Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2008, Article ID 219570, 14 pages
doi:10.1155/2008/219570

Research Article

Energy-Efficient Bandwidth Allocation for Multiuser Scalable Video Streaming over WLAN

Xin Ji,[1,2,3] Sofie Pollin,[1,2,3,4] Gauthier Lafruit,[2] Iole Moccagatta,[2] Antoine Dejonghe,[2,3] and Francky Catthoor[1,2]

[1] *Katholieke Universiteit Leuven, 3000 Leuven, Belgium*
[2] *IMEC, Kapeldreef 75, 3001 Leuven, Belgium*
[3] *Interdisciplinary Institute for Broadband Technology (IBBT), Ghent University, 9000 Gent, Belgium*
[4] *UC Berkeley, CA 94720, USA*

Correspondence should be addressed to Xin Ji, xin.ji@imec.be

Received 27 February 2007; Accepted 9 October 2007

Recommended by Peter Schelkens

We consider the problem of packet scheduling for the transmission of multiple video streams over a wireless local area network (WLAN). A cross-layer optimization framework is proposed to minimize the wireless transceiver energy consumption while meeting the user required visual quality constraints. The framework relies on the IEEE 802.11 standard and on the embedded bitstream structure of the scalable video coding scheme. It integrates an application-level video quality metric as QoS constraint (instead of a communication layer quality metric) with energy consumption optimization through link layer scaling and sleeping. Both energy minimization and min-max energy optimization strategies are discussed. Simulation results demonstrate significant energy gains compared to the state-of-the-art approaches.

1. INTRODUCTION

The demand for multimedia transmission over wireless networks exhibits an ever growing trend. As a result, the transmission of multiple video streams over a single wireless local area network (WLAN) is becoming a key requirement. In this context, quality of service (QoS) provisioning for real-time applications among different users is becoming more and more critical, as wireless networks are affected by extremely error-prone and time-varying conditions. Besides this QoS challenge, low-power consumption is imperative to enable the deployment of broadband wireless connectivity in battery-operated portable devices.

Dynamically, adapting video packet selection and scheduling to achieve appropriate visual quality and energy efficiency for such varying wireless networks is a challenging task. For simplicity, most of the WLAN transmission studies consider throughput as the most important performance metric, while it is not the most appropriate choice for video traffic. Some recent studies try however to improve the transmission performance by exploring the specificities of video traffic. For instance, considering scalable video coding techniques [1–4], different retransmission limits were defined for different MAC priority queues in [5, 6]. These approaches rely on scalable video coding's inherent prioritization in the compressed domain to set MAC priorities. In [7], a solution for scheduling transmission opportunities (referred to as TXOP in the remainder of the present paper) as a function of the data type was proposed.

As far as energy efficiency is concerned, a substantial body of prior work focuses on energy-efficient wireless transmission from the viewpoints of medium access control (MAC) or physical (PHY) layers [8–10]. For energy-efficient wireless media systems, Goel et al. solved an image transmission energy optimization problem subject to distortion and rate constraints [11]. He et al. in [12] developed a power-rate-distortion analysis framework to extend the traditional rate-distortion analysis by including power consumption as a third dimension. Although hardware-specific impacts were appropriately considered in [12], the analysis lacked a sufficient consideration of channel coding and transmission with respect to the time-varying characteristics of the wireless

channel. Focusing on an uplink mobile-to-base station scenario, Lu et al. solved in [13] a power optimization problem subject to an end-to-end distortion constraint relying on H.263 source coding and RS channel coding in conjunction with the Gilbert loss model. In [14], Chandra and Dey presented a technique for enabling real-time video compression and transmission from wireless appliances based on run-time video adaptation, and they estimated the energy consumption based on CPU load. Yousefi'zadeh et al. formulated a set of optimization problems in [15] aimed at minimizing total power consumption of wireless media systems subject to a given level of QoS and an available bitrate relying on multiple antennas. None of the aforementioned power optimization works considered the video scalability influence

on the power consumption of wireless multimedia systems.

In addition, to the best of our knowledge, there is no prior work considering joint optimization of real video quality and energy efficiency for wireless media systems. This optimization requires to take the whole protocol stack into account. Furthermore, only few researches have provided an analysis of the complexity of the proposed optimization problems. To cope with the time-varying QoS, existing methodologies often rely on fixed or nonscalable flow-based optimizations to allocate the available network resources across the various multimedia users. Moreover, previous researches have seldom jointly exploited the adaptation or protection techniques available at the medium access control (MAC) or physical (PHY) layers to enhance the performance of video applications. On the one hand, we can only fully benefit from new technologies if we can analyze the behavior of adaptation processes acting over communication networks, taking into account the intrinsically stochastic nature of communications and observations. On the other hand, adaptation leads to nontrivial tradeoffs among many parameters (i.e., delay, reliability, energy cost, etc.); thus rethinking of the entire communication systems and quality of service must be provided.

The main contribution of this paper is to exploit the application layer peak signal-to-noise ratio (PSNR) scalability enabled by the rate-distortion properties of scalable video bitstreams and to minimize the energy consumption among different users. Instead of using conventional communication layer QoS metrics, such as throughput or packet loss probability, a proper application-level video quality metric is considered in the optimization. Compared to our former work on energy-efficient video transmission over WLAN, the resulting solution enables to further minimize the wireless transceiver energy consumptions by a factor of 2 without degrading the visual quality. The considered setup consists of multiple independent users equipped with mobile terminals (MTs) downloading video streams from the access point (AP) of a WLAN (see Figure 1). The video data are encoded using a scalable video coding scheme and stored on a video server accessed through the AP. Therefore, no real-time encoding is performed. The users receive data over a shared slowly fading wireless channel. It is assumed that different users can require different video qualities. This is a very important and realistic test case. For instance, considering a distance learning system, when a student is studying in real time using a wireless device and facing battery exhausting problems, he/she may be willing to scarify some visual quality to finish the whole studying process.

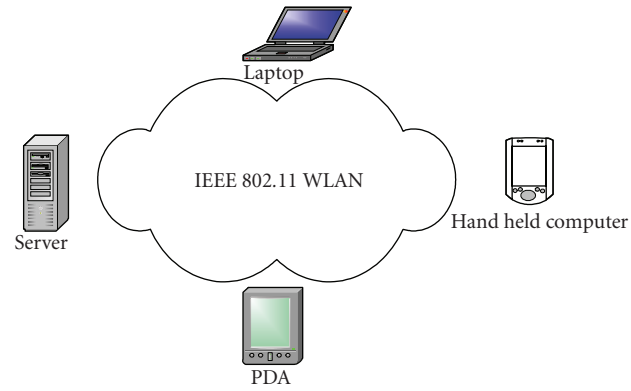

Figure 1: WLAN access point (AP) manages several mobile terminals (MTs) in a centralized network.

The remainder of this paper is organized as follows. Section 2 provides the background for understanding the contributions of this paper. Section 3 briefly reviews the IEEE 802.11 WLAN standards and the deployed 3D wavelet motion-compensated temporal filtering (MCTF) scalable video coding scheme. Section 4 formulates the considered problem statement for energy-efficient video scheduling with rate-distortion awareness. Total energy minimization and fairness optimization are formulated separately. Next, lightweight algorithms are designed to solve the run-time optimization problems for practical use. Appropriate system models are used to instantiate the proposed cross-layer optimization framework given the aforementioned standards. In Section 5, we examine the performance of our framework through simulations. Finally, concluding remarks are provided in Section 6.

2. BACKGROUD AND PRELIMINARY WORK

Compared to the capacity improvements of wireless transmission techniques, there are limited advances in battery capacity. Since more powerful transmission schemes cost more energy, there is an increasing energy gap between the energy requirements of new applications and radio technologies and the energy awareness in the battery. Thus, it is critical to reduce the power consumption or, equivalently, to enhance the energy efficiency of the mobile devices.

The goal of improving the energy efficiency of wireless communication devices has already triggered a lot of researches at various levels, from circuit to communication theories and networking protocols. The energy management problem, in its most general formulation, consists in dynamically controlling the system to minimize the average energy consumption under a performance constraint. Existing researches can be classified into two categories.

