(t)=[P_1(t),...,P_N(t)]$ due to interference. The utility function is defined on the video summarization quality as discussed in Section III.[1]

Compared with problem (1), the objective function of problem (5) is also coupled across users due to interference. This coupling makes solving (5) quite difficult. On the other hand, it is a good idea for the users to avoid transmitting simultaneously in order to improve the transmitting rate and fully exploit the VBR nature of the traffic. To this end, we propose a Time Division Multiplexing (TDM) transmission among video users, where there is no interference among different video users. Whenever a video user is allowed to transmit, it will choose the transmission power such that the received power at the base station equals P_{max}. Denote the corresponding constant transmission rate as denoted as R_{max}, which depends on the several factors such as spreading gain, the interference from voice users, inter-cell interference as well as thermal noise. The total bits user i can transmit during time $[0,T]$ will be the product of R_{max} and t_i, where t_i is the total active transmission time of user i.

As a result, problem (5) can be transformed into a transmission time allocation problem as follows,

$$\max_{\{t_i \geq 0\}, 1 \leq i \leq N} \sum_{i=1}^{N} U_i\left(S_i\left(t_i\right)\right), s.t. \sum_{i=1}^{N} t_i \leq T', \quad (6)$$

where initially we have $T'=T$. We will discuss shortly the case where we need to decrease the value of T' such that $T'<T$. It is not difficult to see that problem (6) is a special case of problem (1), where x_i is replaced by t_i and X_{max} by T'. As a result, the optimal solution of (6) can be found by using the pricing algorithm in Section III.A.

The next step is to determine the transmission schedule of the frames (i.e., how to allocate the transmission time precisely to each frame). This requires us to take into account of the individual deadline requirement of each video summarization frame. In a real-time video streaming application, the deadline of a frame is determined by its position in the original video sequence (before the summarization) plus an initial delay (determined before the start of playback of the entire sequence, which is typically much longer than T). To achieve this, we propose a GREEDY scheduling algorithm as follows.

First, each mobile user reports the sizes and deadlines of its summarization frames to the base station. The base station then schedule the transmissions of the frames from all users one-by-one, in increasing order the deadlines. The transmission time of a particular frame depends on the size of the frame and the constant transmission rate R_{TDM}. Although the GREEDY scheduling is simple, it is optimal among all TDM based schedules.

If all the summary frames can be delivered on or before their individual deadlines, we say the summary frame sequence is schedulable and problem solved. Then we can move on to the next $[0,T]$ time window, and solve a new resource allocation and scheduling problem. However, if one or more frames' deadlines are violated, we say the sequence is nonschedulable and have to resolve problem (6) by decrease T by a small value. In other words, we need to tighten the resource constraint such that the total resource request is less. We will iterate until the resulting summary sequence is schedulable.

B. Video streaming over CDMA downlink

Downlink communications refer to the transmissions from base station to the mobile users, where there is no mutual interference among the users within a cell thanks to the synchronous transmissions. As a result, we can potentially allow simultaneous transmissions to all video users. The optimization variables are the transmission power function of each user i, $P_i(t)$, for a sliding time window $[0, T]$. And we want to choose $P_i(t)$ such that video users' total utility is maximized.

We again take a two-stage approach similar as in the uplink case. In the first stage, we will optimize over the average transmission power of each video user, subject to the total power constraint. This corresponds to solving the following problem,

$$\max_{\{P_i \geq 0\}, 1 \leq P_i \leq N} \sum_{i=1}^{N} U_i\left(S_i\left(P_i\right)\right), s.t. \sum_{i=1}^{N} P_i \leq P'_{max}, \qquad (7)$$

where we temporally assume that each user i will transmit with constant power P_i over the entire time window $[0,T]$, and $P_{max}' = P_{max}$. The average transmission power has a one-to-one correspondence to the total bits a user can transmit during time $[0,T]$, which in turn determines the summarization results and the utility. It is clear that problem (7) is again a special case of problem (1), and can be solved using the pricing algorithm in Section III.A. Then the users can determine the best summarization frames depending on the optimal average transmission power allocation.

