ACCEPTED FROM OPEN CALL

# Traffic Characteristics of H.264/AVC Variable Bit Rate Video

*Geert Van der Auwera, Prasanth T. David, and Martin Reisslein, Arizona State University*


## ABSTRACT

The recently developed H.264/AVC encoder compresses video significantly more efficiently than previous encoders and is expected to be used for compressing the majority of video transported over communication networks. The traffic characteristics of encoded video have a significant impact on the network transport of compressed video, making it very important to study the characteristics of H.264/AVC video traffic. In this article, we examine the bit-rate distortion performance, bit-rate variability, and long-range dependence of the H.264/AVC encoder for long videos up to high-definition resolution. We also explore the impact of smoothing on the H.264/AVC video traffic. We find that compared to the MPEG-2 and MPEG-4 Part 2 encoders, H.264/AVC achieves lower average bit rates for a given video quality at the expense of significantly increased traffic variability that remains at high levels even with smoothing.


## INTRODUCTION

The H.264/MPEG-4 Advanced Video Coding (AVC) standard (also known as H.264/MPEG-4 Part 10) [1] with its fidelity range extensions (FRExt) [2, 3] is expected to have a broad application domain for video transmission and storage up to high definition (HD) resolution. Indications of the growing acceptance of the H.264/MPEG-4 AVC standard, which we refer to as H.264/AVC for brevity, are its recent inclusion in application standards and industry consortia specifications, such as digital video broadcasting (DVB), high definition-digital video disc (HD-DVD), and Blu-Ray. At the same time, there is a growing share of streaming video traffic over the Internet, and the introduction of IPTV over high-speed access network links, such as Ethernet passive optical networks or asymmetric digital subscriber line (ADSL)2+/very high bit-rate DSL (VDSL)2, is ongoing.

In general, video can be encoded with:

- Fixed quantization scales, which results in nearly constant video quality at the expense of variable video traffic (bit rate)
- Rate control, which adapts the quantization scales to keep the video bit rate nearly constant at the expense of variable video quality [4]

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 on variable bit-rate encodings with fixed quantization scales. An additional motivation for the focus on encodings 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]. For a prescribed video quality, the (nearly) constant bit rate of a rate-controlled encoding is generally significantly larger than the average bit rate of the corresponding (with the same prescribed quality) variable bit-rate encoding, and smaller than the peak bit rate of the corresponding variable bit rate encoding. The number of rate controlled video streams that a communication link with fixed transmission rate can simultaneously support is generally the link transmission rate divided by the constant video bit rate. With variable bit rate video, there is the potential to carry more simultaneous video streams on the link because their average bit rate is lower (for the same video quality). Realizing this potential requires the judicious exploitation of statistical multiplexing of the variable bit rate streams such that encoded video frames that are larger than average size (in bits) of one stream coincide with smaller than average frames of another stream. When the small and large frames of the ongoing streams compensate each other through statistical multiplexing, they require (in the long run) a transmission bit rate that is only very slightly above their average rate; whereby, the more variable the video traffic, generally the more challenging it is to achieve the statistical multiplexing for transport with transmission rates close to the average video bit rates. A wide array of video transport mechanisms were developed, striving to realize the potential of variable bit-rate video by efficiently accommodating the varying sizes (in bits) of the encoded video frames while meeting their strict play-out deadlines; see for example, [5, 6].

The bit-rate variability is commonly characterized by the coefficient of variation (CoV) of the frame sizes (in bits), whereby the CoV is defined as the standard deviation of the frame

*This work was supported in part by the National Science Foundation through Grant no. Career ANI-0133252, Grant no. ANI-0136774, and Grant no. CRI-0750927.*

sizes normalized by their mean. (Some studies also use the peak-to-mean frame size ratio, that is, the ratio of the largest encoded frame to the average encoded frame size.) Intuitively, a higher CoV means that the differences between the largest and average size encoded frames are larger and that there are more large frames, making it more difficult to efficiently transmit the video over communication networks [4, 5].

The widespread adoption of the new H.264/AVC video standard with its extensions requires a careful study of the traffic variability characteristics of video coded with the new H.264/AVC codec. Existing studies of the H.264/AVC codec and its extensions, such as [1, 2], focus primarily on the rate-distortion (RD) performance, that is, the video quality as a function of the *average* bit rate, and typically consider only short video sequences up to a few hundred frames. In contrast, in this article, we examine the CoV of the frame sizes as a function of the video quality, that is, we study the bit-rate variability-distortion (VD) curve [7], for H.264/AVC encoded video. We also examine the long range dependence properties [8] of H.264/AVC encoded video. To obtain reliable and meaningful statistical estimates of the traffic variability and long-range dependence properties, it is necessary to examine *long* video sequences with several thousand frames as we do in this study.

Recent studies, for example, [7, 9], have examined the bit-rate variability of video encoded with the older MPEG-4 Part 2 standard and the wide array of existing video transport mechanisms [4–6] that were developed primarily based on the characteristics of MPEG-2 and MPEG-4 Part 2 encoded video. To the best of our knowledge, the bit-rate variability of H.264/AVC encoded video and the bit-rate variability of video up to HD resolution are examined in the present study for the first time. We discover that the H.264/AVC codec produces significantly higher traffic variability than the older MPEG-2 and MPEG-4 Part 2 codecs, resulting in new challenges for efficient network transport.

This article is structured as follows. In the following section, we briefly review the H.264/AVC video codec and FRExt to provide the context for our video traffic studies. We describe the video sequences that we employ and the software that we use for processing and encoding. We also define the video traffic metrics and video quality metrics. We study the bit-rate variability of H.264/AVC and compare with the MPEG-4 Part 2 encoder. We demonstrate and explain the reason for the significantly higher rate variability of H.264/AVC compared to MPEG-4 Part 2. A fundamental technique for mitigating the impact of high video traffic variability is frame size smoothing [10], which averages the sizes of several consecutive frames of a video stream before transmission into the network and is explored for H.264/AVC. Next, we study the long-range dependence characteristics of the H.264/AVC video traffic because long range dependence, in general, has a profound impact on video network transport [11]. Studies of video streaming over communication networks often rely on video traces [12]. Generating traces of HD video encodings is very demanding, and we examine the FRExt video traffic to determine whether lower resolution video traces can be scaled up to simulate HD video traffic. At the end of the article, we summarize our conclusions.

## Overview of H.264/AVC Codec and Extensions

H.264/AVC represents a big leap in video compression technology with typically a 50 percent reduction of average bit rate for a given video quality compared to MPEG-2 and about a 30 percent reduction compared to MPEG-4 Part 2 [13]. The encoding loop from previous standards consisting of block transform in conjunction with motion compensated prediction (MCP) is still in place, but a number of new encoding mechanisms were added that cumulatively provide 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 of the tools for achieving high coding efficiency for high bit-rate applications. The *Extended profile* applies to error-resilient streaming applications. The FRExt amendment adds four *High profiles*: High (HP), High 10 (Hi10P), High 4:2:2 (Hi422P), High 4:4:4 (Hi444P) [2, 3]. The High profile includes improved tools that can result in up to a 10 percent compression gain over the Main profile and up to 59 percent over MPEG-2 for HD video, with only a marginal increase in computational complexity compared to the Main profile. Recently, five additional profiles were added for professional applications, for example, to support intra-only encoding.

