Traffic and Quality Characterization of Single-Layer Video Streams Encoded with the H.264/MPEG-4 Advanced Video Coding Standard and Scalable Video Coding Extension

Geert Van der Auwera, Prasanth T. David, and Martin Reisslein

Abstract—The recently developed H.264/AVC video codec with Scalable Video Coding (SVC) extension, compresses non-scalable (single-layer) and scalable video significantly more efficiently than MPEG-4 Part 2. Since the traffic characteristics of encoded video have a significant impact on its network transport, we examine the bit rate-distortion and bit rate variability-distortion performance of single-layer video traffic of the H.264/AVC codec and SVC extension using long CIF resolution videos. We also compare the traffic characteristics of the hierarchical B frames (SVC) versus classical B frames. In addition, we examine the impact of frame size smoothing on the video traffic to mitigate the effect of bit rate variabilities. We find that compared to MPEG-4 Part 2, the H.264/AVC codec and SVC extension achieve lower average bit rates at the expense of significantly increased traffic variabilities that remain at a high level even with smoothing. Through simulations we investigate the implications of this increase in rate variability on (i) frame losses when transmitting a single video, and (ii) on a bufferless statistical multiplexing scenario with restricted link capacity and information loss. We find increased frame losses, and rate-distortion/rate-variability/encoding complexity tradeoffs. We conclude that solely assessing bit rate-distortion improvements of video encoder technologies is not sufficient to predict the performance in specific networked application scenarios.

Index Terms—Frame loss ratio, H.264/AVC, hierarchical B frames, rate variability-distortion (VD), rate-distortion (RD), single-layer video, statistical multiplexing, SVC, video quality, video traffic.

I. INTRODUCTION

WE STUDY the video traffic generated by the H.264/MPEG-4 Advanced Video Coding standard [1] (H.264/AVC for brevity), also known as H.264/MPEG-4 Part 10, and its recently standardized Scalable Video Coding extension (SVC) [2], [3]. This new video technology is expected to have a broad application domain for wired and wireless video transmission, and storage up to high definition (HD) resolution. Indications of the growing acceptance of H.264/AVC are its adoption in application standards and industry consortia

specifications, such as DVB, ATSC, 3GPP, 3GPP2, MediaFLO, DMB, DVD Forum (HD-DVD), and Blu-Ray Disc Association (BD-ROM). At the same time, the introduction of IPTV over high speed access network links is ongoing, e.g., over Ethernet Passive Optical Networks (EPONs) or ADSL2+/VDSL2, and mobile TV technologies are made widely available. IPTV, mobile TV, and satellite TV are considered key applications that can make H.264/AVC the dominant video encoder in the broadcasting and consumer market.

In general, video can be encoded *(i)* with fixed quantization scales, which results in nearly constant video quality at the expense of variable video traffic (bit rate), or *(ii)* with rate control, which adapts the quantization scales to keep the video bit rate nearly constant at the expense of variable video quality [4]. In order to examine the fundamental traffic characteristics of the H.264/AVC video coding standard, which does not specify a normative rate control mechanism, we focus primarily on encodings with fixed quantization scales (and provide a brief study of encodings with rate control in Section V-D). An additional motivation for the focus on variable bit rate video encoded with fixed quantization scales is that the variable bit rate streams allow for statistical multiplexing gains that have the potential to improve the efficiency of video transport over communication networks [4]. The development of video network transport mechanisms that meet the strict playout deadlines of the video frames and efficiently accommodate the variability of the video traffic is a challenging problem. A wide array of video transport mechanisms has been developed and evaluated, based primarily on the characteristics of MPEG-2 and MPEG-4 Part 2 encoded video [5], [6]. The widespread adoption of the new H.264/AVC video standard necessitates the careful study of the traffic characteristics of video coded with the new H.264/AVC codec and its extensions. Therefore, it is necessary to examine the new video encoder's statistical characteristics and compression performance from a communication network perspective. We study the Main profile of the H.264/AVC encoder using long Common Intermediate Format (CIF) 352 × 288 pixel resolution sequences. Our study of the newest H.264 SVC extension analyzes single-layer (non-scalable) video traffic characteristics of long CIF videos, i.e., although the H.264 SVC single-layer encoding supports temporal scalability, we group the individual temporal layers and consider the aggregate stream.

The bit rate-distortion characteristics of H.264/AVC, H.264 SVC, and MPEG-4 Part 2 have been extensively studied in the literature [1], [7], [8]. In contrast, in the present study, we research the *joint characterization* of bit rate-distortion and higher order bit rate statistics, such as the variability of the bit rate, as a function of the distortion. First, we perform a detailed analysis of elementary statistics of the video traffic. We study statistics of frame sizes, group of picture (GoP) sizes, frame and GoP qualities, and correlations between frame sizes and qualities. We use bit rate-distortion (RD) and bit rate variability-distortion (VD) curves to compare H.264/AVC and SVC single-layer traffic to the traffic of the MPEG-4 Part 2 [9] encoder, which is the predecessor of H.264/AVC. In addition, we study several GoP structures (including classic B frame prediction and hierarchical B frame prediction) and analyze the impact of frame size smoothing on the video traffic variability.

Our main findings are that H.264/AVC and H.264 SVC single-layer video traffic is significantly more variable than MPEG-4 Part 2 traffic under similar encoding conditions. At the same time, we confirm the significant average bit rate savings. The increased bit rate variability is observed over a wide range of average qualities of the encoded streams and for all tested video sequences. This makes the transport of H.264/AVC and H.264 SVC single-layer traffic more challenging than MPEG-4 Part 2 traffic. Even when frame size smoothing is employed to mitigate the effect of the increased variability, we find that the smoothed traffic is still significantly more variable compared to MPEG-4 Part 2 traffic when the same smoothing is applied. We simulate two streaming scenarios to quantify the effect of the increased bit rate variability on *(i)* the frame loss ratio when transmitting a single video stream over a fixed-bandwidth bottleneck link, and *(ii)* on a basic real-time bufferless statistical multiplexing model. We observe that the increased bit rate variability results in significantly higher frame losses for H.264/AVC encoded streams compared to MPEG-4 Part 2 encoded streams. Secondly, we observe that a significant improvement in bit rate-distortion efficiency does not suffice to conclude that there is an equal gain in the number of supported streams on a link with constrained bandwidth and information loss probability. We find that the increased bit rate variability can lead to insignificant gains in number of supported streams when the additional encoding complexity is taken into consideration. Therefore, we conclude that solely assessing bit rate-distortion improvements of video encoder technologies is not sufficient to predict the performance in certain networked application scenarios, such as statistical multiplexing of streams.

All encodings presented in this study are publicly available as from the video traces library at:http://trace.eas.asu.edu. Frame size video traces [10] are files mainly containing video frame time stamps, frame types (e.g., I, P, or B), encoded frame sizes (in bits), and frame qualities (PSNR). Video traces are employed in simulation studies of the transport of video over communication networks, see e.g., [11]–[15], and as a basis for video traffic models, as for instance in [16]–[24]. Advantages over using regular encoded bit streams in simulations, are the availability of a large number of traces of long and real video sequences, the fact that video traces are not copyrighted, and that only knowledge of basic concepts of video encoding are required. We also provide tools that interface with popular network simulators, resulting in fast and reliable network simulation results, otherwise only available to networking researchers with in-depth video coding expertise and large computational resources for the encoding of many long video sequences with numerous encoding parameters.

This paper is structured as follows. In Section II, we review related work. In Section III, we present a brief overview of the examined video coding standards. In Section IV, we describe the employed video test sequences, encoding tools, and video traffic metrics. In Section V, we study the video traffic statistics for the different encoders and GoP structures considering frame and GoP size statistics, autocorrelations, and frame size smoothing. In Section VI, we examine the implications of the higher traffic variability with the new H.264 and SVC codecs for basic single video stream and multiplexed multiple video stream network transport. We summarize our conclusions in Section VII.

II. RELATED WORK

The traffic characterizations of MPEG-1 and MPEG-4 Part 2 [9] encoded video, examined e.g., in [25]–[30], have formed the basis for a plethora of studies addressing the challenges of modeling the video traffic, see e.g., [16]–[24], and of efficiently transporting the variable bit rate video traffic over networks to meet the playout deadlines of the video frames, see for instance [5], [6], [11]–[15], [31]. To the best of our knowledge, the bit rate variability of H.264/AVC and SVC are for the first time examined in the present study.

Existing studies of the H.264/AVC codec and its extensions, such as [1], [7], [8], focus primarily on the rate-distortion (RD) performance, i.e., the video quality (PSNR) as function of the *average bit rate*, and typically consider only short video sequences up to a few hundred frames. In contrast, for the transport over communication networks, the traffic variability is also a key concern. Therefore, we study the *bit rate variability* as a function of the video quality or distortion, which we express in the bit rate variability-distortion (VD) curve. In order to obtain reliable and meaningful statistical estimates of the traffic variability and other properties, it is necessary to examine *long* video sequences with several thousand frames as we do in this study.

We note that for one fixed GoP pattern, a preliminary study [32] briefly compared the bit rate variability-distortion of the H.264/AVC encoder with the variability of the MPEG-4 Part 2 and MPEG-2 encoders. In contrast, in this study we comprehensively compare the H.264/AVC encoder, the H.264 SVC encoder, and the MPEG-4 Part 2 encoder for a range of GoP patterns. In addition, we compare hierarchical B frames with classical B frames, examine the impact of rate control on the traffic

variability, and explore the implications of the increased variabilities on network transport in this study.

III. MPEG-4 VIDEO STANDARDS

We briefly introduce the state-of-the-art video codecs (encoder/decoder) in the MPEG-4 family and their applications. MPEG-4 is a family of open international standards that provide tools for the delivery of multimedia. The tools include codecs for the compression of audio and video, graphics and interactive features. MPEG-4's latest video codec is Part 10 or AVC, the Advanced Video Codec, which is also identically standardized as ITU H.264. The latest standardization effort addressing scalability is the extension of H.264/AVC called Scalable Video Coding (SVC). In the following sections we briefly introduce the following video codecs: MPEG-4 Part 2, H.264/AVC, and H.264 SVC.

A. MPEG-4 Part 2

The MPEG-4 Part 2 [9] standard combines tools in profiles, and levels provide a way to limit computational complexity, e.g., by specifying the bit rate. For applications where hardware cost or power considerations make implementing H.264/AVC difficult, MPEG-4 Part 2 offers the Simple and Advanced Simple Profile specifications.

The most used profile for streaming video is the *Simple Profile* (SP). This profile is defined for two-way and very low complexity receivers, such as wireless videophones. Therefore, the tools are selected by giving priority to low-delay and low-complexity. SP includes the compression tools to encode I frames and P frames, 1/2 pixel motion compensation, AC/DC prediction, 4 motion vectors per macroblock (4-MV) and Unrestricted MV. Furthermore, error-resilience tools are supported.

The *Advanced Simple Profile* (ASP) was defined with Internet and streaming video in mind. For these applications the delay is less of an issue and the targeted platforms have high processing power. Therefore, ASP has tools that allow to improve the quality of video over SP. For example, the ASP profile contains 1/4 pixel motion compensation, B frames, and global motion compensation.

B. H.264/AVC

H.264/AVC represents a big leap in video compression technology with typically a 50% reduction of average bit rate for a given video quality compared to MPEG-2 and about a 30% reduction compared with MPEG-4 Part 2 [33]. Block transforms in conjunction with motion compensation and prediction are still the core of the encoder as in previous standards, but a number of new encoding mechanisms have been added which give a much better performance over previous standards [1].

The H.264/AVC standard defines several profiles. The *Baseline profile* is intended for low-delay applications, low processing power platforms, and for high packet loss environments. The *Main profile* encompasses all tools for achieving high coding efficiency for high bit rate applications. The

Extended profile is meant for error-resilient streaming applications. The FRExt amendment adds four *High profiles*: High (HP), High 10 (Hi10P), High 4:2:2 (Hi422P), and High 4:4:4 (Hi444P) [7], [34]. The High profile has improved tools which can result in up to 10% compression gains over the Main profile and up to 59% over MPEG-2 for High Definition video with only a small increase in computational complexity compared to the Main profile. Recently, five additional profiles have been added for professional applications, e.g., supporting intra-only encoding.

A major improvement is the introduction of the entropy coding scheme Context Adaptive Binary Arithmetic Coding (CABAC), which typically gives 10–15% bit rate savings [33] over previous variable length coding schemes used in MPEG-2/4. Since arithmetic coding is compute intensive, the Main profile also supports a scheme called Context Adaptive Variable Length Coding (CAVLC), which is an improved version of older variable length coding schemes. Other new normative tools include spatial intra frame prediction which predicts a region of a given frame from other regions of the same frame, a new integer transform which significantly reduces ringing artifacts, and an adaptive in-loop deblocking filter which reduces artifacts [33]. H.264/AVC also introduces a new tool called *Variable Block sizes* which introduce a different number of square and rectangular macroblock sizes, such as (4×4), (8×8), and (16×8) pixels. These different block sizes permit selecting the optimal block size for motion compensation and prediction. H.264/AVC also uses Lagrangian based rate-distortion optimization [33].

In previous standards, one reference frame (I or P) from the past for prediction of P frame blocks was allowed, and one reference frame (I or P) from the past and one reference frame (I or P) from the future for prediction of B frame blocks were allowed, whereby the blocks from these past and future reference frames were weighted equally to form the predicted B frame block. Similarly, for prediction of a B frame block in H.264/AVC, two blocks are selected from the reference frames; however, there are two lists that each can contain *multiple* reference frames. One block is selected from a frame in each of the two reference lists and these blocks can be weighted *unequally* [35].

C. H.264 SVC

In 2007, the SVC scalability extension [2], [3] has been added to the H.264/AVC standard. The SVC extension provides temporal scalability, coarse (CGS), medium (MGS), and SNR scalability in general, spatial scalability, and combined spatio-temporal-SNR scalability (restricted set of spatio-temporal-SNR points can be extracted from a global scalable bit stream).

