H.264 Coarse Grain Scalable (CGS) and Medium Grain Scalable (MGS) Encoded Video: A Trace Based Traffic and Quality Evaluation

Rohan Gupta, Akshay Pulipaka, Patrick Seeling, Lina J. Karam, and Martin Reisslein

***Abstract*—The scalable video coding (SVC) extension of the H.264/AVC video coding standard provides two mechanisms, namely coarse grain scalability (CGS) and medium grain scalability (MGS), for quality scalable video encoding, which varies the fidelity (signal-to-noise ratio) of the encoded video stream. As H.264/AVC and its SVC extension are expected to become widely adopted for the network transport of video, it is important to thoroughly study their network traffic characteristics, including the bit rate variability. In this paper, we report on a large-scale study of the rate-distortion (RD) and rate variability-distortion (VD) characteristics of CGS and MGS. We found that CGS achieves low bit rate overheads in the 10–30% range compared to H.264 SVC single-layer encodings only for encodings with a total of up to three quality levels; more quality levels result in substantially higher overheads. The traffic variabilities of CGS are generally lower than for single-layer streams. We found that in the low to mid range of the MGS quality scalability, MGS can achieve the same or even slightly higher RD efficiency than corresponding single-layer encoding; toward the upper end of the MGS quality scalability range the RD efficiency drops off significantly. MGS layer extraction following the hierarchical B frame structure gives nearly as high RD performance as RD-optimized extraction. In the range of high RD efficiency, MGS streams have significantly higher traffic variabilities than single-layer streams at the frame time scale. At the group-of-pictures (GoP) time scale, MGS has similar or lower levels of traffic variability compared to single-layer streams. Generally, MGS layer extraction over the time horizon of individual GoPs gives significantly lower traffic variability than extraction over the time horizon of the full video sequence.**

***Index Terms*—Coarse grain scalability, H.264 SVC, medium grain scalability, rate-distortion, rate variability-distortion, traffic variability.**

I. Introduction

THE flexible adaptation of video traffic bit rates benefits many video transport systems, including IPTV systems [1]–[3], satellite distribution systems [4], [5], and wireless networks [6]–[13]. The scalable video coding (SVC) extension [14] of the H.264/AVC video coding standard seeks to fulfill the need for flexible rate adaptation through temporal, spatial, and quality scalability modes. The traffic characteristics of the temporal and spatial scalability modes have been examined in [15] and we focus on the quality scalability in this article. (The study [15] briefly examined the traffic of the complete MGS enhancement layer, but did not consider the medium grain scalability achieved by partitioning the complete enhancement layer, which is the focus of the present MGS study.) The SVC scalability extension, which we refer to as H.264 SVC, provides two forms of quality scalability, namely coarse grain scalability (CGS) and medium grain scalability (MGS). In this article, we present traffic and quality analyses based on CGS and MGS encodings of 30-minute long videos from a wide range of content genres.

Generally, a thorough understanding of the traffic and quality characteristics of encoded video is the basis for traffic modeling and the development of video transport mechanisms. For MPEG-4 single-layer video and MPEG-4 scalable video (which was RD-inefficient) as well as single-layer H.264 video, extensive traffic modeling, see e.g., [16]–[26], and transport mechanism development, see for instance [27]–[31], have been conducted. Similarly, the network transport of H.264 SVC scalable video has begun to attract significant research interest, see for instance [1], [32]–[37]. A traffic model for H.264 SVC temporal scalability of the base layer and the complete enhancement layer has been proposed in [38]. Furthermore, an RD model of H.264 SVC quality scalability through dropping complete enhancement layers has been studied in [39], [40]. Similar to [15], the studies [38]–[40] did not consider medium grain scalability through partitioning of the complete enhancement layer.

To the best of our knowledge, no prior study of the traffic variability, which is a key concern for video transport [41], has been conducted for H.264 SVC quality scalable video. In this article, we report on a large-scale study of the fundamental RD performance and the traffic variability characteristics of H.264 SVC quality scalable video for long (over ten thousand frames) video sequences. We compare the RD performance and traffic variability of H.264 SVC CGS and MGS encodings with the corresponding H.264 SVC single-layer encodings. We note that the rate-distortion (RD) characteristics of H.264 SVC quality scalable encoded video has been examined in [14], [42], [43] for short video sequences up to a few hundred frames. In contrast, we examine both the RD and traffic variability

Manuscript received February 10, 2011; revised February 29, 2012; accepted March 15, 2012. Date of publication May 04, 2012; date of current version August 17, 2012. This work was supported in part by the National Science Foundation under Grant CRI-0750927.

R. Gupta was with Arizona State University. He is now with Qualcomm Inc., San Diego, CA 92121 USA (e-mail: Rohan.Gupta@asu.edu).

A. Pulipaka, L. J. Karam, and M. Reisslein are with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287-5706 USA (e-mail: Akshay.Pulipaka@asu.edu; karam@asu.edu; reisslein@asu.edu).

P. Seeling is with the Department of Computer Science, Central Michigan University, Mount Pleasant, MI 48859 USA (e-mail: pseeling@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2012.2191702

characteristics for long video sequences, which are needed for reliable evaluation.

All video traffic and quality data from this study are publicly available in the form of video traces [44] from the video trace library at http://trace.eas.asu.edu. A video trace characterizes an encoded video stream by providing time stamp, frame type (e.g., I, P, or B), frame size (in byte), and PSNR quality for each encoded frame (and layer of a scalable encoding). Video traces can be readily fed into simulation models of video transport systems; thus, facilitating the evaluation of novel transport mechanisms.

The paper is organized as follows. Section II gives a brief introduction to the quality scalable modes of H.264 SVC. Section III provides the evaluation set-up, including an overview of the video test sequences, the encoder settings, as well as video quality and traffic metrics. Sections IV and V present and discuss the CGS and MGS video quality and traffic characteristics. Section VI summarizes the article.

II. Overview of H.264 Quality Scalability

The SVC extension builds on the well-designed core coding tools of the H.264/AVC standard [45]–[47] by adding features for efficiently supporting scalability. Similar to H.264/AVC, H.264 SVC organizes the encoded video data into network abstraction layer units (NALUs). The bi-directionally predicted (B) frames in H.264 SVC have a hierarchical structure; whereby, the B frames in a given layer of the hierarchy form a temporal layer [14]. In particular, the I and P frames form the temporal base layer $T = 0$ and the β, $\beta = 2^{\tau} - 1$ B frames between successive I and P frames are organized into τ temporal enhancement layers, $T = 1, 2, \ldots, \tau$. Throughout, the B frames in a given temporal enhancement layers T are predictively encoded with respect to the frames in the lower temporal layers, i.e., the temporal base layer frames and the B frames in temporal enhancement layers $1, 2, \ldots, T-1$.

The lowest video quality that can be decoded in SVC is called the *base layer* (which can also be decoded by a non-scalable single layer decoder). Successive layers are referred to as *enhancement layers*. The process of encoding an enhancement layer from the lower layer(s) is referred to as *inter-layer prediction*. While H.264 SVC supports up to 128 layers, the actual number of layers in an encoding depends on the application needs. With the currently specified profiles, the maximum number of enhancement layers is limited to 47 layers [14].

In this study, we focus on the quality scalability in H.264 SVC. Quality scalable layers have the same spatio-temporal resolution but differ in fidelity. The H.264 SVC extension supports two quality scalable modes, namely coarse grain scalability (CGS) and medium grain scalability (MGS).

A. Overview of Coarse Grain Scalability (CGS)

Coarse grain scalability (CGS) can be viewed as a special case of spatial scalability in H.264 SVC, in that similar encoding mechanisms are employed but the spatial resolution is kept constant. More specifically, similar to spatial scalability, CGS employs inter-layer prediction mechanisms, such as prediction of macroblock modes and associated motion parameters and prediction of the residue signal [14]. CGS differs from spatial scalability in that the up-sampling operations are not performed. In CGS, the residual texture signal in the enhancement layer is re-quantized with a quantization step size that is smaller than the quantization step size of the preceding CGS layer. SVC supports up to eight CGS layers, corresponding to eight quality extraction points [48], i.e., one base layer and up to seven enhancement layers.

We use $B\_E_1\_E_2\_\ldots$ to denote the quantization parameter (QP) values of the base layer, first enhancement layer, second enhancement layer, and so on. Commonly, these QP values are equally spaced, and we define Delta QP (DQP) as $\mathrm{DQP} = B - E_1 = E_n - E_{n+1}$ for $n = 1, 2, \ldots$.

The current H.264 SVC software reference (JSVM 9.16) constrains the inter-layer prediction to three dependency layers [49], whereby one layer has to be the base layer. To improve the RD performance we have extended the reference software to provide inter-layer prediction for more than three dependency layers and report results for both the original and modified reference software in Section IV.

B. Overview of Medium Grain Scalability (MGS)

While CGS provides quality scalability by dropping complete enhancement layers, MGS provides a finer granularity level of quality scalability by partitioning a given enhancement layer into several MGS layers [14]. Individual MGS layers can then be dropped for quality (and bit rate) adaptation.