We now briefly discuss the main new features of H.264/AVC and refer to [1] for more details. A major improvement is the introduction of the entropy coding scheme, context-adaptive binary arithmetic coding (CABAC), which typically gives 10–15 percent bit rate savings [13] over previous variable-length coding schemes used in MPEG-2/4. Because 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 that significantly reduces ringing artifacts, and an adaptive in-loop deblocking filter that reduces blocking artifacts [13]. H.264/AVC also introduces a new tool called *variable block sizes* that introduce a different number of square and rectangular block sizes, such as $(4 \times 4)$, $(8 \times 4)$, and $(16 \times 8)$ pixels. These different block sizes permit selecting the optimal block size for MCP.

Video compression, in general, is a trade-off in an RD sense, between the removal of redundancies by the encoding tools (including the reduction of the bit rate through the quantiza-

*H.264/AVC represents a big leap in video compression technology with typically a 50 percent reduction of average bit rate for a given video quality compared to MPEG-2 and about a 30 percent reduction compared to MPEG-4 Part 2.*

The bit rate variability is characterized by the coefficient of variation. A high CoV makes it more difficult to efficiently transport video over communication networks.

tion of prediction errors) and the introduced visual distortion. Previous codecs worked primarily toward optimizing either one of these two goals. H.264/AVC uses Lagrangian-based RD optimization to jointly optimize both goals [13]. This RD optimization can be applied to individual encoding mechanisms, as well as to the entire encoder. For instance, the RD optimization helps in making macroblock mode decisions, that is, deciding whether a given macroblock should be intra-coded (using the block transform) or inter-coded with MCP from a macroblock of a different frame. Similarly, in MCP, the RD optimization can be used to find the optimal motion vectors. These Lagrangian RD optimizations can improve the compression efficiency by up to 9 percent [13], but significantly increase the complexity of the encoding. Therefore, these optimization features may or may not be used, depending on the target application.

The MCP in previous standards allowed for one reference frame (I or P) from the past, for predicting P frame blocks, and one reference frame (I or P) from the past and one reference frame (I or P) from the future, for predicting B frame blocks, whereby the blocks from these past and future reference frames were weighted equally to form the predicted B frame block. Similarly, for predicting 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* [14].

## Video Sequences, Encoding Tools, and Video Traffic Metrics

The common intermediate format (CIF) (352 × 288 pixels) video sequences used for the encodings presented in this study are the ten-minute *Sony Digital HD video camera recorder* demo sequence, which we refer to as the *Sony Demo* sequence and the first half-hour of the *Silence of the Lambs* movie, which is a drama/thriller.

The *Sony Demo* sequence is originally an HD video sequence with 1280 × 720 pixels. The sequence consists of 29 scenes with complex texture and a wide range of low-to-high motion activity. We also use ten minutes of the *Terminator 2* HD sequence with the same resolution. These two HD sequences were originally encoded in *Windows Media 9* format at very high quality (perceptually perfect). We decoded the sequences into uncompressed YUV format using the *MEncoder* tool (http://www.mplayerhq.hu). We also used this tool to downsample the original sequences to CIF resolution.

More experiments showing the same trends as presented here are reported in [15] for the long sequences *Star Wars 4*, *Tokyo Olympics*, and *NBC 12 News*, which can respectively be described as science fiction/action, sports, and news video.

We employ the JM reference software (http://iphome.hhi.de/suehring/tml/, version 10.2), which is the official MPEG and International Telecommunication Union (ITU) reference implementation, for the H.264/AVC Main profile and FRExt encodings, the MPEG-4 Part 2 *Microsoft v2.3.0* software, and the *FFmpeg* MPEG-2 implementation (http://ffmpeg.mplayerhq.hu/).

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 frame with $N_x \times N_y$ pixels and 8-bit pixel values is computed as a function of the mean squared error (MSE) as:

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

$$PNSR = 10 \cdot \log_{10} \frac{255^2}{MSE}. \tag{2}$$

For a video sequence consisting of $M$ frames encoded with a given quantization scale, let $X_m$, $m = 1, \ldots, M$, denote the sizes (in bit) of the encoded video frames. The mean frame size of the encoded video sequence is defined as

$$\bar{X} = \frac{1}{M} \sum_{m=1}^{M} X_m, \tag{3}$$

while the variance $\sigma^2$ (square of the standard deviation) of the frame sizes is defined as

$$\sigma^2 = \frac{1}{(M-1)} \sum_{m=1}^{M} (X_m - \bar{X})^2. \tag{4}$$

The coefficient of variation is defined as

$$CoV = \frac{\sigma}{\bar{X}} \tag{5}$$

and is widely employed as the measure of the variability of the frame sizes, that is, 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 VD curve [7].

## Bit-Rate Variability of H.264/AVC vs MPEG-4 Part 2

We first study the bit rate VD relationship of the H.264/AVC encoder using the Main profile. We choose the H.264/AVC encoder settings such that the bit-rate distortion is optimized, and we compare the resulting bit-rate variability with that of the MPEG-4 Part 2 encoder using the advanced simple profile (ASP). We will demonstrate that the rate variability of H.264/AVC video traffic is substantially higher. The reason for this increased traffic variability with H.264/AVC are the improved compression tools of H.264/AVC. When we disable key new H.264/AVC encoding tools to obtain equivalent encoding mechanisms as employed by MPEG-4 Part 2 ASP, we observe a sharp drop in rate variability to the level of MPEG-4 Part 2.

## ENCODING SET UP

For the initial VD comparison between H.264/AVC and MPEG-4 Part 2, we present encodings over a large bit-rate range for the *Sony Demo* and *Silence of the Lambs* sequences. We employ the H.264/AVC encoder in the Main profile with all compression tools enabled, as specified earlier, that is, using variable block sizes, three reference frames from the past and the future, referenced B frames, P and B frame-weighted prediction, CABAC, and rate-distortion optimization (RDO). We designate these settings as "Full-RDO." We also encode with RDO disabled, denoted as "Full-noRDO."

We used the MPEG-4 Part 2 encoder with the ASP profile, which adds B frames to the *Simple* profile, as well as quarter-pixel (sample) accurate MCP. Quarter pixel MCP refines motion vectors that are estimated with half-pixel accuracy in the Simple profile, to quarter-pixel accuracy. Half (resp. quarter)-pixel accurate MCP allows motion vectors to point to blocks that are offset (interpolated) by a half-pixel (resp. quarter pixel) distance from the pixels of a reference video frame. We do not employ RDO with the MPEG-4 Part 2 encoder. The settings with half-pixel (resp. quarter) accurate MCP are designated as "ASP-Hpel" (resp. "ASP-Qpel"). For comparison purposes, we also encode with full-pixel accuracy, and we refer to these settings as "ASP-Fpel," meaning that no pixel interpolation is performed. The MPEG-4 Part 2 encoder with the ASP-Fpel, ASP-Hpel, and ASP-Qpel settings uses one reference frame for the past and the future respectively, and $16 \times 16$ blocks (pixels) for MCP that are potentially split into $8 \times 8$ blocks by the MCP process.

Additionally, we switch off some key new H.264/AVC encoding tools, and we refer to these encoding settings as "Sparse." The Sparse encodings are obtained with the CAVLC entropy coder, only one reference frame for the past and the future, only MCP block sizes $16 \times 16$ and $8 \times 8$ are used, no referenced B frames, no weighted prediction, and no RDO. We distinguish between two Sparse encoding settings: with quarter-pixel accurate MCP, denoted by "Sparse-Qpel," and with full-pixel accurate MCP, denoted by "Sparse-Fpel." The H.264/AVC reference implementation does not support half-pixel accuracy.

For all these encodings, the group of pictures (GoP) structure is set to *IBBPBBPBBPBB* (12 frames, with two B frames per I/P frame), which we denote by G12-B2.