Next we need to determine the exact values of $P_i(t)$ for each user i. Typically a constant power transmission will not be able to meet the deadlines, since the frames could come in bursts. Thus we need to determine time varying functions of $P_i(t)$ to meet the deadlines such that the total power constraint $\sum_{j} P_j(t) \leq P_{max}$ is satisfied for all values

of t belongs to $[0, T]$. To this end, we design a greedy water-filling type of scheduling solution as follows.

First, the base station sorts the summary frames of all users in increasing order of delivery deadlines as in the uplink case. Assuming the l-th frame in the sorted sequence belongs to user i, and we denote its size, available time and delivery deadline as $\{B_l^i, t_l^i, T_l^i\}$, where $t_l^i < T_l^i$. Let $P(t)$ be the committed total power function for processed frames so far (the 1st to the $l-1$-th frames), we look at $P(t)$ in time $[t_l^i, T_l^i]$, and search on a water filling level L, such that the transmit power function available for frame l is,

$$P_l^i(t, L) = \begin{cases} L - P(t), & t \in [t_l^i, T_l^i] \\ 0, & otherwise \end{cases} \qquad (8)$$

We need to choose a water-level L_l^* such that the corresponding power allocation will enable the successful transmission of B_l^i bits within time $[t_l^i, T_l^i]$. The corresponding power function for the l-th frame will be $P_l^i(t, L_l^*)$. The process is illustrated in Fig. 7, where $P_l^i(t,L_l)$ is the shaded area bounded by $P(t)$ and L, between t_l^i and T_l^i.

The algorithm schedules each frame in the order of delivery deadlines, until the last frame's power function is computed. Then each user's transmitting power function is computed as,

$$P_i(t) = \sum_{k \in K_i} P_l^j\left(t; L_l^*\right) \qquad (9)$$

where K_i denotes the set of frames belonging to user i. Notice that the water-filling scheduler tries to utilize the minimum power possible to meet the deadline constraints.

Similar as in the uplink case, the summary frames may not be schedulable (i.e., the deadlines are always met but the total power constraint might be violated at some t) In that case, we need to reduce the value of P_{max}' and resolve problem (7).
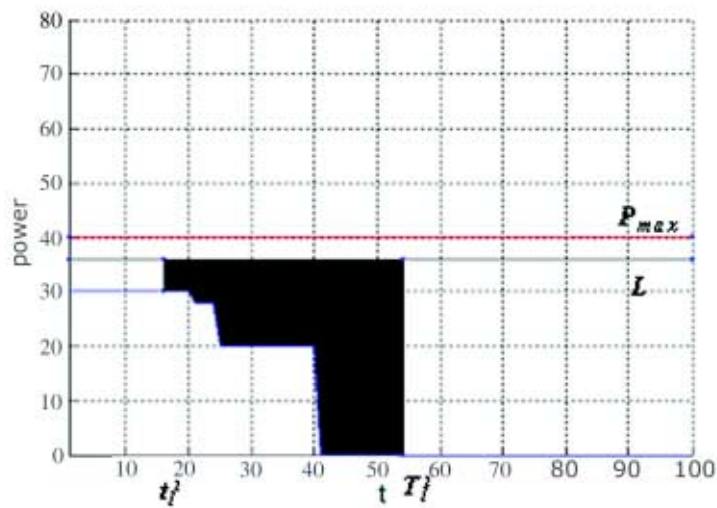

Fig.7 Water filling scheduling example

---

[1] Here we assume that each video user can achieve a peak received power equal to Pmax, thus there is no individual power constraint.