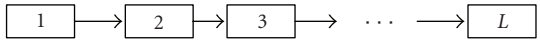

Figure 2: Directed acyclic graph of embedded bitstream.

(i) Top-down approaches: approaches that are intrinsically utilization- and hardware-aware but communication-unaware are categorized as top-down. The communicating device is treated as any electronic circuit, and general-purpose techniques like dynamic power management and energy-aware design are applied. The first technique is defined as dynamically reconfiguring an electronic system to provide the requested performance levels with a minimum number of active components and minimum loads on those components [16, 17]. The second technique can be defined as designing systems that present a desirable energy-performance behavior for energy management [8, 18, 19].

(ii) Bottom-up approaches: approaches that are intrinsically communication-aware but hardware- and utilization-unaware are categorized as bottom-up. They rely on the fundamentals of information and communication theories to derive energy-aware transmission techniques and communication protocols. We find here, for instance, the transmission scaling techniques which exploit the fundamental tradeoff that exists between transmission rate/power and energy [20, 21]. Network power management techniques also fall in this category, targeting the minimization of the transmission power under QoS constraints [22].

(iii) Top-down and bottom-up approaches can easily result in a fundamental contradiction. A good example is the conflict between transmission scaling at the physical (PHY) layer (bottom-up) and sleeping schemes at the MAC layer [23] (top-down). Scaling tends to minimize transmission energy consumption by transmitting with the lowest power over the longest feasible duration, whereas sleeping tends to minimize the duty cycle of the radio circuitry by transmitting as fast as possible. Clearly, the two techniques are contradictory when it comes to defining the optimal transmit rate and power allocation.

In [24, 25], we showed that a cross-layer combination of both approaches can significantly decrease the energy consumption in a multiuser scenario. A framework was proposed for allocating the network resources energy efficiently. The framework is subdivided into two steps and it focuses on the PHY and MAC layers for which the energy, packet error rate (PER), and transmission time are considered. First, during the design-time phase, the performance-energy scalability resulting from the available controllable parameters of the system is analyzed. Cost-resource-quality tradeoffs, taking into account energy cost, PER quality, and transmission time resource requirements of each user, are fully characterized for each possible system state (i.e., a finite set of possible realizations of external variables tracking system dynamics). Second, during the run-time phase, knowing the current system state and relying on the tradeoff characterized in the design-time phase, the server/access point searches the tradeoff curves of the different users in order to minimize the total energy cost subject to a fixed and bounded transmission time delay. It then allocates the corresponding configuration to the different user devices.

In this paper, we introduce the rate-distortion property of the video bitstreams into the proposed cross-layer framework and show that significant energy gains can be achieved by exploiting this property. Besides all the scalability existing in the PHY and MAC layers, a significant amount of scalability is available in the video bitstream. A directed acyclic graph is often used to express the interdependencies between the different data units. A typical dependence graph of an embedded coded bitstream is sequential, as shown in Figure 2 [26]. The arrow directions show that a data unit can be correctly decoded only when the dependent data units are also correctly decoded. From the graph, we know that the loss of different data units can result in varying decoded visual qualities. Many unequal error protection schemes have been developed based on this observation. By introducing this property into the proposed cross-layer framework, we show that significant energy gains can be achieved. The proposed scheme is practical and can be integrated within existing wireless and multimedia standards.

3. WLAN VIDEO STREAMING SYSTEM OVERVIEW AND ENERGY-PERFORMANCE MODELING

The use of IEEE 802.11 WLANs is growing at a rapid pace. With the substantial increase in the available bitrates, the transmission of real-time audio/video applications over WLANs becomes a reality. In this section, we first briefly introduce the IEEE 802.11 standard and the scalable video coding scheme that are considered in the present work. It is however important to emphasize that the cross-layer algorithms proposed in this paper can be deployed with any video coding scheme where the bitstream can be organized into data units with embedded structure (see Section 3.3). Based on this description, we show how to calculate the energy consumption, the transmission delay, the error probability of the data, and the expected quality of the received decoded video.

3.1. PHY modes of 802.11a OFDM and channel model

The IEEE 802.11a [27] PHY layer is based on orthogonal frequency division multiplexing (OFDM), and it provides eight different modes with different modulation schemes and code rates resulting in data transmission rates ranging from 6 to 54 Mbps. The corresponding data rate and the associated power level requirements are provided in Table 1, where N_{DBPS} denotes the number of data bits per symbol.

3.1.1. PHY layer performance model

We consider a direct-conversion radio transceiver architecture [28]. Four control dimensions have significant impact on energy and performance for these OFDM transceivers: the modulation order (N_{mod}), the code rate (B_c), the power

Table 1: Multiple PHY modes for IEEE 802.11a.

Mode	Data rate (Mbps)	Min (dBm)	Modulation	Code rate (R)	N_{DBPS}
1	6	−82	BPSK	1/2	24
2	9	−81	BPSK	3/4	36
3	12	−79	QPSK	1/2	48
4	18	−77	QPSK	3/4	72
5	24	−74	16-QAM	1/2	96
6	36	−70	16-QAM	3/4	144
7	48	−66	64-QAM	1/2	192
8	54	−65	64-QAM	3/4	216

amplifier transmit power (P_{TX}), and its linearity specified by the backoff (b). For a given data rate, communication performance is determined by the bit error rate (BER) at the receiver. Adding nonlinearity distortion to the received signal power, the BER can be expressed as a function of the received signal-to-noise and distortion ratio (SINAD) which can be expressed as

$$\text{SINAD} = \frac{P_{\text{TX}} \times A}{A \times D(b) + kT \times W \times N_f}, \tag{1}$$

where A denotes the channel attenuation, the constants k, T, W, and N_f are the Boltzman constant, working temperature, channel bandwidth, and noise figure of the receiver, respectively, and the relation between the power amplifier backoff b and the distortion $D(b)$ has been characterized empirically for the Microsemi LX 5506 [29] 802.11a PA. The considered BER-SINAD relation follows the model provided in [30]. The BER-SINAD curves for different channel states for all the considered PHY modes have been shown in Figure 3.

3.1.2. PHY layer energy model

Our energy model assumes the implementation detailed in [31]. The corresponding parameters are provided in Table 2. The time needed to wake up the system is assumed to be 100 microseconds. Denoting P_{PA} as the power consumption of the power amplifier, P_{FE} as the power consumption of the front end (FE), P_{BB} as the power consumption of the baseband, and E^R_{DSP} as the digital signal processor energy consumption for decoding a single bit of a turbo-coded packet, we obtain the following expressions for the energy needed to send or receive a MAC service data unit (MSDU) of length L_{MSDU} under bit rate B_{bit}:

$$\begin{aligned} E_{\text{TX}} &= \left(\frac{P^T_{\text{PA}} + P^T_{\text{FE}} + P^T_{\text{BB}}}{B_{\text{bit}}}\right) \times L_{\text{MSDU}}, \\ E_{\text{RX}} &= \left(\frac{P^R_{\text{FE}} + P^R_{\text{BB}}}{B_{\text{bit}}} + E^R_{\text{DSP}}\right) \times L_{\text{MSDU}}. \end{aligned} \tag{2}$$

3.2. Error probability, energy consumption, and transmission delay of the IEEE 802.11 MAC

Considering a possible transmission configuration vector K (each specific control dimension listed in Table 2 corresponds to an entry in this vector), the energy and time needed to send an MSDU can be, respectively, expressed as $E_{\text{MSDU}}(K)$ and $\text{TXOP}_{\text{MSDU}}(K)$ [24, 25]. The energy cost and time of transmitting an application layer packet p are then, respectively, defined as $E_p(K)$ and $\text{TXOP}_p(K)$, and these values depend on the number of fragmented data units that need to be transmitted or retransmitted for successful packet transmission. The retransmission scheme details of 802.11 MAC can be found in [32]. As the total energy and time needed to transmit a packet p are the sum of the energy and time needed to transmit its fragments, $E_p(K)$ and $\text{TXOP}_p(K)$ can be, respectively, expressed as

$$\begin{aligned} E_p(K) &= (m + y)E_{\text{MSDU}}(K), \\ \text{TXOP}_p(K) &= (m + y)\text{TXOP}_{\text{MSDU}}(K), \end{aligned} \tag{3}$$

where m denotes the number of MSDU fragments for the considered packet p, and y denotes the allowed number of MSDUs that can be retransmitted for the given packet p.