Splitting transform coefficients into MGS layers: Medium grain scalability (MGS) splits a given enhancement layer of a given video frame into up to 16 MGS layers (also referred to as quality layers). In particular, MGS divides the transform coefficients, obtained through transform coding of a given macroblock, into multiple groups. Each group is assigned to a prescribed MGS layer.

We initially consider a 4×4 macroblock. We let w_m, $m = 1, 2, \ldots, 16$, denote the number of transform coefficients in MGS layer m within an enhancement layer, whereby

$$\sum_{m=1}^{16} w_m = 16. \tag{1}$$

The number of transform coefficients w_m is also referred to as the "weight" of MGS layer m. An MGS encoding can be represented by giving the weights in the vector form $\mathbf{W} = [w_1, w_2, w_3, \ldots, w_{16}]$, whereby a $w_i = 0$ if it is not specified. Fig. 1 illustrates the splitting of the transform coefficients of a 4×4 macroblock into three MGS layers with the weights $\mathbf{W} = [3, 3, 10]$, i.e., $w_1 = 3$, $w_2 = 3$, and $w_3 = 10$ while $w_4, \ldots, w_{16} = 0$. As another example, consider the weights $\mathbf{W} = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$, which result in sixteen MGS layers, each containing one transform coefficient.

When extending this approach of splitting transform coefficients into layers to 8×8 macroblocks, there are two approaches in H.264 MGS. One approach is to divide a given 8×8 macroblock into four 4×4 submacroblocks and to split the coefficients of each 4×4 submacroblock according to the above approach illustrated in Fig. 1. This submacroblock approach is

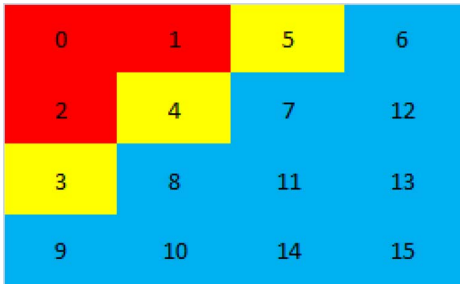


Fig. 1. Illustration of allocation of transform coefficients of a 4×4 macroblock to MGS layers for weight vector $\mathbf{W} = [3, 3, 10]$. Coefficients with indices 0–2 constitute MGS layer 1, while coefficients with indices 3–5 constitute MGS layer 2 and coefficients with indices 6–15 constitute MGS layer 3.

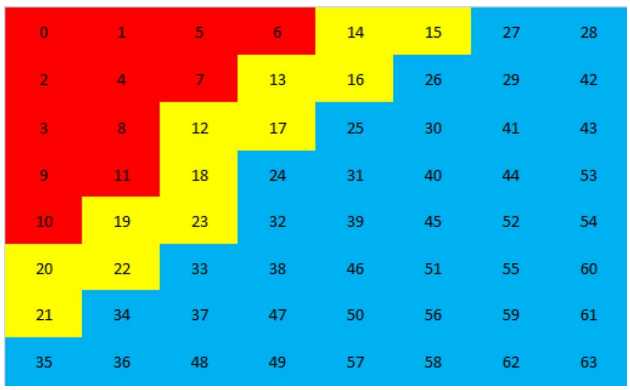


Fig. 2. Illustration of extension of splitting of transform coefficients into MGS layers without subdivision of an 8×8 macroblock. For the example weights $\mathbf{W} = [3, 3, 10]$, the first $4 \cdot w_1 = 4 \cdot 3$ coefficients form MGS layer 1, the next $4 \cdot 3$ coefficients form MGS layer 2, and the remaining $4 \cdot 10$ coefficients form MGS layer 3.

usually employed in conjunction with context-adaptive variable length coding (CALVC) entropy coding [50].

When the other main entropy encoding scheme, context-based adaptive binary arithmetic coding (CABAC) [51], is used, the 8×8 macroblock is not subdivided. Instead, the above approach for splitting the transform coefficients of a 4×4 macroblock is extended to the 8×8 macroblock by multiplying each weight w_i by a factor of four. That is, the coefficients are considered in the conventional zigzag order and $4 \cdot w_m$ coefficients are assigned to MGS layer m as illustrated in Fig. 2. Throughout the remainder of this paper, we consider CABAC, which is widely used in H.264 encodings.

Each MGS layer of a given video frame (picture) forms a single NALU [14]. In our example with $\mathbf{W} = [3, 3, 10]$, the enhancement layer of a given video frame is divided into three NALUs, one for each MGS layer.

Bit rate extraction: With MGS encoding, the video bit rate is adjusted by dropping enhancement layer NALUs, one at a time, until the target bit rate is achieved. No NALUs are dropped from the base layer. We consider the following common approaches for dropping NALUs:

(i) MGS layer approach: The NALUs from the highest indexed MGS layer are dropped first. For instance, with three MGS layers, the MGS layer approach first drops NALUs from MGS layer 3; then, if further rate reduction is needed, NALUs from MGS layer 2 are dropped, and so on.

(ii) Priority ID approach: The priority ID approach, also referred to as MLQL Assigner & Ordered\TopLayer Extractor approach in [52] and as JSVM QL in [53], [54] and implemented in the reference Joint Scalable Video Model (JSVM) software [55], employs RD optimization strategies. A priority ID in the range 0 (lowest importance)—63 (highest importance) is assigned to each NALU. For bitstream extraction, first the NALUs with the highest priority ID are selected, followed by the NALUs with lower priority IDs, until the target bit rate is reached. We conduct the priority ID assignment and NALU extraction over the full video sequence, i.e., all M frames of a given video, so as to give the RD optimization strategies the maximal time horizon.

(iii) MGS-temporal layer approach: The MGS-temporal layer approach [50], [56] prioritizes the NALU extraction according to the temporal layers $T = 0, 1, 2, \ldots, \tau$, followed by prioritization according to the MGS layers. Specifically, the highest MGS layer of the frames in the highest temporal layer $T = \tau$ have the lowest priority, the second highest MGS layer of these frames in temporal layer $T = \tau$ have the next lowest priority, and so on. Thus, for rate adaptation, the MGS-temporal layer approach first drops the MGS layers (from highest to lowest) from the highest temporal layer $T = \tau$. For further rate reduction, the MGS layers (from highest to lowest) are dropped from the second highest temporal layer $T = \tau - 1$, and so on. For a prescribed target bit rate, we conduct the MGS-temporal layer extraction over the time horizon of the full video sequence as well as over the time horizon of individual GoPs.

III. Evaluation Set Up

A. Video Sequences

We present evaluation results for the following representative videos with the CIF (352×288 pixels) resolution and a frame rate of 30 frames/second.

- The ten minute Sony Digital Video Camera Recorder demo sequence (17,682 frames), which we refer to as *Sony* sequence, a documentary style video with a mixture of detailed scenes with high texture content and wide a range of motion activities.
- The first half-hour of the movie *Silence of the Lambs* (54,000 frames), a drama/thriller genre video.
- The first half-hour of the movie *Star Wars IV* (54,000 frames), a science fiction/action genre video.
- 30 minutes of *NBC 12 News* (49,523 frames), an evening news cast, including the commercials.
- The first half-hour of the movie *Citizen Kane* (54,000 frames), a drama/mystery genre video.
- The first half-hour of the movie *Die Hard* (54,000 frames), an action/crime/drama/thriller genre video.

These video sequences represent a wide range of motion and texture levels. (Results for additional videos are available at http://trace.eas.asu.edu.) These sequences were obtained with the MEncoder tool through decoding the original DVD sequences into the YUV format and subsampling to CIF resolution.

B. H.264 SVC Encoding Set-Up

We used the SVC JSVM reference software encoder (version 9.16). We set the GoP pattern to G16B15 (IBBBBBBBBBBBBBBB, 16 frames with 15 B frames per I frame) with hierarchical B frames, as we found through additional evaluations that the G16B15 GoP pattern gives better RD performance compared to the G16B7, G16B3, and G16B1 GoP patterns. We set the MeQP values, which determine the Lagrangian parameters for motion estimation and mode decision of key pictures, to values smaller than the QP values as this setting resulted in RD-efficient coding [57]. We used the macroblock adaptive inter-layer prediction, which employs a rate-distortion optimization framework. We used the CABAC coding scheme and enabled the 8×8 transform.

Following the recommendations of [58] on block matching metrics, we employ a combination of sum of absolute difference (SAD) for full pixel and Hadamard for sub pixel motion estimation. Similarly, following [58], we employ fast search block matching with a search range of 16.

1) CGS Encoding Set-Up: For the encodings with the original reference software with the three dependency layer restriction, we employ inter-layer prediction from the base layer and the first enhancement layer. We set the base layer quantization parameter to $B = 48$ and consider $\mathrm{DQP} = 15$, 10, and 6 to cover a wide quality adaptation range.