In the following, we discuss the concept of hierarchical B frames in more detail, since our study refers to this concept repeatedly. SVC's temporal scalability is built on the hierarchical prediction concept for B frames. The introduction of hierarchical B frames has allowed the H.264 SVC encoder to achieve
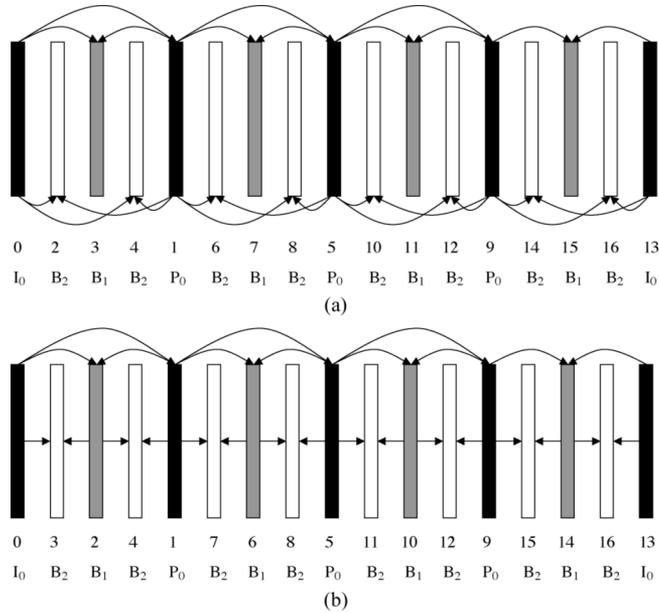
Fig. 1. B frame prediction structures. (a) Classical B frame prediction structure. (b) Hierarchical B frame prediction structure.

temporal scalability while at the same time improving RD efficiency compared to the classical B frame prediction method employed by the older MPEG standards (MPEG-1/2/4-Part 2) and by default in H.264/AVC. In Fig. 1, we illustrate both concepts for predicting B frames.

Hierarchical B frames are an important new concept that was first introduced in H.264/AVC using generalized B frames and was later found to be the best method to build the Scalable Video Coding (SVC) extension on. Hence, the H.264 SVC encoded single-layer stream is decodable by existing H.264/AVC decoders. The scalability modes do require new SVC capability, with the supported modes depending on the applications or equivalently on the H.264 SVC profiles. In this description we do not go into detail about low-delay or constrained delay B frame prediction structures. We refer to [3] for a detailed discussion and further reading.

Fig. 1(a) depicts the classical B frame prediction structure, where each B frame is predicted only from the preceding I or P frame and from the subsequent I or P frame. Other B frames are not referenced since this is not allowed by video standards preceding H.264/AVC. This restriction is lifted in the generalized B frame paradigm that was first introduced in the H.264/AVC standard. Fig. 1(b) depicts the hierarchical B frame structure which uses B frames for the prediction of B frames. The illustrated case is the dyadic hierarchy of B frames, meaning that the number of B frames n in between the key pictures (I or P frames) equals $n = 2^k - 1$, $k = 1, 2, 3, \ldots$. The hierarchy with 3 B frames (I frame period is 16) is depicted in Fig. 1(b). In this example, the frame sequence is $I_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2$, where the index represents the temporal layer number.

The coding efficiency of hierarchical B frames depends on the number of hierarchical B frames (temporal levels) and on the choice of quantization parameters for each B frame.

Therefore, H.264 SVC introduces cascading quantizers which assign a higher quantization parameter value (lower quality) to B frames belonging to higher temporal layers. This concept is based on the insight that the lowest temporal layer 0 requires higher quality than the next temporal layer, since all other predictions depend on it. The quality of each subsequent temporal layer can be gradually reduced since fewer layers depend on it. Apparently the quality fluctuation that is introduced within a GoP is not subjectively noticeable by human observers, as studied by the standard committee.

IV. VIDEO SEQUENCES, ENCODING TOOLS, AND VIDEO TRAFFIC METRICS

A. Video Sequences

The CIF video sequences used for the statistics presented in this study are the ten minute *Sony Digital Video Camera Recorder* demo sequence (17,682 frames at 30 frames/sec), which we refer to as *Sony Demo* sequence, the first half hour of the *Silence of the Lambs* movie (54,000 frames at 30 frames/sec), the *Star Wars IV* movie (54,000 frames at 30 frames/sec), and the first hour of the *Tokyo Olympics* video (133,128 frames at 30 frames/sec). We also use about 30 minutes of the *NBC 12 News* (49,523 frames at 30 frames/sec), including the commercials. The video sequences *Silence of the Lambs*, *Star Wars IV*, *Tokyo Olympics*, and *NBC 12 News* can respectively be described as drama/thriller, science fiction/action, sports, and news video. Due to space constraints, we present in Sections V-B and V-C only illustrative plots for encodings with *Silence of the Lambs* and in Section V-E only illustrative plots for *Silence of the Lambs* and *Star Wars IV*. The corresponding plots for the other video sequences are available in [36].

B. Encoding Tools

We decoded the original DVD sequences into the uncompressed YUV format using the MEncoder tool and used this tool to downsample to the CIF resolution (352×288 pixels). We employ the JM reference software (version 10.2), which is the official MPEG and ITU reference implementation for the H.264/AVC Main profile, the MPEG-4 Part 2 Microsoft v2.3.0 software, and the SVC reference software named JSVM (version 5.9).

C. Video Traffic Metrics

Here we provide a brief overview of essential video traffic metrics. For a video sequence consisting of M frames encoded with a given quantization scale, we let $X_m(m = 1, \ldots, M)$ denote the sizes [bits] of the encoded video frames. The mean frame size $\bar{X}$ [bits] of the encoded video sequence is defined as

$$\bar{X} = \frac{1}{M} \sum_{m=1}^{M} X_m, \qquad (1)$$

while the variance S_X^2 of the frame sizes (S_X is the standard deviation [bits]) is defined as

$$S_X^2 = \frac{1}{(M-1)} \sum_{m=1}^{M} (X_m - \bar{X})^2. \qquad (2)$$

The coefficient of variation of frame sizes [unit free] is defined as

$$CoV_X = \frac{S_X}{\bar{X}} \qquad (3)$$

and is widely employed as a measure of the variability of the frame sizes, i.e., the bit rate variability of the encoded video. Plotting the CoV as a function of the quantization scale (or equivalently, the PSNR video quality) gives the rate variability-distortion (VD) curve [30]. Alternatively, the peak-to-mean (Peak/Mean or PtM) ratio of the frame sizes is commonly used to express the traffic variability. If $X_{\max}$ is the maximum size of all M frames, then the peak-to-mean frame size ratio PtM_X [unit free] is defined as

$$PtM_X = \frac{X_{\max}}{\bar{X}}. \qquad (4)$$

If each video frame is transmitted during one frame period T (e.g., 33 ms for 30 frames/s), then the bit rate R_m [bits/s] required to transmit frame X_m is

$$R_m = \frac{X_m}{T}, \qquad (5)$$

and analogously, the peak bit rate $R_{\max}$ [bits/s] is defined as

$$R_{\max} = \frac{X_{\max}}{T}. \qquad (6)$$

We define a *Group of Pictures* (GoP) of an encoded video stream as one I frame and all subsequent P and B frames before the next I frame in the stream. The size of GoP n is denoted by $Y_n(n = 1, \ldots, M/N)$ [bits] and equals the sum of the N frames that belong to the GoP. The mean GoP size $\bar{Y}$ [bits] is defined as

$$\bar{Y} = \frac{1}{M/N} \sum_{n=1}^{M/N} Y_n, \qquad (7)$$

and if $Y_{\max}$ is the maximum of all GoP sizes Y_n, then the peak-to-mean GoP size ratio PtM_Y [unit free] is defined as

$$PtM_Y = \frac{Y_{\max}}{\bar{Y}}. \qquad (8)$$

The coefficient of variation of GoP sizes is

$$CoV_Y = \frac{S_Y}{\bar{Y}}. \qquad (9)$$

We use the Peak Signal-to-Noise Ratio (PSNR) as the objective measure of the quality of a reconstructed video frame $R(x,y)$ with respect to the uncompressed video frame $F(x,y)$. The larger the difference between $R(x,y)$ and $F(x,y)$, or equivalently, the lower the quality of $R(x,y)$, the lower the PSNR value. The PSNR is expressed in decibels [dB] to accommodate the logarithmic sensitivity of the human visual system. The PSNR is typically obtained for the luminance video frame and in case of a $N_x \times N_y$ frame consisting of 8-bit pixel values, it is computed as a function of the mean squared error (MSE) as

$$MSE = \frac{1}{N_x \cdot N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} [F(x,y) - R(x,y)]^2, \quad (10)$$

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{MSE}. \qquad (11)$$

We denote the PSNR quality of a video frame m by Q_m and define the average PSNR quality $\bar{Q}$ of a video sequence as

$$\bar{Q} = \frac{1}{M} \sum_{m=1}^{M} Q_m. \qquad (12)$$

The coefficient of quality variation is defined as

$$CoQV = \frac{S_Q}{\bar{Q}}. \qquad (13)$$

For a detailed definition of all statistics used in this study, we refer to [28].

V. TRAFFIC ANALYSIS

We compare H.264/AVC, H.264 SVC, and MPEG-4 Part 2 single-layer video traffic using several GoP structures. We note that the B frame prediction structure of SVC, named hierarchical B frames (see Section III-C), differs from the classic B frame prediction structure which is by default employed by H.264/AVC and MPEG-4 Part 2. However, H.264/AVC also supports hierarchical B frames and therefore, SVC single-layer
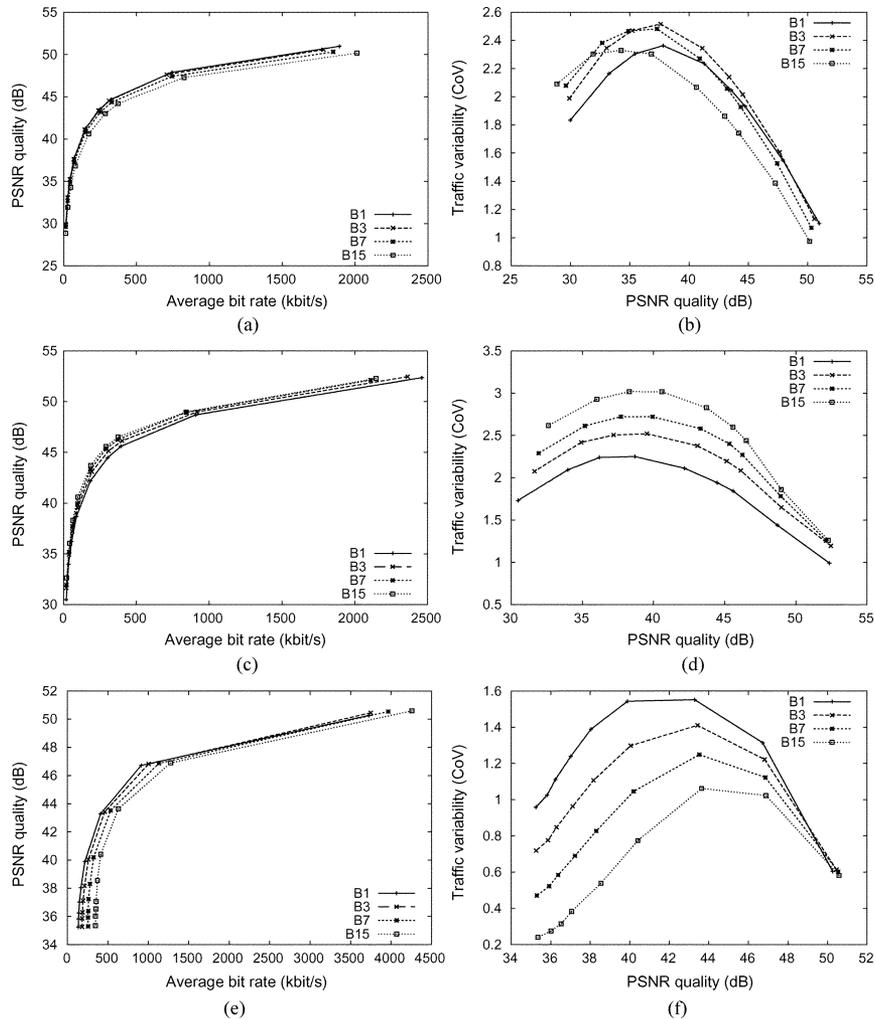
Fig. 2. RD and VD curves comparing GoP structures *G16-B1*, *G16-B3*, *G16-B7*, and *G16-B15* for *Silence of the Lambs*. (a) H.264/AVC RD curves (b) H.264/AVC VD curves (c) H.264 SVC RD curves (d) H.264 SVC VD curves (e) MPEG-4 Part 2 RD curves (f) MPEG-4 Part 2 VD curves.

encoding is compatible with H.264/AVC encoding. Hence, our single-layer comparison between H.264/AVC and SVC is equivalent to a comparison between the classical B frame prediction and hierarchical B frames.

A. Encoding Setup

In the subsequent experiments, we employ four different GoP structures, namely *IBPBPBPBPBPBPBPB* (16 frames, with 1 B frame per I/P frame), which we denote by *G16-B1*, *IBBBPBBBPBBBPBBB* (16 frames, with 3 B frames per I/P frame) denoted by *G16-B3*, *IBBBBBBBPBBBBBBB* (16 frames, with 7 B frames per I/P frame) denoted by *G16-B7*, and *IBBBBBBBBBBBBBBB* (16 frames, with 15 B frames per I frame) denoted by *G16-B15*. In the context of SVC, these four GoP structures are respectively designated by their "GoP size" which is the number of hierarchical B frames plus one key picture, either of type I or P. Hence, *G16-B1* has GoP size 2, *G16-B3* has GoP size 4, *G16-B7* has GoP size 8, and *G16-B15* has GoP size 16. In the following, we employ our own GoP structure notation to emphasize the repetitive I-P-B frame type patterns in the encodings and to avoid confusion. These four GoP structures are natural structures for hierarchical B frames

and allow us to compare the three encoders based on identical underlying GoP patterns.

We employ the H.264/AVC encoder in the Main profile with all compression tools enabled, as specified in Section III-B, i.e., using variable block sizes, three reference frames for the past and the future, referenced B frames, P and B frame weighted prediction, CABAC, and rate-distortion optimization (RDO). We designate these settings by "Full-RDO". The H.264 SVC settings are similar.

We use the MPEG-4 Part 2 encoder in the *Advanced Simple* profile (ASP) to encode the sequences, for comparison with the H.264/AVC encodings. This ASP profile adds B frames to the *Simple* profile. We employ half pixel motion compensated prediction; RDO is not supported by the reference encoder implementation. The MPEG-4 Part 2 encoder uses one reference frame for the past and one for the future, and 16×16 blocks for motion estimation that can be split into 8×8 blocks.

B. GoP Structure Comparison

Selected RD graphs for the *Silence of the Lambs* sequence encoded with H.264/AVC, H.264 SVC, and MPEG-4 Part 2 are depicted in Figs. 2(a), 2(c), and 2(e). Each figure depicts

the RD curves for all GoP structures for a particular encoder. We observe that the H.264/AVC encoder achieves the best RD performance for GoP structure *G16-B3* with almost coinciding RD curves. For the MPEG-4 Part 2 encoder the RD efficiency decreases significantly with increasing number of B frames in the GoP structures. Contrary to these two encoders, the H.264 SVC encoder achieves best RD performance for the *G16-B15* GoP structure and lowest for *G16-B1*. From RD comparison plots between all three encoders, not included due to space constraints, we find that for GoP structure *G16-B1*, H.264/AVC and H.264 SVC have comparable RD performance. However, H.264 SVC increasingly outperforms H.264/AVC for GoP structures *G16-B3* to *G16-B15*.