Similarly, the loss probability of a single MSDU is denoted as $P_{\text{MSDU}}(K)$, and it is computed based on the PHY performance model introduced before. Since the probability that a given packet p is received correctly depends on the probabilities that each of its fragments is received correctly, We compute the packet error rate $\text{PER}^m_y(K)$ at application layer according to

$$\begin{aligned} \text{PER}^m_y(K) &= 1 - \sum_{j=0}^{y} P^m_{ej}(K), \\ P^m_{ey}(K) &= \sum C^m_i (P_{\text{MSDU}})^i (1 - P_{\text{MSDU}})^{(m-i)} P^i_{y-i}(K), \\ P^m_{e0}(K) &= (1 - P_{\text{MSDU}})^m. \end{aligned} \tag{4}$$

We refer to [24, 25] for more details on the wireless channel model and the link layer scaling (adapting the modulation order and code rate to spread the transmission over time) and sleeping (introducing as much as possible transmission idle period) optimization schemes.

3.3. Distortion, energy, and delay of scalable video bitstream

Embedded scalable video coding has been an active research topic in recent years. It has the attractive capability of reconstructing lower resolution or lower quality videos from a single bitstream, hence providing simple and flexible solutions for transmission over heterogeneous network conditions and easier adaptation to a variety of storage devices and

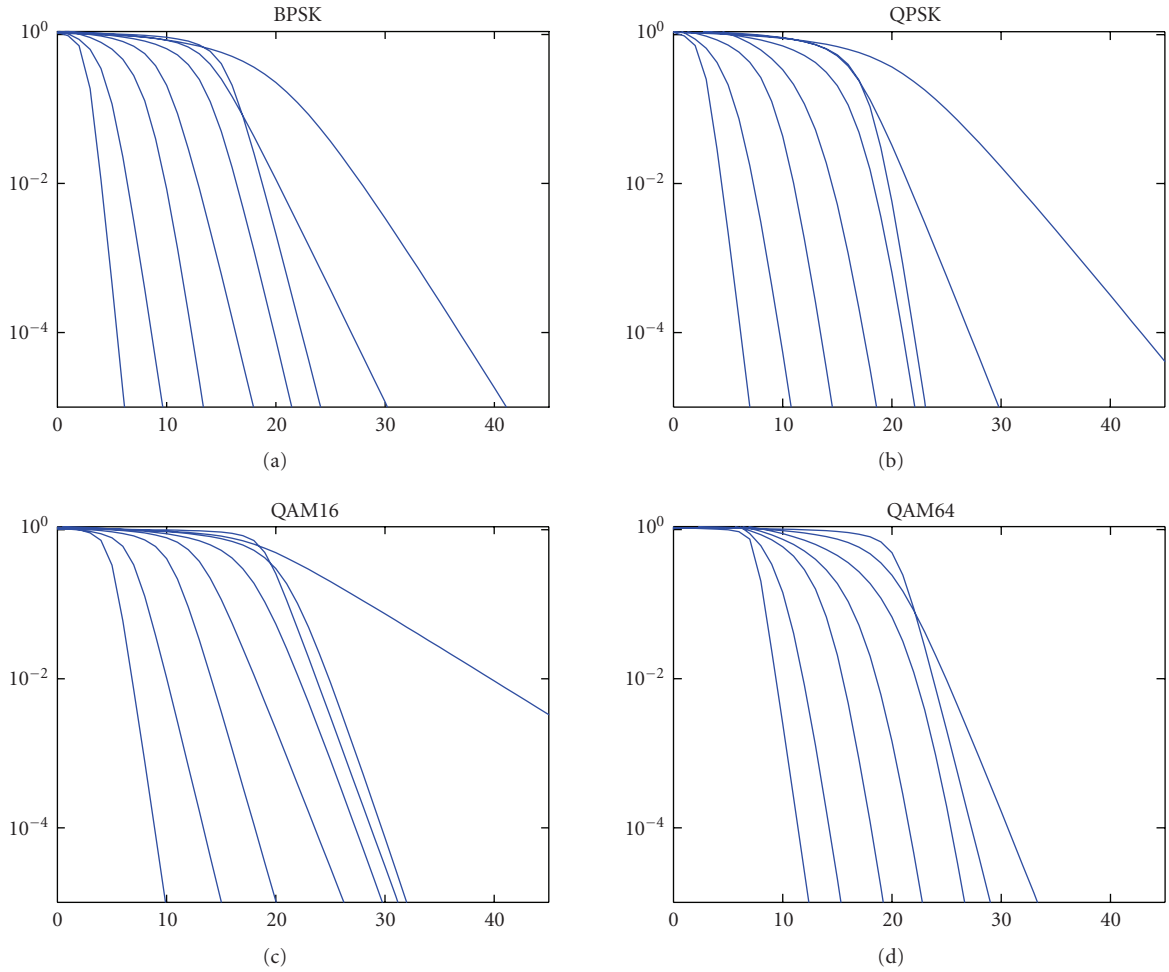

Figure 3: BER-SINAD relations.

Table 2: Parameters of the energy model.

Performance model	Energy model	MAC model	Control dimensions
W = 20 MHz	$P_{\text{FE}}^{T} = 200$ mW	$L_{\text{MSDU}} = 1500$ B	Backoff (dB) 6 to 6
B = 250 kBaud	$P_{\text{FE}}^{R} = 200$ mW	$T_{\text{ACK}} = 52\,\mu$s	P_{out} (dBm) 0 to 20
$N_c = 48$	$P_{\text{BB}}^{T} = 50$ mW	$T_{\text{PLCP}} = 20\,\mu$s	Modulation BPSK, QPSK, 16-QAM, 64-QAM
T = 198 K	$P_{\text{BB}}^{R} = 50$ mW	$T_{\text{SIFS}} = 16\,\mu$s	Code rates 1/2, 2/3, 3/4
$N_f = 10$ dB	$E_{\text{DSP}}^{R} = 8.7\,nJ/b$	Block = 288	PSNR

terminals. Accordingly, many recent video codecs, such as the Scalable Video Coding (SVC) extensions of H.264/AVC [3], MPEG-4 FGS [4], and so forth, enable embedded scalable coding.

3.3.1. Architecture of the considered scalable video encoder

We consider a scalable video codec based on motion-compensated temporal filtering (MCTF) and a wavelet transform [2]. MCTF aims at removing the temporal redundancies of video sequences. Unlike predictive coding schemes, it does not employ a closed-loop prediction scheme. Instead, it uses an open-loop pyramidal decomposition to remove both long-term and short-term temporal dependencies in an efficient manner [33]. After the removal of the temporal redundancies, the produced low-pass and high-pass frames are decomposed spatially by discrete wavelet transform (DWT). In a typical MCTF-based video compression, the rate allocation of the scalable bitstream is possible for a maximum granularity of one group of pictures (GOP). Encoder and decoder thus process the video sequence on a GOP-by-GOP basis, which creates naturally independent data units group.

An important feature of wavelet transforms is the inherent support of scalability in the compressed domain. Coupled with the embedded coding techniques, wavelet video coding achieves continuous rate scalability. After applying the wavelet transform, the resulting subband coefficients are coded using bitplane coding and a global rate-distortion

optimization. As a result, the final bitstream is constructed to satisfy the bitrate constraint and minimize the overall distortion [2].

To achieve quality scalability, a multilayer bitstream is formed where each layer represents a quality-level improvement. The fractional bitplane coding ensures that the bitstream is embedded with fine granularity. In this work, we distribute the rate of the layers inside a GOP in a way that every enhancement quality layer contributes to a similar distortion decrease. The resulting embedded bitstream has a sequential dependency; each layer can only be decoded under the condition that all the previous layers have been received. Note that in our simulations, no error concealment is used. In the next section, we will explain in detail how to estimate the distortion in the case of packet losses for these coding assumptions.