## RESULTS AND DISCUSSION

The RD graphs obtained for the CIF resolution *Sony Demo* and *Silence of the Lambs* sequences are depicted in Fig. 1a and Fig. 1c. We observe that the RD results for H.264/AVC with Full-RDO are a clear improvement over all MPEG-4 Part 2 RD results. We also provide the H.264/AVC Full-noRDO curve, as it allows for an interesting comparison with the MPEG-4 Part 2 RD curves. When the RDO feature is disabled, the H.264/AVC encoder still outperforms all MPEG-4 Part 2 RD results by a large bit-rate margin. Overall, the bit rate savings with H.264/AVC vary roughly from more than 50 percent in the low quality range to more than 30 percent in the high quality range compared to the best MPEG-4 Part 2 RD results for these two sequences.

The RD properties of both encoders already were elaborately studied, for example, in [16]. Conversely, this study focuses on the VD properties. Therefore, we depict the corresponding VD graphs in Fig. 1b and Fig. 1d. We observe that the VD curve values are significantly higher for H.264/AVC with Full-RDO than the values of the MPEG-4 Part 2 VD curves, especially in the low-to-medium quality range. For the H.264/AVC encodings with Full-noRDO, the rate variability drops slightly compared to when using RDO, but the rate variability is still significantly higher than for the MPEG-4 Part 2 encodings.

Where does this substantial variability increase with H.264/AVC stem from? By conducting two encoding experiments, we demonstrate that the new and improved MCP tools of H.264/AVC are mainly responsible. We switch off H.264/AVC tools and encode with the Sparse-Qpel and Sparse-Fpel settings. These settings employ comparable MCP tools for both H.264/AVC and MPEG-4 Part 2, that is, similar variable block sizes, pixel accuracy for MCP, and number of reference frames.

First, we discuss the RD results of the experiments. The H.264/AVC Sparse-Qpel RD curves represent a significant drop in RD efficiency compared to the Full-RDO and Full-noRDO curves, but there is still a large improvement over all MPEG-4 Part 2 RD curves, including ASP-Qpel. When quarter-pixel accurate MCP is switched off and full-pixel accuracy is used for H.264/AVC (Sparse-Fpel), the RD efficiency drastically drops and becomes comparable to the MPEG-4 Part 2 ASP-Fpel RD curves. We also note that on the scale of the RD curves in Fig. 1a and Fig. 1c, the difference in RD efficiency between ASP-Qpel, ASP-Hpel, and ASP-Fpel is small. Overall, the RD analysis illustrates the importance of the improved MCP tools of H.264/AVC for outperforming the MPEG-4 Part 2 encoder because with the full-pixel accurate MCP tool configuration, the RD efficiency of both encoders is generally equivalent. The RD analysis also indicates that the quarter-pixel MCP tool of H.264/AVC achieves a significant RD improvement compared to the quarter-pixel MCP of MPEG-4 Part 2.

Turning to the VD curves, we observe sharp drops of the rate variability for H.264/AVC when using the Sparse-Qpel and Sparse-Fpel settings, compared to the Full-RDO encodings. The maximum variability of H.264/AVC Sparse-Fpel is comparable to the maximum of the MPEG-4 Part 2 ASP-Fpel VD curves for both sequences. This indicates that the large differ-

*Overall, the RD analysis illustrates the importance of the improved MCP tools of H.264/AVC for outperforming the MPEG-4 Part 2 encoder because with the full-pixel accurate MCP tool configuration, the RD efficiency of both encoders is generally equivalent.*

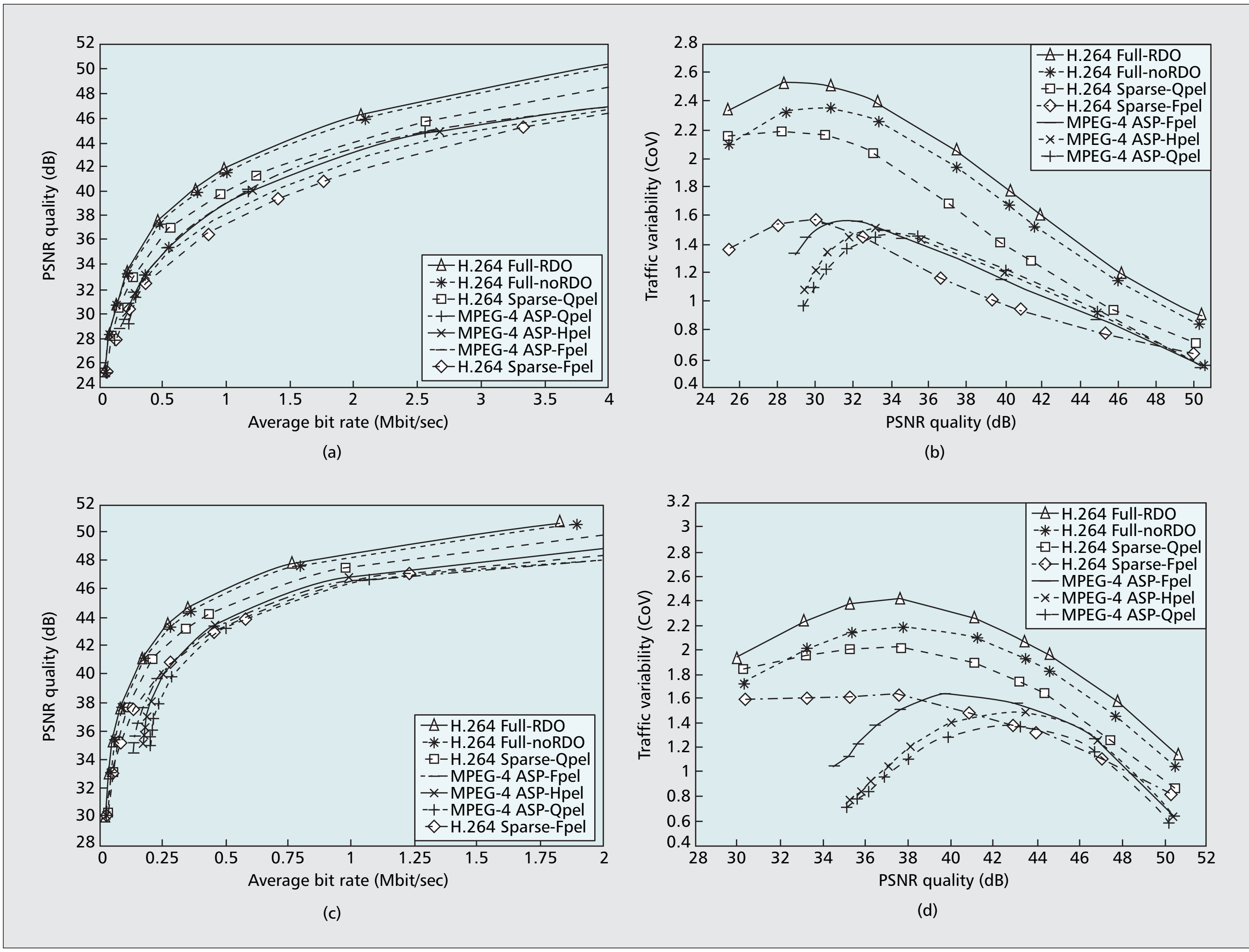
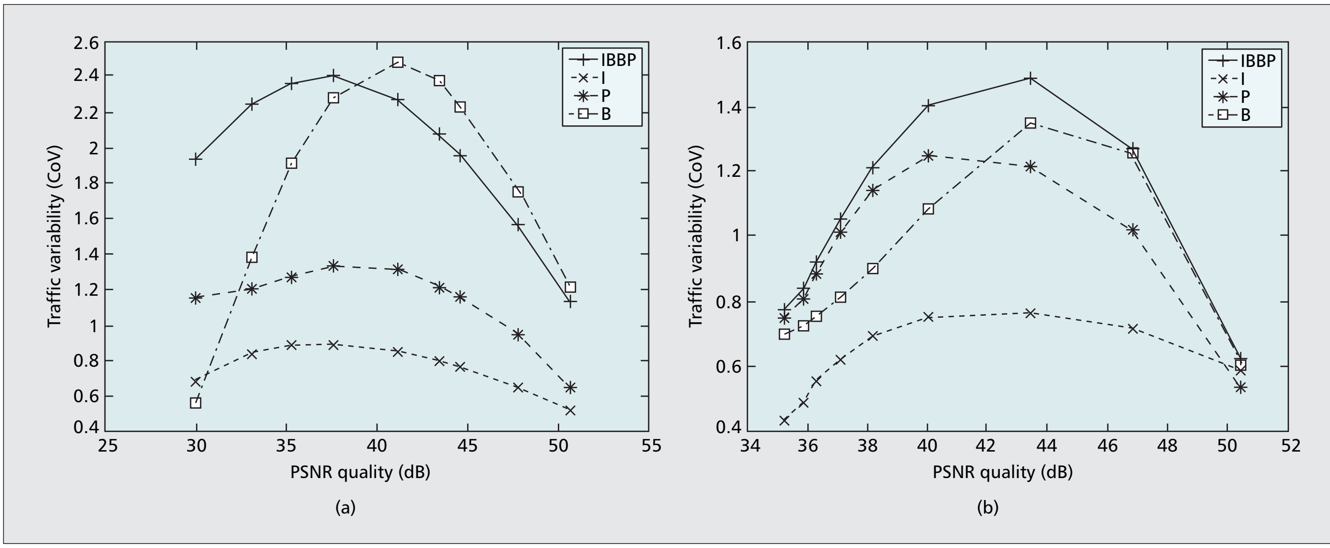