V. NUMERICAL RESULTS

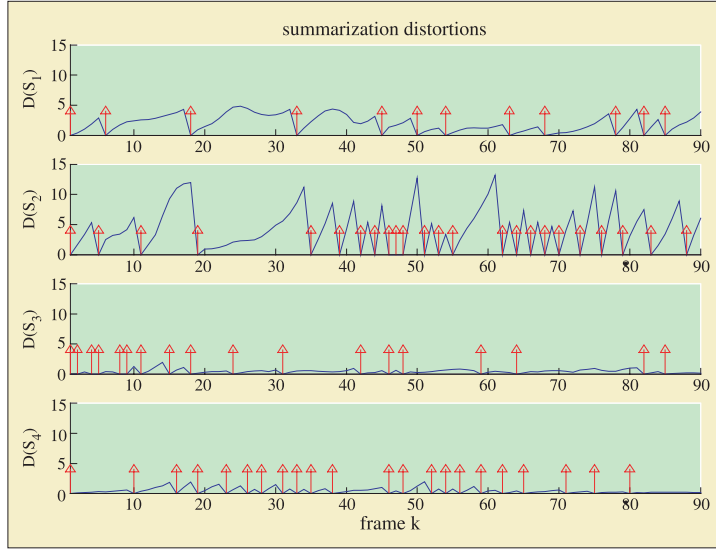In this section, we show simulation results for the algo-

Fig.8 Resulting video summary distortion for $P_{max}=2.4$

rithm presented in Section IV.B. For detailed simulations results of both downlink and up-link transmissions, we refer readers to [Huang06c], [Li06].

We choose four video clips with different content activity levels. Clips 1 and 2 are frames 150-239 and frames 240-329 from the "foreman" sequence, and clips 3 and 4 are frames 50-139 and 140-229 from the "mother-daughter" sequence, respectively. Clips 1 and 2 are coded at an average PSNR of 27.8dB, and clips 3 and 4 at 31.0dB. There are 90 frames within each video clip at a frame rate of 30Hz, which corresponds to a time segment of T=3 secs. Here the channels gains of four users are given as H=[0.75, 1.00, 0.80, 0.65], which will determine the transmission rate together with the allocated transmission power for each video frame.

At the summarization-power allocation phase, a total transmitting power threshold of $P_{max}=2.4$ is given, and the optimal price is found as $\lambda^*=101.45$ through the price iteration. The resulting video summary distortions are plotted in Fig. 8. The resulting average bit rates for 4 clips are *20.1, 43.3, 8.1* and *9.4* kbps, respectively.

With an initial delay of *1* sec, the joint water-filing scheduler achieves maximum total transmission power of *2.45*. This slightly violates the total power constraint of $P_{max}=2.4$. As a result, we need to iterate at least once more between the two stages in Algorithm 3 by tightening the power

constraint in problem (7).

The power allocation results, $P_1(t)\sim P_4(t)$, for the video summaries generated in Fig. 8 are shown in Fig. 9a. The dotted line is the total power function $P(t)$. Notice that each user's power function is not constant at all but the total power function is rather flat and achieves efficient utilization of the power resource. As a comparison, the results based on the single user Earliest Deadline First Serve (EDFS) scheduling are plotted in Fig. 9b, which leads to a maximum power of P_{max}=7.56. The efficiency of joint power scheduling is clearly demonstrated in this case.

The computational burden of the pricing solution is distributed among video sources and the base station. The amount of information need to be communicated for the pricing scheme is kept to a minimum.