In addition to the RD graphs, the VD graphs are provided in Figs. 2(b), (d), and (f). From the H.264/AVC figure, we observe that the bit rate variability increases from GoP structure *G16-B1* to *G16-B3*, and then decreases for *G16-B7* and *G16-B15*, with the latter having a lower variability than *G16-B1*. For the MPEG-4 Part 2 encodings, the highest rate variability occurs for *G16-B1* and decreases with increasing number of B frames. On the contrary, for the H.264 SVC encoder the highest variability occurs for the *G16-B15* GoP structure and gradually decreases with decreasing number of B frames. For the GoP structures *G16-B3* to *G16-B15*, the variabilities of the SVC encodings are significantly higher than for H.264/AVC, with values around 3.0 for the *Silence of the Lambs* and even surpassing this high level for *Sony Demo* [36].

These observed RD and VD behaviors as a function of GoP structures, are explained as follows. First, there is some influence of the choices of quantization parameters for each frame type (I, P, or B). For the H.264/AVC encodings, the quantization parameter of the B frames is set two units larger than the parameters for the I and P frames (which are equal), while for the MPEG-4 Part 2 encodings we set all quantization parameters equal for all frame types. H.264 SVC employs a complex, but deterministic assignment of quantization parameters to frames belonging to the temporal layers (cascading of quantization parameters), with the lowest QPs (highest quality) assigned to frames belonging to the temporal base layer and gradually higher QPs (lower quality) assigned to frames of higher temporal layers. Second, H.264 SVC uses a hierarchical reference frame structure (dyadic) inside each GoP that is completely different from the reference frame structure employed by the other two encoders. Both reasons, cascading QP assignments and hierarchical B frame structure, are the cause of the significantly different behavior of the RD and VD curves of the H.264 SVC encoder as a function of the GoP structures compared to the other encoders. Furthermore, we observe that the better the RD performance of a particular GoP structure, the higher the corresponding traffic variability.

C. Frame Size and GoP Size Statistics

We summarize key frame size and GoP size statistics in Table I by reporting the minimum, mean, and maximum across

the five considered video sequences for the various statistical measures. We group H.264/AVC, H.264 SVC, and MPEG-4 Part 2 results for selected quantization scales that provide similar minimum to maximum ranges (across the five sequences) of the mean PSNR frame qualities $\bar{Q}$ (30–35 dB, 35–40 dB, and 40–45 dB) to facilitate the comparison of the statistical measures across encoders. (We refer to [36] for the detailed statistics for all sequences.) We provide statistics for GoP structure *G16-B3* and also for GoP structure *G16-B15* for the SVC encoder, since SVC has best RD efficiency for *G16-B15* among all four GoP structures. The SVC statistics for *G16-B3* allow for a comparison across encoders between classical and hierarchical B frames based on identical GoP structures, eliminating influences of different numbers of P and B frames within the GoPs. In the first column of each table the encoding mode is specified as the GoP structure, e.g., *G16B3*, followed by a code representing the encoder (F for H.264/AVC with *Full-RDO*, SV for H.264 SVC, and Mp for MPEG-4 Part 2), and ending with the quantization scale.

For each average PSNR quality range, we observe the much higher compression ratios, or equivalently smaller average frame sizes and bit rates, obtained with the H.264/AVC, and H.264 SVC encoders compared to the MPEG-4 Part 2 encoder, as well as the significantly higher coefficient of variation CoV and peak-to-mean PtM values. The CoV and PtM values of the GoP sizes are significantly lower than the values of the frame sizes. We provide a detailed analysis of smoothing on frame size statistics in Section V-E. In the following, we provide plots to illustrate the statistical properties of the *G16-B3* encodings of *Silence of the Lambs* for relatively high quality settings ($QP = 24$ for H.264/AVC, $QP = 28$ for H.264 SVC, and $q = 4$ for MPEG-4 Part 2) and relatively low quality settings ($QP = 38$ for H.264/AVC, $QP = 42$ for H.264 SVC, and $q = 28$ for MPEG-4 Part 2). We have chosen these particular settings, because the corresponding average video qualities of the *Silence of the Lambs* encodings are very close for all three encoders.

Fig. 3 depicts frame sizes as a function of frame number m. We observe that the frame sizes have similar behaviors for all encodings with peaked and smoothed traffic for approximately the same indices, which is related to the video content, with peak values occurring for frames that are harder to compress. The MPEG-4 Part 2 traces overall have larger frame sizes than the H.264/AVC and SVC encodings, except for a few peaks in the H.264/AVC and H.264 SVC plots that exceed the corresponding peaks in the MPEG-4 Part 2 plots. The coefficient of variation is harder to observe visually, but one can estimate the observed average frame sizes and compare with the peak values. The average frame size values of the MPEG-4 Part 2 encodings appear to be higher compared to the peaks than for H.264/AVC and SVC encodings, hence the higher variability of the latter two. For each encoder, we observe that the variability is higher for the low video quality compared to the high quality.

In Fig. 4 we present histograms of the frame sizes which are plotted up to the maximum frame sizes, which are 31,061, 8,291,

TABLE I
OVERVIEW OF FRAME SIZE, GoP SIZE, BIT RATE, AND QUALITY STATISTICS OF SINGLE-LAYER ENCODINGS WITH H.264/AVC (F), H.264 SVC (SV), AND MPEG-4 PART 2 (Mp)

Encoding Mode		Compr. ratio	Frame Size			Bit Rate		GoP Size		Frame Quality	
			Mean $\bar{X}$ [kbyte]	CoV $S_X/\bar{X}$	Peak/M. $X_{max}/\bar{X}$	Mean $\bar{X}/T$ [Mbps]	Peak X_{max}/T [Mbps]	CoV $S_Y/\bar{Y}$	Peak/M. $Y_{max}/\bar{Y}$	Mean $\bar{Q}$ [dB]	CoV $CoQV$
G16B3F22	Min	33.566	1.296	1.057	7.994	0.311	4.585	0.546	2.814	39.918	0.034
G16B3F22	Mean	71.524	2.718	1.523	15.216	0.652	8.415	0.731	6.338	42.650	0.072
G16B3F22	Max	117.303	4.530	2.016	27.627	1.087	10.514	1.108	12.798	44.621	0.097
G16B3SV24	Min	33.691	1.291	1.244	8.559	0.310	4.943	0.479	2.611	40.082	0.043
G16B3SV24	Mean	72.598	2.678	1.729	16.420	0.643	8.944	0.654	6.091	43.138	0.076
G16B3SV24	Max	117.819	4.513	2.197	29.304	1.083	11.219	0.997	10.868	45.189	0.098
G16B15SV24	Min	37.381	1.216	1.456	11.588	0.292	6.406	0.481	2.675	39.965	0.045
G16B15SV24	Mean	77.872	2.450	2.058	21.480	0.588	10.815	0.652	6.091	43.388	0.078
G16B15SV24	Max	125.064	4.068	2.598	37.374	0.976	13.474	0.971	11.527	45.802	0.110
G16B3Mp04	Min	26.030	1.896	0.738	6.753	0.455	5.684	0.476	3.024	39.234	0.032
G16B3Mp04	Mean	50.845	3.723	1.076	11.751	0.894	9.182	0.681	5.779	41.485	0.064
G16B3Mp04	Max	80.215	5.842	1.411	18.466	1.402	11.244	0.986	10.970	43.424	0.094
G16B3FRC1	Min	33.606	1.297	1.007	9.420	0.311	10.230	0.248	3.895	39.613	0.078
G16B3FRC1	Mean	71.847	2.693	1.524	32.330	0.646	16.439	0.494	10.964	42.729	0.146
G16B3FRC1	Max	117.273	4.525	1.906	54.497	1.086	27.922	0.732	19.334	44.635	0.249
G16B3MpRC1	Min	22.863	1.896	0.757	7.635	0.455	7.998	0.024	1.104	35.900	0.071
G16B3MpRC1	Mean	48.489	4.147	1.234	19.166	0.995	15.587	0.671	8.725	39.951	0.159
G16B3MpRC1	Max	80.196	6.651	1.863	38.580	1.596	22.089	1.617	15.383	42.970	0.315
G16B3F28	Min	83.141	0.601	1.478	12.474	0.144	2.520	0.522	3.053	36.630	0.046
G16B3F28	Mean	156.962	1.191	1.877	21.301	0.286	5.387	0.749	7.401	39.047	0.088
G16B3F28	Max	252.882	1.829	2.345	38.578	0.439	6.687	1.130	15.060	41.114	0.111
G16B3SV34	Min	131.208	0.406	1.773	15.427	0.097	1.993	0.499	2.650	35.000	0.053
G16B3SV34	Mean	235.991	0.779	2.108	24.919	0.187	4.187	0.684	6.787	37.429	0.091
G16B3SV34	Max	374.488	1.159	2.521	43.248	0.278	5.333	1.008	12.726	39.534	0.120
G16B15SV34	Min	139.496	0.411	2.193	20.707	0.099	2.517	0.451	2.521	35.618	0.056
G16B15SV34	Mean	239.866	0.754	2.592	32.459	0.181	5.339	0.641	6.719	38.258	0.095
G16B15SV34	Max	369.869	1.090	3.103	55.639	0.262	6.635	0.941	12.459	40.586	0.121
G16B3Mp08	Min	58.234	0.993	0.954	9.189	0.238	3.502	0.525	2.777	35.408	0.046
G16B3Mp08	Mean	99.445	1.775	1.152	13.557	0.426	5.319	0.636	5.681	37.729	0.079
G16B3Mp08	Max	153.091	2.611	1.312	19.208	0.627	6.323	0.831	10.021	40.046	0.099
G16B3FRC2	Min	83.069	0.602	1.442	16.527	0.144	7.261	0.393	6.166	36.403	0.103
G16B3FRC2	Mean	157.067	1.187	1.948	47.483	0.285	12.291	0.670	13.012	39.168	0.178
G16B3FRC2	Max	252.737	1.831	2.642	73.719	0.439	27.887	1.316	24.333	41.595	0.308
G16B3MpRC2	Min	58.229	0.994	0.975	10.610	0.239	5.710	0.052	1.394	32.569	0.091
G16B3MpRC2	Mean	99.705	1.766	1.407	22.890	0.424	8.461	0.684	10.667	36.701	0.169
G16B3MpRC2	Max	153.006	2.612	2.536	37.476	0.627	13.315	2.266	16.223	39.308	0.302
G16B3F38	Min	308.086	0.178	1.810	19.962	0.043	1.041	0.498	2.863	30.648	0.065
G16B3F38	Mean	544.005	0.331	2.170	28.957	0.079	2.129	0.671	7.869	32.936	0.111
G16B3F38	Max	854.575	0.494	2.667	46.594	0.118	2.710	0.953	14.833	35.216	0.148
G16B3SV42	Min	338.892	0.161	1.823	19.924	0.039	0.957	0.483	2.810	30.191	0.066
G16B3SV42	Mean	598.676	0.299	2.149	28.039	0.072	1.870	0.636	6.846	32.565	0.099
G16B3SV42	Max	941.786	0.449	2.630	43.998	0.108	2.312	0.884	11.984	34.933	0.129
G16B15SV42	Min	334.204	0.175	2.230	25.525	0.042	1.257	0.448	2.395	31.426	0.065
G16B15SV42	Mean	567.396	0.314	2.633	36.694	0.075	2.564	0.591	6.188	33.722	0.103
G16B15SV42	Max	871.274	0.455	3.252	58.425	0.109	3.287	0.824	10.513	36.029	0.124
G16B3Mp20	Min	126.739	0.628	0.752	9.835	0.151	1.752	0.439	2.596	30.550	0.066
G16B3Mp20	Mean	173.512	0.922	0.944	10.687	0.221	2.339	0.485	4.127	33.377	0.094
G16B3Mp20	Max	242.029	1.200	1.210	11.619	0.288	2.832	0.538	5.612	36.298	0.107
G16B3FRC3	Min	307.709	0.178	1.836	32.933	0.043	1.930	0.410	2.863	30.810	0.130
G16B3FRC3	Mean	543.046	0.331	2.487	86.024	0.079	6.782	0.882	17.950	33.286	0.209
G16B3FRC3	Max	854.164	0.494	3.986	170.161	0.119	17.228	2.412	50.069	36.050	0.331
G16B3MpRC3	Min	126.417	0.663	0.895	24.789	0.159	5.737	0.306	2.903	30.320	0.096
G16B3MpRC3	Mean	168.103	0.943	1.105	41.489	0.226	9.327	0.418	9.603	33.272	0.156
G16B3MpRC3	Max	229.454	1.203	1.271	54.371	0.289	13.737	0.599	17.651	36.631	0.230

29,044, 7,104, 35,555, and 5,702 Bytes in Figs. 4(a)–(f), respectively. We observe that H.264/AVC and SVC encodings have narrower histograms with longer tails than the MPEG-4 Part 2 encodings. This is the case both for low and high qualities. This resembles the higher energy compaction property of the H.264 encoders, or equivalently, their better compression efficiency. The GoP size histograms of the H.264/AVC and SVC encoders, not included due to space constraints, exhibit similar narrowness compared to MPEG-4 Part 2.

In Fig. 5, we plot the autocorrelation coefficient of the frame sizes as a function of the lag in frames. The frame size autocorrelation is a "comb of spikes" superimposed on a slowly decaying curve. The larger peaks occur for lags that are multiples of 16, i.e., the I frame period, and are the result of the correlation of the large I frames with each other and also the P frames, and to a lesser extent the B frames. The three smaller peaks in between the larger peaks are the result of the correlation of the I and the P frames with each other. For other lag values, the I or P frames are correlated with the B frames, resulting in relatively small autocorrelation. We observe that the decay of the autocorrelation curves is somewhat faster for the high qualities than for the low qualities. The decay of the MPEG-4 Part 2 encodings is much faster than for the H.264/AVC and SVC au-

tocorrelations. Small negative autocorrelation values appear for large lags and are the result of signal symmetries around the average frame size. Representative GoP size sequence autocorrelation plots are provided in Fig. 6. None of the curves have an exponential decay, indicating the presence of long range dependencies.