3.3.2. Distortion, energy, and delay calibrations of video bitstream

Commonly used quality measurements of reconstructed images and videos are mean squared error (MSE) and peak signal-to-noise ratio (PSNR). Typical PSNR values should range from 30 to 40 dB. Taking only quality scalability into account and assuming a stable channel during one GOP time period, it is possible to calculate the expected distortion contribution of each quality layer on a GOP-per-GOP basis. We focused on a GOP-based approach instead of the more fine granular ones to limit overhead and complexity.

Let us assume that each GOP is encoded into L quality layers and that a quality layer is the smallest application layer data unit. Let D_l denote the distortion corresponding to the reception of layers 1 to l ($1 < l < L$), and let D_0 denote the distortion associated with losing the first layer. Denoting the error probability of layer l under transmission configuration K_l as PER_{K_l}, the probability of correctly receiving the quality layers until layer l is $\prod_{j=1}^{l}(1-\mathrm{PER}_{K_j})$. Relying on the sequential dependency of the embedded bitstream structure, the expected average distortion D_e over one GOP can then be calculated as

$$D_e = \mathrm{PER}_{K_l} \times D_0 + \sum_{i=1}^{L-1}\left[\prod_{j=1}^{i}(1-\mathrm{PER}_{K_j})\right] \times \mathrm{PER}_{K_{i+1}} \times D_i + \left[\prod_{i=1}^{L}(1-\mathrm{PER}_{K_i})\right] \times D_L. \tag{5}$$

The energy E_{GOP} of the whole GOP can be expressed as the sum of its layers:

$$E_{\mathrm{GOP}} = \sum_{i=1}^{L} E_{p_i}(K_i), \tag{6}$$

where $E_{p_l}(K_l)$ denotes the associated energy cost under configuration K_l.

Similarly, the transmission time $\mathrm{TXOP}_{\mathrm{GOP}}$ of the whole GOP is

$$\mathrm{TXOP}_{\mathrm{GOP}} = \sum_{i=1}^{L} \mathrm{TXOP}_{p_i}(K_i), \tag{7}$$

where $\mathrm{TXOP}_{p_i}(K_i)$ denotes the transmission time under configuration K_l.

4. ENERGY-EFFICIENT MULTIUSER CROSS-LAYER OPTIMIZATION

4.1. Problem formulation

In this paper, we focus on techniques that efficiently adapt the transmission strategy in order to minimize the transceiver energy cost while meeting the required end-to-end distortion and delay. Most of the existing solutions do not take into account the rate-distortion properties of video bitstreams, and therefore they often lead to inferior network efficiency and suboptimal qualities for the video users.

As we operate in a very dynamic environment, the system behavior will vary over time. Both the energy cost function and the resources required for transmission will depend on this run-time behavior. In the considered wireless video streaming environment, the system state is determined by the current channel state and the rate-distortion property of the video bitstream. Each GOP can then be associated with a set of possible system states S, which determines the mapping of the transmission strategies K to the energy cost ($K \rightarrow E_{\mathrm{GOP},S}$) and the required bandwidth resource ($K \rightarrow \mathrm{TXOP}_{\mathrm{GOP},S}$). Each user experiences different channel and rate-distortion dynamics, resulting in different system states over time, which may or may not be correlated with other users. It is this important characteristic which makes it possible to exploit multiuser diversity for energy efficiency.

From the former analysis, and under the assumption that all video users can require their own end-to-end quality, the optimization problem is formulated with video quality as one of the constraints. We consider two different objectives: minimizing the total energy cost of all users, and the maximum energy cost among all users (fairness rule). For both objectives, we provide a low-complexity run-time optimization algorithm. The advantage of the proposed solutions will be analyzed and discussed in Section 5.

4.1.1. Optimization towards total energy minimization

The optimization consists in finding for each user u, $u \in (1,\ldots,N)$, the configuration K_u^* that minimizes the overall energy cost, subject to radio resource and video distortion constraints. Such configuration is applied at the beginning of every GOP transmission interval, considering the current channel conditions and video rate-distortion properties:

$$K_u^* = \min\left(\sum_{u=1}^{N} E_{\mathrm{GOP}_u}(K_u)\right), \tag{8}$$

subject to

$$D_{e_u} \le D_u^r, \quad u \in (1,\ldots,N), \qquad \sum_{u=1}^{N} \mathrm{TXOP}_{\mathrm{GOP}_u} \le T_r, \tag{9}$$

where D_u^r and T_r denote the distortion and time constraints, respectively.

4.1.2. Optimization towards fairness

In this approach, we consider how to allocate the bandwidth and transmission strategies to achieve more fair energy cost among all the users and formulate this problem as a min-max problem. For N users inside the network, the optimization problem is formulated as a min-max problem to find for each of the users u the configuration K_u^* such that

$$K_u^* = \text{argmin}(\max E_{\text{GOP}_u}(K_u)), \quad u = 1, \ldots, N, \tag{10}$$

subject to

$$\begin{aligned} D_{e_u} &\le D_u^r, \quad u \in (1, \ldots, N), \\ \sum_{u=1}^{N} \text{TXOP}_{\text{GOP}_u} &\le T_r, \end{aligned} \tag{11}$$

where D_u^r and T_r denote the distortion and time constraint, respectively.

4.2. Two-phase solution approach

Each of the above formulated problems is a multidimensional assignment problem, which is known to be nonpolynomial (NP) time hard problem. To obtain a tractable run-time complexity, we proposed a two-phase solution approach; at design time, for each possible system state, the optimal operating points (namely, Pareto sets) are determined according to their minimal energy cost and resource (TXOP) consumption. At run time, a low-complexity algorithm is provided for the formulation of each of the problems relying on the design time calibration.

To solve the optimization towards total energy minimization, we convert the problem into a Lagrangian relaxation problem. The main steps are as follows.

(i) At the *design time*, the optimal operational points are determined for each possible system state according to their minimal energy cost, resource (TXOP) consumption, and distortion. The operational points are generated to reduce the search space from the initial problem.

(ii) At the *run time*, the bisection algorithm is used to solve the optimization problem.

To solve the min-max problem, the main steps are as follows.

(i) At the *design time*, the derivation of the optimal operational points is performed for the original Min-Max problem after the system states of all users are known.

(ii) At *run time*, a lightweight water-filling scheme is proposed to assign the optimal system configuration to each user.

In the Sections 4.2.1–4.2.4, the design-time and run-time approaches will be introduced, respectively. The two proposed algorithms will be detailed in the following.

4.2.1. Design-time phase

The goal of the design-time phase is to determine, for each possible system state, the optimal operating points according to their minimal cost and resource consumption. In this paper, the system states are denoted by different channel statuses and the dynamic rate-distortion properties of video traffic loads. To that end, we consider the Pareto concept for multi-objective optimization [34].

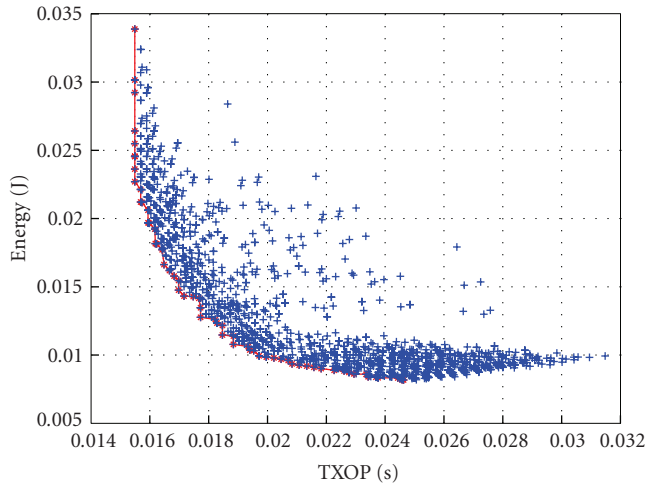

Figure 4: Energy versus TXOP Pareto curve example.