VI. CONCLUSION

In this paper, we first review the state-of-art results on video source adaptations and collaborations in wireless communica-
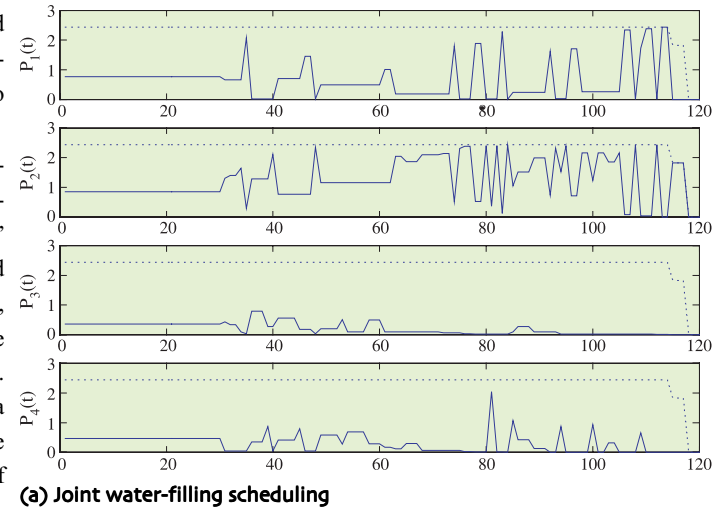
(a) Joint water-filling scheduling
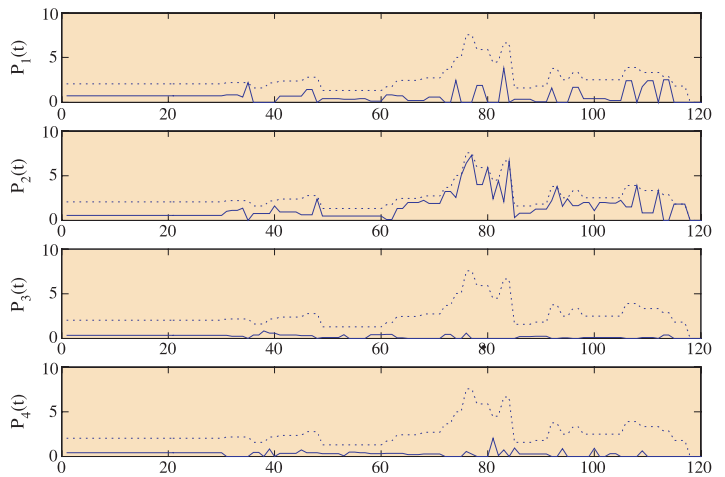
(b) Single user based scheduling

Fig.9 Frames scheduling results

tion networks. For source adaptations, we discuss video transcoding, scalable coding and summarization. For multi-user collaborations, we discuss both pricing and auction schemes. We further apply video summarization and pricing scheme to study the problems of uplink and downlink video streaming over multiple access CDMA networks. After formulating a Network Utility Maximization (NUM) problem, we develop two-stage joint resource allocation and scheduling algorithms to solve the problem. We first use a dual decomposition based technique to perform average resource allocations, then schedule the video frames such that the deadlines constraints are met and the VBR nature of the traffic are fully utilized. Compared with existing heuristics, the proposed algorithms achieve better efficiency in resource utilization and better over all received video quality.

REFERENCES

[Assuncao98] P. Assuncao and M. Ghanbari, "A frequency-domain video transcoder for dynamic bit-rate reduction of MPEG-2 bitstreams", *IEEE Trans. on CSVT*, vol. 8, Dec. 1998

[Boyd04] S. Boyd and L. Vandenberghe, chapter 5, pp. 215-273, *Convex Optimization*, Cambridge Press, 2004.

[CDMA2K] TIA/EIA Interim Standard for CDMA 2000 Spread Spectrum System, May, 2002.

[Chang95] S.-F. Chang and D. G. Messerschmitt, "Manipulation and compositingof MC-DCT compressed video," *IEEE J. Sel. Areas Communications (J-SAC)*., vol. 13, no. 1, pp. 1-11, Jan. 1995.

[Chang05] S. F. Chang and A. Vetro, "Video adaptation: concepts, technologies, and open issues", Proc of IEEE, pp.148-158, Jan. 2005.