D. Impact of Rate Control on Rate Variabilities

So far we have focused on open-loop variable bit rate encoding, which allows us to examine the pure impact of video encoding technologies on traffic statistics. Nevertheless, often rate control algorithms are used to adapt the bit rate of a video stream towards a specified target bit rate. Studying rate controlled video traffic implies the selection of a particular algorithm [37], and hence dependency of the traffic analysis on this algorithm. With these limitations in mind, we provide rate control results for comparison with the variable bit rate statistics of MPEG-4 Part 2 and H.264/AVC encodings provided in Table I.

We consider the *TM5* rate control technique for MPEG-4 Part 2 encodings and the rate control algorithm of the *JM 12.2* reference software for H.264/AVC encodings [37]. We set the target bit rates for each sequence equal to the mean bit rates of the corresponding variable bit rate encodings with GoP structure
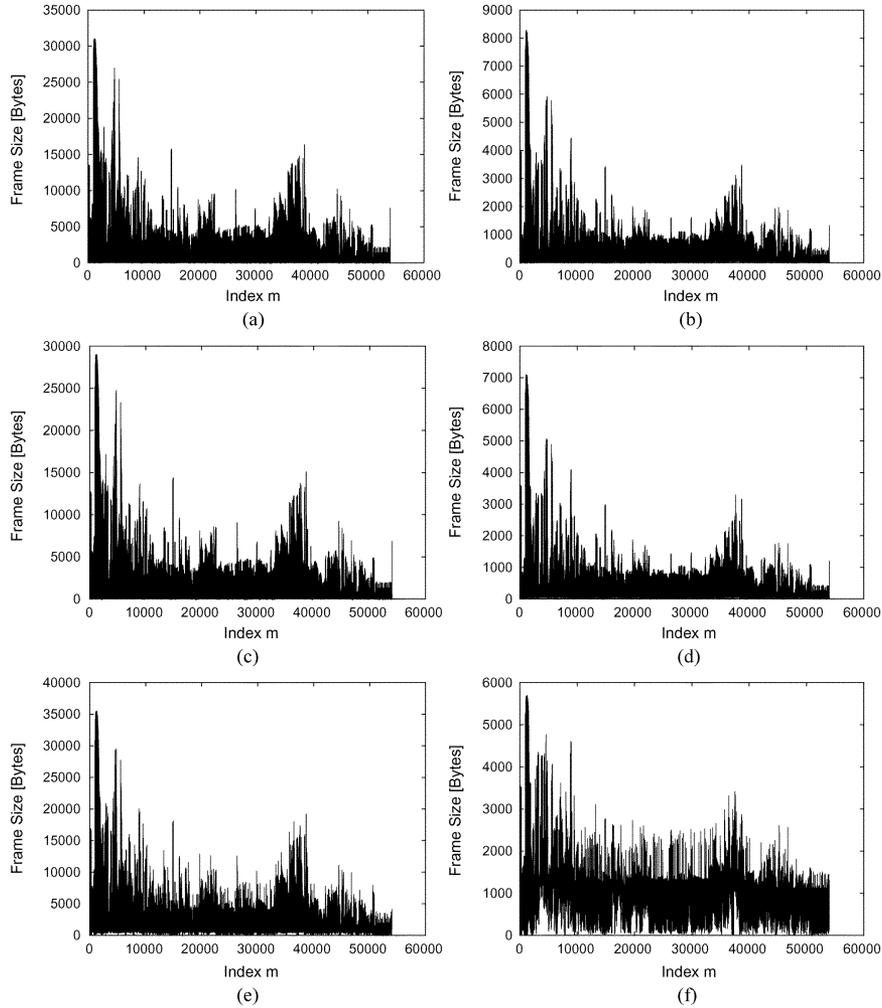
Fig. 3. Frame size plots of *Silence of the Lambs* G16-B3 encodings. (a) H.264/AVC ($QP = 24$) (b) H.264/AVC ($QP = 38$) (c) H.264 SVC ($QP = 28$) (d) H.264 SVC ($QP = 42$) (e) MPEG-4 Part 2 ($q = 4$) (f) MPEG-4 Part 2 ($q = 28$).

G16-B3. Table I summarizes the traffic statistics, whereby *FRC* means H.264/AVC with rate control and *MpRC* means MPEG-4 Part 2 with rate control. The H.264/AVC rate control achieved all target rates quite accurately for all sequences, while *TM5* mostly achieved its target rates within a small margin.

We first observe from Table I that the mean CoV and PtM of the frame sizes as well as the $CoQV$ values with rate control are typically larger than the corresponding metrics without rate control. On the other hand, the mean CoV of the GoP sizes with rate control is typically smaller than without rate control. Furthermore, the maximum CoV and PtM values for frame and GoP sizes, are typically significantly larger for the rate controlled traffic, while the minimum CoV and PtM values are smaller for GoP sizes with rate control. These observations can be explained by the long video sequences with many scene changes that make prediction of rates by the control algorithm more challenging, resulting in larger maximum CoV and PtM. Moreover, the larger time horizons, such as GoP lengths, that the rate control algorithms work on to achieve the target bit rate, and the different treatment of I, P, and B frames to maintain compres-

sion efficiency, result in widely varying individual frame sizes and qualities.

From this brief rate control experiment, we conclude that rate control has very limited effectiveness in mitigating the observed increases of the bit rate variabilities between MPEG-4 Part 2 and H.264/AVC. We leave a detailed analysis of rate control for future work.

E. Frame Size Smoothing

In order to mitigate the effect of variable video frame sizes on network transport, a wide variety of frame size smoothing mechanisms have been developed and studied in the context of the MPEG-4 Part 2, H.263, and preceding codecs, see for instance [38]–[45]. In this section, we examine the fundamental impact of frame size smoothing on H.264/AVC, H.264 SVC, and MPEG-4 Part 2 traffic by considering the elementary smoothing of the frames over non-overlapping blocks of a frames each. More specifically, with the aggregation level a, the sizes of a consecutive frames are averaged, and transmitted at the corresponding average bit rate across a
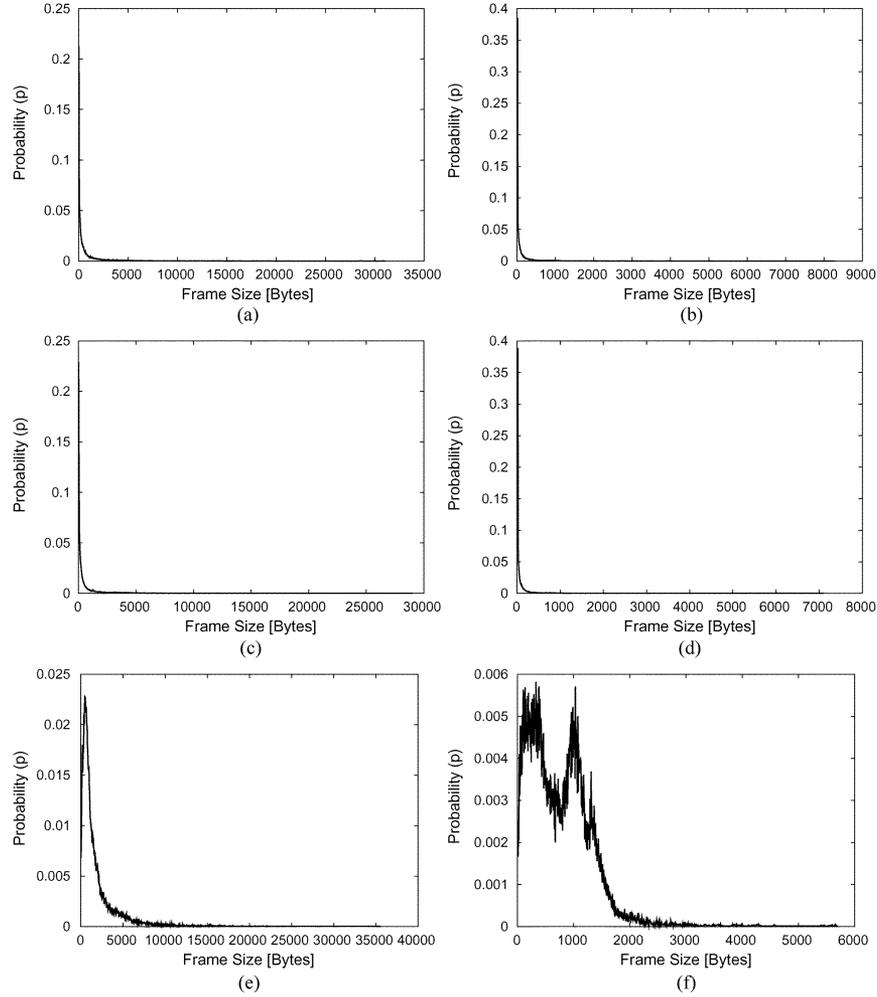
Fig. 4. Frame size histogram plots of *Silence of the Lambs* G16-B3 encodings. (a) H.264/AVC ($QP = 24$) (b) H.264/AVC ($QP = 38$) (c) H.264 SVC ($QP = 28$) (d) H.264 SVC ($QP = 42$) (e) MPEG-4 Part 2 ($q = 4$) (f) MPEG-4 Part 2 ($q = 28$).

network. Given the original (unsmoothed) frame size sequence X_m $(m = 1, \ldots, M)$, we obtain the smoothed frame sizes

$$Y_n = \frac{1}{a} \sum_{m=(n-1)a+1}^{na} X_m \qquad (14)$$

for $n = 1, \ldots, M/a$ and examine their CoV.

To illustrate the effect of frame size smoothing on the bit rate variability, we plot the VD curves of both the unsmoothed and the smoothed (denoted by sm in the figures) H.264/AVC, SVC, and MPEG-4 Part 2 video traffic of selected *Silence of the Lambs* and *Star Wars IV* encodings in Figs. 7 and 8. The traffic is smoothed over respectively $a = 2$ and $a = 8$ frames. From Figs. 7 and 8, and VD plots of other encodings [36], we observe that the variability of the H.264/AVC and SVC traffic smoothed over two frames is significantly higher than the unsmoothed MPEG-4 Part 2 traffic for all sequences and all GoP structures, except for *G16-B1* [36]. For the latter, the variability of the smoothed traffic is partially higher and partially lower than the unsmoothed MPEG-4 Part 2 traffic. However, it is always higher than the variability of the MPEG-4 traffic smoothed

over two frames. More smoothing (achieved with larger a) of the H.264/AVC and SVC traffic lowers the variability, however, for the same smoothing the MPEG-4 traffic variability also drops and stays well below the smoothed H.264/AVC, and SVC traffic. In some cases, such as for the *Silence of the Lambs* sequence with GoP structure *G16-B15* [36], the variability of the H.264/AVC, and SVC traffic smoothed over eight frames is still higher than or comparable to the unsmoothed MPEG-4 Part 2 traffic.

These encoding results illustrate the significantly higher bit rate variability of H.264/AVC and H.264 SVC video traffic compared to MPEG-4 Part 2 video traffic, even when frame size smoothing is applied. This increased rate variability must be taken into account and its impact evaluated when using existing network protocols and mechanisms for streaming H.264/AVC and H.264 SVC encoded video.

F. Quality and Correlation Statistics

Next, we analyze the video quality of our encodings. We use the PSNR as our quality metric, which is overall a good measure of video frame quality and is easy to compute for large numbers
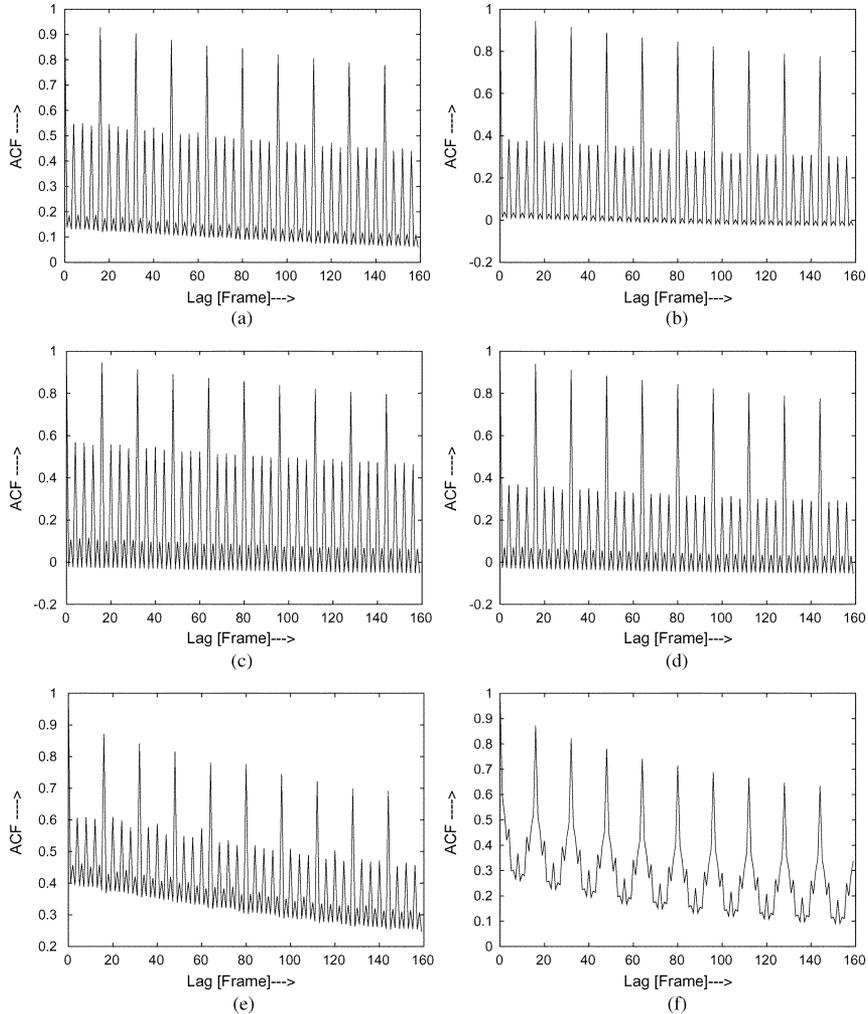
Fig. 5. Frame size autocorrelation plots of *Silence of the Lambs* G16-B3 encodings. (a) H.264/AVC ($QP = 24$) (b) H.264/AVC ($QP = 38$) (c) H.264 SVC ($QP = 28$) (d) H.264 SVC ($QP = 42$) (e) MPEG-4 Part 2 ($q = 4$) (f) MPEG-4 Part 2 ($q = 28$).

of long video encodings. For a detailed specification of the statistics used in this section, we refer to [28]. We focus on the luminance component in our analysis.