**Figure 1.** *RD and rate VD characteristics for CIF* Sony Demo *and* Silence of the Lambs *encoded with H.264/AVC with and without new encoding tools and MPEG-4 Part 2 with range of pixel accuracy in MCP: a) RD graph for* Sony Demo*; b) VD graph for* Sony Demo*; c) RD graph for* Silence of the Lambs*; d) VD graph for* Silence of the Lambs.

ence in rate variability between H.264/AVC and MPEG-4 Part 2 is due to the improved compression tools of H.264/AVC. The improved MCP tools, in particular, including the variable block sizes and multiple reference frames allowed in H.264/AVC, generally result in smaller P and B frames compared to the MPEG-4 Part 2 encoded P and B frames. Because the compression improvement of the I frames (due to spatial intra-prediction) is relatively smaller, the combination of all compression tool improvements results in the observed higher bit-rate variability for the H.264/AVC encoder. Due to space restrictions, we cannot show the impact of switching off intra-prediction. It results in a slightly increased rate variability because the I frame sizes increase compared to the other frame types.

When comparing the RD and VD curves of the above encoding settings for both encoders, we observe a direct relationship between increasing the RD efficiency and an increase in rate variability. For instance, as the RD efficiency of the H.264/AVC codec increases from Sparse-Fpel via Sparse-Qpel and Full-noRDO to Full-RDO, the rate variability increases. Even slight RD curve differences are represented by VD curve differences but somewhat amplified. In subsequent experiments with the H.264/AVC Main profile encodings, we apply the Full-RDO encoding options. For the MPEG-4 Part 2 encodings, we employ the ASP-Hpel settings because there are only relatively small bit rate savings (if any) associated with quarter-pixel accuracy, whereas the encoding times are increased.

To provide further insight into the VD behavior, in Fig. 2 we plot the VD curves separately for I frames, P frames, and B frames, as well as the overall sequence for *Silence of the Lambs* encoded using H.264/AVC (Full-RDO) and MPEG-4 Part 2 (ASP-Hpel) with GoP structure *G12-B2*. We observe from the figure that for both encoders, the B frames have the highest variability, followed by the P frames and the I frames. We also observe that the variability of the overall frame sequence (`IBBP`) is dominated by the B frame variability. The dominating effect of the B frame variability can be explained by the high number of B frames in the *G12-B2* GoP structure, which has eight B frames out of a total of 12 frames, and by the general analytical relationship between the I, P, and B frame variabilities and the sequence variability derived in [9].




■ **Figure 2.** *VD curves for I, P, and B frames, and overall sequence (*`IBBP`*), for* Silence of the Lambs *encoded with a) H.264/AVC "Full-RDO"; b) MPEG-4 Part 2 "ASP-Hpel."*

Comparing Fig. 2a and Fig. 2b, we observe that whereas the I and P frame variabilities are only somewhat higher with H.264/AVC, the B frame variability is sharply higher with H.264/AVC compared to MPEG-4 Part 2. This indicates that the improved MCP tools of H.264/AVC have a particularly profound effect on the B frames. The corresponding RD results underscore this strong effect of H.264/AVC on the B frames: For an overall average PSNR frame quality of 43.5 dB, the mean sizes for I, P, and B frames are 5658 bytes, 1634 bytes, and 348 bytes, respectively, for H.264/AVC — compared to 6721 bytes, 2234 bytes, and 1186 bytes for MPEG-4 Part 2. For an average PSNR of 35.2 dB, the mean sizes for I, P, and B frames are 1214 bytes, 279 bytes, and 53 bytes for H.264/AVC — compared to 1614 bytes, 744 bytes, and 604 bytes for MPEG-4 Part 2. These RD results in conjunction with the VD results in Fig. 2 indicate that the improved coding tools in H.264/AVC give significantly higher compression ratios and higher variabilities for I and P frames compared to MPEG-4 Part 2; yet, the B frames experience higher compression ratio gains and variability increases by far with H.264/AVC.

## Frame Size Smoothing

To mitigate the effect of variable video frame sizes on network transport, a wide variety of frame size smoothing mechanisms were developed and studied in the context of the MPEG-4 Part 2, H.263, and preceding codecs [10]. In this section, we examine the fundamental impact of frame size smoothing on H.264/AVC 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 network. Given the original (unsmoothed) frame size sequence $X_m$, $m = 1, \ldots, M$, we obtain the smoothed frame sizes

$$Y_n = \frac{1}{a} \sum_{m=(n-1)a+1}^{na} X_m \tag{6}$$

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

### Encoding Set Up

In the subsequent encoding experiments, we employ two different GoP structures, namely *IBPBPBPBPBPBPBPB* (16 frames, with one B frame per I/P frame), denoted by *G16-B1*, and *IBBBPBBBPBBBPBBB* (16 frames, with three B frames per I/P frame), denoted by *G16-B3*. These two GoP structures are employed for the *Sony Demo* CIF sequence encodings and the *Silence of the Lambs* CIF encodings. The RD graphs for both sequences are depicted in Fig. 3a and Fig. 3c. The significant rate-distortion efficiency improvement of the H.264/AVC encoder over the MPEG-4 Part 2 encoder observed in the preceding section also is apparent here.