2) MGS Encoding Set-Up: Our default weight vector is $\mathbf{W} = [1, 2, 2, 3, 4, 4]$. We employ inter-layer prediction with RD optimization from the highest available quality (MGS) layer. We employ one enhancement layer and use the default quantization parameters $B = 35$ for the base layer and $E = 25$ for the enhancement layer.

C. Video Quality and Traffic Metrics

We employ the average of the peak signal-to-noise ratio (PSNR) values of the frames of a video sequence as objective video quality measure. For a given frame with $N_x \times N_y$ pixels with 8-bits per pixel, the PSNR is calculated from the mean squared error

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

$$\mathrm{PSNR} = 10 \cdot \log_{10} \frac{255^2}{\mathrm{MSE}} \quad (3)$$

For a video sequence consisting of M frames, let X_m, $m = 1, 2, \ldots, M$, denote the sizes (in bit) of the encoded video frames. We only consider the size of the actual encoded video data and the size of the H.264 network adaptation layer overheads that are incurred by encoding enhancement layers and by splitting of an enhancement layer into MGS layers; other types of overhead, for example, streaming protocol encapsulation overheads, are not considered. In this study, we mainly focus on the traffic measures:

$$\text{Mean framesize } \bar{X} = \frac{1}{M} \sum_{m=1}^{M} X_m \quad (4)$$

$$\text{Variance of frame size } \sigma^2 = \frac{1}{M-1} \sum_{m=1}^{M} (X_m - \bar{X})^2 \quad (5)$$

$$\text{Coefficient of variation of frame size CoV} = \frac{\sigma}{\bar{X}}. \quad (6)$$

The rate-distortion (RD) curve is the plot of the average of the PSNR values of the frames in an encoded video sequence as a function of the mean bit rate $\bar{X}/T$, whereby $T = 1/30$ seconds. The CoV of the frame size is a common measure of the traffic variability (fluctuation); plotting the CoV as a function of the average PSNR video quality gives the rate variability-distortion (VD) curve [24], [59]. Analogously to these frame time scale traffic measures, we define the corresponding group of pictures (GoP) time scale traffic measures based on the sizes (in bit) of the frames in each GoP of an encoded video sequence.

IV. CGS Traffic and Quality Characteristics

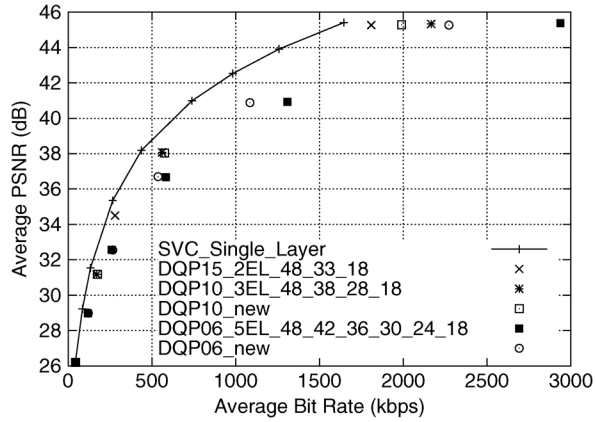
From our extensive studies, we include representative results for *Sony, NBC News*, and *Star Wars* in this section. We note that preliminary results considering only the encoder with the three dependency layer restriction were presented in [60].

A. CGS Rate-Distortion (RD) Performance

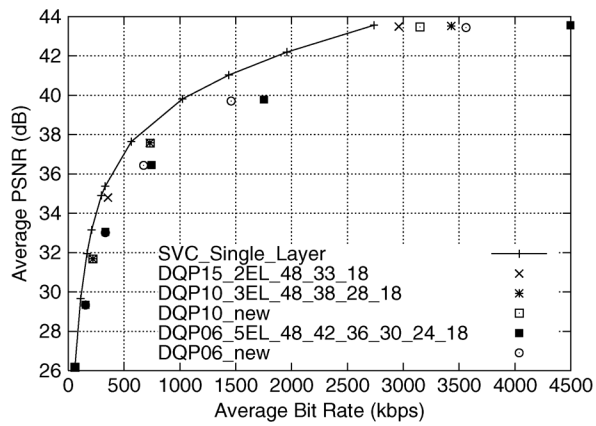
In Fig. 3, we plot the RD points for H.264 CGS encodings (which we do not connect in RD curves as CGS provides only these discrete RD points) and compare with the RD curve of the single-layer H.264 SVC encodings. The plotted average PSNR video quality and the bit rate values represent the aggregate of the base layer and applicable enhancement layer(s). For instance, for the $\mathrm{DQP} = 15$ encodings, the bottom-left point corresponds to the base layer only, the middle point to the aggregate of the base layer and first enhancement layer, and the upper right point to the aggregate of the base layer and the two enhancement layers. We observe from Fig. 3 that encodings with $\mathrm{DQP} = 15$ have the highest RD performance among the CGS encodings. With decreasing DQP, and correspondingly more layers, the RD performance is reduced. We also observe that for a given (fixed) number of four or more layers, the encoder modification that permits more than three dependency layers substantially improves the RD performance.

For a closer comparison of the CGS encodings with the single-layer encodings, we give in Table I the percentage increase in the average bit rate of the CGS aggregate stream (with the modification to permit more than three dependency layers) up to and including a prescribed layer with respect to the single-layer encoding with the same average PSNR video quality. For $\mathrm{DQP} = 15$, we observe a bit rate increase of around 8–20% for *Sony* and *NBC News*, while the bit rate increase is 19–31% for *Star Wars*. For smaller DQP values, the bit rate increases reach 30–60% and even around 80% for *Star Wars*.

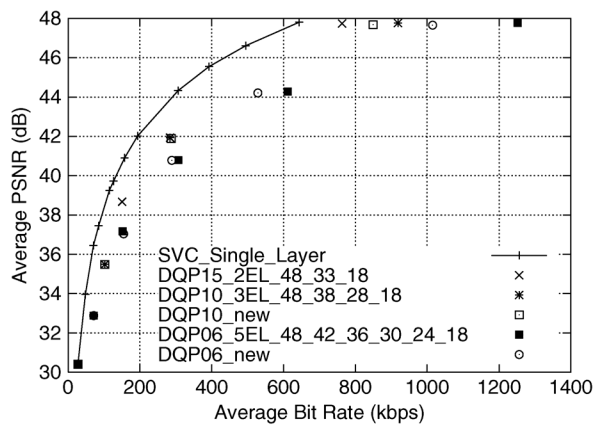
Overall, these results confirm the observations made in [14] for short test sequences in that the H.264 SVC CGS bit rate overhead compared to single-layer encodings increases with decreasing DQP values and correspondingly larger numbers of enhancement layers. A bit rate overhead within 10% to 30% can be achieved for relatively large DQP values and correspondingly



(a)



(b)



(c)

Fig. 3. RD curves of H.264 SVC CGS (modification permitting more than three dependency layers is denoted by "new") and H.264 SVC single-layer encodings. (a) *Sony*. (b) *NBC News*. (c) *Star Wars*.

few quality layers that provide streams with relatively large differences in PSNR video quality.

B. CGS Rate Variability-Distortion (VD) Performance

In Fig. 4, we compare the VD curves of H.264 SVC single-layer encodings with the curves obtained by connecting the individual PSNR quality–CoV points of the H.264 SVC CGS encodings; more specifically, the PSNR quality–CoV points are for the aggregate of the base layer and the applicable enhancement layer(s). In contrast, in Table II we give the CoV values for the individual layers.

TABLE I
AVERAGE BIT RATE INCREASE [IN PERCENT] OF H.264 SVC CGS ENCODING WITH MODIFICATION PERMITTING MORE THAN THREE DEPENDENCY LAYERS RELATIVE TO H.264 SVC SINGLE LAYER ENCODING WITH SAME AVERAGE PSNR QUALITY

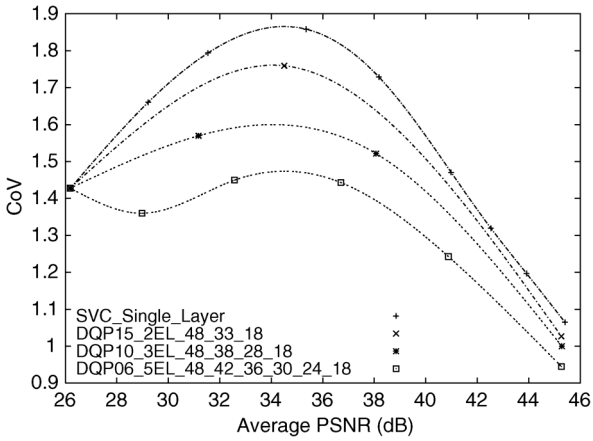
SONY						
DQP	B	E_1	E_2	E_3	E_4	E_5
15	0	15.0	10.0	-	-	-
10	0	28.0	28.2	18.17	-	-
06	0	43.5	51.6	49.9	43.4	34.93
NBC News						
DQP	B	E_1	E_2	E_3	E_4	E_5
15	0	20.1	8.1	-	-	-
10	0	31.8	29.3	15.1	-	-
06	0	39.4	59.01	56.03	42.74	30.09
Star Wars						
DQP	B	E_1	E_2	E_3	E_4	E_5
15	0	30.83	18.54	-	-	-
10	0	44.58	46.55	31.98	-	-
06	0	48.15	78.21	83.5	72.52	57.69