We observe from Table I for all three encoders that the mean PSNR $\bar{Q}$ decreases as the quantization parameter used in the encodings increases. This is expected for decreasing bit rates. Conversely, the coefficient of quality variation $CoQV$ increases when the video quality decreases. This means that the relative quality fluctuations are larger and more visible when the video quality is low. The same observations are valid on the GoP level. (The GoP quality metrics are not included in Table I due to space constraints.) Furthermore, we found that the values of the coefficient of quality variation on the GoP level are close to the values on the frame level. However, from an examination of the quality ranges (difference between highest and lowest PSNR frame quality) we found a distinction between the frame level and the GoP level, with the latter ranges being consistently smaller. These trends are independent of the GoP structures.

The report [36] also presents the frame size-PSNR quality correlation coefficients ρ_{XQ}, as well as the corresponding correlation coefficient $\rho_{XQ}^{(G)}$ for the GoP aggregation level. In sum-

mary, we found that there exists a general trend that the magnitude of ρ_{XQ} on the frame level decreases as the quality decreases. On the GoP level, the magnitude of $\rho_{XQ}^{(G)}$ tends to be higher than on the frame level and tends to increase with decreasing quality for the H.264/AVC encodings. Conversely, for the MPEG-4 Part 2 encodings, the GoP level magnitudes tend to decrease with decreasing video quality as do the frame level magnitudes. This is an interesting distinction between both encoders.

VI. IMPLICATIONS OF INCREASED RATE VARIABILITIES

In the previous sections, we focused on the statistical characterization of the single-layer (non-scalable) video traffic as generated by the H.264/AVC (classical B), H.264 SVC (hierarchical B), and MPEG-4 Part 2 (classical B) encoders. We observed the improved rate-distortion (RD) efficiency of hierarchical B frames (H.264 SVC) compared to the classical B frames (H.264/AVC), and a tremendous RD improvement over MPEG-4 Part 2. However, together with this increase in RD efficiency, the bit rate variability, measured in the coefficient of variation and the peak-to-mean ratio of the frame and GoP
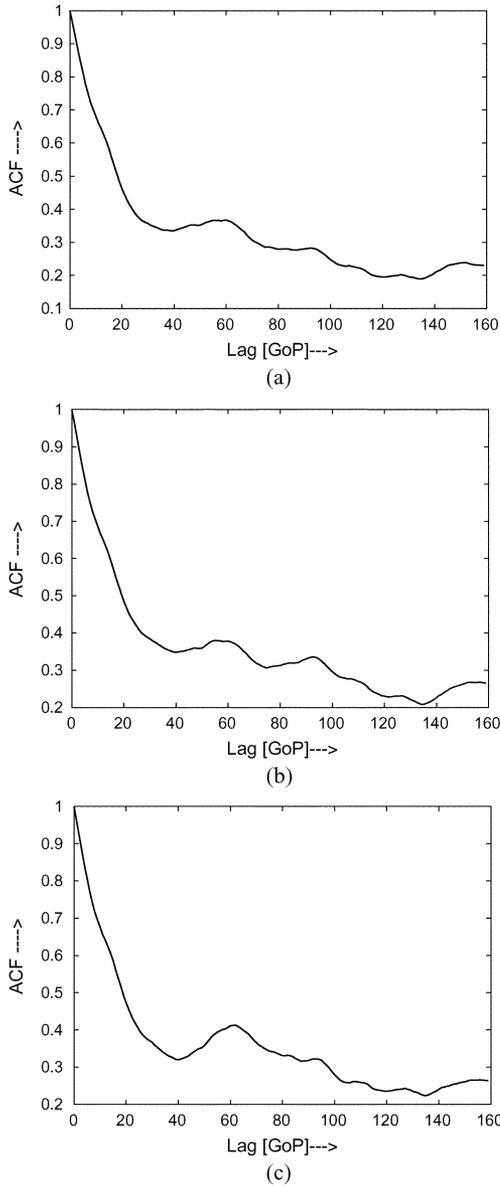
Fig. 6. GoP size autocorrelation plots of *Silence of the Lambs* G16-B3 encodings. (a) H.264/AVC ($QP = 24$) (b) H.264 SVC ($QP = 28$) (c) MPEG-4 Part 2 ($q = 4$).

Fig. 7. VD curves for *Silence of the Lambs* with GoP structure G16-B3, unsmoothed and smoothed (sm). (a) Unsmoothed, smoothed (sm) $a = 2$ (b) Unsmoothed, smoothed (sm) $a = 8$.
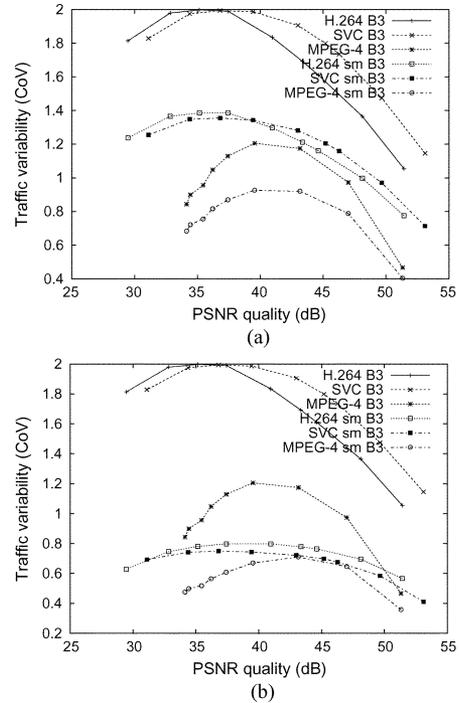
Fig. 8. VD curves for *Star Wars IV* with GoP structure G16-B3, unsmoothed and smoothed (sm). (a) Unsmoothed, smoothed (sm) $a = 2$ (b) Unsmoothed, smoothed (sm) $a = 8$.

sizes, increases significantly. Therefore, in this section we investigate the implications of this increase in rate variability with two simulation studies. In the first study, we examine the frame losses by comparing the transport of a single stream encoded with H.264/AVC or MPEG-4 Part~2. In our second study, we assess the impact of increased rate variability in a bufferless statistical multiplexing scenario.

A. Implications for Frame Loss Ratio

1) Encoding Setup: Ten different half hour video sequences, namely *Silence of the Lambs*, *Star Wars IV*, *Indiana Jones*, *Citizen Kane*, *Die Hard*, *The Firm*, *Terminator 1*, *Gandhi*, *Tokyo Olympics*, and *NBC News* were encoded with H.264/AVC and MPEG-4 Part 2 with GoP structure of *G16-B3* in CIF resolution as in the previous sections.

2) Results and Discussion: We evaluate the frame loss ratio, i.e., number of frames dropped in the network to the number of transmitted frames, through NS-2 [46] simulations. We consider
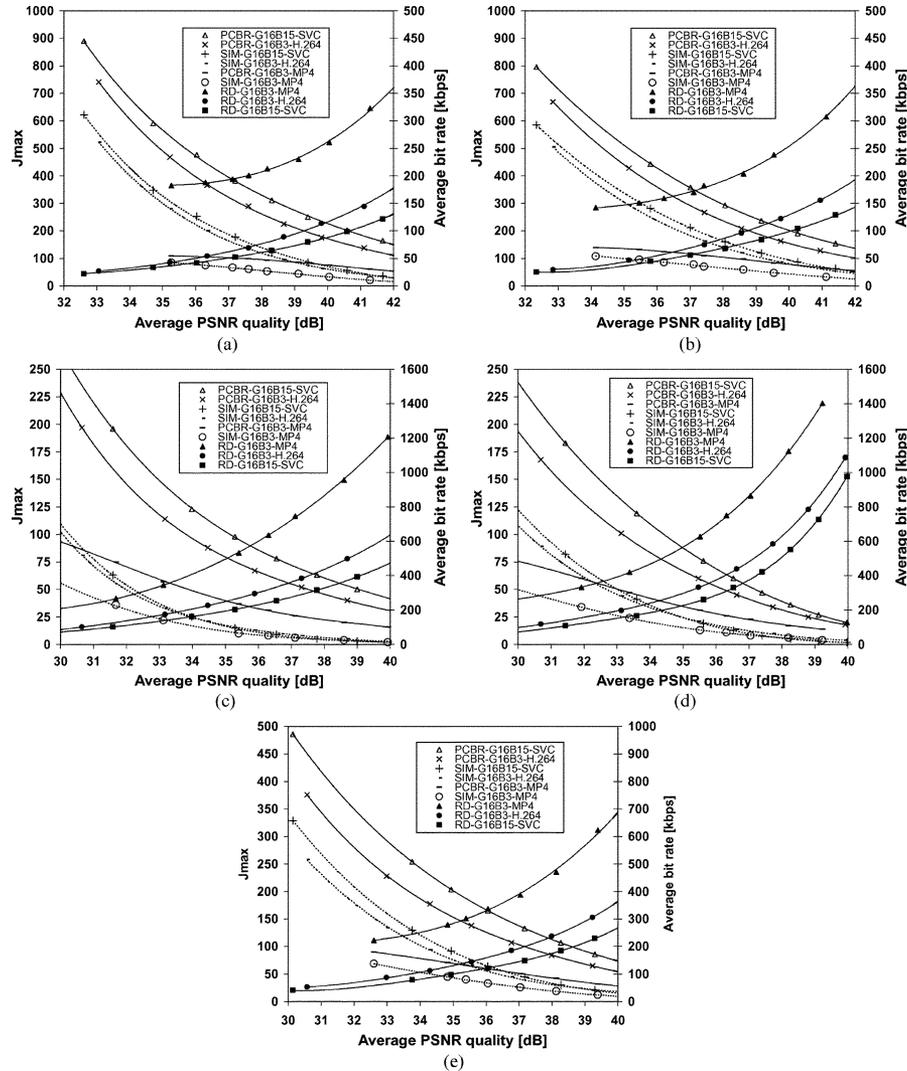
Fig. 9. $J_{\max}$ simulation (SIM) and RD curves for five long CIF sequences encoded with H.264/AVC (*G16-B3*), H.264 SVC (*G16-B15*), and MPEG-4 Part 2 (*G16-B3*). The channel capacity is $C = 20$ Mbps and the bit loss probability is $\epsilon = 10^{-5}$. Perfect CBR (PCBR) $J_{\max}$ curves are included for reference. (a) *Silence of the Lambs* (b) *Star Wars IV* (c) *Sony Demo* (d) *NBC 12 News* (e) *Tokyo Olympics*.

TABLE II
FRAME LOSS RATIOS FOR TRANSMITTING SINGLE STREAM OVER BOTTLENECK LINK WITH RATE SET TO LINK FACTOR TIMES AVERAGE BIT RATE OF ENCODED STREAM

Link Factor	H.264/AVC Mean	H.264/AVC Var.	MPEG-4 Part 2 Mean	MPEG-4 Part 2 Var.	Difference Mean	Difference Var.	HalfWidth 90% CI
1	0.158	0.0018	0.108	0.0003	0.050	0.0021	0.020
1.1	0.127	0.0017	0.079	0.0003	0.048	0.0021	0.020
1.2	0.102	0.0017	0.056	0.0003	0.045	0.0020	0.019
1.3	0.084	0.0014	0.041	0.0002	0.042	0.0017	0.017
1.4	0.070	0.0011	0.030	0.0001	0.039	0.0013	0.016
1.5	0.058	0.0009	0.021	0.0000	0.037	0.0010	0.014

a basic network configuration consisting of a video source and a sink connected by two routers in series. Each router had a queue set to 100 packets in NS-2 and employed drop tail queueing. There was a 10 ms propagation delay on all links. Each video frame was transmitted in one packet, which was dropped if it did not fit into the remaining free router buffer space. For a

given video sequence, streams of approximately the same average PSNR quality were selected. The bandwidth of the bottleneck link between the two routers was set to a link factor times the average bit rate of the selected streams, to normalize the difference in bit rates which are significantly different for the two standards. In Table II, a link factor of 1 corresponds to the bandwidth of the bottleneck link being equal to the average bit rate of the considered stream. Link factors 1.1, 1.2 etc., correspond to increasing bandwidth of the bottleneck link.

Table II presents the mean of the frame loss ratio across all ten sequences for H.264/AVC and MPEG-4 Part 2 as well as the mean difference between these two frame loss ratios and the 90% confidence interval on the mean difference. We observe that in the considered transmission scenarios, H.264/AVC encoded streams experience larger frame loss ratios than MPEG-4 encoded streams and that this difference between H.264/AVC and MPEG-4 Part 2, which is statistically significant, decreases with increasing link factor.
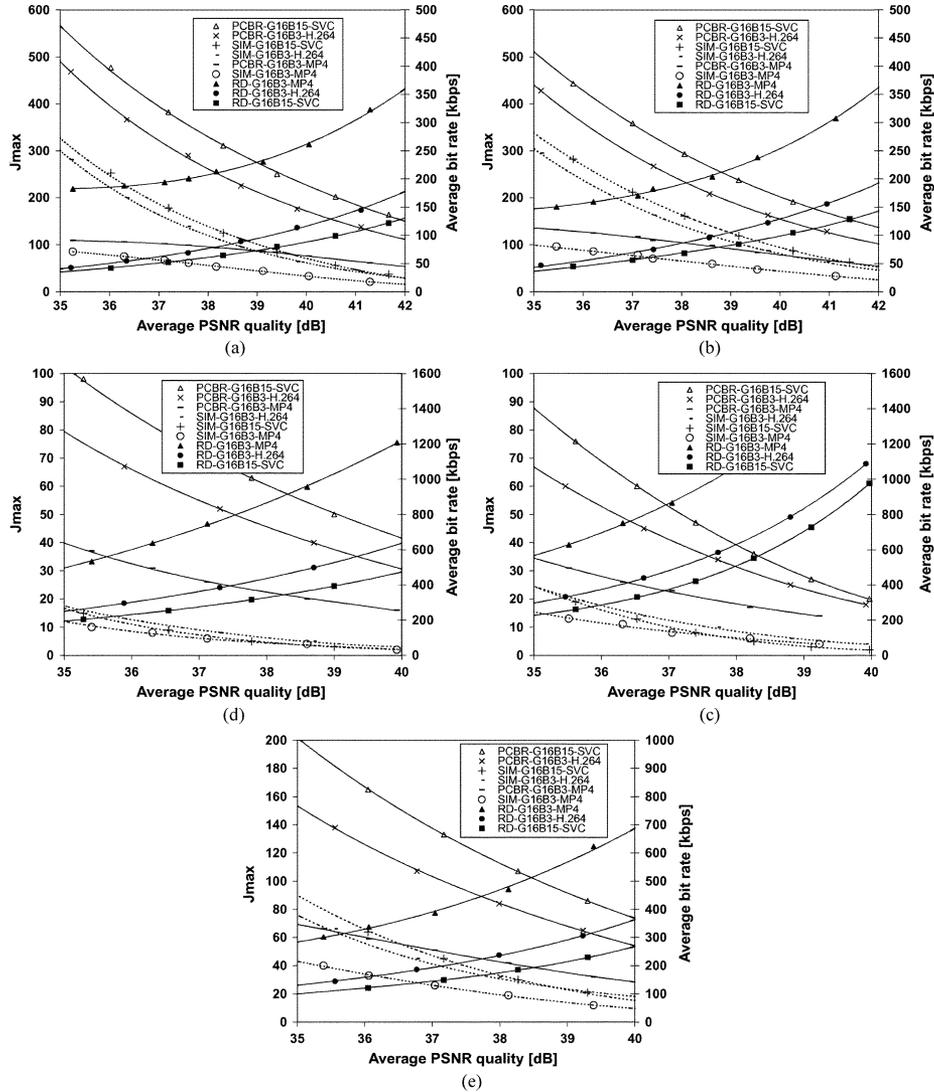
Fig. 10. Detail plots for the quality range 35–40/42 dB of the $J_{\max}$ simulations in Fig. 9. Perfect CBR (PCBR) $J_{\max}$ curves are included for reference. (a) *Silence of the Lambs* (b) *Star Wars IV* (c) *Sony Demo* (d) *NBC 12 News* (e) *Tokyo Olympics*.