### Results and Discussion

To illustrate the effect of frame size smoothing on the bit-rate variability, we plotted the VD curves of both the unsmoothed and the smoothed (denoted by sm in the figures) H.264/AVC and MPEG-4 Part 2 video traffic in Fig. 3b and Fig. 3d. The *G16-B1* traffic is smoothed over $a = 2$ frames and the *G16-B3* type traffic is smoothed over $a = 4$ frames. We observe that the bit-rate variability of the smoothed H.264/AVC video traffic is significantly higher or comparable to the rate variability of the unsmoothed MPEG-4 Part 2 over a wide PSNR range, such as in Fig. 3b, over the full PSNR range, and in Fig. 3d, from small PSNR values until about 41dB. Throughout, the smoothed H.264/AVC video traffic is much more variable than the smoothed MPEG-4 Part 2 video traffic.

These encoding results (along with more extensive experiments in [15]) illustrate the sig-

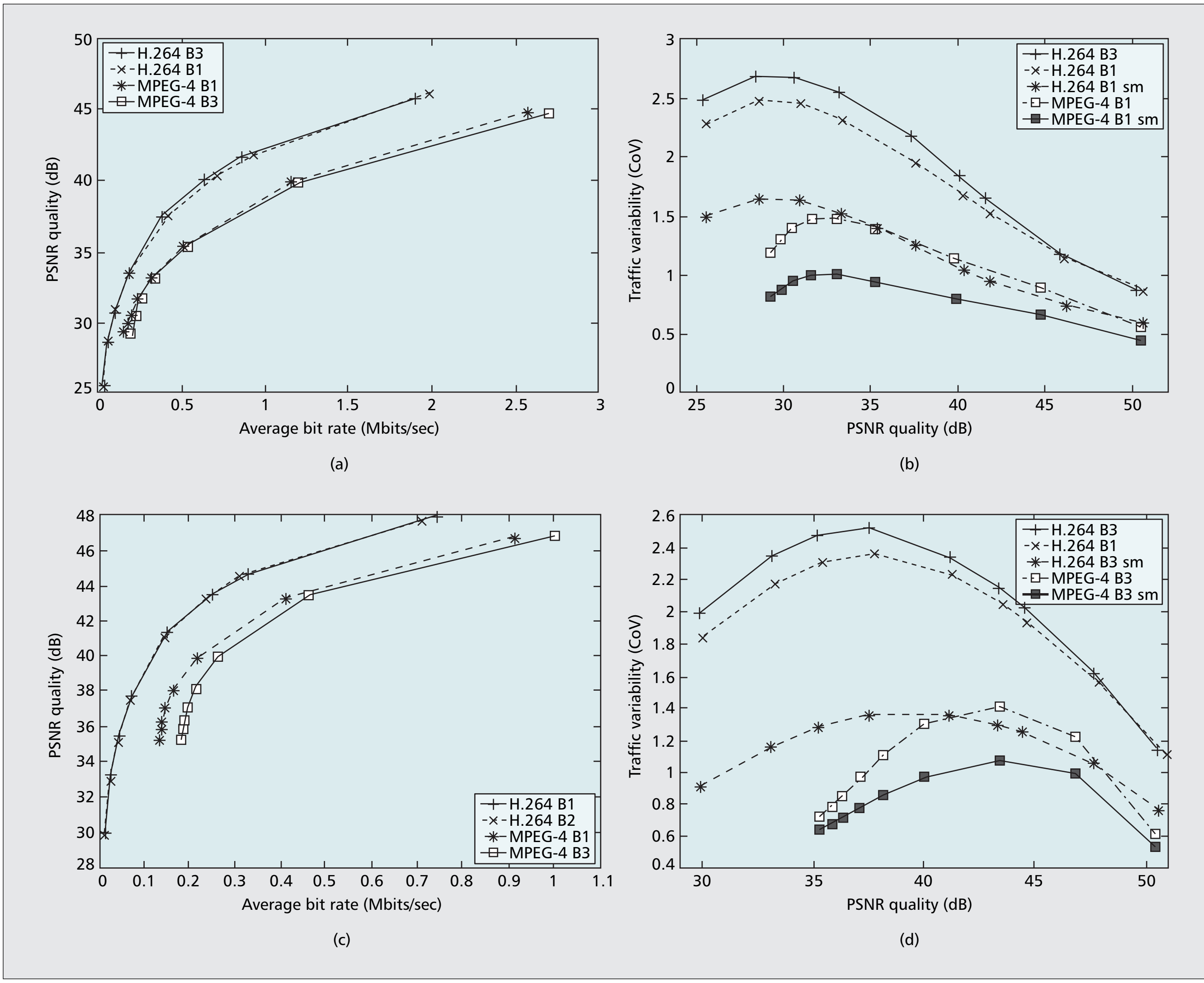



**Figure 3.** *RD and VD graphs encoded with H.264/AVC "Full-RDO" and MPEG-4 Part 2 "ASP-Hpel" with GoP structures G16-B1 and G16-B3 without and with frame size smoothing (*`sm`*): a) RD graph for CIF* Sony Demo *sequence; b) VD graph for CIF* Sony Demo *sequence; c) RD graph for CIF* Silence of the Lambs*; d) VD graph for CIF* Silence of the Lambs.

nificantly higher bit-rate variability of H.264/AVC 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 encoded video.

## Long-Range Dependence

It is well-known that long-range dependence in video traffic can have a significant impact on the performance of packet-switched networks [11]. The losses and delays of queuing systems are considerably larger for video traffic with a high degree of long range dependence than for traffic with low long-range dependence. Intuitively, long-range dependent traffic is bursty (highly variable) over a wide range of timescales.

The Hurst parameter is a metric for the degree of long-range dependence [8]. In general, time series without long-range dependence have a Hurst parameter of 0.5. Hurst parameters between 0.5 and 1.0 indicate long-range dependence, with larger Hurst parameters indicating a higher degree of long-range dependence. We estimate the Hurst parameters of the video traffic from *pox* diagrams of the R/S statistic [8]. For each frame size sequence, we generate pox diagrams of R/S for different aggregation levels $a$, that is, we average the frame sizes over non-overlapping blocks of $a$ frames and then plot the pox diagram of R/S. Hurst parameters larger than 0.5 for all aggregation levels are a strong indication of long-range dependence.

Table 1 presents the Hurst parameters estimated from the H.264/AVC and MPEG-4 Part 2 encodings of the *Silence of the Lambs* sequence (G16-B3 GoP) for approximately equal average high qualities (quantization scales $QP$ = 24 for H.264/AVC and $q$ = 4 for MPEG-4 Part 2) and low qualities ($QP$ = 38 for H.264/AVC and $q$ = 28 for MPEG-4 Part 2). The table covers aggregation levels ranging from $a$ = 1 to 800 frames. Fig. 4 depicts the pox diagrams for the aggregation level $a$ = 48 for the high qualities.

| (a) H.264/AVC | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Agg. Level *a* | 1 | 16 | 32 | 48 | 96 | 192 | 304 | 400 | 496 | 608 | 704 | 800 |
| $QP = 24$ | 0.890 | 0.881 | 0.868 | 0.876 | 0.858 | 0.848 | 0.860 | 0.898 | 0.856 | 0.852 | 0.813 | 0.872 |
| $QP = 38$ | 0.840 | 0.880 | 0.868 | 0.863 | 0.863 | 0.840 | 0.851 | 0.894 | 0.874 | 0.871 | 0.866 | 0.883 |
| (b) MPEG-4 Part 2 | | | | | | | | | | | | |
| Agg. Level *a* | 1 | 16 | 32 | 48 | 96 | 192 | 304 | 400 | 496 | 608 | 704 | 800 |
| $q = 4$ | 0.936 | 0.897 | 0.889 | 0.893 | 0.874 | 0.885 | 0.888 | 0.921 | 0.919 | 0.900 | 0.880 | 0.896 |
| $q = 28$ | 0.885 | 0.847 | 0.829 | 0.849 | 0.826 | 0.822 | 0.796 | 0.822 | 0.763 | 0.807 | 0.852 | 0.796 |

**Table 1.** *Hurst parameters for* Silence of the Lambs *encoded with H.264/AVC and MPEG-4 Part 2 as a function of aggregation level* a *(in frames) and quantization parameter of encoding.*

We observe from the tables that the *Silence of the Lambs* encodings with H.264/AVC and MPEG-4 Part 2 have similar large values (> 0.75) for all aggregation levels. This indicates a high degree of long-range dependence. It is interesting to note that the Hurst parameter estimates are similar for both encoders despite the improved performance and higher rate variability of the H.264/AVC encoder compared to MPEG-4 Part 2.