Let us consider the following multi-objective programming problem:

$$\underset{X \in \Omega}{\text{MIN}}\, f_1(X), f_2(X), \ldots, f_M(X) \tag{12}$$

$$\begin{aligned} f_i(X_1) &\le f_i(X_2), \quad \forall\ i \in 1, 2, \ldots, M, \\ f_j(X_1) &< f_j(X_2), \quad \forall\ j \in 1, 2, \ldots, M, \end{aligned} \tag{13}$$

A solution X_1 is strictly better than a solution X_2 if X_1 is at least as good as X_2 with respect to all the M objectives (the first condition of (13)), and X_1 is strictly better than X_2 with respect to at least one objective (the second condition of (13)). A *Pareto optimal solution* is defined as if there is no other solution strictly better than X_1. A multi-objective optimization problem may have multiple Pareto optimal solutions, and different decision makers with different preferences may select different Pareto optimal solutions. The set of all possible Pareto optimal solutions constitutes a *Pareto frontier* in the objective space. A two-dimensional Pareto frontier is also called a *Pareto curve*. Figure 4 shows an example of Pareto curve considering energy and network resources as objectives.

At design time, for each possible system state, we compute the three-dimensional Pareto frontier, considering the optimization objectives of the distortion D, the network resource TXOP, and the energy E. The Pareto frontier can be found by any global optimal algorithm since complexity is not the concern at the design-time step.

From the video side, the design-time calibration can be provided for different GOP sizes. This enables adaptation to channel violations by choosing a smaller GOP size for the next coherence period. Depending on the channel state (how long it is stable), we may adapt the number of frames without influencing the later part of the video bitstream—thanks

(1) Initialization:
(a) allocate to each of the u users the lowest cost possible for the given state $E^{\min}_{\mathrm{GOP}_u}$.
(2) If $\sum_{u=1}^{N} \mathrm{TXOP}^0_u > T_r$,
initialize $\lambda_{\max}$, $\lambda_{\min}$, and $\lambda_{\mathrm{trying}}$.
Do
restore previous $\lambda_{\mathrm{trying}}$:
$\lambda_{\mathrm{trying}} = (\lambda_{\max} + \lambda_{\min})/2$.
For each user,
if $\lambda_{\mathrm{trying}}$ is higher than the highest or lower than the lowest,
jump out of the loop
or else find $\lambda > \lambda_{\mathrm{trying}} > \lambda_{\mathrm{next}}$.
If the total delay is lower than the constraint
$\lambda_{\max} = \lambda_{\mathrm{trying}}$,
restore the difference between the total delay and the constraint
or else $\lambda_{\min} = \lambda_{\mathrm{trying}}$.
While (the difference between total delay and constraint converges to the same point),
(3) for each of the users,
the Pareto energy TXOP set will be searched
till finding the configurations which have a λ just lower or equal to the resulting λ,
and these configurations are the optimal output settings.

Algorithm 1: Run-time bisection search algorithm to find the Lagrange multiplier and the optimal configuration.

to the open-loop temporal decomposition of the MCTF scheme.

4.2.2. Run-time phase

Once the system states of all users have been known at run time, lightweight schemes are proposed to assign the best system configuration to each user.

The 3D *Pareto frontier* is firstly converted to a 2D *Pareto curve* according to the QoS constraint. This step can also be incorporated in the design-time phase by providing several QoS constraint levels (2D *Pareto curve*) for the run-time choice. The *Pareto frontier* is first pruned by deleting those settings that cannot satisfy the QoS constraint. The remaining cost-resource tradeoffs are further explored to extract a *Pareto curve*.

After the Pareto pruning, n Pareto cost-resource sets are available for each user. The run-time algorithms for both problem formulations are discussed in the next sections.

4.2.3. Proposed algorithm for minimizing total energy

The optimization problem expressed in (9)–(10) is reformulated so as to introduce a Lagrangian multiplier λ [35]:

$$\text{minimize } J_{\mathrm{tot}} = \sum_{u=1}^{N} E_{\mathrm{GOP}_u}(K_u) + \lambda \sum_{u=1}^{N} \mathrm{TXOP}_{\mathrm{GOP}_u}, \tag{14}$$

subject to

$$\sum_{u=1}^{N} \mathrm{TXOP}_{\mathrm{GOP}_u} \le T_r. \tag{15}$$

The conventional solution consists in constructing a convex hull of the operational points first, and then searching the slope (λ sets) of the convex hull. In contrast, we define the λ sets to be the slope $E_{\mathrm{GOP}_u}/\mathrm{TXOP}_{\mathrm{GOP}_u}$ of each operational point, and we find the lowest λ^* which satisfies the constraint. From the definition of λ, we know that it represents the energy cost compared to the resource. And from the Pareto property, for each specific user, the $\lambda(K_u)$ is increasing with $E_{\mathrm{GOP}_u}(K_u)$. The lower the λ is, the lower the E_{GOP_i} will be. Thus, if all the users choose configurations with a λ lower than λ^*, the constraint will not be satisfied. And if all the users choose configurations with a λ larger than λ^*, the energy cost will be more than the resulting one.

A bisection search is proposed to find the appropriate λ^*. The first step of the initial solution is to include the lowest cost point from each user (the highest resource requirement according to Pareto property). The amount of the resources used by this initial solution is $\mathrm{TXOP}^0 = \sum_{u=1}^{N} \mathrm{TXOP}^0_u$. In the next step, if TXOP^0 is higher than the resource constraint T_r, we use the bisection search until finding the appropriate λ satisfying the resource constraint. Without loss of generality, we assume that each of these u users maintains a U cost-resource Pareto setting. In this case, the complexity of this step is O(NU log (NU)). From the Pareto property, λ is strictly increasing with the energy. After finding the appropriate λ for each user, the Pareto set will be searched. The configurations which have a λ just lower or equal to the resulting λ are the optimal output settings. The complexity of this step is O(NU). The pseudocode of the algorithm is shown in Algorithm 1.

4.2.4. Proposed algorithm for minimizing the maximum energy

A greedy water-filling algorithm is proposed to solve the run-time searching for this problem. The first step of the

(1) Initialization:
allocate to each of the N users the lowest cost possible for the given state $E^{\min}_{\mathrm{GOP}_u}$.
Construct an N-value energy level vector,
with each of these values corresponding to the energy cost of one of the users.
(2) If $\sum_{u=1}^{N} \mathrm{TXOP}_u^0 > T_r$,
for the user who requires the lowest energy cost in this step,
sort out its energy TXOP tradeoff curve,
until a setting whose energy cost exceeds the second lowest energy cost level
is found or the resource constraint is satisfied.
(3) If the resource constraint is not satisfied,
update the energy level vector and repeat step 2 until the resource constraint is satisfied.

Algorithm 2: Run-time greedy water-filling algorithm.

initial solution is also to include the lowest cost point from each user (the highest resource requirement according to the Pareto property). Suppose that there are N users and the resource requirement of each of these N users composes U water-filling level vectors. The amount of the resources used by this initial solution is $\mathrm{TXOP}^0 = \sum_{u=1}^{N} \mathrm{TXOP}_u^0$.

In the next step, if TXOP^0 is higher than the resource constraint T_r, for the user which achieves the lowest energy cost among others, we reallocate the setting until its energy cost exceeds the second energy cost level or the resource constraint is satisfied. If the resource constraint is not satisfied by this step, we update the water-filling level vector and repeat the last step until the resource constraint is satisfied. The resulting outputs are the optimal settings for all users. The complexity of the water-filling algorithm is $O(\mathrm{NU}^2)$. The pseudocode of the algorithm is shown in Algorithm 2.

If the step sizes of the Pareto curve axes are infinitesimally small, the attentative reader might indeed observe that the K_u^* we find is the optimum configuration to achieve the min-max energy cost among users.

Proof. For configuration set K_u^*, for all u, $E_u^* \leq \max E_u^*$.

If there exists a configuration set $\overline{K}_u$, which results for all u in $\max \overline{E}_u < \max E_u^*$, then for all u, $\overline{K}_u < \max E_u^*$.

From the descending searching style of step 2, we have for all u, $\overline{E}_u \leq E_u^*$, and there exists at least one u' such that $\overline{E}_{u'} < E_{u'}^*$.