B. Implications for Statistical Multiplexing

1) Experimental Setup: We investigate a basic real-time frame-based video streaming scenario modeled by a bufferless statistical multiplexer [30], [47]–[50]. In this model, a channel with bandwidth capacity C connects a video server with a bufferless statistical multiplexer to J receivers. Each video frame is transmitted during one frame period T (e.g., 33 ms for a frame rate of 30 frames/s). If the frame size equals $X_m(j)$ bits, with m denoting the frame index and j the stream index, then the bit rate required to transmit frame m of stream j is given by $X_m(j)/T$. If frame m of each stream j $(j = 1, \ldots, J)$ is statistically multiplexed onto the channel, then the aggregated bit rate is given by $R = \sum_{j=1}^{J} X_m(j)/T$.

In each experiment, we stream J identical video sequences, however, for each stream the starting phase is randomly selected according to a uniform distribution over all frames m $(m = 1, \ldots, M)$ of this one sequence [10], [48]. The streams are wrapped around to obtain streams of equal lengths.

We define this basic real-time video streaming scenario to provide a "ground truth" for studying the implications of the bit rate variabilities. We could have chosen a complex streaming scenario with several routers, buffers, aggregated traffic consisting of diverse video streams (content) and data cross traffic, etc., however, this would introduce "arbitrary" parameters that influence the outcome of the experiment.

In our simulations, we measure the information loss probability [48], [49], i.e., the information loss (bits) that occurs when the aggregated bit rate R exceeds the channel capacity C, and is given by:

$$P_{\text{loss}}^{\text{info}} = \frac{E\left[(R-C)^+ \times T\right]}{E[R \times T]} = \frac{E\left[(R-C)^+\right]}{E[R]}, \qquad (15)$$

where $[x]^+ = \max(0, x)$. The goal of the first set of simulations is to estimate the maximum number of video streams $J_{\max}$ that can be accommodated by the link capacity C, while constraining the information loss probability to a value smaller than ϵ. We consider $\epsilon = 10^{-5}$ and $\epsilon = 10^{-3}$, and set $C = 20$

Mbps. Many independent replications of each simulation were run until the 90% confidence interval of the information loss probability estimate was less than 10% of the corresponding sample mean. In the second set of simulations, we estimate the minimum link capacity $C_{\min}$ that accommodates a prescribed number of streams J while keeping the information loss probability smaller than a specified information loss probability ϵ, which we set to $\epsilon = 10^{-5}$. For each $C_{\min}$ estimate we perform 500 runs, each consisting of 1000 independent video streaming simulations.

We consider the five long CIF sequences described in Section IV, and encode them with H.264/AVC using GoP structure *G16-B3*, with H.264 SVC using GoP structure *G16-B15*, and with MPEG-4 Part 2 using GoP structure *G16-B3*. The chosen quantization parameters correspond to the range of average PSNR qualities from approximately 30 dB (acceptable quality) to at least 40 dB (high quality). We selected the GoP structures so that overall the highest RD efficiency is achieved for each encoder, as we observed in Section V. This way we are able to study the implications of the higher rate variability of hierarchical B frames which result in higher RD efficiency at the expense of a significant increase in computational complexity.

2) Results for $J_{\max}$: Fig. 9 depicts the RD curves and $J_{\max}$ simulation curves that are obtained with $\epsilon = 10^{-5}$ for the five sequences. Next, we discuss the $J_{\max}$ simulation curves for H.264/AVC encodings (*SIM-G16B3-H.264*), for H.264 SVC encodings (*SIM-G16B15-SVC*), and for MPEG-4 Part 2 encodings (*SIM-G16B3-MP4*).

For each sequence, the average bit rate difference between the three encoders is immediately clear. The $J_{\max}$ curves in the quality range up to 35 dB demonstrate a significant increase in the number of streams that the link supports for the H.264 SVC and H.264/AVC encodings compared to MPEG-4 Part 2. However, the $J_{\max}$ values are affected by the rate variability of the video traffic. To illustrate this effect, we additionally plot the $J_{\max}$ curves corresponding to the multiplexing of perfect constant bit rate (*PCBR*) traffic, denoted by PCBR in Fig. 9. We define *PCBR* video traffic as the sequence of identical frame size values that are all equal to the average frame size of the video stream. Hence, the rate variability of a *PCBR* video stream is zero and $J_{\max}$ is determined by dividing the link capacity C by the stream's average bit rate, resulting in the theoretical maximum value for $J_{\max}$. Comparing the $J_{\max}$ curves of the unsmoothed VBR traffic with those of the *PCBR* video traffic, we observe large differences that are attributable to the rate variability. The VBR traffic clearly results in fewer supported streams on the link than the *PCBR* video traffic. We also observe for all sequences that the gap between the *PCBR* $J_{\max}$ curves of the H.264 SVC and the H.264/AVC encodings is much wider than the gap between the corresponding H.264 SVC and H.264/AVC VBR traffic curves. This indicates that the relatively large reduction in the average bit rate with H.264 SVC compared to H.264/AVC translates generally into only relatively small increases in the number of supported H.264 SVC VBR traffic streams compared to the number of H.264/AVC streams, pro-

viding evidence of the profound impact of the rate variability increase of H.264 SVC on $J_{\max}$ compared to H.264/AVC. We remark that there is no data available for the MPEG-4 Part 2 curves in Figs. 9(a), (b), and (e) below respectively 35 dB, 34 dB, and 32.5 dB. The reason is that for these sequences the lowest quality is achieved around these qualities.

In Fig. 10, we zoom in on the high quality range (>35 dB) of Fig. 9. In the high quality range, the average bit rates of the streams are quite large compared to C. Therefore, the number of streams that can be supported by the link is relatively small and as a result the statistical multiplexing effect that mitigates the rate variability of the streams is reduced. We observe that the gap between the $J_{\max}$ curves of the unsmoothed H.264/AVC and H.264 SVC traffic is quite narrow for all sequences. Very interesting is that the $J_{\max}$ curves of the *Sony Demo* and *NBC 12 News* intersect around 35 dB when the number of multiplexed streams drops below approximately 20. This is a very important observation, because this means that the RD efficiency gain of H.264 SVC is completely compensated by the associated increased rate variability. For very high quality (>38 dB) the H.264 SVC $J_{\max}$ curve for the *Sony Demo* sequence even approaches the MPEG-4 Part 2 $J_{\max}$ curve, and surprisingly, for the *NBC 12 News* sequence the H.264 SVC $J_{\max}$ curve is below the MPEG-4 Part 2 $J_{\max}$ curve.

While the RD gains of H.264 SVC compared to the other two encoders are generally the largest for high qualities, this advantage does not necessarily translate into effective gains in the number of supported streams. On the contrary, the very advanced H.264 SVC encoder is outperformed by the MPEG-4 Part 2 encoder in case of the complex *NBC 12 News* sequence at very high qualities. The above observations are clearly dependent on the video content, but also on the number of multiplexed streams. Therefore, in Section VI-B-3 we perform a second set of simulations that estimate the required link capacity given the number of supported streams subject to a maximum information loss probability.

Next, we provide in Fig. 11 the detailed analysis of the $J_{\max}$ gains of H.264 SVC with respect to H.264/AVC for the information loss probability $\epsilon = 10^{-3}$. Each plot shows two $J_{\max}$ curves and two RD curves for respectively H.264/AVC (*G16-B3*) and H.264 SVC (*G16-B15*). A quick survey of the $J_{\max}$ curves shows a significant increase of the number of streams that the link supports for the H.264 SVC encodings compared to H.264/AVC, especially in the lower half of the quality range in each figure.

To obtain insight into the stream gains versus the bit rate gains, we fit fourth order polynomials (least mean squares fit) through each set of $J_{\max}$ simulation points and corresponding RD points. These fitted polynomials allow for the resampling of the curves and for the computation of relative gains (%) based on samples with corresponding average qualities (PSNR). We define the RD gain or average bit rate gain as the difference between the bit rates of two streams with identical average qualities, respectively encoded with H.264/AVC and SVC. We divide this difference by the H.264/AVC bit rate and express it as
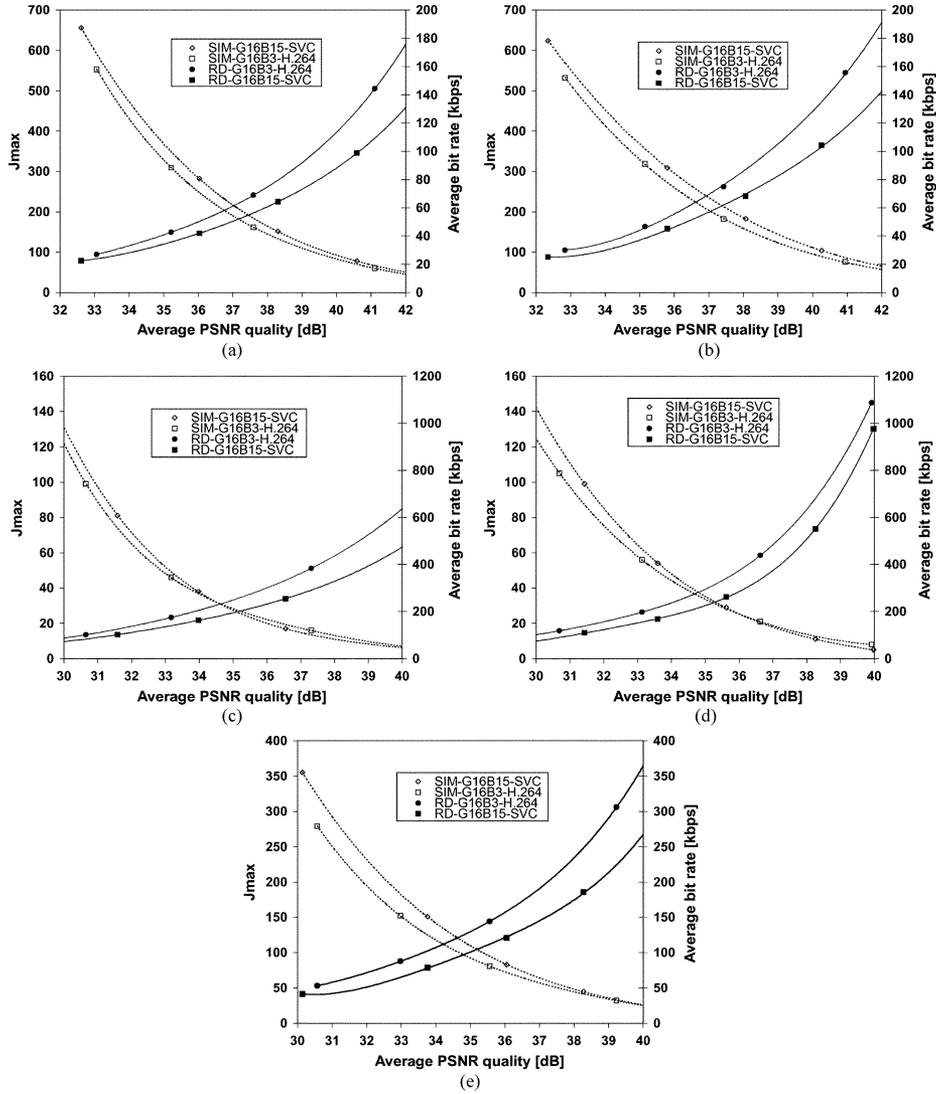
Fig. 11. $J_{\max}$ simulation ($\mathtt{SIM}$) and RD curves for five long CIF sequences encoded with H.264/AVC (*G16-B3*) and H.264 SVC (*G16-B15*). The channel capacity is $C = 20$ Mbps and the bit loss probability is $\epsilon = 10^{-3}$. (a) *Silence of the Lambs* (b) *Star Wars IV* (c) *Sony Demo* (d) *NBC 12 News* (e) *Tokyo Olympics*.

a percentage in Fig. 12. We define the $J_{\max}$ gain or supported stream gain as the difference between the $J_{\max}$ value of the SVC stream and the $J_{\max}$ value of the H.264/AVC stream, again for identical qualities. We divide this difference by the $J_{\max}$ value of the SVC stream and express it as a percentage in Fig. 12. Positive gains correspond to an increase in the number of supported streams and a reduction in the average bit rate of H.264 SVC. The reason for this choice of gain definitions is explained next.

In Fig. 12, the linear trend curves represent the average bit rate gains as a function of the average quality and the parabolic trend curves represent the supported stream gains. We observe that the average bit rate gains exceed 10% and reach values of more than 25%. In the perfect constant bit rate streaming scenario (*PCBR*), the average bit rate determines the number of supported streams on the link. This implies that the supported stream gains equal the bit rate gains according to our gain definitions above. However, in our streaming scenario, this is clearly not the case as we observe supported stream gains that are significantly smaller than the bit rate gains for the entire range of

average qualities. Secondly, the supported stream gain curves reach a maximum and are parabolic in contrast to the linear bit rate gain curves. Therefore, the observed gain differences between the bit rate gains and the supported stream gains depend on the average quality of the stream. Since supported stream gains differ strongly with the video sequence, there is a strong content dependency as well. All these observations point to the strong implications of the bit rate variability of the stream under test, which results in significant supported stream losses compared to the *PCBR* scenario. There can even be negative gains of supported streams or, equivalently, fewer streams are supported by the link with H.264 SVC than with H.264/AVC even though there is a significant bit rate gain (i.e., average bit rate reduction) with H.264 SVC compared to H.264/AVC. This is the case for high quality encodings of the sequences *NBC 12 News* (> 36 dB), *Sony Demo* (> 35 dB), and *Tokyo Olympics* (> 40.5 dB).