This similarity of the long-range dependence properties may be due to the fact that the new coding mechanisms responsible for the increased compression gains in H.264/AVC operate primarily on a time scale on the order of tens of frames, namely, seconds of video run time. Thus, the traffic characteristics over very long time scales, say hundreds or thousands of video frames, or equivalently, minutes or tens of minutes of video run time, which govern to a large extent the long-range dependence properties, may not be affected significantly.

The technical report [15] contains a more extensive long-range dependence analysis of all encodings, incorporating Hurst parameters estimates using pox diagrams of the R/S statistics, periodograms, and variance time plots. The long-range dependence properties appear consistently strong for all quality levels (determined by quantization parameter) of the videos.

## High Definition Video

In this section, we examine the traffic of HD video encoded with the H.264/AVC encoder with FRExt [2, 3] for the purposes of:

- Comparing the H.264/AVC FRExt traffic with the traffic of MPEG-2 encoded HD video
- Exploring the scaling of video traces of CIF format video to traces of HD video

Video traces, which are widely used in studies on video network transport and in video traffic modeling are files containing video frame timestamps, frame types (e.g., I, P, or B), encoded frame sizes (in bits), and frame qualities (PSNR) [12]. The motivation for exploring the upscaling of video traces is that small video formats, such as CIF, are encoded relatively fast, whereas HD video requires very long encoding times, limiting the generation of HD video traces for networking studies.

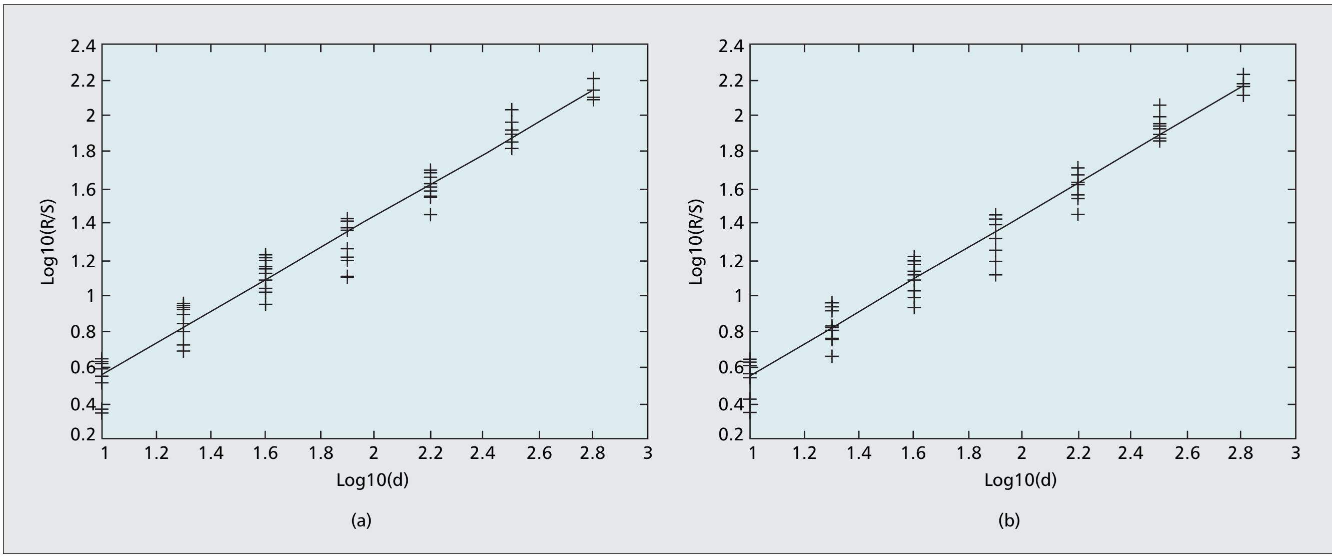
### Encoding Set Up

For HD video encoding, we employ the H.264/AVC encoder with FRExt [2, 3] to optimally compress the high definition video footage. The profile is set to High, the number of reference frames is set to two for both the past and the future, fast rate-distortion optimization is enabled, P- and B-weighted prediction is disabled, referenced B pictures is disabled, and the CABAC arithmetic coder is chosen. Our encoding tests indicate that more reference frames do not significantly improve compression performance for the *Sony Demo* sequence, but significantly increase encoding time. The employed GoP structure is *G12-B2*, namely, *IBBPBBPBBPBB*.

Since most of the legacy HD video is currently encoded in MPEG-2, we employ the *FFmpeg* MPEG-2 encoder implementation (*mpeg2video* setting) with GoP structure *G12-B2* to encode the HD sequence for comparison with H.264/AVC FRExt.

### Results and Discussion

The rate-distortion and the rate variability-distortion graphs for the *Sony Demo* and *Terminator 2* sequences are depicted in Fig. 5. The encoding results for these HD sequences with the H.264/AVC FRExt and MPEG-2 encoders show interesting distinctions between the two encoders. The bit rates obtained with the H.264/AVC FRExt encoder are clearly much smaller than those obtained with the MPEG-2 encoder. Also the rate variability is significantly different for both encoders. The rate variability without smoothing is up to two times higher for the H.264/AVC FRExt encoder than for MPEG-2. Smoothing over $a = 3$ frames reduces the traffic variability significantly, but the smoothed H.264/AVC FRExt traffic has much higher variability than the smoothed MPEG-2 traffic. These observations are consistent with our earlier observations from the CIF encoding experiments.




■ **Figure 4.** *Pox diagrams of R/S with aggregation level* a = *48 for* Silence of the Lambs *G16-B3 encodings: a) H.264/AVC,* QP = *24,* H = *0.876; b) MPEG-4 Part 2,* q = *4,* H = *0.893.*

## Investigation of Obtaining HD Video Traces Through Scaling

High definition frame size video traces of these two sequences are available in our video trace library at http://trace.eas.asu.edu. However, the encoding times on a contemporary PC are extremely long, limiting the generation of a large set of HD video traces that would be required for network simulations [12]. Therefore, we investigate if a simple relationship exists between the frame sizes (in bits) of the encoded HD video and the frames sizes of the corresponding video when downsampled to CIF resolution and then encoded.

Since similarly high bit-rate variabilities are obtained for the HD resolution as for the CIF resolution, one might be tempted to upscale CIF video frame sizes (in bits), encoded with H.264/AVC using the Main profile, to HD video frame sizes by multiplying with the factor obtained by dividing the HD resolution by the CIF resolution. This way, HD frame size video traces could be obtained with less computational effort because only the CIF resolution video would be required to be encoded, which requires significantly less computation time than HD video encoding. From a purely mathematical perspective, this scaling would leave the coefficient of variation unchanged because both the standard deviation and the mean are scaled by the same value. Although this may seem a simple solution, enabling the reuse of CIF video traces, the reality of frame-size scaling is much more complex.