We observe from Fig. 4 that the VD curves of the CGS encodings with $\mathrm{DQP} = 15$ exhibit the same trends as the single layer encodings (previously examined in [24]) of first increasing, peaking, and then decreasing CoV values. On the other hand, for $\mathrm{DQP} = 6$, i.e., the relatively RD-inefficient encoding with five enhancement layers, we observe initially decreasing, then increasing, and finally decreasing trends. These CoV trends for $\mathrm{DQP} = 6$ for the aggregate stream indicate that the first CGS enhancement layer reduces the traffic variability compared to the base layer stream. Indeed, we observe in Table II that the first enhancement layer (the first two for *Star Wars*) has significantly lower variability than the base layer. Adding the first CGS enhancement layer to the base layer thus smoothes the traffic somewhat, resulting in overall reduced variability for the aggregate stream. We further observe from Table II for $\mathrm{DQP} = 6$ that the second enhancement layer has the highest CoV (the third for *Star Wars*) leading to the increase in the variability in the aggregate stream in Fig. 4. The highest enhancement layers have relatively low CoV values, resulting in the drop of the aggregate stream variability.

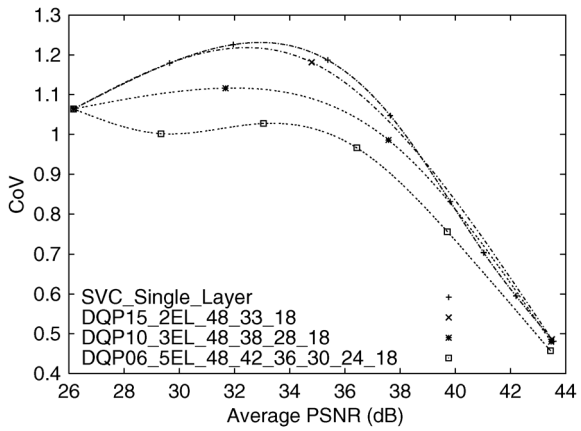
In summary, we conclude that H.264 SVC CGS encodings with relatively few enhancement layers (two in our studies) that span a wide quality and bit rate range result in a bit rate overhead of 10–30% compared to single layer H.264 SVC encodings. The traffic variabilities of these CGS layers and the resulting aggregate streams are slightly lower than the traffic variabilities of the single layer streams. For three or more enhancement layers the bit rate overhead of CGS increases substantially, while the traffic variability of the CGS layers decreases only slightly. In additional evaluations with videos in the full HD (1920×1080 pixel) format, which are not included in detail due to space constraints, we found that these conclusions hold similarly for CGS encodings of HD video.

V. MGS TRAFFIC AND QUALITY CHARACTERISTICS

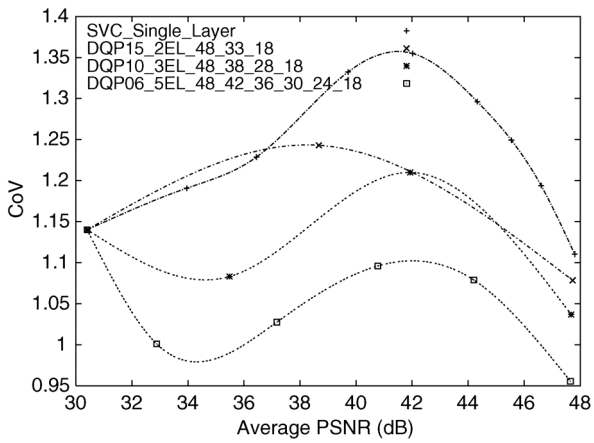
In this section, we examine the traffic and quality characteristics of H.264 SVC MGS encodings. We initially study the impact of the MGS weights, the base and enhancement layer quantization parameters, and the extraction method. We then examine the rate-distortion and rate variability-distortion characteristics of MGS. Throughout, we present curves from represen-



(a)



(b)



(c)

Fig. 4. VD curves of H.264 SVC CGS (with modification permitting more than three dependency layers) and H.264 SVC single-layer encodings. (a) *Sony*. (b)*NBC News*. (c) *Star Wars*.

tative video sequences from our extensive encoding and traffic studies.

A. MGS Weights

Fig. 5 shows the RD curve of the *Sony* sequence with different MGS weights for fixed quantization parameters of $B = 35$ for the base layer and $E = 25$ for the enhancement layer with priority ID based extraction over the full sequence. We observe from Fig. 5 that the RD performance for the different

TABLE II
CoV Values of Individual CGS Layers (Encoded With Modification Permitting More Than Three Dependency Layers)

CoV for *SONY demo*						
DQP	B	E_1	E_2	E_3	E_4	E_5
15	1.43	1.83	0.92	-	-	-
10	1.43	1.63	1.52	0.84	-	-
06	1.43	1.33	1.54	1.45	1.08	0.74
CoV for *NBC News*						
DQP	B	E_1	E_2	E_3	E_4	E_5
15	1.06	1.22	0.41	-	-	-
10	1.06	1.15	0.95	0.36	-	-
06	1.06	0.97	1.06	0.92	0.61	0.3
CoV for *Star Wars*						
DQP	B	E_1	E_2	E_3	E_4	E_5
15	1.14	1.28	1.05	-	-	-
10	1.14	1.07	1.29	0.97	-	-
06	1.14	0.93	1.06	1.19	1.07	0.84

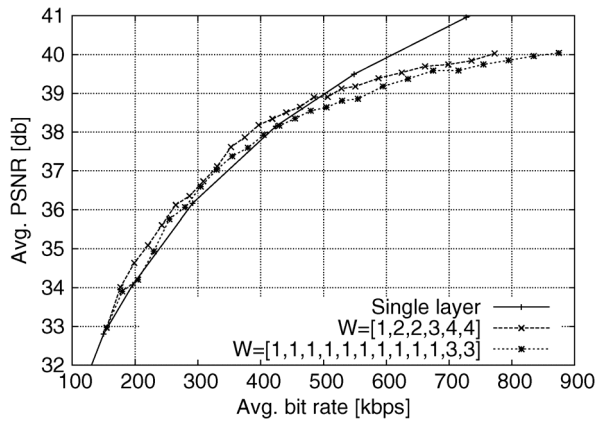


Fig. 5. RD curve for *Sony* sequence for different MGS weights $\mathbf{W}$; with priority ID extraction over full sequence and $B = 35, E = 25$, fixed.

MGS weights is nearly the same with the MGS weights $\mathbf{W} = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 3]$ giving slightly lower RD performance in the range from moderate to high bit rates compared to the $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ MGS weights.

We also observe from Fig. 5 that the MGS RD curves are very close to the RD curve of the single-layer encoding, and even slightly exceed the single-layer RD curve in the range from low to moderate bit rates up to around 450 kbps. The slightly better RD performance of MGS is primarily due to the RD prioritized bistream extraction. For small to moderate additions to the base layer stream, MGS provides those groups of low-frequency transform coefficients (from the upper left of the illustrations in Figs. 1 and 2 that are most RD efficient. Adding these most RD-efficient transform coefficients of select video frames to the base layer stream can slightly improve the RD efficiency over the single-layer stream that is encoded with fixed quantization scales across all video frames. This characteristic of MGS is examined in more detail in Section V-D.

As the quality increases to approach 40 dB, all MGS layers from all video frames are needed and we observe a significant drop in RD efficiency compared to the single-layer encoding. This drop in RD efficiency is due to the overhead of MGS encoding.

We briefly note that the different MGS encodings start at the same point on the RD curve. This is because the base layer of H.264 SVC is compliant with the AVC single layer and all the

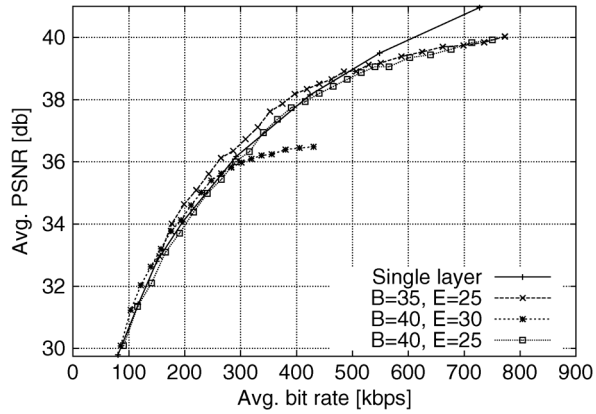


Fig. 6. RD curve for *Sony* sequence with $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ MGS weights and different base (B) and enhancement (E) layer QP values with priority ID extraction over full sequence.

