# 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. 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 and single-layer streams have similar levels of traffic variability.**

***Index Terms*—Coarse grain scalability, H.264 SVC, medium grain scalability, rate-distortion performance, rate variability-distortion performance, 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]–[25], and transport mechanism development, see for instance [26]–[30], have been conducted. Similarly, the network transport of H.264 SVC scalable video has begun to attract significant research interest, see for instance [1], [31]–[36]. A traffic model for H.264 SVC temporal scalability of the base layer and the complete enhancement layer has been proposed in [37]; similar to [15] the study [37] did not consider medium grain scalability through partitioning of the complete enhancement layer.

To the best of our knowledge, no prior analysis of the traffic variability, which is a key concern for video transport [38], 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], [39], [40] for short video sequences up to a few hundred frames. In contrast, we examine both the RD and traffic variability 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 [41] 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

Please direct correspondence to M. Reisslein.

This work was supported in part by the National Science Foundation through Grant No. CRI-0750927.

R. Gupta was with Arizona State University when this work was conducted and is now with Qualcomm Inc., San Diego, CA.

P. Seeling is with the Dept. of Computing and New Media Technologies, University of Wisconsin—Stevens Point, email: `pseeling@ieee.org`.

A. Pulipaka, M. Reisslein, and L.J. Karam are with the School of Electrical, Computer, and Engineering Eng., Arizona State University, Goldwater Center MC 5706, Tempe AZ 85287-5706, phone: (480)965–8593, fax: (480)965–8325, email: {`Rohan.Gupta, Akshay.Pulipaka, reisslein, karam`}`@asu.edu`, web: `http://mre.faculty.asu.edu`.

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. Section IV presents and discusses the CGS and MGS video quality and traffic characteristics. Section V 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 [42]–[44] 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 lowest video data 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 [45], 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 [46], 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-A1.

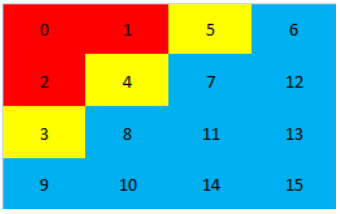



Fig. 1. Illustration of allocation of transform coefficients of a $4 \times 4$ macroblock to MGS layers for weight vector $\mathbf{W}$ = [3,3,10]. Coefficients with indices 0–2 constitute MGS layer 1, while coefficients with indices 3–5 constitute MGS layer 2, and coefficients with indices 6–15 constitute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

### 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.

*a) 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 $4 \times 4$ 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 $4 \times 4$ 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 $8 \times 8$ macroblocks, there are two approaches in H.264 MGS. One approach is to divide a given $8 \times 8$ macroblock into four $4 \times 4$ submacroblocks and to split the coefficients of each $4 \times 4$ submacroblock according to the above approach illustrated in Fig. 1. This submacroblock approach is usually employed in conjunction with context-adaptive variable length coding (CALVC) entropy coding [47].

When the other main entropy encoding scheme, context-based adaptive binary arithmetic coding (CABAC) [48], is used, the $8 \times 8$ macroblock is not subdivided. Instead, the above approach for splitting the transform coefficients of a $4 \times 4$ macroblock is extended to the $8 \times 8$ macroblock by multiplying each weight $w_i$ by a factor of four. That is, the coefficients are considered in the conventional zig-zag order and $4 \cdot w_m$

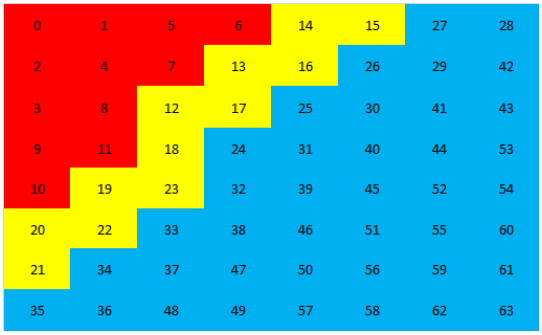



Fig. 2. Illustration of extension of splitting of transform coefficients into MGS layers without subdivision of an $8 \times 8$ macroblock. For the example weights $\mathbf{W} = [3, 3, 10]$, the first $4 \cdot w_1 = 4 \cdot 3$ coefficients form MGS layer 1, the next $4 \cdot 3$ coefficients form MGS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

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.