[Chiang06] Chiang, M., Low, S.H., Calderbank, A.R., Doyle, J.C., "Layering As Optimization Decomposition: A Mathematical Theory of Network Architectures", *Proceedings of the IEEE*, Jan. 2007

[Cover91] Cover, T., Thomas, J., 1991. Elements of Information Theory, Wiley and Sons, New York.

[Gong01] Y. Gong and X. Liu, "Video Summarization with Minimal Visual Content Redundancies", *Proc. of Int'I Conference on Image Processing*, 2001.

[Hanjalic99] A. Hanjalic and H. Zhang, "An Integrated Scheme for Automated Video Abstraction Based on Unsupervised Cluster-Validity Analysis", IEEE Trans. CSVT, vol.9, December 1999.

[Hanjalic02] A. Hanjalic, "Shot-Boundary Detection: Unraveled and Resolved?", IEEE Trans. on CSVT, vol.12, No. 2, Feburary 2002.

[Huang05] Huang., J. Wireless Resource Allocation: Auctions, Games and Optimization, Ph.D. dissertation, Northwestern University, 2005.

[Huang06] Huang, J., Berry, R., Honig, M.L., 2004. Auction-

based Spectrum Sharing. ACM/Springer Mobile Networks and Applications Journal (MONET), 11(3): 405-418, 2006

[Huang06b] Huang, J., Berry, R., Honig, M.L., "Distributed Interference Compensation for Wireless Networks", IEEE Journal on Selected Areas in Communications, 24(5): 1074-1084, May 2006

[Huang06c] Huang, J., Li, Z., Chiang, M., Katsaggelos, A.K., "Joint Rate Control and Scheduling for Wireless Uplink Video Streaming", Proc of IEEE International Packet Video Workshop, April 2006

[IS95] Mobile Station-Base Station compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System, TIA/EIA Interim Standard 95(IS-95-A), Washington, DC: Telecommunications Industry Association, May 1995.

[JVT05] Joint Scalable Video Model JSVM-4, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-Q202, 2005, Nice, France.

[Kam01]Kam, A.C., Minn, T., Siu, K.-Y., Supporting rate guar-antee and fair access for bursty data traffic in W-CDMA. IEEE Journal on Selected Areas in Communications, 19(11): 2121-2130, 2001.

[Kim02] C. Kim, and J.-N. Hwang, "Object-based video abstraction for video surveillance systems" IEEE Trans. CSVT, 12(12):1128-1138, Dec., 2002.

[Lee02] Lee, J.W., Mazumdar, R.R., Shroff, N.B, Downlink power allocation for multi-class CDMA wireless networks. Proceedings of IEEE INFOCOM, 2002.

[Li01] W. Li, "Overview of FGS in MPEG-4 Video Standard", IEEE Trans. CSVT, 11(3):301-307, Mar. 2001.

[Li05] Li, Z., Katsaggelos, A.K., Schuster G., Gandhi, B., Rate-Distortion Optimal Video Summary Generation. IEEE Trans. on Image Processing, 14(10): 1550-1560.

[Li05b] Li, Z., Cheng, Q., Katsaggelos, A.K., Ishtiaq, F.,. Video Summarization and Transmission Adaptation for Very Low Bit Rate Multiuser Wireless Uplink Video Com-munication. Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP).

[Li06] Z. Li, J. Huang, and A. K. Katsaggelos, "Pricing Based Collaborative Multi-User Video Streaming Over Power Constrained Wireless Downlink", Proc. IEEE Int'I Conference on Audio, Speech and Signal Processing (ICASSP), Toulouse, France, May 2006.

[Lienhart01] R. Lienhart, "Reliable Transition Detection in Videos: A Survey and Practioner's Guide", International Journal of Image and Graphics, 1(3):469-486, 2001.

[Liu05]Liu, S., Kuo, C.J. "Joint temporal-spatial bit allocation for video coding with dependency". IEEE Trans. on Cir-cuits & System for Video Tech, 15(1): 15-26, 2005.