From these observations, we could conclude that the average bit rate efficiency improvements of H.264 SVC, using the
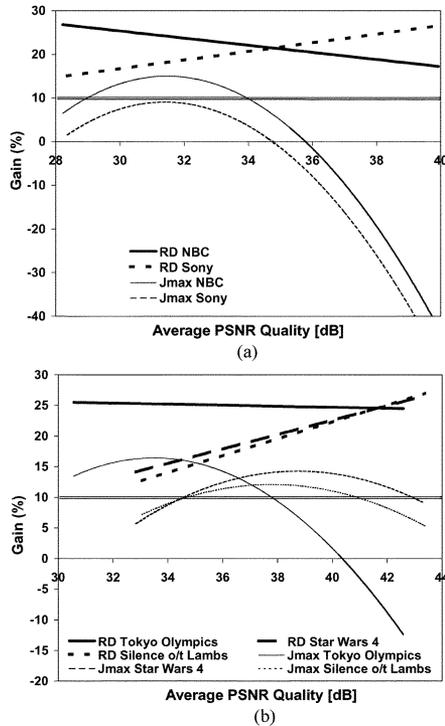
Fig. 12. J_{max} and RD gain curves (%) corresponding to Fig. 11. (a) *NBC 12 News*, *Sony Demo* (b) *Silence o/t Lambs, Star Wars IV, Tokyo Olympics*.

G16-B15 GoP structure and hierarchical B frames, result in significant gains in the number of streams supported with bufferless statistical multiplexing over a wide average quality range compared to H.264/AVC, using the *G16-B3* GoP structure and classical B frames. However, this conclusion does not consider the complexity increase of the *G16-B15* GoP structure based on hierarchical B frames. The *G16-B15* structure contains 15 B frames (per I frame), while the *G16-B3* structure contains 12 B frames and 3 P frames (per I frame). If we consider P frames about half as complex as B frames, then a basic complexity estimate suggests that the *G16-B15* structure is 12.5% more complex than the *G16-B3* structure. We consider hierarchical B frames to have the same complexity as classical B frames. If we take this complexity increase into account and expect at least a 10% increase in the number of supported streams to warrant the extra complexity cost, then we observe from Fig. 12 (10% line is double line in graphs) that the supported stream gain curve for the *Sony Demo* sequence falls completely below the 10% gain increase line. Also a large portion of the *NBC 12 News*, and *Tokyo Olympics* sequences' quality range would not benefit from the extra encoding complexity. Hence, in these circumstances the extra complexity may not be warranted, even though there is a clear rate-distortion benefit.

3) Results for C_{min}: In the second set of simulations, we examine the effect of the rate variability increase on statistical multiplexing from the perspective of the minimum link capacity C_{min} that accommodates a prescribed number of unsmoothed streams J. Fig. 13 depicts the C_{min} curves for the five long CIF sequences. The simulations are provided for $J = 4$, 16, and

64, for $\epsilon = 10^{-5}$. The C_{min} curves allow for examining the impact of the rate variability increase of H.264 SVC for a pre-determined number of streams J over the entire quality range. The study of the statistical multiplexing behavior for a small number of streams is important in the scenario of a low bandwidth link with a small number of streams, and in the scenario of a high bandwidth link with a few very high quality streams. The smaller C_{min}, the smaller the network resource requirements. We have not included the 90% confidence intervals in Fig. 13, because they are very small and would make the figures very cluttered.

For $J = 64$ we observe that the C_{min} values are lower for the H.264 SVC streams than for H.264/AVC streams. The link capacity difference is particularly significant in the high quality range (>35 dB). An exception is the *NBC 12 News* sequence which exhibits a reversal of the curves, i.e., lower C_{min} for H.264/AVC than for H.264 SVC in the very high quality range (>38 dB). Generally, in the quality range below 35 dB, the C_{min} differences become relatively small. However, both encoders have a clear advantage over MPEG-4 Part 2.

When we reduce the number of streams to $J = 16$, the statistical multiplexing effect is less able to compensate bit rate variabilities, as is evidenced in Fig. 13. For all sequences, the H.264/AVC streams are accommodated by C_{min} values that are smaller than or nearly equal to the values for the H.264 SVC streams, despite the lower average bit rates of the SVC streams. H.264 SVC still outperforms MPEG-4 Part 2 over the entire quality range and for all sequences.

For a small number of multiplexed streams, such as $J = 4$, the increased rate variability of H.264 SVC results in C_{min} values that are overall comparable to those of multiplexed MPEG-4 Part 2 streams. For the *Sony Demo* sequence, the H.264/AVC streams require the lowest link capacity values in the high quality range, while H.264 SVC has comparable performance to MPEG-4 Part 2. Below 35 dB, the C_{min} values are comparable for all encoders. For the *Star Wars 4* sequence, the MPEG-4 Part 2 streams require significantly higher minimum link capacities than H.264 SVC, which in turn is outperformed by H.264/AVC. For the *Silence of the Lambs* and *Tokyo Olympics* sequences, we observe the surprising result that H.264 SVC requires the highest C_{min} values over the entire quality range, and MPEG-4 Part 2 even outperforms H.264/AVC below, respectively, 38 dB and 36 dB. For the *NBC 12 News* sequence, H.264 SVC also has worst performance in the quality range above 35 dB.

From these minimum link capacity estimates, we conclude that the rate variability increase of H.264 SVC compared to H.264/AVC, which is due to hierarchical B frames with cascading quantizers, seriously affects the bufferless statistical multiplexing behavior. Even so, that for a relatively small number of multiplexed streams (<16), H.264/AVC generally results in lower C_{min} requirements, while depending on the video sequence, H.264 SVC can even be outperformed by MPEG-4 Part 2 streams. We briefly explain these observation next.
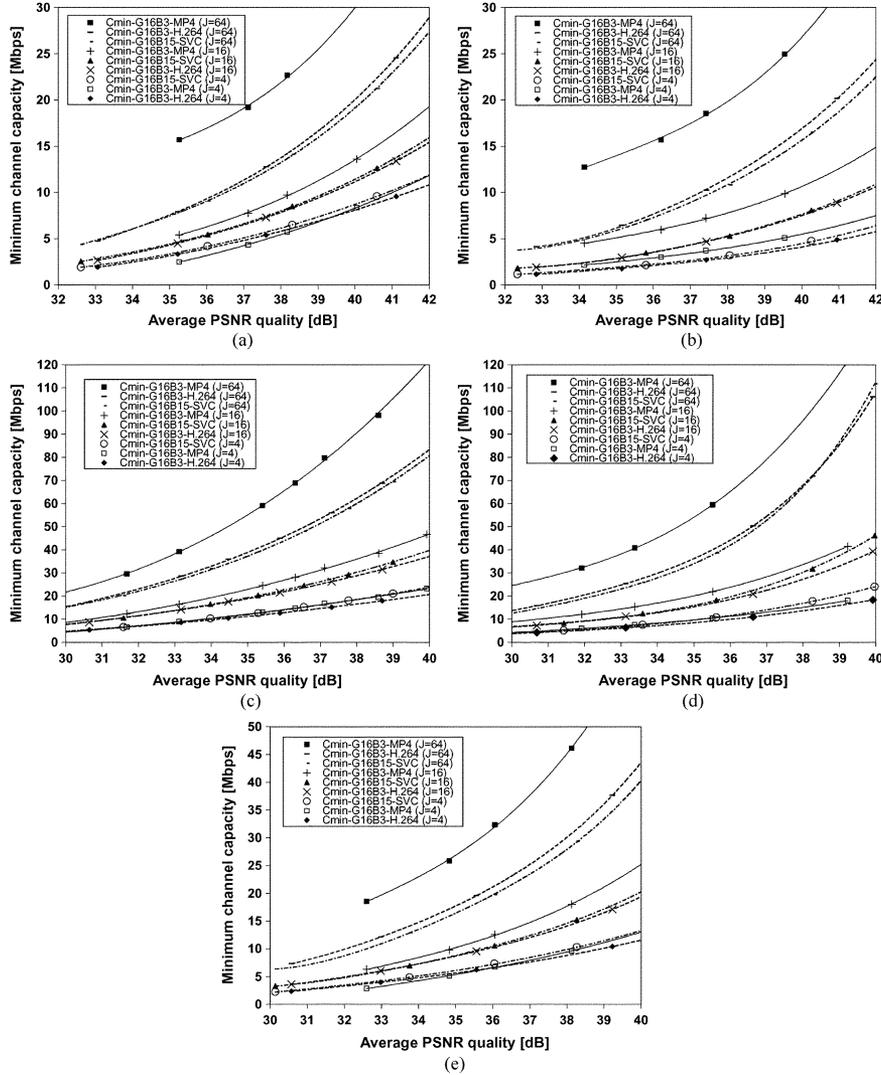
Fig. 13. Minimum channel capacity $C_{\min}$ simulation results for five long CIF sequences encoded with H.264/AVC (*G16-B3*), H.264 SVC (*G16-B15*), and MPEG-4 Part 2 (*G16-B3*). The bit loss probability is $\epsilon = 10^{-5}$. $C_{\min}$ curves are provided for the number of streams $J = 4$, 16, and 64. (a) *Silence of the Lambs* (b) *Star Wars IV* (c) *Sony Demo* (d) *NBC 12 News* (e) *Tokyo Olympics*.

$C_{\min}$ can theoretically be determined by building the probability density distribution of the aggregated streams (aggregated frame sizes) and determining the capacity that results in a specified loss ϵ. Since these distributions are long-tailed, for a small loss probability, such as $\epsilon = 10^{-5}$, the $C_{\min}$ value is determined by the probability situated in the tail of the distribution. The distribution's tail length (maximum aggregated frame size) is determined by the maximum frame size of the multiplexed stream. We have observed that H.264 SVC streams (*G16-B15*) have larger maximum frame sizes than H.264/AVC streams (*G16-B3*) for approximately the same average video quality. The larger maximum frame sizes of H.264 SVC are caused by the cascading quantizers of H.264 SVC that assign a different quality or quantizer parameter to frames of different temporal layers. Since I and P frames need the highest quality, they are also assigned the smallest quantizer parameter, while subsequent temporal layers are assigned larger quantizers and, hence, lower quality, as explained in Section III-C. For H.264 SVC with GoP

structure *G16-B15*, i.e., with five temporal layers, this means that to achieve a similar average quality as for H.264/AVC with classical B frames, the qualities of the H.264 SVC I frames must be higher than the qualities of the H.264/AVC I frames. Consequently, the H.264 SVC I frame sizes will be larger than the H.264/AVC I frame sizes. This means that the aggregated stream distributions of H.264 SVC streams will have longer tails than the distributions of H.264/AVC streams. Since the $C_{\min}$ value for small losses is dependent on the probability in the tail and the tail length, H.264 SVC can have higher $C_{\min}$ values than H.264/AVC, even though H.264 SVC has smaller average frame sizes than H.264/AVC.

VII. CONCLUSIONS

We have examined in detail the network traffic characteristics of variable bit rate H.264/AVC and H.264 SVC single-layer (non-scalable) encoded video. We have focused on a set of long video test sequences with a wide range of typical texture and

motion features. In summary, we found the following distinct characteristics of the H.264/AVC and H.264 SVC video traffic:

- From our *joint characterization* of the average bit rate and bit rate variability for a fixed desired video quality, we confirmed that H.264/AVC, and H.264 SVC codecs lead to significant average bit rate savings with respect to the MPEG-4 Part 2 codec. At the same time, the variability of the H.264/AVC and H.264 SVC video traffic is significantly higher than the variability of the MPEG-4 Part 2 video traffic. Whereas the coefficient of variation (standard deviation normalized by mean) of the frame sizes reaches levels above 2.4 for H.264/AVC, and even above 3.0 for H.264 SVC, it does generally not exceed 1.5 with MPEG-4 Part 2 [24], [30].
- The comparison between classical B frames (default in H.264/AVC) and hierarchical B frames (H.264 SVC), based on four GoP structure patterns that are supported by the encoders (*G16-B1*, *G16-B3*, *G16-B7*, and *G16-B15*), indicates that hierarchical B frames outperform classical B frames at the expense of higher rate variability. From the four tested GoP structures, *G16-B3* results in the best RD efficiency for H.264/AVC with classical B frames and *G16-B15* results in the best RD efficiency for H.264 SVC with hierarchical B frames.
- Depending on the application scenario, it may be possible to smooth the video traffic before sending it into the network, thus reducing the traffic variability at the expense of introducing smoothing delay. We observed that the smoothed H.264/AVC and H.264 SVC video traffic exhibits variabilities at the same level or above the unsmoothed MPEG-4 Part 2 video traffic, indicating that even when smoothing is employed, the transport mechanisms for the new H.264/AVC (and extensions) video will need to be designed to accommodate substantial traffic variabilities.
- Our streaming simulation studies demonstrated *(i)* that the increased bit rate variability results in significantly higher frame losses for H.264/AVC encoded video compared to MPEG-4 Part 2 encoded video when transmitting a single video stream over a bottleneck link, and *(ii)* that for bufferless statistical multiplexing, a significant improvement in bit rate-distortion efficiency does not suffice to conclude that there is an equal gain in the number of streams that can be statistically multiplexed onto a link subject to an information loss probability constraint. We have thus demonstrated the relevance and importance of investigating the implications of increased video traffic rate variabilities for video network transport, and that solely focusing on rate-distortion efficiency improvements may not necessarily lead to optimal operating points for all networking scenarios.

There are several directions for important future work. One direction is to examine the suitability of existing traffic models and video transport mechanisms for H.264/AVC and SVC video traffic. The existing traffic models, such as [16]–[23], and

video transport mechanisms for a wide range of communication networks, including general IP networks, see e.g., [51]–[54], wireless networks, see e.g., [55]–[57], and peer-to-peer network [58]–[60], were primarily developed based on MPEG-4 Part 2 video traffic. It is therefore necessary to examine how well these existing traffic models describe and how efficiently the existing mechanisms can transport the significantly more variable H.264/AVC and SVC video traffic. If necessary the existing traffic models and transport mechanisms need to be extended to accommodate the unprecedented variability of the H.264/AVC and SVC video traffic.

ACKNOWLEDGMENT

REFERENCES

[1] D. Marpe, T. Wiegand, and G. Sullivan, "The H.264/MPEG-4 advanced video coding standard and its applications," *IEEE Communications Magazine*, vol. 44, no. 8, pp. 134–143, Aug. 2006.