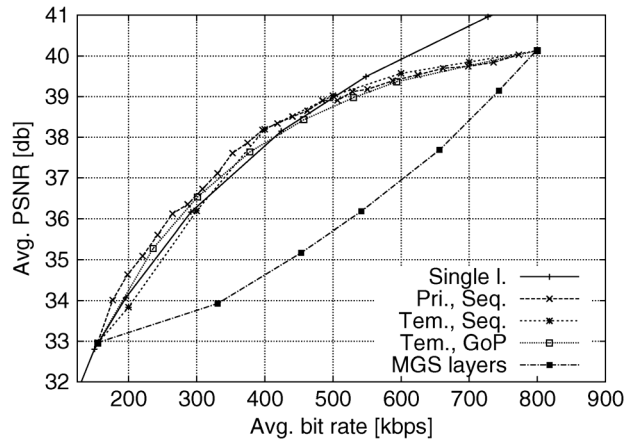


Fig. 7. RD curve for *Sony* sequence for extraction by: priority ID over full sequence (Pri., Seq.), MGS-temporal layer over full sequence (Tem., Seq.), MGS-temporal layer over individual GoPs (Tem., GoP), and MGS layers.

MGS encodings have the same QP value for the base layer. We present results for the $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ MGS weights in the remainder of this paper.

B. Quantization Parameter

Fig. 6 shows the RD curve for the *Sony* sequence for different quantization parameters for the base layer B and enhancement layer E. Comparing first the $B = 40, E = 30$ and $B = 40, E = 25$ encodings, we observe that for low bit rates up to around 300 kbps, the $B = 40$, $E = 30$ encoding has slightly higher RD performance than the $B = 40, E = 25$ encoding. This is mainly because $E = 30$ results in relatively stronger quantization and thus fewer bits required for the encoding at this lower end of the quality range. For bit rates above 300 kbps, the $E = 30$ encoding approaches the upper end of its quality range resulting in the observed drop in RD performance.

Next, turning to the comparison between the $B = 35, E = 25$ and $B = 40$, $E = 25$ encodings, we observe that the $B = 35$, $E = 25$ encoding gives higher RD performance in the mid bit rate range from about 200–500 kbps; while for higher bit rates, the RD curves closely approach each other. In the mid bit rate range, the $B = 35$, $E = 25$ encoding benefits from the relatively higher quality base layer encoding, which provides a higher quality reference for encoding the enhancement layer, and thus higher RD efficiency in the encoding of the enhancement layer. As we approach the upper end of the quality range of the enhancement layer, the advantage due to the higher quality base layer diminishes.

Overall, we observe from Fig. 6 that a wider spread between base and enhancement layer quantization parameters (e.g., $B = 40$, $E = 25$) provides a wider range of quality (and corresponding bit rate adaptation) at the expense of slightly reduced RD performance compared to encodings with a narrower quantization parameter spread. Unless stated otherwise we use the setting $B = 35$, $E = 25$ in the remainder of this article.

C. Extraction Method

In Fig. 7 we compare the RD curves of H.264 SVC MGS with priority ID based extraction over the full video sequence (denoted by "Pri., Seq."), MGS-temporal layer based extraction conducted over the full video sequence (denoted by "Tem., Seq.") and conducted over individual GoPs (denoted by "Tem., GoP"), and MGS layer based extraction with the RD curves of H.264 SVC single-layer encoding. Notice that the MGS layer curve for the weights $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ has seven points, corresponding to the base layer only for all frames of the video sequence, the base layer with one MGS layer for all frames, the base layer with two MGS layers, and so on, until all six MGS layers are added to the base layer. We observe that the RD performance obtained using priority ID and MGS-temporal layer based extraction is significantly higher than the MGS layer based extraction for the entire span between the points corresponding to the base layer only and the base layer plus the full enhancement layer. The priority ID based approach, which slightly outperforms the single layer encoding for low to moderate bit rates, selects the most RD efficient MGS layers (NALUs) for select frames to add to the base layer stream and thus provides excellent RD performance. On the other hand, the MGS layer based extraction adds the same number of MGS layers for each video frame. This approach thus ignores the contributions to the average PSNR video quality of a given MGS layer (NALU) relative to its size (in bits).

Turning to the MGS-temporal layer based extraction, we observe that for both time horizons, i.e., dropping NALUs to meet the prescribed target bit over the full sequence or for each individual GoP results in RD performance that closely approximates the RD performance with priority ID based extraction. The MGS-temporal layer extraction exploits the hierarchical B frame structure as well as the MGS encoding structure by first dropping the MGS layers from highest to lowest from the B frames with no dependent frames in temporal layer $T = \tau$. Then, the MGS layers are dropped from the B frames with the fewest dependent frames in temporal layer $T = \tau - 1$. The MGS layer dropping proceeds with the B frames with the next smallest number of dependent B frames in temporal layer $T = \tau - 2$, and so on. The results in Fig. 7 indicate that this MGS-temporal layer extraction approach, which has very low computational complexity, gives nearly the same RD performance as the computationally demanding priority ID based

extraction. In the following sections we further examine the MGS-temporal layer extraction over individual GoPs, which is suitable for rate-adaptations over short time scales as they may be necessary in transport networks with varying available bandwidth.

Contrasting the results for the extraction methods with the results for the base and enhancement layer quantization parameters B and E in Section V-B, we observe that the extraction method has a relatively large impact on the RD performance. In particular, the extraction method strongly affects the RD performance across the entire range of PSNR qualities (and corresponding bit rates) covered by the enhancement layer. In contrast, the B setting typically has a relatively small impact and the E settings mainly affects the positioning of the upper end of the quality range covered by the enhancement layer (whereby the RD performance drops when approaching the upper end of the covered quality range). Comparing with the results in Section V-A, we observe the relatively minor impact of the MGS weights $\mathbf{W}$.

D. MGS Rate-Distortion (RD) Performance

In Fig. 8, we compare the H.264 SVC MGS and single-layer RD curves for the considered test sequences from a wide range of video content genres. We observe that the RD curve behaviors of the *Sony* sequence, which we have focused on in the presentation so far, are quite representative for a wide range of video content genres. In particular, we observe that for this wide set of test videos, the RD curve of MGS with priority ID based extraction is very close or slightly exceeds the RD curve of the single layer encoding at low to moderate PSNR video qualities and corresponding bit rates. The improved RD performance with MGS with priority ID extraction is achieved through the RD optimization which selectively discards MGS layer NALUs from selected frames if the NALUs provide relatively small PSNR improvements for their size (in bits).

We further observe from Fig. 8 that the RD curve of MGS-temporal layer extraction over individual GoPs is generally just slightly below the curve for priority ID extraction over the full sequence and quite close to the curve of the single layer encoding. These results indicate that even with the low-complexity MGS-temporal layer extraction, H.264 SVC MGS can provide flexible adaptation over a wide video quality and bit rate range while achieving nearly the same RD performance as single-layer encoding. Nevertheless, further research on computationally efficient extraction mechanisms that maximize the RD performance, such as [61]–[63], has the potential to close the gap between priority ID and MGS-layer based extraction. In further evaluations that are not included in detail due to space constraints, we found that MGS-temporal layer extraction over individual GoPs gives similar variability of the individual frame PSNR qualities as priority ID extraction over the full sequence.

At the upper end of the MGS RD curves we observe for all videos a significant drop in RD efficiency compared to the single layer encoding. At the upper end of the quality (and bit rate) range, all MGS layer NALUs are included (even the least RD efficient NALUs). As a result, the overhead of the MGS encoding can not be offset by selecting the most RD efficient MGS layer NALUs and the full effect of the overhead becomes visible. Clearly, these results suggest to select the enhancement layer quantization parameter E such that the upper end of the RD curve is sufficiently higher (about 1–2 dB for the considered test sequences) than the targeted highest streaming video quality.

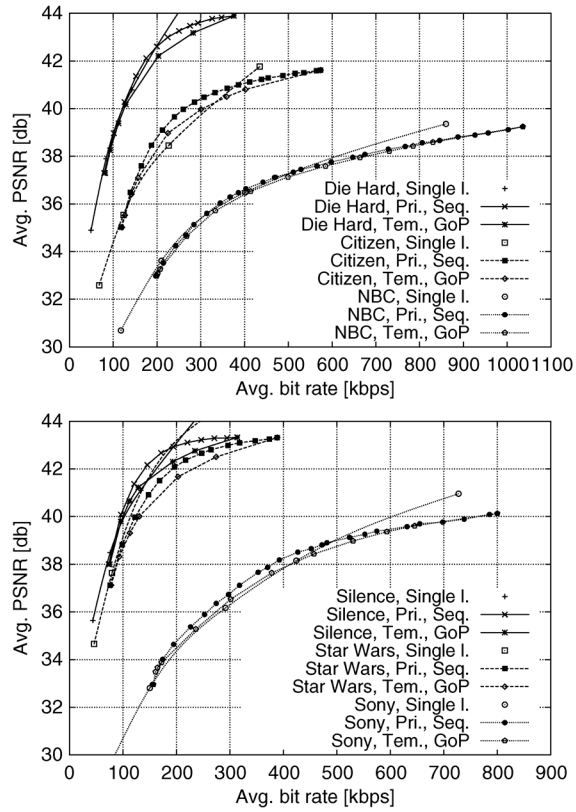


Fig. 8. RD curves for a wide range of test sequences with MGS weights $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ and quantization parameters $B = 35, E = 25$ with priority ID based extraction over full sequence and MGS-temporal layer based extraction over individual GoPs.