In Fig. 6, we depict the histograms of the real scaling factors for the case where the *Sony Demo* and *Terminator 2* CIF sequence frame sizes, encoded with H.264/AVC in the Main profile with GoP structure *G12-B2* and quantization parameter *QP* = 24, are compared to the corresponding HD sequence frame sizes, encoded in the High profile employing the same GoP structure and quantization parameter *QP* = 28. We chose these quantization parameters because they have rate variabilities that are very close.

We conclude from the histograms of scaling factors in Fig. 6 that the actual scaling factors are spread over a wide range and are far from the theoretical value of 9.09 suggested by the ratio of the HD resolution (1280 × 720 pixels) to the CIF resolution. For the *Sony Demo* sequence, the actual average scaling factor is 5.4, and the maximum actual scaling factor is as large as 493. For *Terminator 2*, the actual average scaling factor is 2.9, and the maximum is 188. This deviation from the theoretical scaling factor is caused by differences in coding tools enabled by both H.264/AVC profiles, as well as video content detail differences between both resolutions. This observation illustrates the necessity of encoding actual HD sequences or the necessity of building a complex frame-size scaling model to obtain traces of HD video for network performance studies.

## Conclusions

We examined the network traffic characteristics of variable bit rate H.264/AVC encoded video. We focused on a long test video sequence with a wide range of typical texture and motion features and a long excerpt of the thriller movie *Silence of the Lambs* in this article but have found similar characteristics for long sequences from the following genres: science fiction movies, action movies, and sports videos [15]. In summary, we found the following distinct characteristics of the H.264/AVC video traffic:

- We confirm that for a fixed, desired video quality, the H.264/AVC encoder cuts the average bit rate typically to up to a half of the average bit rate achieved by the older MPEG-2 and MPEG-4 Part 2 encoders. This underscores the significant improvements in coding technology of H.264/AVC over the older standards and likely will drive the popularity of the H.264/AVC standard for video streaming over bandwidth constrained networks.
- The variability of the H.264/AVC video traffic is significantly higher than the traffic variabil-

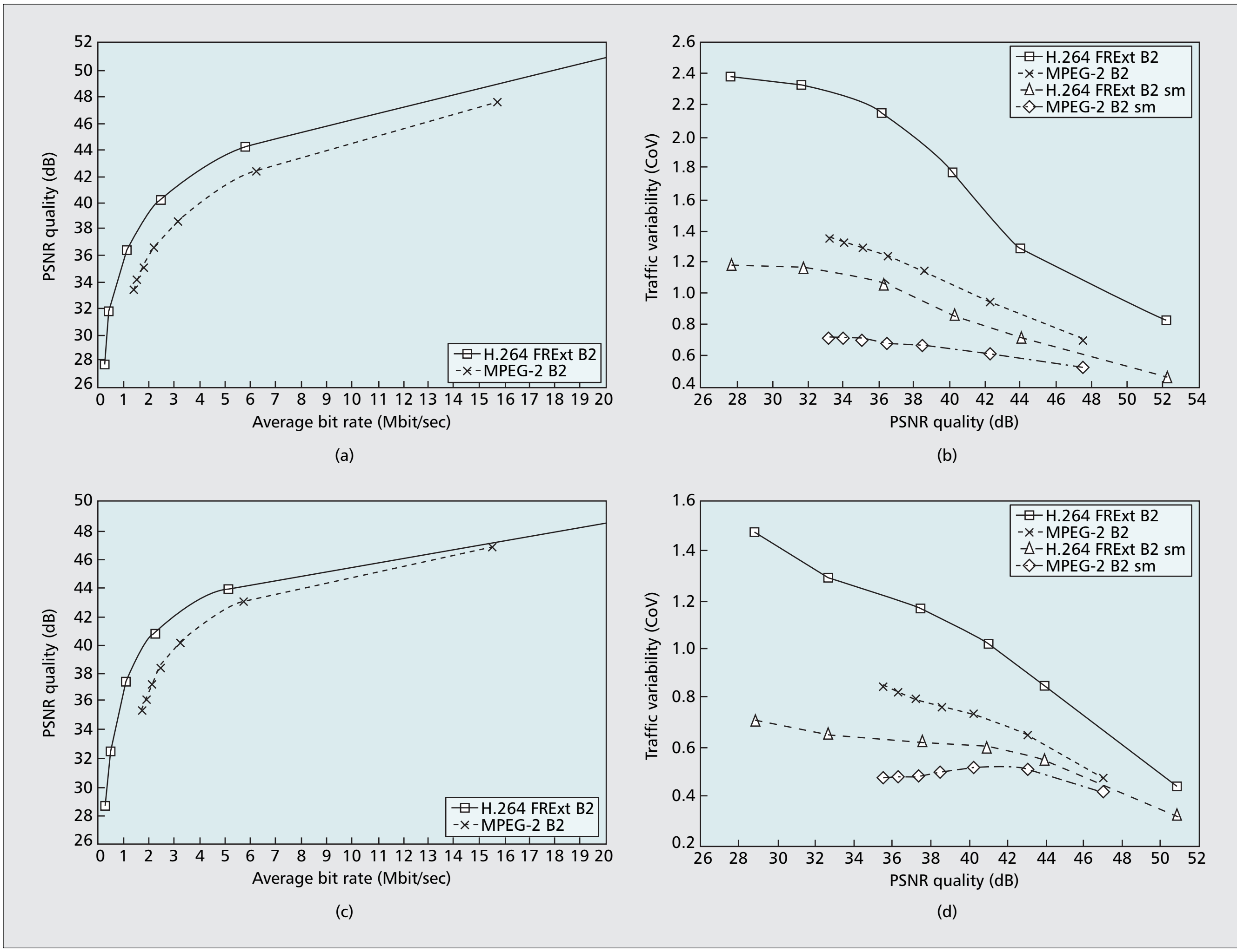



**Figure 5.** *RD and VD graphs for two 10 min HD sequences encoded with H.264/AVC FRExt and MPEG-2 with and without smoothing: a) RD graph for* Sony Demo*; b) VD graph for* Sony Demo*; c) RD graph for* Terminator 2*; and d) VD graph for* Terminator 2*.*

ity of the older MPEG codecs, as demonstrated for MPEG-2-encoded HD video and MPEG-4 Part 2-encoded CIF video. Whereas the CoV (standard deviation normalized by mean) of the frame sizes reaches levels above 2.4 for H.264/AVC, it does not exceed 1.5 with MPEG-4 Part 2. The levels of the CoV of the frame size above 1.5 are unprecedented; with MPEG-4 Part 2, the CoV did not exceed 1.5, and the levels were typically in the range from 0.9 to 1.4 [7, 12].

• 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 [10]. We observed that the smoothed H.264/AVC video traffic can exhibit 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 video must be designed to accommodate substantial traffic variabilities.

• The long-range dependence characteristics of the H.264/AVC video traffic are similar to the long-range dependence characteristics of MPEG-4 Part 2 encoded video.