[2] R. Schafer, H. Schwarz, D. Marpe, T. Schierl, and T. Wiegand, "MCTF and scalability extension of H.264/AVC and its application to video transmission, storage and surveillance," in *Proceedings of Visual Communications and Image Processing (VCIP), Proceedings of SPIE—Volume 5960*, Beijing, China, July 2005, pp. 596 011-1–596 011-12.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[4] T. Lakshman, A. Ortega, and A. Reibman, "VBR video: Tradeoffs and potentials," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 952–973, May 1998.

[5] A. R. Reibman and M. T. Sun, *Compressed Video Over Networks*. New York: Marcel Dekker, 2000.

[6] D. Wu, Y. Hou, W. Zhu, Y.-Q. Zhang, and J. Peha, "Streaming video over the internet: Approaches and directions," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 282–300, Mar. 2001.

[7] D. Marpe, T. Wiegand, and S. Gordon, "H.264/MPEG-4 AVC Fidelity Range Extensions: Tools, profiles, performance, and application areas," in *Proc. IEEE Int. Conf. on Image Proc. (ICIP)*, Sept. 2005, pp. 593–596.

[8] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1194–1203, Sept. 2007.

[9] ISO/IEC JTC 1/SC 29/WG 11 N2802, Information Technology-Generic Coding of Audio-Visual Objects-Part 2: Visual, Final Proposed Draft Amendment 1.   Geneva, July 1999.

[10] P. Seeling, M. Reisslein, and B. Kulapala, "Network performance evaluation with frame size and quality traces of single-layer and two-layer video: A tutorial," *IEEE Communications Surveys and Tutorials* vol. 6, no. 3, pp. 58–78, Third Quarter, 2004 [Online]. Available: http://trace.eas.asu.edu, video traces

[11] S. Bakiras and V. O. K. Li, "Maximizing the number of users in an interactive video-on-demand system," *IEEE Trans. Broadcasting*, vol. 48, no. 4, pp. 281–292, Dec. 2002.

[12] P. Koutsakis and M. Paterakis, "Policing mechanisms for the transmission of videoconference traffic from MPEG-4 and H.263 video coders in wireless ATM networks," *IEEE Trans. Vehicular Technology*, vol. 53, no. 5, pp. 1525–1530, 2004.

[13] B. Nikolaus, J. Ott, C. Borrmann, and U. Borrmann, "Generalized greedy broadcasting for efficient media-on-demand transmissions," *IEEE Trans. Broadcasting*, vol. 51, no. 3, pp. 354–359, 2005.

[14] J. Roberts, "Internet traffic, QoS, and pricing," *Proceedings of the IEEE*, vol. 92, no. 9, pp. 1389–1399, 2004.

[15] Y. Xu and R. Guerin, "Individual QoS versus aggregate QoS: A loss performance study," *IEEE/ACM Trans. Networking*, vol. 13, no. 2, pp. 370–383, 2005.

[16] A. Alheraish, S. Alshebeili, and T. Alamri, "A GACS modeling approach for MPEG broadcast video," *IEEE Trans. Broadcasting*, vol. 50, no. 2, pp. 132–141, June 2004.

[17] N. Ansari, H. Liu, Y. Q. Shi, and H. Zhao, "On modeling MPEG video traffics," *IEEE Trans. Broadcasting*, vol. 48, no. 4, pp. 337–347, Dec. 2002.

[18] M. Dai and D. Loguinov, "Analysis and modeling of MPEG-4 and H.264 multi-layer video traffic," in *Proc. of IEEE INFOCOM*, Miami, FL, Mar. 2005, pp. 2257–2267.

[19] X.-D. Huang, Y.-H. Zhou, and R.-F. Zhang, "A multiscale model for MPEG-4 varied bit rate video traffic," *IEEE Trans. Broadcasting*, vol. 50, no. 3, pp. 323–334, Sept. 2004.

[20] M. M. Krunz and A. M. Makowski, "Modeling video traffic using $M/G/\infty$ input processes: A compromise between Markovian and LRD models," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 733–748, June 1998.

[21] C. H. Liew, C. K. Kodikara, and A. M. Kondoz, "MPEG-encoded variable bit-rate video traffic modelling," *IEE Proceedings Communications*, vol. 152, no. 5, pp. 749–756, Oct. 2005.

[22] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar, "Modeling full-length video using Markov-modulated gamma-based framework," *IEEE/ACM Trans. Networking*, vol. 11, no. 4, pp. 638–649, Aug. 2003.

[23] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar, "Study of long duration MPEG-trace segmentation methods for developing frame size based traffic models," *Computer Networks*, vol. 44, no. 2, pp. 177–188, 2004.

[24] G. Van der Auwera, M. Reisslein, and L. J. Karam, "Video texture and motion based modeling of rate variability-distortion (VD) curves," *IEEE Trans. Broadcasting*, vol. 53, no. 3, pp. 637–648, Sept. 2007.

[25] W.-C. Feng, *Buffering Techniques for Delivery of Compressed Video in Video-on-Demand Systems*.   : Kluwer Academic Publisher, 1997.

[26] M. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," in *Proceedings of ACM Sigcomm*, London, UK, Sept. 1994, pp. 269–280.

[27] M. Krunz, R. Sass, and H. Hughes, "Statistical characteristics and multiplexing of MPEG streams," in *Proceedings of IEEE Infocom '95*, April 1995, pp. 455–462.

[28] M. Reisslein, J. Lassetter, S. Ratman, O. Lotfallah, F. Fitzek, and S. Panchanathan, "Traffic and quality characterization of scalable encoded video: A large-scale trace-based study, Part 1: Overview and definitions," ASU, Tempe, AZ, Dec. 2003 [Online]. Available: http://trace.eas.asu.edu

[29] O. Rose, "Simple and efficient models for variable bit rate MPEG video traffic," *Performance Evaluation*, vol. 30, no. 1–2, pp. 69–85, 1997.

[30] P. Seeling and M. Reisslein, "The rate variability-distortion (VD) curve of encoded video and its impact on statistical multiplexing," *IEEE Trans. Broadcasting*, vol. 51, no. 4, pp. 473–492, Dec. 2005.

[31] P. Cuenca, A. Garrido, F. Quiles, and L. Orozco-Barbosa, "An efficient protocol architecture for error-resilient MPEG-2 video communications over ATM networks," *IEEE Trans. Broadcasting*, vol. 45, no. 1, pp. 129–140, Mar. 1999.

[32] G. Van der Auwera, P. T. David, and M. Reisslein, "Traffic characteristics of H.264/AVC variable bit rate video," *IEEE Communications Magazine* 2008 [Online]. Available: http://www.fulton.asu.edu/mre/h264CommMag07.pdf, in print

[33] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: Tools, performance and complexity," *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 7–28, First Quarter, 2004.

[34] G. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," in *Proc. of SPIE 5558, Conference on Applications of Digital Image Processing XXVII, Special Session on Advances in New Emerging Standard: H.264/AVC I*, Denver, CO, Aug. 2004, pp. 454–474.

[35] A. Puri, X. Chen, and A. Luthra, "Video coding using the H.264/MPEG-4 AVC compression standard," *Journal of Visual Communication and Image Representation*, vol. 19, no. 9, pp. 793–849, Oct. 2004.

[36] G. Van der Auwera, P. T. David, and M. Reisslein, "Video traffic analysis of H.264/AVC and extensions: Single-layer statistics," ASU, Tempe, AZ, Tech. Rep., July 2007, at [Online]. Available: http://www.fulton.asu.edu/mre/h264_traffic_single_layer_ext.pdf

[37] Z. Chen and K. Ngan, "Recent advances in rate control for video coding," *Signal Processing: Image Communication*, vol. 22, no. 1, pp. 19–38, Jan. 2007.

[38] C. Bewick, R. Pereira, and M. Merabti, "Network constrained smoothing: Enhanced multiplexing of MPEG-4 video," in *Proceedings of IEEE International Symposium on Computers and Communications*, Taormina, Italy, July 2002, pp. 114–119.

[39] H.-C. Chao, C. L. Hung, and T. G. Tsuei, "ECVBA traffic-smoothing scheme for VBR media streams," *International Journal of Network Management*, vol. 12, pp. 179–185, 2002.

[40] W.-C. Feng and J. Rexford, "Performance evaluation of smoothing algorithms for transmitting prerecorded variable-bit-rate video," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 302–312, Sept. 1999.

[41] T. Gan, K.-K. Ma, and L. Zhang, "Dual-plan bandwidth smoothing for layer-encoded video," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 379–392, Apr. 2005.

[42] M. Krunz, W. Zhao, and I. Matta, "Scheduling and bandwidth allocation for distribution of archived video in VoD systems," *Journal of Telecommunication Systems, Special Issue on Multimedia*, vol. 9, no. 3/4, pp. 335–355, Sept. 1998.

[43] H. Lai, J. Y. Lee, and L.-K. Chen, "A monotonic-decreasing rate scheduler for variable-bit-rate video streaming," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 2, pp. 221–231, Feb. 2005.

[44] A. Solleti and K. J. Christensen, "Efficient transmission of stored video for improved management of network bandwidth," *International Journal of Network Management*, vol. 10, pp. 277–288, 2000.

[45] J. C. H. Yuen, E. Chan, and K.-Y. Lam, "Real time video frames allocation in mobile networks using cooperative pre-fetching," *Multimedia Tools and Applications*, vol. 32, no. 3, pp. 329–352, Mar. 2007.

[46] "NS-2 The Network Simulator," 2007, available from [Online]. Available: http://www.isi.edu/nsnam/ns/

[47] S. Racz, T. Jakabfy, J. Farkas, and C. Antal, "Connection admission control for flow level QoS in bufferless models," in *Proc. IEEE INFOCOM*, 2005, pp. 1273–1282.

[48] M. Reisslein and K. W. Ross, "Call admission for prerecorded sources with packet loss," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 1167–1180, Aug. 1997.

[49] J. Roberts, U. Mocci, and J. Virtamo, *Broadband Network Traffic: Performance Evaluation and Design of Broadband Multiservice Networks, Final Report of Action COST 242, (Lecture Notes in Computer Science, Vol. 1155)*.   New York: Springer Verlag, 1996.

[50] Z. Zhang, J. Kurose, J. Salehi, and D. Towsley, "Smoothing, statistical multiplexing and call admission control for stored video," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1148–1166, Aug. 1997.

[51] T. Ahmed, A. Mehaoua, R. Boutaba, and Y. Iraqi, "Adaptive packet video streaming over IP networks: A cross-layer approach," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 385–401, Feb. 2005.

[52] T. Kim and M. H. Ammar, "Optimal quality adaptation for scalable encoded video," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 344–356, Feb. 2005.

[53] M. Krunz, "Bandwidth allocation strategies for transporting variable-bit-rate video traffic," *IEEE Communications Magazine*, vol. 37, no. 1, pp. 40–46, Jan. 1999.

[54] G.-M. Muntean, P. Perry, and L. Murphy, "A new adaptive multimedia streaming system for all-IP multi-service networks," *IEEE Trans. Broadcasting*, vol. 50, no. 1, pp. 1–10, Mar. 2004.

[55] L. Haratcherev, J. Taal, K. Langendoen, R. Lagendijk, and H. Sips, "Optimized video streaming over 802.11 by cross-layer signaling," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 115–121, Jan. 2006.

[56] S. Khan, Y. Peng, E. Steinbach, M. Sgroi, and W. Kellerer, "Application-driven cross-layer optimization for video streaming over wireless networks," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 122–130, Jan. 2006.

[57] F. Yang, Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Bit allocation for scalable video streaming over mobile wireless internet," in *Proc. IEEE INFOCOM*, 2004, pp. 2142–2151.

[58] H.-Y. Hsieh and R. Sivakumar, "Accelerating peer-to-peer networks for video streaming using multipoint-to-point communication," *IEEE Communications Magazine*, vol. 42, no. 8, pp. 111–119, Aug. 2004.

[59] E. Kim and J. Liu, "Design of HD-quality streaming networks for real-time content distribution," *IEEE Trans. Consumer Electronics*, vol. 52, no. 2, pp. 392–401, May 2006.

[60] J. Liang and K. Nahrstedt, "DagStream: Locality aware and failure resilient peer-to-peer streaming," in *Proc. SPIE/ACM Multimedia Computing and Networking, Proceedings of SPIE—Volume 6071*, Jan. 2006, pp. 60 710L-1–60 710L-15.

Geert Van der Auwera received the Ph.D. degree in Electrical Engineering from Arizona State University, Tempe, USA, in 2007, and the Belgian MSEE degree from Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 1997. His research interests are video traffic and quality characterization, video streaming mechanisms and protocols, and video coding. Until the end of 2004, he was a scientific advisor for IWT-Flanders, the Institute for the Promotion of Innovation by Science and Technology in Flanders, Belgium. In 2000, he joined IWT-Flanders after researching wavelet video coding at the Electronics and Information Processing Department (ETRO), VUB. In 1998, Mr. Van der Auwera's thesis on motion estimation in the wavelet domain received the Barco and IBM prizes by the Fund for Scientific Research of Flanders, Belgium.

Prasanth T. David received his B.Tech. degree from the University of Kerala, Trivandrum, India in 2001. From 2001, he was a Systems Design Engineer with G4 Matrix Technologies, India, where he worked on video processing for a media processing engine. From 2005 to 2007 he was with the Electrical Engineering Dept. at Arizona State University, Tempe, from where he obtained his M.S. degree. Since then, he is with Intel Corp., CA. His fields of interest are multimedia coding, media processors, and video streaming technologies.

Martin Reisslein is an Associate Professor in the Department of Electrical Engineering at Arizona State University (ASU), Tempe. He received the Dipl.-Ing. (FH) degree from the Fachhochschule Dieburg, Germany, in 1994, and the M.S.E. degree from the University of Pennsylvania, Philadelphia, in 1996. Both in electrical engineering. He received his Ph.D. in systems engineering from the University of Pennsylvania in 1998. During the academic year 1994–1995 he visited the University of Pennsylvania as a Fulbright scholar. From July 1998 through October 2000 he was a scientist with the German National Research Center for Information Technology (GMD FOKUS), Berlin and lecturer at the Technical University Berlin. From October 2000 through August 2005 he was an Assistant Professor at ASU. He served as editor-in-chief of the *IEEE Communications Surveys and Tutorials* from January 2003 through February 2007 and has served on the Technical Program Committees of *IEEE Infocom*, *IEEE Globecom*, and the *IEEE International Symposium on Computer and Communications*. He has organized sessions at the *IEEE Computer Communications Workshop (CCW)*. He maintains an extensive library of video traces for network performance evaluation, including frame size traces of MPEG-4 and H.264 encoded video, at http://trace. eas.asu.edu. His research interests are in the areas of Internet Quality of Service, video traffic characterization, wireless networking, optical networking, and engineering education. (web: http://www.fulton.asu.edu/~mre).