E. MGS Rate Variability-Distortion (VD) Performance

1) Frame Time Scale: In Fig. 9, we compare the frame time scale rate variability-distortion (VD) curves of the MGS streams extracted based on priority IDs (over the full sequence) and MGS-temporal layers (over individual GoPs). We also examined the VD curves for the MGS streams extracted based on MGS layers (not plotted here to avoid clutter) and found that their CoV values are low (in the range 0.03–0.2 and with a mean of 0.11 for our test videos) and, for a given video, constant across the range of PSNR values.

We observe from Fig. 9 that in the mid range of the PSNR qualities, e.g., 35–37 dB for *NBC*, the MGS streams with priority ID based extraction have somewhat higher CoV values than the MGS-temporal layer extraction streams. In additional evaluations that are not included so as to avoid clutter, we found that in the mid range of the PSNR qualities, both MGS streams have significantly higher CoV values than the single-layer encodings. For instance, for *Die Hard*, the maximum CoV is increased from approximately 1.4 for the single-layer stream to about 2.4 for the MGS stream with priority ID extraction, while for *Sony*, the maximum CoV increases from approximately 2.1

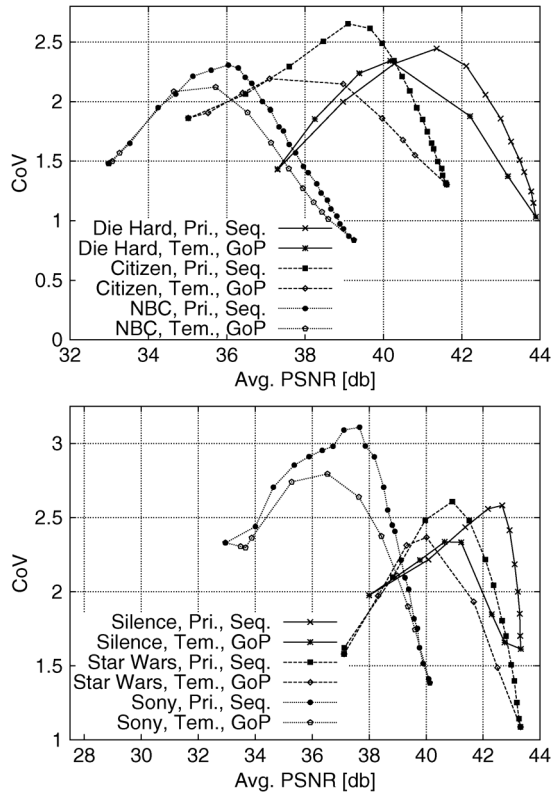


Fig. 9. Comparison of frame time scale VD curves for MGS streams with $B = 35$, $E = 25$, and $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ using priority ID extraction over the full sequence and MGS-temporal layer extraction over individual GoPs.

to 3.1. Thus, we conclude that the MGS encoding and stream extraction adds substantially to the frame time scale traffic variability. More specifically, for the low to moderate PSNR quality ranges, the selective addition of the most RD efficient MGS layer NALUs adds to frame sizes (in bit) such that their variability is increased, i.e., relatively more bits are added to frames that are already large.

In further evaluations we also found that for the upper end of the PSNR quality range of the MGS streams, their CoV values drop below the corresponding CoV values of the single-layer streams by about 0.15–0.3. At the upper end of the quality range, all the MGS layer NALUs are added in for all frames and the overhead of the MGS encoding leads to the pronounced drop in RD efficiency observed in Section V-D. As the CoV is defined as the standard deviation of the frame sizes normalized by their mean, the pronounced increase in the mean frame size is mainly responsible for the relatively steep drop of the CoV values.

We observe from Fig. 9 that generally videos with a high degree of heterogeneity in the levels of motion and texture complexity result in higher variability in the streamed frame sizes. For instance, *Sony* and *Citizen Kane*, which have a very wide range of motion and texture levels in their scenes, give high CoV values. On the other hand, videos with consistently high levels of motion, such as *Die Hard*, give relatively lower CoV values.

Inspecting Fig. 9 closer, we observe that the priority ID extraction conducted over the full sequence has higher CoV values than the MGS-temporal extraction over individual GoPs primarily in the mid-range of the VD curves. The VD curves for both extraction methods have the same endpoints. (In further evaluations, which are not included due to space constraints, we observed that MGS-temporal layer extraction over the full sequence gives similarly high bit rate variabilities as priority ID extraction over the full sequence.) For both extraction methods, the left (lowest PSNR video quality) endpoint corresponds to streaming the base layer (i.e., the lowest possible bit rate); whereas, the right (highest PSNR video quality) endpoint corresponds to streaming the base layer plus the full MGS enhancement layer (i.e., the highest possible bit rate). For target bit rates between the lowest and highest possible rates, the extraction for individual GoPs strives to meet the prescribed target bit rate when averaging over the frames in each individual GoP, i.e., strives to equalize the sizes (in bit) of the individual GoPs. Equal GoP sizes reduce the variability of the frames sizes across different GoPs compared to the extraction over the full sequence, which only strives to meet the prescribed target bit rate when averaging over all frames in the sequence.

2) GoP Time Scale: Turning to the traffic variability at the GoP time scale, we observe from Fig. 10 that aggregating (i.e., effectively smoothing) the frames over each GoP is quite effective in reducing the traffic variability of the extracted MGS streams compared to the frame time scale considered in Fig. 9. We found in additional evaluations that the CoV values of the GoP sizes of the MGS streams (both with priority ID and MGS-temporal layer extraction) are close to the corresponding CoV values of single-layer encodings in the range of low PSNR qualities. For moderate to high PSNR qualities, the MGS streams have typically lower CoV values than the single-layer streams. For instance, for *Die Hard*, the single-layer GoP size CoV values are above 0.52, whereas priority ID extraction gives CoV values as low as 0.31 and MGS-temporal layer extraction gives CoV values as low as 0.1. Similarly, for *Sony*, the single-layer CoV values stay above 0.51, while priority ID and MGS-temporal layer extraction achieve CoV values as low as 0.35 and 0.25, respectively. Thus, the added traffic variability of the MGS streams compared to the single-layer streams at the frame time scale has effectively been eliminated. This implies that the added variability that was introduced by the selective inclusion of MGS layer NALUs for select frames has mainly added variability among the frames within a GoP.

We observe from Fig. 10 that in the mid range of PSNR video qualities, the MGS-temporal layer extraction over individual GoPs reduces the GoP size CoV values considerably compared to the priority ID extraction conducted over the entire sequence. The CoV values of MGS-temporal layer extraction drop to values close to and even below 0.2. For very low target bit rates, both extraction mechanisms stream only the base layer, while for very high target bit rates they stream the base layer plus full MGS enhancement layer. For mid range target bit rates, the MGS-temporal layer extraction over individual GoPs meets the target bit rates for each individual GoP, except for GoPs with a base layer rate above the target bit rate and GoPs with the base layer plus full enhancement layer rate below the target bit rate. In contrast, the priority ID based extraction over

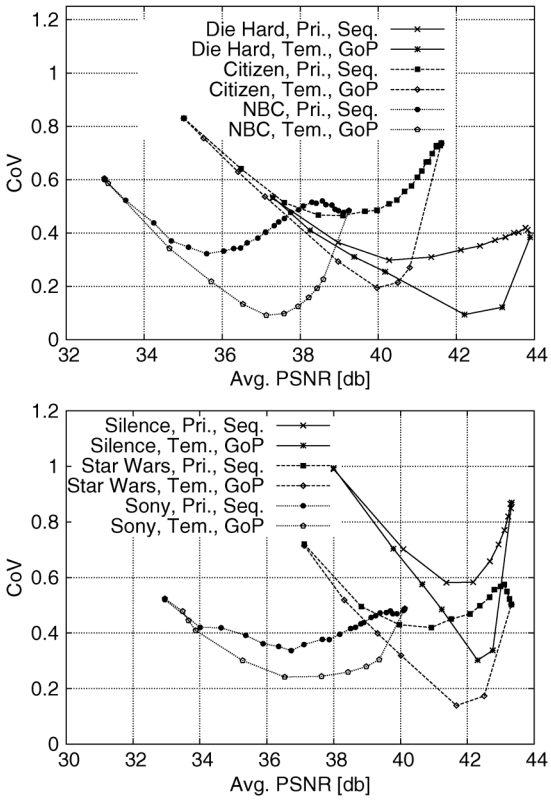


Fig. 10. Comparison of GoP time scale VD curves for MGS streams with $B = 35$, $E = 25$ and $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ using priority ID extraction over the full sequence and MGS-temporal layer extraction over individual GoPs.

the full sequence, strives to meet the target bit rate only over the time horizon of the full sequence, allowing for significantly higher variations of the GoP sizes.