[Liu04] Liu, P., Zhang, P., Jordan, S., Honig, M.L., "Single-Cell Forward Link Power Allocation Using Pricing in Wire-less Networks", IEEE Transactions on Wire-

less Communications: 3(2): 533-543, 2004

[Low2003] Low, S.H. " Duality Model of TCP and Queue Management Algorithms". IEEE/ACM Transactions on Networking, 11(4):525-536, Aug. 2003.

[Marbach03] P. Marbach and R. Berry. " Downlink resource allocation and pricing for wireless networks," in Proc. IEEE INFOCOM, New York, Jun. 2002, pp. 1470-1479.

[Merhav97] N. Merhav and V. Bhaskaran, "Fast algorithms for DCT-domain image downsampling and for inverse motion compensation", IEEE Trans. Circuits Syst. Video Tech., vol. 7, no. 3, pp. 468-476, Jun. 1997.

[Nas50] Nash, J. " Equilibrium points in n-person games", Proceedings of the National Academy of Sciences, 36: 48-49, 1950.

[Ohm94] J-R Ohm, " Three-dimensional subband coding with motion compensation," IEEE Trans. Image Processing, vol. 3, pp. 559-571, Sept. 1994.

[Ohm05] J.-R. Ohm, "Advances in Scalable Video Coding", Proc. of IEEE, pp. 42-56, Jan. 2005.

[Palomar06] Palomar, D.P., Chiang, M., A Tutorial on Decomposition Methods for Network Utilty Maxmization. To appear in IEEE Journal on Selected Areas of Communications, 2006.

[Pitas06] C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot detection and condensed representation", IEEE Signal Processing Magazine, 23(2):28-37, Mar. 2006.

[Ramchandran06] R. Puri, A. Majumdar, P. Ishwar, and K. Ramchandran, "Distributed Video Coding in Wireless Sensor Networks", IEEE Signal Processing Magazine, 23 (4), pp. 94-106., 2006.

[Ras01] Rasmusen, E. Games and Information: An Introduction to Game Theory, Blackwell Publishers, 3 edition, 2001.

[Rei03] Reichl, P., Bessler, S., Stiller, B. "Secondchance Auctions for Multimedia Session Pricing". Proc. MIPS 2003, Naples, Italy, Nov. 2003

[Saraydar02] Saraydar, C., Mandayam, N.B., Goodman, D., "Efficient Power Control via Pricing in Wireless Data Networks", IEEE Trans. on Wireless Communications, 50 (2): 291-303, 2002.

[Schaar03] M. van der Schaar, S. Krishnamachari, S. Choi, X. Xu, 2003. Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs. IEEE Journal on Selected Areas in Communications, 21(10):1752-1763.

[Shanableh00] T. Shanableh and M. Ghanbari, "Heterogeneous video transcoding to lower spatial-temporal resolutions and different encoding formats,", IEEE Trans. Multimedia, vol. 2, no. 2, pp. 101-110, Jun. 2000.

[Song00] J. Song and B.-L. Yeo, "A fast algorithm for DCT-domain inverse motion compensation based on shared information in a macroblock", IEEE Trans. Circuits

Syst. Video Techl., vol. 10, no. 5, pp. 767-775, Aug. 2000.

[Srikant04] Srikant, R., The Mathematics of Internet Congestion Control, Birkhauser, Boston, 2004.

[Sun96] H. Sun, W. Kwok, and J. W. Zdepski, "Architectures for MPEG compressed bitstream scaling," IEEE Trans. Circuits Syst. Video Tech., vol. 6, no. 2, pp. 191-199, Apr. 1996.

[Sun06] Sun, J., Zheng, L., Modiano, E. "Wireless Channel Allocation Using An Auction Algorithm", IEEE Journal of Selected Areas of Communications, 24(5):1085-1096, 2006

[Tse05] Tse, D., Viswanath, P., Fundamentals of Wireless Communication, Cambridge University Press, 2005.