There are several directions for important work in the future. One direction is to examine the suitability of existing traffic models and video transport mechanisms for H.264/AVC video traffic. The existing traffic models, such as [17], and video transport mechanisms for a wide range of communication networks, including general IP networks, wireless networks, and peer-to-peer networks, were primarily developed based on MPEG-4 Part 2 video traffic [4, 5]. Therefore, it is 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 video traffic. If necessary, the existing traffic models and transport mechanisms must be extended to accommodate the unprecedented variability of the H.264/AVC video traffic. Another important direction is to study the traffic characteristics of scalable encoded H.264/AVC video. The H.264/AVC codec and its extensions provide a number of novel scalability paradigms, such as combined temporal-spatio-SNR scalability. The impact of these novel scalability techniques on video transport over communication networks is largely uncharted territory.

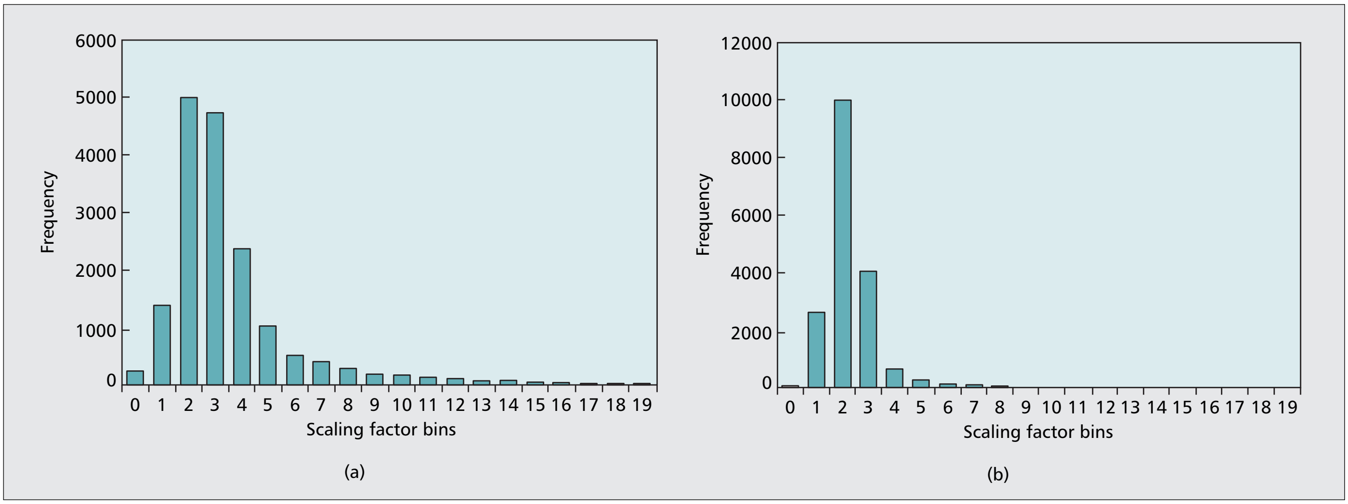



**Figure 6.** *CIF-to-HD frame size scaling factor histograms: a)* Sony Demo*; b)* Terminator 2.

## REFERENCES

[1] D. Marpe, T. Wiegand, and G. Sullivan, "The H.264/MPEG4 Advanced Video Coding Standard and Its Applications," IEEE Commun. Mag., vol. 44, no. 8, Aug. 2006, pp. 134–43.
[2] D. Marpe, T. Wiegand, and S. Gordon, "H.264/MPEG-4 AVC Fidelity Range Extensions: Tools, Profiles, Performance, and Application Areas," Proc. IEEE Int'l. Conf. Image Proc. (ICIP), Genoa, Italy, Sept. 2005, pp. 593–96.
[3] G. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions," Proc. SPIE 5558, Conf. Applications of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard: H.264/AVC I, Denver, CO, Aug. 2004, pp. 454–74.
[4] T. Lakshman, A. Ortega, and A. R. Reibman, "VBR Video: Trade-Offs and Potentials," Proc. IEEE, vol. 86, no. 5, May 1998, pp. 952–73.
[5] A. R. Reibman and M. T. Sun, Compressed Video over Networks, Marcel Dekker, New York, 2000.
[6] D. Wu et al., "Streaming Video over the Internet: Approaches and Directions," IEEE Trans. Circuits and Systems for Video Technology, vol. 11, no. 3, Mar. 2001, pp. 282–300.
[7] 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, Dec. 2005, pp. 473–92.
[8] J. Beran, Statistics for Long-Memory Processes, Chapman & Hall, 1994.
[9] 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, Sept. 2007, pp. 637–48.
[10] 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, Sept. 1999, pp. 302–12.
[11] J. Beran et al., "Long-Range Dependence in Variable-Bit-Rate Video Traffic," IEEE Trans. Commun., vol. 43, no. 2–4, Feb.–Apr. 1995, pp. 1566–79.
[12] 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 Commun. Surveys and Tutorials, vol. 6, no. 3, 2004, pp. 58–78; video traces; http://trace.eas.asu.edu
[13] J. Ostermann *et al.*, "Video Coding with H.264/AVC: Tools, Performance and Complexity," IEEE Circuits and Systems Mag., vol. 4, no. 1, 2004, pp. 7–28.
[14] A. Puri, X. Chen, and A. Luthra, "Video Coding Using the H.264/MPEG-4 AVC Compression Standard," J. Visual Communication and Image Representation, vol. 19, no. 9, Oct. 2004, pp. 793–849.
[15] G. Van der Auwera, P. David, and M. Reisslein, "Video Traffic Analysis of H.264/AVC and Extensions: Single-Layer Statistics," ASU, Department of Electrical Engineering, Tempe, AZ, Tech. Rep., Feb. 2007; http://trace.eas.asu.edu/publications/h264trafficsinglelayerext.pdf
[16] T. Wiegand et al., "Rate-Constrained Coder Control and Comparison of Video Coding Standards," IEEE Trans. Circuits and Systems for Video Tech., vol. 13, no. 7, July 2003, pp. 688–703.
[17] M. Dai and D. Loguinov, "Analysis and Modeling of MPEG-4 and H.264 Multi-Layer Video Traffic," Proc. IEEE INFOCOM, Miami, FL, Mar. 2005, pp. 2257–67.

## BIOGRAPHIES

GEERT VAN DER AUWERA (geert.vanderauwera@asu.edu) received a Ph.D. degree in electrical engineering from Arizona State University, Tempe, in 2007, and an M.S.E.E. degree from Vrije Universiteit Brussel (VUB), Belgium, in 1997. His research interests are video traffic and quality characterization, video streaming mechanisms and protocls, and video coding. Until the end of 2004 he was a scientific advisor for IWT-Flanders, the Insitute for the Promotion of Innovation by Science and Technology, Flanders, Belgium. In 2000 he joined IWT-Flanders after researching wavelet video coding at the Electronics adn Information Processing Department, VUB. In 1998 his thesis on motion estimation in the wavelet domain received the Barco and IBM prizes from the Fund for Scientific Research of Flanders.

PRASANTH T. DAVID (prasanth.david@asu.edu) received his B.Tech degree from the University of Kerala, Trivandrum, India in 2001. From 2001 to 2005 he was a systems design engineer at 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 Department at Arizona State University, from which he received his M.S. degree. Since then he has been with Intel Corporation, California. His fields of interest are multimedia coding, media processors, and video streaming technologies.

MARTIN REISSLEIN (reisslein@asu.edu) received his Ph.D. in systems engineering in 1998 from the University of Pennsylvania in Philadelphia. He is an associate professor in the Department of Electrical Engineering at Arizona State University, Tempe. From July 1998 through October 2000, he was a scientist with the German National Research Center for Information Technology (GMD FOKUS), Berlin and a lecturer at the Technical University Berlin. 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.