VI. Conclusion

We have examined the traffic and quality characteristics of H.264 SVC quality-scalable video encodings, considering both coarse grain scalability (CGS) and medium grain scalability (MGS). For a test set of long videos from a wide range of content genres, we have studied the rate-distortion (RD) and rate variability-distortion (VD) characteristics. We have found that for encodings with two enhancement layers, i.e., three possible stream qualities, CGS is 10–30% less RD efficient than single-layer H.264 SVC encoding. The corresponding individual CGS layer and aggregate streams have slightly lower traffic variability than single-layer SVC streams while having a similar bell-shaped VD curve. For a larger number of enhancement layers, the RD efficiency drops significantly while the traffic variability of the individual CGS layers is only slightly reduced.

For H.264 SVC MGS, we found that the mechanism for extracting the MGS enhancement layers for each frame from the encoded bit stream has a relatively large impact on the RD and VD characteristics. We considered extraction based on MGS layers, extraction based on priority IDs assigned by an RD optimization approach conducted over a full video sequence, as well as extraction based on MGS-temporal layers conducted over a full video sequence or individual GoPs. We found that extraction by MGS layers gives poor RD performance. On the other hand, the RD curves with priority ID based extraction are very close and sometimes even slightly above the RD curves of the corresponding single-layer encodings for the low to moderate quality range. Toward the upper end of the quality range, the RD efficiency drops below the single-layer RD curve. The low-complexity MGS-temporal layer extraction achieves RD performance very slightly below the high-complexity priority ID approach.

In the range where the MGS RD efficiency is close to the single-layer RD efficiency, the MGS streams have significantly higher traffic variability than the corresponding single-layer streams at the frame time scale. This result has important implications for network transport mechanisms of H.264 SVC MGS video that operate at the frame time scale as these frame level transport mechanisms need to accommodate significantly larger traffic variability than previously experienced for single-layer streams. We also found that streams obtained with MGS-temporal layer extraction over individual GoPs have significantly lower traffic variability than streams obtained with extraction conducted over the full sequence.

Smoothing the video traffic to the GoP time scale effectively reduces the variability of MGS traffic to levels near or below those experienced for single-layer video smoothed over GoPs. Thus, traffic smoothing is highly recommended when streaming MGS streams. In particular, GoP smoothing of streams extracted on the GoP time scales gives very low traffic variability in the mid quality range.

There are many directions for future research on the traffic and quality characteristics as well as the network transport of H.264 SVC quality scalable video. One important direction is to develop and validate mathematical traffic models of CGS and MGS layer traffic, including models for the rate adaptation achieved through partitioning the MGS layer and the related traffic variability. Another direction is to examine how transport mechanisms for both wired and wireless networks can efficiently transport the highly variable MGS streaming traffic.

Acknowledgment

The authors are grateful to Dr. H. Schwarz from the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany, for sharing detailed insights into H.264 MGS video coding.

References

[1] H. Joo and H. Song, "Wireless link state-aware H.264 MGS coding-based mobile IPTV system over WiMAX network," *Journal of Visual Communication and Image Representation*, vol. 21, no. 2, pp. 89–97, Feb. 2010.

[2] I. Kofler, R. Kuschnig, and H. Hellwagner, "Improving IPTV services by H.264/SVC adaptation and traffic control," in *Proc. IEEE Int. Symp. Broadband Multimedia Systems and Broadcasting (BMSB)*, 2009, pp. 1–6.

[3] T. Wiegand, L. Noblet, and F. Rovati, "Scalable video coding for IPTV services," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 527–538, Jun. 2009.

[4] S. Mirta, T. Schierl, T. Wiegand, P. Inigo, C. LeGuern, C. Moreau, L. Guarnieri, and J. Tronc, "HD video broadcasting using scalable video coding combined with DVB-S2 variable coding and modulation," in *Proc. Adv. Satellite Multimedia Syst. Conf. (ASMA) Signal Process. Space Commun. Workshop (SPSC)*, 2010, pp. 114–121.

[5] A. Morell, G. Seco-Granados, and M. Vazquez-Castro, "Cross-layer design of dynamic bandwidth allocation in DVB-RCS," *IEEE Syst. J.*, vol. 2, no. 1, pp. 62–73, Mar. 2008.

[6] M. Bocus, J. Coon, C. Canagarajah, J. McGeehan, S. Armour, and A. Doufexi, "Joint call admission control and resource allocation for H.264 SVC transmission over OFDMA networks," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, 2010, pp. 1–5.

[7] B. Ciubotaru and G.-M. Muntean, "SASHA—a quality-oriented handover algorithm for multimedia content delivery to mobile users," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 437–450, Jun. 2009.

[8] A. Detti, P. Loreti, N. Blefari-Melazzi, and F. Fedi, "Streaming H.264 scalable video over data distribution service in a wireless environment," in *Proc. IEEE Symp. World Wireless Mobile Multimedia Netw. (WoWMoM)*, 2010, pp. 1–3.

[9] O. Hillestad, A. Perkis, V. Genc, S. Murphy, and J. Murphy, "Adaptive H.264/MPEG-4 SVC video over IEEE 802.16 broadband wireless networks," in *Proc. Packet Video*, 2007, pp. 26–35.

[10] D. Hu and S. Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3501–3511, Nov. 2010.

[11] P. Koutsakis, "Dynamic versus static traffic policing: A new approach for videoconference traffic over wireless cellular networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 9, pp. 1153–1166, Sep. 2009.

[12] A. Kholaif, T. Todd, P. Koutsakis, and A. Lazaris, "Energy efficient H.263 video transmission in power saving wireless LAN infrastructure," *IEEE Trans. Multimedia*, vol. 12, no. 2, pp. 142–153, Feb. 2010.

[13] M. Martini, R. Istepanian, M. Mazzotti, and N. Philip, "Robust multilayer control for enhanced wireless telemedical video streaming," *IEEE Trans. Mobile Comput.*, vol. 9, no. 1, pp. 5–16, Jan. 2010.

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

[15] G. Van der Auwera, P. T. David, M. Reisslein, and L. Karam, "Traffic and quality characterization of the H.264/AVC Scalable Video Coding extension," *Adv. Multimedia, Article ID 164027*, pp. 1–27, 2008.

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

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

[18] S. Colonnese, S. Rinauro, L. Rossi, and G. Scarano, "H.264 video traffic modeling via hidden Markov process," in *Proc. EUSIPCO*, 2009.

[19] M. Dai, Y. Zhang, and D. Loguinov, "A unified traffic model for MPEG-4 and H.264 video traces," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 1010–1023, Aug. 2009.

[20] A. Lazaris and P. Koutsakis, "Modeling multiplexed traffic from H.264/AVC videoconference streams," *Comput. Commun.*, vol. 33, no. 10, pp. 1235–1242, Jun. 2010.

[21] N. M. Markovich, A. Undheim, and P. J. Emstad, "Classification of slice-based VBR video traffic and estimation of link loss by exceedance," *Comput. Netw.*, vol. 53, no. 7, pp. 1137–1153, May 2009.

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

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

[24] G. Van der Auwera, P. T. David, and M. Reisslein, "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," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 698–718, Sep. 2008.

[25] J.-A. Zhao, B. Li, and I. Ahmad, "Traffic model for layered video: An approach on markovian arrival process," in *Proc. Packet Video*, Apr. 2003.

[26] W. Zhou, D. Sarkar, and S. Ramakrishnan, "Traffic models for MPEG-4 spatial scalable video," in *Proc. IEEE Globecom*, Dec. 2005, pp. 256–260.

[27] M. Etoh and T. Yoshimura, "Advances in wireless video delivery," *Proc. IEEE*, vol. 93, no. 1, pp. 111–122, Jan. 2005.

[28] M. Fidler, V. Sander, and W. Klimala, "Traffic shaping in aggregate-based networks: implementation and analysis," *Comput. Commun.*, vol. 28, no. 3, pp. 274–286, Feb. 2005.

[29] E. Gurses and O. Akan, "Multimedia communication in wireless sensor networks," *Annals Telecommun.*, vol. 60, no. 7/8, pp. 799–827, Jul./Aug. 2005.

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

[31] Z. Wang, H. Xi, G. Wei, and Q. Chen, "Generalized PCRTT offline bandwidth smoothing based on SVM and systematic video segmentation," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 998–1009, Aug. 2009.

[32] J. Casasempere, P. Sanchez, T. Villameriel, and J. Del Ser, "Performance evaluation of H.264/MPEG-4 Scalable Video Coding over IEEE 802.16e networks," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, 2009, pp. 1–6.