[Vetro03] A. Vetro, C. Christopulos, and H. Sun, "Video Transcoding Architectures and Techniques", IEEE Signal Processing Magazine, vol. 20, no. 2, Mar., 2003.

[von47] von Neumann, J., Morgenstern O. Theory of Games and Economic Behavior. Princeton University Press, 1947.

[Wang05] Wang, L., Aghvami, A., Chambers, W., "Design issues of uplink media access control (MAC) protocols for multi-media traffic over DS-CDMA systems", IEEE Trans. on Multimedia, 7(3): 551-562, 2005.

[Wang00] Y. Wang, Z. Liu and J-C. Huang, "Multimedia Content Analysis", IEEE Signal Processing Magazine, vol. 17, November 2000

[Wu00] F. Wu, S. Li, and Y.-Q. Zhang, "DCT-prediction based progressive fine granularity scalable coding," in Proc. IEEE Int. Conf. Image Processing, 2000, pp. 1903-1906.

[Xin05] J. Xin, C-W. Lin, and M-T. Sun, "Digital Video Transcoding", Proceedings of IEEE, vol.93, no. 1, January, 2005.

[Xu04] Xu, L., Shen, X., Mark, J, "Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks", IEEE Trans. on Wireless Communications, 3(1): 60-73, 2004.

[Yoo04] Yoo, T., Setton, E., Zhu, X., Goldsmith, A., Girod, B,. Cross-layer design for video streaming over wireless ad hoc networks. Proceedings of IEEE Multimedia Signal Processing Workshop (MMSP), 2004.

[Youn99] J. Youn, M.-T. Sun, and J. Xin, "Video transcoder architectures for bit rate scaling of H.263 bit streams", Proc ACM Multimedia, Nov, 1999.

[Zhang05] Zhang, Q., Zhu, W., Zhang, Y.-Q., 2005. End-to-end QoS for video delivery over wireless internet. Proceedings of the IEEE, 93(1):123-134.

[Zhao02] Zhao, S., Xiong, Z., Wang, X., 2002. Joint error control and power allocation for video transmission over CDMA networks with multiuser detection. IEEE Trans. on Cir-cuits & System for Video Tech, 12(6): 425-437.

[Zheng03] Zheng, H., Optimizing wireless multimedia transmissions through cross layer design. Proceedings of IEEE International Conference on Multimedia & Expo (ICME), 2003.

BIOGRAPHIES

Zhu Li *received his B.S. in Computer Science from Sichuan University, Chengdu, in 1992, and M.S in Computer Science from University of Louisiana at Lafayette, in 1997, and his PhD in ECE from Northwestern University in 2004. He has been with the Multimedia Research Lab, Motorola Labs, since 2000.*

His research interests include video coding and communication, game theory and optimization decomposition techniques in multimedia networking, as well as manifold modeling and machine learning in multimedia retrieval and mining. He has 6 issued or pending patents, 23 publications in book chapter, journals and conference proceedings in these areas.

He is an IEEE member. He is a guest co-editor for Journal of Advances in Multimedia special issue on "Collaboration and Optimization in Multimedia Communication", a track co-chair for "Wireless Multimedia Streaming and Networking" for IEEE IC3N '07. He has also served as TPC member for IEEE Int'l Conf on Communication (ICC)'06, '07, IEEE Int'l Conf on Multimedia & Expo (ICME)'06, PacketVideo'06, IEEE GLOBECOM '06 and IEEE CCNC '07.