*b) 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. Two common approaches for dropping NALUs are as follows:

$(i)$ 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 [49] and as JSVM QL in [50], [51] and implemented in the reference Joint Scalable Video Model (JSVM) software [52], 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 for all video frames, followed by the NALUs with lower priority IDs until the target bit rate is reached.

## III. Evaluation Set Up

### A. Video Sequences

We present evaluation results for the following representative videos with 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 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) including the commercials.
- The first half our of the movie *Citizen Kane* (54,000 frames), a drama/mystery genre video.
- The first half hour of the movie *Gandhi* (54,000 frames), a biography/drama/history genre video.
- The first half hour of the movie *Indiana Jones* (54,000 frames), an action/adventure genre video.
- The first half hour of the movie *Die Hard* (54,000 frames), an action/crime/drama/thriller genre video.

(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), 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 values of MeQP, which are used for determining the Lagrangian parameters for motion estimation and mode decision of key pictures, to values smaller than the QP values. We used the CABAC coding scheme and enabled the $8 \times 8$ transform.

Following the recommendations of [53] 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 [53], 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 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

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

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

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:

- Mean frame size $\bar{X} = \frac{1}{M}\sum_{m=1}^{M} X_m$
- Variance of frame size $\sigma^2 = \frac{1}{M-1}\sum_{m=1}^{M}\left(X_m - \bar{X}\right)^2$
- Coefficient of variation (CoV) of frame size $\mathrm{CoV} = \sigma/\bar{X}$

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. Plotting the CoV as a function of the average PSNR video quality gives the rate variability-distortion (VD) curve [23]. 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. Results and Discussion

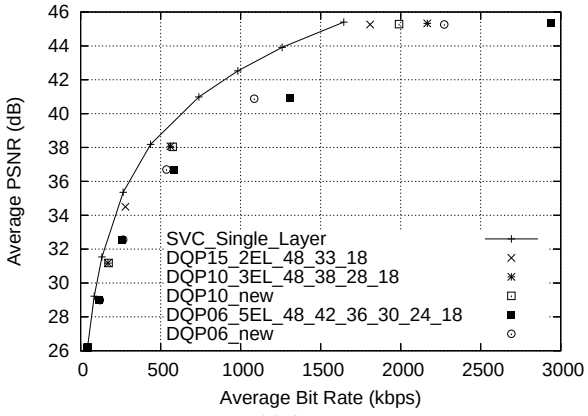
### A. CGS Traffic and Quality Characteristics

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 [54].

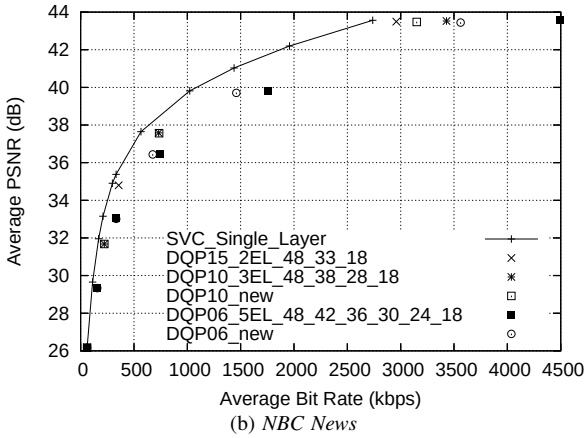
*1) CGS Rate-Distortion (RD) Performance:* In Fig. 3, we plot the RD performance for H.264 CGS encodings and compare with single-layer H.264 SVC encodings. The plotted average PSNR video quality and the bit rate values represent the aggregate of the base layer and applicable enhancement layer(s). For instance, for the 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 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 the modification of the reference encoder 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 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*.