[33] D. Gomez-Barquero, K. Nybom, D. Vukobratovic, and V. Stankovic, "Scalable video coding for mobile broadcasting DVB systems," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2010, pp. 510–515.

[34] M. Jubran, M. Bansal, and L. Kondi, "Low-delay low-complexity bandwidth-constrained wireless video transmission using SVC over MIMO systems," *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 1698–1707, Dec. 2008.

[35] H. Mansour, Y. Fallah, P. Nasiopoulos, and V. Krishnamurthy, "Dynamic resource allocation for MGS H.264/AVC video transmission over link-adaptive networks," *IEEE Trans. Multimedia*, vol. 11, no. 8, pp. 1478–1491, Dec. 2009.

[36] J. Xu, R. Hormis, and X. Wang, "MIMO video broadcast via transmit-precoding and SNR-scalable video coding," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 456–466, Apr. 2010.

[37] B. Zhang, M. Wien, and J.-R. Ohm, "A novel framework for robust video streaming based on H.264/AVC MGS coding and unequal error protection," in *Proc. Int. Symp. Intelligent Signal Processing and Communication Systems (ISPACS)*, 2009, pp. 114–121.

[38] D. Fiems, B. Steyaert, and H. Bruneel, "A genetic approach to Markovian characterisation of H.264/SVC scalable video," *Multimedia Tools and Applications*, 2012, in press.

[39] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos, "Rate and distortion modeling of medium grain scalable video coding," in *Proc. IEEE Int. Conf. Image Process.*, 2008, pp. 2564–2567.

[40] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, "Rate and distortion modeling of CGS coded scalable video content," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 165–180, Apr. 2011.

[41] T. Lakshman, A. Ortega, and A. Reibman, "VBR video: tradeoffs and potentials," *Proc. IEEE*, vol. 86, no. 5, pp. 952–973, May 1998.

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

[43] C. Mazataud and B. Bing, "A practical survey of H.264 capabilities," in *Proc. Commun. Netw. Services Res. Conf. (CNSR)*, 2009, pp. 25–32.

[44] 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 Tutorials*, vol. 6, no. 3, pp. 58–78, Third Quarter, 2004.

[45] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 560–576, Jul 2003.

[46] 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 Syst. Mag.*, vol. 4, no. 1, pp. 7–28, 2004.

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

[48] J. D. Cock, S. Notebaert, P. Lambert, and R. Van de Walle, "Architectures for fast transcoding of H.264/AVC to quality-scalable SVC streams," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1209–1224, Nov. 2009.

[49] X. Li, P. Amon, A. Hutter, and A. Kaup, "Performance analysis of inter-layer prediction in scalable video coding extension of H.264/AVC," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 66–74, Mar. 2011.

[50] H. Kirchhofffer, D. Marpe, H. Schwarz, and T. Wiegand, "A low complexity approach for increasing the granularity of packet based fidelity scalability in scalable video coding," presented at the Picture Coding Symp. (PCS), Lisbon, Portugal, Nov. 2007.

[51] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 620–636, Jul. 2003.

[52] Joint Video Team, Doc. JVT-S043, "Multi layer quality layers," Apr. 2006.

[53] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, "Optimized rate-distortion extraction with quality layers in the scalable extension of H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1186–1193, Sep. 2007.
[54] E. Maani and A. K. Katsaggelos, "Optimized bit extraction using distortion modeling in the scalable extension of H.264/AVC," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2022–2029, Sep. 2009.
[55] J. Reichel, H. Schwarz, and M. Wien, "Joint scalable video model 11 (JSVM 11)," *Joint Video Team, Doc. JVT-X202*, Jul. 2007.
[56] B. Gorkemli, Y. Sadi, and A. Tekalp, "Effects of MGS fragmentation slide mode and extraction strategies on the performance of SVC with medium-grained scalability," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2010, pp. 4201–4204.
[57] *"Joint Scalable Video Model (JSVM) Software Manual,"* Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG), 2011.
[58] F. Niedermeier, M. Niedermeier, and H. Kosch, *Quality assessment of the MPEG-4 scalable video codec* 2009 [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:0906.0667
[59] M. Reisslein, J. Lassetter, S. Ratnam, 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," Tech. Rep. Arizona State Univ., 2002.
[60] A. Pulipaka, P. Seeling, M. Reisslein, and L. Karam, "Overview and traffic characterization of coarse-grain quality scalable (CGS) H.264 SVC encoded video," in *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2010.
[61] H. Lee, Y. Lee, D. Lee, J. Lee, and H. Shin, "Implementing rate allocation and control for real-time H.264/SVC encoding," in *Dig. Tech. Papers Int. Conf. Consum. Electron. (ICCE)*, Jan. 2010, pp. 269–270.
[62] R. Li, J. Sun, and W. Gao, "Fast weighted algorithms for bitstream extraction of SVC Medium-Grain scalable video coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2010, pp. 249–254.
[63] L. Wei, G. Sen, C. Xu, and Z. Jihong, "SVC bitstream extraction based on the importance of MGS slice," in *2nd Int .Conf. Industrial Inf. Syst. (IIS)*, Jul. 2010, vol. 1, pp. 148–151.

Rohan Gupta received the Bachelor's Degree in Communications and Computer engineering in from LNMIIT University, Jaipur, India, in 2007 and his Master's Degree in Electrical Engineering from Arizona State University, Tempe, in 2009. Since 2010, Rohan is with the Multimedia group at Qualcomm Inc. His fields of interests are video streaming technologies, scalable video coding, video processing, wireless communications, and graphics.

Akshay Pulipaka is a Ph.D. student in the School of Electrical, Computer, and Energy Engineering at Arizona State University (ASU), Tempe. He received the M.S. degree in Electrical Engineering from ASU, in 2009, and the B.E. degree in Electronics and Communications Engineering from Osmani University, Hyderabad, India, in 2007. During the year 2011 he visited the Fraunhofer Institute-HHI, Berlin, Germany. He is a student member of IEEE/ACM. His research interests are in the areas of video processing, networking, streaming, and quality assessment.

Patrick Seeling is an Assistant Professor in the Department of Computer Science at Central Michigan University, Mount Pleasant, MI. He received his Dipl._Ing. Degree in Industrial Engineering and Management from the Berlin Institute of Technology, Germany, in 2002 and his Ph.D. in Electrical Engineering from Arizona State University, Tempe, in 2005. He was a Faculty Research Associate with ASU from 2005 to 2007, and an Assistant Professor with the University of Wisconsin-Stevens Point from 2008-2011. His areas of interest are multimedia, optical, and wireless networking and engineering education. Patrick Seeling is a Senior Member of the ACM and IEEE.

Lina J. Karam received the Bachelor of Engineering degree in computer and communications engineering from the American University in Beirut in 1989 and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1992 and 1995, respectively.

She is currently a Professor in the School of Electrical, Computer and Energy Engineering (ECEE) at Arizona State University, Tempe, where she directs the Image, Video, and Usability (IVU) and the Real-Time Embedded Signal Processing (RESP) Laboratories. She worked at Schlumberger Well Services (Austin, Texas) on problems related to data modeling and visualization, and in the Signal Processing Department of AT&T Bell Labs (Murray Hill, New Jersey) on problems in video coding during 1992 and 1994, respectively. Prof. Karam is the recipient of an NSF CAREER Award.

Dr. Karam served as the Chair of the IEEE Communications and Signal Processing Chapters in Phoenix in 1997 and 1998. She also served as an Associate Editor of the IEEE Trans. Image Processing from 1999 to 2003 and of the IEEE Signal Processing Letters from 2004 to 2006, as a member of the IEEE Signal Processing Society's Conference Board from 2003 to 2005, and as a member of the IEEE Signal Processing Society's Technical Direction Board from 2008 to 2009. Prof. Karam served as the lead guest editor of the IEEE Journal on Selected Topics in Signal Processing, Special Issue on Visual Media Quality Assessment and as a Technical Program Chair of the 2009 IEEE International Conference on Image Processing. She co-founded the International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM) and the International Workshop on Quality of Multimedia Experience (QoMEX). She currently serves on the editorial boards of the IEEE Trans. Image Processing and the Foundations and Trends in Signal Processing journals. She is the General Chair of the 2011 IEEE Signal Processing Society's DSP and SPE Workshops and of the 2016 IEEE International Conference on Image Processing (IEEE ICIP). She is an elected member of the IEEE Circuits and Systems Society's DSP Technical Committee, the IEEE Signal Processing Society's IVMSP Technical Committee, and the IEEE Signal Processing Society's Education Technical Committee.

Martin Reisslein 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. He received the Ph.D. in systems engineering from the University of Pennsylvania in 1998.

He is a Professor in the School of Electrical, Computer, and Energy Engineering at Arizona State University (ASU), Tempe. 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. He currently serves as Associate Editor for the IEEE/ACM Trans. Networking and for Computer Networks. His research interests are in the areas of multimedia networking, optical access networks, and engineering education. He is a senior member of the ACM and IEEE.