Jianwei Huang *received the B.S. degree in Electrical Engineering from Southeast University, Nanjing, China, in 2000, and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Northwestern University, Evanston, IL, in 2003 and 2005, respectively. He has been working as a Postdoctoral Research Associate in the Department of Electrical Engineering, Princeton University, Princeton, NJ, since November 2005. During the summer of 2004 and 2005, he worked as a software engineer in the Mathematics of Communication Networks Group at Motorola, Arlington Heights, IL. During the summer of 1999, he worked as an intern in the Department of Change Management at GKN Westland Aerospace Co. Ltd., Isle of Wight, UK.*

His current research interests lie in the areas of wireless and wireline communications networks, with emphases on game theoretical models, resource allocation, network pricing, dynamic spectrum sharing, mobile ad hoc and sensor networks, multimedia communications, stochastic

and nonconvex optimizations.

He was awarded a 2001 Walter P. Murphy Fellowship at Northwestern University, and a 1999 Chinese National Excellent Student Award. He is the Lead Guest Editor for Journal of Advances in Multimedia special issue on Collaboration and Optimization in Multimedia Communication, and a TPC member for IEEE CCNC'07.

Aggelos K. Katsaggelos *received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees both in electrical engineering from the Georgia Institute of Technology, in 1981 and 1985, respectively. In 1985 he joined the Department of Electrical and Computer Engineering at Northwestern University, where he is currently professor. He was the holder of the Ameritech Chair of Information Technology (1997-2003). He is also the Director of the Motorola Center for Seamless Communications and a member of the Academic Affiliate Staff, Department of Medicine, at Evanston Hospital.*

Dr. Katsaggelos is a member of the Publication Board of the IEEE Proceedings, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, the Editorial Board of Academic Press, Marcel Dekker: Signal Processing Series, Applied Signal Processing, and Computer Journal. He has served as editor-in-chief of the IEEE Signal Processing Magazine (1997-2002), a member of the Publication Boards of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, an Associate editor for the IEEE Transactions on Signal Processing (1990-1992), an area editor for the journal Graphical Models and Image Processing (1992-1995), a member of the Steering Committees of the IEEE Transactions on Image Processing (1992-1997) and the IEEE Transactions on Medical Imaging (1990-1999), a member of the IEEE Technical Committee on Image and Multi-Dimensional Signal Processing (1992-1998), and a member of the Board of Governors of the IEEE Signal Processing Society (1999-2001). He is the editor of Digital Image Restoration (Springer-Verlag 1991), co-author of Rate-Distortion Based Video Compression (Kluwer 1997), co-editor of Recovery Techniques for Image and Video Compression and Transmission, (Kluwer 1998), co-author of Super-resolution for Images and Video (Claypool, 2006) and Joint Source-Channel Video Transmission (Claypool, 2006). He is the co-inventor of twelve international patents, a Fellow of the IEEE, and the recipient of the IEEE Third

Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), an IEEE Signal Processing Society Best Paper Award (2001), and an IEEE International Conference on Multimedia and Expo Paper Award (2006). He is a Distinguished Lecturer of the IEEE Signal Processing Society (2007-08).

Mung Chiang *is an Assistant Professor of Electrical Engineering, and an affiliated faculty of the Program in Applied and Computational Mathematics at Princeton University. He received the B.S. (Hons.) degree in electrical engineering and in mathematics, and the M. S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1999, 2000, and 2003, respectively. He conducts research in the areas of optimization of communication systems, theoretical foundations of network architectures, wireless and DSL broadband access networks, and stochastic theory of communications and networking. He has been awarded a Hertz Foundation Fellowship, and received the Stanford University School of Engineering Terman Award for Academic Excellence, the SBC Communications New Technology Introduction contribution award, the National Science Foundation CAREER Award, and the Princeton University Howard B. Wentz Junior Faculty Award. He is a co-editor of Springer book series on Optimization and Control of Communication Systems, the Lead Guest Editor of the IEEE Journal of Selected Areas in Communications special issue on Nonlinear Optimization of Communication Systems, a Guest Editor of the IEEE Transactions on Information Theory and IEEE/ACM Transactions on Networking joint special issue on Networking and Information Theory, and the Program Co-Chair of the 38th Conference on Information Sciences and Systems.*