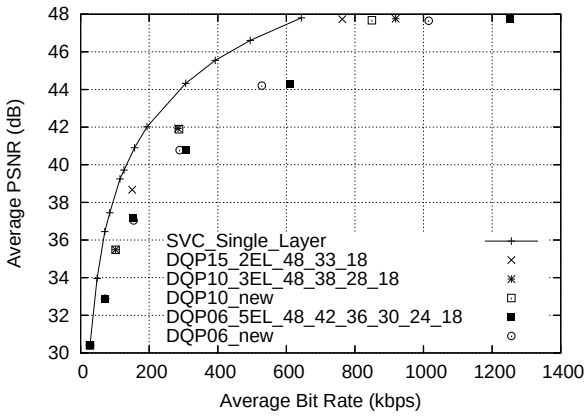
Overall, these results confirm the observations made in [14] for short test sequences in that the H.264 SVC CGS bit rate overhead compared to single-layer encodings increases with decreasing DQP values and correspondingly larger numbers of enhancement layers. A bit rate overhead within 10% to 30% can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




(a) *Sony*




(b) *NBC News*




(c) *Star Wars*

Fig. 3. RD curves of H.264 SVC CGS (with modification permitting more than three dependency layers) and H.264 SVC single-layer encodings.

### B. CGS Rate Variability-Distortion (VD) Performance

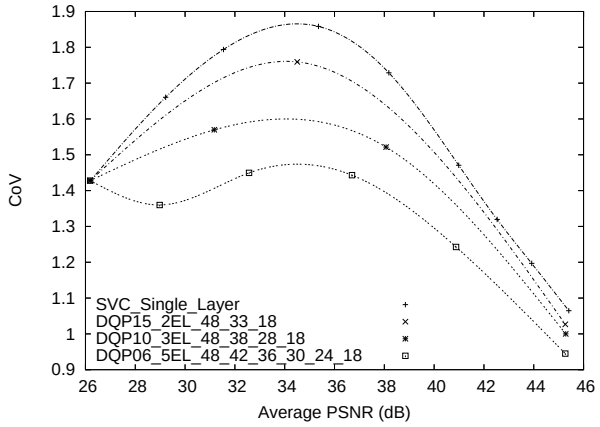
In Figure 4, we compare the VD curves of H.264 SVC single-layer encodings with the curves obtained by connecting

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.

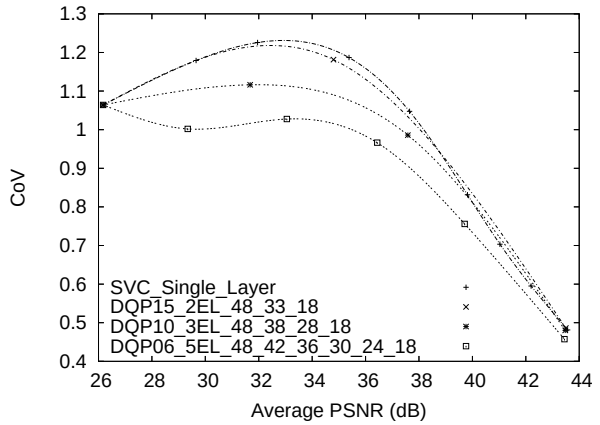
| *SONY demo* | | | | | | |
|---|---|---|---|---|---|---|
| DQP | $B$ | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ |
| 15 | 0 | 15.0 | 10.0 | - | - | - |
| 10 | 0 | 28.0 | 28.2 | 18.17 | - | - |
| 06 | 0 | 43.5 | 51.6 | 49.9 | 43.4 | 34.93 |
| *NBC News* | | | | | | |
| DQP | $B$ | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ |
| 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 |

TABLE II
COV VALUES OF INDIVIDUAL CGS LAYERS (ENCODED WITH MODIFICATION PERMITTING MORE THAN THREE DEPENDENCY LAYERS).