From the definition of Pareto property, we have $\overline{\mathrm{TXOP}}_u \geq \mathrm{TXOP}_u^*$, and there exists at least one u' such that $\overline{\mathrm{TXOP}}_{u'} > \mathrm{TXOP}_{u'}^*$. Hence, $\sum \overline{\mathrm{TXOP}}_u > \sum \mathrm{TXOP}_u^*$.

From the water-filling searching of step 2, we know that for all the resulting $\sum \mathrm{TXOP}$ higher than the $\sum \mathrm{TXOP}_u^*$, the constraint cannot be satisfied. Thus, there is no configuration set that can satisfy the constraint while achieving a max energy cost lower than that of K_u^*. □

Due to the discrete step size of the possible configurations that form the Pareto curve, there might exist other configurations that achieve less max energy cost. This is especially likely to happen if the step sizes are very irregular. This is a problem inherent to the discrete nature of the system, and it is well known that for such problems finding the optimal solution can be very hard. This is similar to the known Knapsack problem where the goal is to pack different discrete items with different resource constraints and values to the user. If an infinite set of items would be present, with infinitesimally small differences in terms of resource cost and value, the problem would be easy to solve. The discrete nature however makes it NP-hard. Many approximations however exist that allow to find a close-to optimal solution that works well enough in practice.

The difference between the maximum energy cost achieved by the algorithm and the optimum one lies however for sure between the maximum and minimum energy cost achieved by the last adaptation. In theory, this difference is hence bounded by the largest step size found in the Pareto curves that are the possible optimal configurations for the system. Practically, the convergence of the algorithm provides a solution close to the optimal solutions with reasonable complexity. The reason is that in practice, the step sizes between the different points on the curve are small enough.

5. NUMERICAL RESULTS

Based on the proposed two-phase approaches and the considered transceiver system models, we now verify the energy savings over a range of practical scenarios.

5.1. Simulation setup

In the experiments, a GOP size of 16 is assumed. Four sequences (bus, city, foreman, mother and daughter) are considered here as examples of video with different levels of motion activities, thus resulting in different bitrate versus distortion. All the sequences have CIF (352×288, $4:2:0$) resolution and 30 frames per second. The number of quality layers is set to 5. Empirically, for an image/video of CIF size, PSNR value of 25–35 dB corresponds to an acceptable visual quality for most of the users. We therefore encoded every sequence with a visual quality of approximately 35 dB for the full-length bitstream and 25 dB for the base layer portion of the bitstream. The intermediate bitstream rates of every quality layer of each video sequence are shown in Table 3.

Since network congestion influence on the performance-energy tradeoff is not the focus of the current paper, we limit

Table 3: Bitrate settings of a different video sequence.

Sequence name	Bus	City	Foreman	Mother and daughter
Bitrate of the 1st layer	256 kbps	64 kbps	96 kbps	64 kbps
Bitrate of the 2nd layer	384 kbps	128 kbps	112 kbps	80 kbps
Bitrate of the 3rd layer	448 kbps	256 kbps	192 kbps	96 kbps
Bitrate of the 4th layer	512 kbps	384 kbps	288 kbps	112 kbps
Bitrate of the 5th layer	1028 kbps	448 kbps	384 kbps	128 kbps

Table 4: Different PER's influence to received video quality.

PER	BUS encoded at 32.4 dB	Mobile encoded at 32.7 dB	Foreman encoded at 34.3 dB
0.05	31.0406	31.8462	32.8051
0.01	32.1733	32.3564	34.1034
0.005	32.1845	32.695	34.1606
0.001	32.3422	32.7484	34.2982

the number of users in the network so that their requirement can be satisfied. In the real-time variant channel simulation, we use the best possible quality transmission configuration when the channel state is very bad and therefore the quality requirement can almost never be reached. The average quality results turn out to always match the requirement well.

Each quality layer of the bitstream is encapsulated into a separate network packet. Thus, if one network packet is dropped, the corresponding quality layer is lost. Every network packet is further fragmented in MAC Service Data Units (MSDU) of 1500 Bytes at link level. The maximal number of retransmission is limited to 10 times.

An indoor channel model for the 5 GHz band [36] was used assuming a terminal moving uniformly at a speed between 0 to 5.2 km/h (walking speed). A set of 1000 time-varying frequency channel response realizations (sampled every 2 ms over one minute) were generated and normalized in power. The bitstream was modulated using a turbo-coded 802.11a OFDM PHY. The resulting PHY dynamics were then mapped to an 8-state Markov model, as detailed in [28].

In Table 4, we show the network packet error rate (PER)'s influence to the received video quality. From the comparison of values in this table, we reach the conclusion that with PER lower than 1e-2, the video can be regarded as correctly received. When calculating the configuration at design-time, to further assure the stable visual quality, the first quality layer is always given as a configuration with error probability lower than 1e-2. The sequence has been iteratively transmitted more than 10 times to get relevant statistics.

We consider in the sequel the following three transmission strategies.

(i) *SoA reference point:* the server uses the highest feasible modulation in addition to code rate that enables to transmit the packets with a loss probability lower than 1e-2 (transmit as fast as possible). After successfully receiving and decoding the required video bitstream, the mobile devices are switched to sleep mode. This approach aims at maximizing sleep duration. It is proposed in commercial 802.11 interfaces [37].

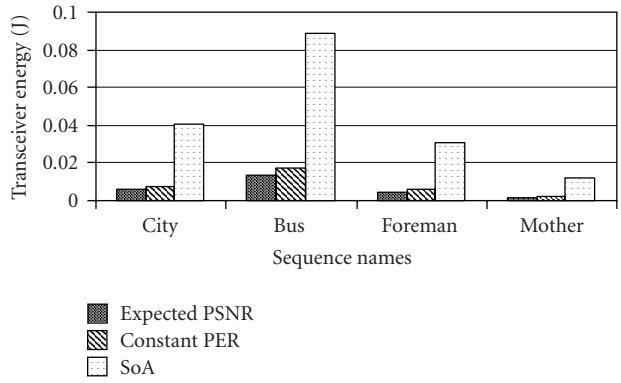

Figure 5: Impact of video content.

(ii) *Design-time + run-time approach 1: "constant PER":* this is the approach introduced in [24, 25]. With this strategy, every video packet is transmitted with a configuration resulting in a PER lower than 1e-2 until the transmitted bitstream reaches the quality required by the users. Instead of always transmitting the packets with the highest possible data rate, an optimal schedule exploring the tradeoff brought by link layer scaling and sleeping is introduced.

(iii) *Design-time + run-time approach 2: "expected PSNR":* this is the approach introduced in this paper. In this transmission strategy, we introduce the expected visual distortion into the design-time Pareto frontier calculation. By emphasizing differently the total energy minimization and fairness improvement for the run-time algorithm, this transmission strategy can be further differentiated into the following two schemes.

  (a) Min total energy: the total energy consumption of the users' terminal transceivers is minimized.

  (b) Min-max energy: the maximum energy consumption among the users' terminal transceivers is minimized.

  The detailed results for the two proposed run-time schemes are discussed in Section 5.2.4.

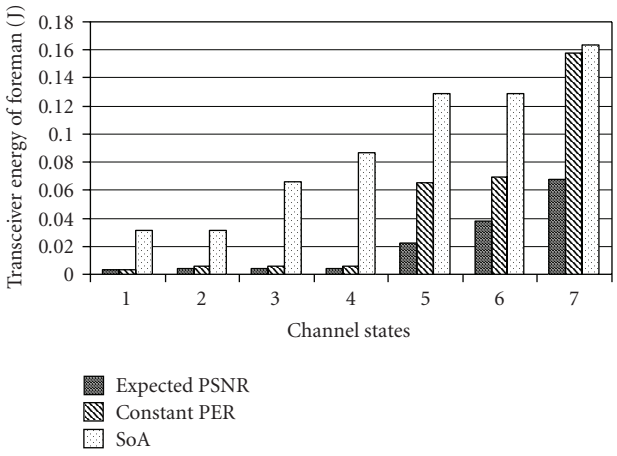

Figure 6: Impact of channel status.

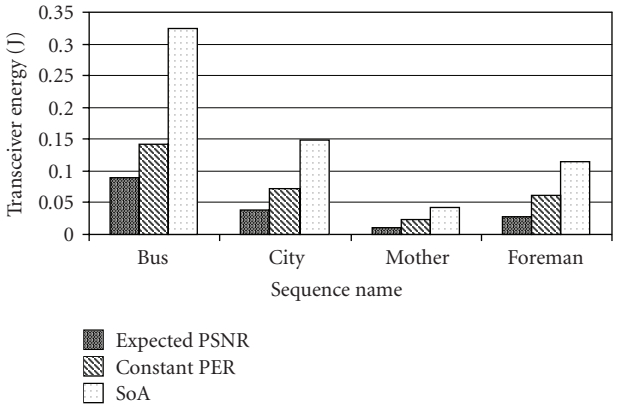

Figure 7: Impact of a time variant channel and multiple user to the transceiver energy of different schedulers.

5.2. Results analysis

The simulation results show that a significant energy decrease can already be achieved with the "constant PER" approach compared to the state-of-the-art approach. When the "expected PSNR" approach is used, simulation proves that energy savings up to a factor of 2 can be achieved while maintaining a uniform visual quality, thus significantly improving QoS. In Sections 5.2.1–5.2.4, we show the detail results from the aspects of different video content, channel status, user requirements together with the fairness discussion.

5.2.1. Impact of the video content

Figure 5 shows the influence of the video content on the energy cost for the different approaches listed above. The PER constraint is fixed to 1e-2 for the "constant PER" approach. The QoS constraint is fixed to 35 dB for the "expected PSNR" approach. As expected, the higher the bitrate of a sequence is, the higher the transceiver energy cost will be. Those results are provided by delivering video on a fixed channel state (channel state 2, with 40% occurrence probability).

5.2.2. Impact of the channel status

The impact of the channel state is shown in Figure 6. The channel state is again assumed to be constant during the whole transmission. The foreman sequence is considered here. The PER constraint is fixed to 1e-2 for the "constant PER" approach. The QoS constraint is fixed to 35 dB for the "expected PSNR" approach. Seven channels are used in this test, where channel 1 is the best and 7 is the worst. From the results, it is clear that the "constant PER" approach outperforms the SoA approach for all the channel states. The "expected PSNR" approach enables further energy savings, because video packets are treated differently based on their visual relevance.

For bad channel states, the energy cost of the "constant PER" approach tends to be similar to that of the SoA scheme. But for bad channel states, the "expected PSNR" approach provides the biggest savings. This is because the "expected PSNR" approach relaxes the PER requirement of low-importance video packets.

In Figure 7, we consider a time-varying channel and evaluate the energy cost of several video sequences (with different rate-distortion properties) transmitted simultaneously. The channel varies independently over all the users on a GOP-by-GOP basis. For the "constant PER" approach, the PER constraint is 1e-2. For the "expected PSNR" approach, the QoS requirement used is 35 dB. Total energy consumption after transmission is shown in Figure 7. Compared to the static channels (see Figure 6), the time-varying channels cause a further increase in energy cost. In addition, the energy cost further increases because of the multiuser scenario. Each packet has to be transmitted with lower TXOP to share the bandwidth with other users, thus increasing the energy consumption to maintain the QoS requirement. Nevertheless, even under these conditions, the energy cost for each user has been reduced approximately by a factor 2, compared to the SoA approach. The "Expected PSNR" outperforms the "Constant PER" approach by another factor of 2.

5.2.3. Impact of the different user requirements.

In this section, we present the impact of the different user QoS requirements on the energy cost. The four different sequences with QoS requirements of 35 dB, 33 dB, and 31 dB are tested simultaneously on the time-varying channel. The energy cost of all these sequences after transmission is presented in Figure 8. It is clear that the lower the QoS requirement is, the lower the energy consumption will be. The reason is straightforward; the lower the quality is, the smaller the bitrate will be, and hence the lower the energy cost. This shows once again that by taking into account the rate-distortion properties into the optimization system, we can obtain more energy gains.

5.2.4. Fairness discussion

So far we have considered only one of the two proposed solutions, that is, the total energy minimization. In this final section we present the performance of the second proposed

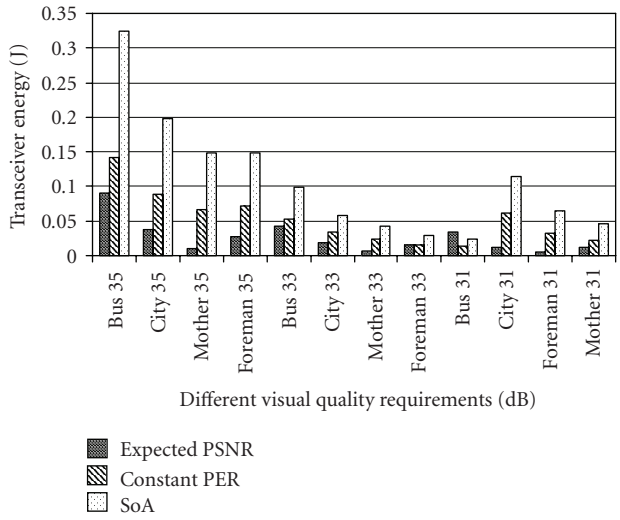

Figure 8: Impact of a time variant channel and multiple user with different quality requirements to the transceiver energy.

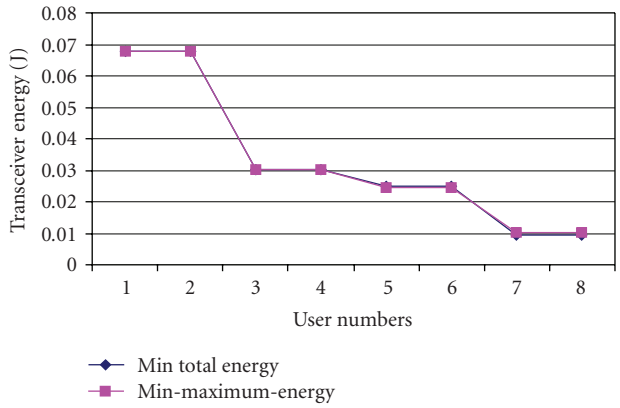

Figure 9: Energy comparison for 8 users requiring video bitstreams simultaneously.

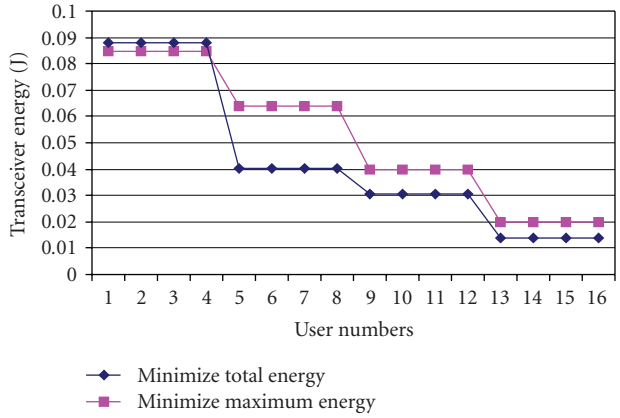

Figure 10: Energy comparison for 16 users requiring video bitstreams simultaneously.

solution, that is, the fairness solution. In particular, we compare the impacts of the two run-time algorithms on the energy consumption.

In Figure 9, we consider 8 users simultaneously requiring video streams with an expected PSNR of 35 dB. Each of the 4 videos is required by 2 users. No significant difference is measured in terms of energy cost.

Figure 10 shows 16 users receiving video bitstreams simultaneously with an expected PSNR equal to 35 dB. Each of the 4 videos is required by 4 users. Under this setup, bandwidth requirements are much more stringent than those in the former setup, and the difference between the two run-time approaches becomes significant. The results show that the energy consumption of those users who require the maximum energy cost decreases by about 3.5%, with the energy consumption of other users increasing by more than 30%.

From the results presented in the previous sections, we conclude that in general the minimal total energy approach reduces the energy cost by a factor of 2. Based on the results presented in this section, we, therefore, recommend the min-max energy approach unless the users are facing energy exhausting issues.

6. CONCLUSIONS AND FUTURE WORK

We have introduced a cross-layer optimization framework to minimize the wireless transceiver energy consumption for downloading multiple video streams over a WLAN. Relying on the IEEE 802.11 standard and scalable video coding, the proposed solutions schedule the packets transmission by both exploiting link layer scaling and sleeping trade-offs, and integrating rate-distortion properties of the video sequences into the optimization scheme. Results have shown that in comparison with state-of-the-art approaches, the proposed expected PSNR approach achieves stable visual quality according to the user QoS requirements, while largely decreasing the energy cost. Compared to link layer optimization, the proposed expected PSNR approach achieves energy gains by a factor of 2. Fairness and total energy minimization approaches have also been discussed in this paper.

The work presented in this paper offers an insight view of cross-layer optimization for energy-efficient bandwidth allocation using scalable video coding. Our future work will concentrate on the scalability provided by the different video coding schemes. In particular, future work will focus on the scalable video coding (SVC) extension of H.264/AVC, the upcoming state-of-the-art scalable video coding, which uses different temporal and spatial decomposition schemes compared to the one used in this paper. Also, in this paper, the distortion calculation was GOP-based. Though calculating the distortion based on packets will increase overhead and complexity, it would be interesting to see how the proposed approaches perform under these conditions, despite the fact that we expect that the general trends presented here will be maintained. Additionally, we plan to investigate the performance of the proposed approaches in uplink streaming, P2P video transmission, and so forth. These scenarios have very different timing constraints, hence requiring more interesting optimization schemes and solutions. Finally, it would be

interesting to investigate multimedia applications over ad-hoc networks and identify the energy and congestion controls needed to optimize the performance of these systems.

ACKNOWLEDGMENT

The work presented in the paper is partly based on results published in ICASSP 2007.

REFERENCES

[1] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1374–1387, 2000.

[2] J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 290–315, 2001.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable H.264/MPEG4-AVC extension," in *Proceedings of IEEE International Conference on Image Processing (ICIP '06)*, pp. 161–164, Atlanta, Ga, USA, October 2006.

[4] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 301–317, 2001.

[5] Q. Li and M. van der Schaar, "Providing adaptive QoS to layered video over wireless local area networks through real-time tetry limit adaptation," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 278–290, 2004.

[6] A. Ksentini, M. Naimi, and A. Gueroui, "Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 107–114, 2006.

[7] D. Skyrianoglou, N. Passas, and A. Salkintzis, "Traffic scheduling for multimedia QoS over wireless LANs," in *Proceedings of IEEE International Conference on Communications (ICC '05)*, vol. 2, pp. 1266–1270, Seoul, Korea, May 2005.

[8] A. Sinha, A. Wang, and A. Chandrakasan, "Energy-scalable system design," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 10, no. 2, pp. 135–145, 2002.

[9] A. J. Goldsmith, "The capacity of downlink fading channels with variable rate and power," *IEEE Transactions On Vehicular Technology*, vol. 46, no. 3, pp. 569–580, 1997.

[10] D. Qiao, S. Choi, A. Jain, and K. Shin, "MiSer: an optimal low-energy transmission strategy for IEEE 802.11a/h," in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking*, pp. 161–175, San Diego, Calif, USA, September 2003.

[11] M. Goel, S. Appadwedula, N. R. Shanbhag, K. Ramchandran, and D. L. Jones, "A low-power multimedia communication system for indoor wireless applications," in *Proceedings of IEEE Workshop on Signal Processing Systems (SiPS '99)*, pp. 473–482, Taipei, Taiwan, October 1999.

[12] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 5, pp. 645–658, 2005.

[13] X. Lu, Y. Wang, and E. Erkip, "Power efficient H.263 video transmission over wireless channels," in *Proceedings of IEEE International Conference on Image Processing (ICIP '02)*, vol. 1, pp. 533–536, Rochester, NY, USA, September 2002.

[14] S. Chandra and S. Dey, "Addressing computational and networking constraints to enable video streaming from wireless appliances," in *Proceedings of the 3rd Workshop on Embedded Systems for Real-Time Multimedia (STMED '05)*, pp. 27–32, New York, NY, USA, September 2005.

[15] H. Yousefi'zadeh, H. Jafarkhani, and M. Moshfeghi, "Power optimization of wireless media systems with space-time block codes," *IEEE Transactions on Image Processing*, vol. 13, no. 7, pp. 873–884, 2004.

[16] T. Simunic, H. Vikalo, P. Glynn, and G. De Micheli, "Energy efficient design of portable wireless systems," in *Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED '00)*, pp. 49–54, Rapallo, Italy, July 2000.

[17] S. Cui, A. J. Goldsmith, and A. Bahai, "Modulation optimization under energy constraints," in *Proceedings of IEEE International Conference on Communications (ICC '04)*, vol. 4, pp. 2805–2811, Paris, France, June 2004.

[18] J. T. Ludwig, S. H. Nawab, and A. P. Chandrakasan, "Low-power digital filtering using approximate processing," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 3, pp. 395–399, 1996.

[19] A. Wang, W. R. Heinzelman, and A. P. Chandrakasan, "Energy-scalable protocols for battery-operated microsensor networks," in *Proceedings of IEEE Workshop on Signal Processing Systems (SiPS '99)*, pp. 483–492, Taipei, Taiwan, October 1999.

[20] E. Uysal-Biyikogly, B. Prabhakar, and A. El Gamal, "Energy-efficient packet transmission over a wireless link," *ACM/IEEE Transactions on Networking*, vol. 10, no. 4, pp. 487–499, 2002.

[21] C. Schurgers, *Energy-aware communication systems*, Ph.D. thesis, University of California, Los Angeles, Calif, USA, 2002.

[22] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 641–646, 1993.

[23] W. Ye, J. Heidemann, and D. Estrin, "Medium access control with coordinated adaptive sleeping for wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 3, pp. 493–506, 2004.

[24] R. Mangharam, R. Rajkumar, S. Pollin, et al., "Optimal fixed and scalable energy management for wireless networks," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, vol. 1, pp. 114–125, Miami, Fla, USA, March 2005.

[25] S. Pollin, R. Mangharam, B. Bougard, et al., "MEERA: cross-layer methodology for energy efficient resource allocation in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 617–628, 2007.

[26] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390–404, 2006.

[27] IEEE Std 802.11a, "Standard for telecommunications and information exchange between systems—LAN/MAN specific requirements—part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE, 1999.

[28] B. Bougard, S. Pollin, G. Lenoir, et al., "Energy-scalability enhancement of wireless local area network transceivers," in *Proceedings of the 5th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC '04)*, pp. 449–453, Lisboa, Portugal, July 2004.

[29] Microsemi LX5506 InGaP HBT 4.5–6.5GHz Power Amplifier.

[30] J. G. Proakis, *Digital Communications*, McGraw-Hill, Berkshire, UK, 4th edition, 2001.

[31] L. Benini, A. Bogliolo, and G. De Micheli, "A survey of design techniques for system-level dynamic power management,"

IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 8, no. 3, pp. 299–316, 2000.

[32] IEEE Standard 802.11-1999, Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, 1999.

[33] M. van der Schaar, Y. Andreopoulos, and Z. Hu, "Optimized scalable video streaming over IEEE 802.11a/e HCCA wireless networks under delay constraints," *IEEE Transactions on Mobile Computing*, vol. 5, no. 6, pp. 755–768, 2006.

[34] J. Lin, "Multiple-objective problems: pareto-optimal solutions by method of proper equality constraints," *IEEE Transactions on Automatic Control*, vol. 21, no. 5, pp. 641–650, 1976.

[35] H. Everett III, "Generalized lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, no. 3, pp. 399–417, 1963.

[36] European Telecom Standards Institue, "Channel models for HIPERLAN/2 in different indoor scenarios," ETSI 3ERI085B, 1998.

[37] Atheros White Paper, "Power consumption and energy efficiency comparisons of WLAN products," 2003.