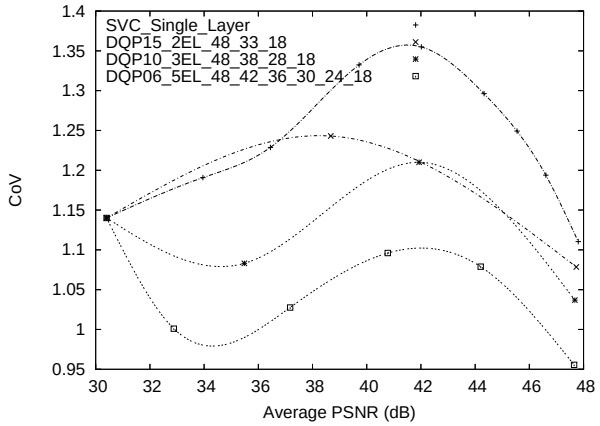
| CoV for *SONY demo* | | | | | | |
|---|---|---|---|---|---|---|
| DQP | $B$ | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ |
| 15 | 1.43 | 1.83 | 0.92 | - | - | - |
| 10 | 1.43 | 1.63 | 1.52 | 0.84 | - | - |
| 06 | 1.43 | 1.33 | 1.54 | 1.45 | 1.08 | 0.74 |
| CoV for *NBC News* | | | | | | |
| DQP | $B$ | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ |
| 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 |



(a) *Sony*



(b) *NBC News*



(c) *Star Wars*

Fig. 4. VD curves of H.264 SVC CGS (with modification permitting more than three dependency layers) and H.264 SVC single-layer encodings.

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.

We observe from Fig. 4 that the VD curves of the CGS encodings with DQP = 15 exhibit the same trends as the single layer encodings (previously examined in [23]) of first increasing, peaking, and then decreasing CoV values. On the other hand, for 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 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 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 CoVs, 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

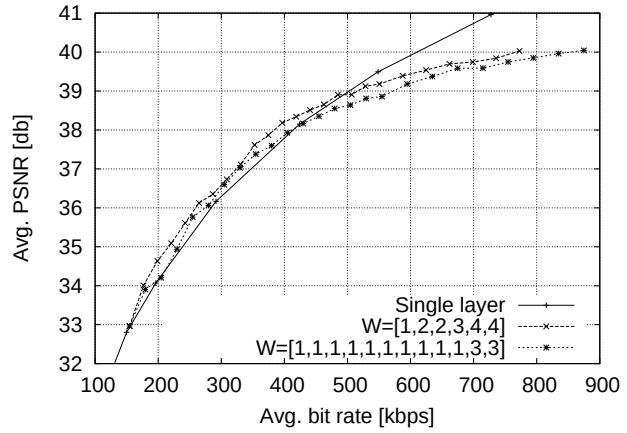



Fig. 5. RD curve for *Sony* sequence for different MGS weights $\mathbf{W}$; $B = 35$, $E = 25$, fixed.

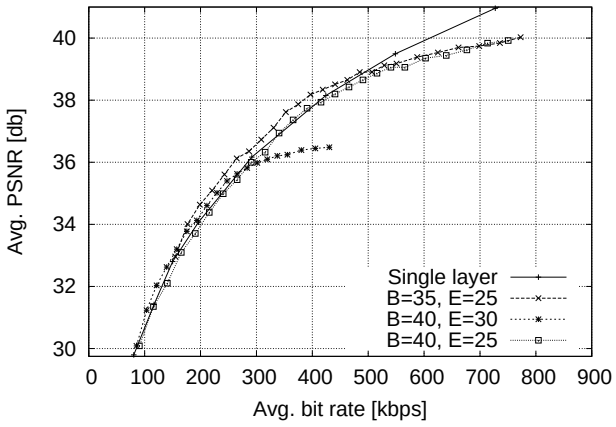



Fig. 6. RD curve for *Sony* sequence with $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ MGS weights and different base and enhancement layer QP values.

decreases only slightly.

### C. MGS Traffic and Quality Characteristics

In this section, we examine the traffic and quality characteristics of H.264 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 representative video sequences from our extensive encoding and traffic studies.

### D. MGS weights

Figure 5 shows the RD performance 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. We observe from Fig. 5 that the RD performance for the different MGS weights is nearly the same with the MGS weights $\mathbf{W} = [1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 3]$ 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 IV-G.

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

### E. Quantization Parameter

Figure 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 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

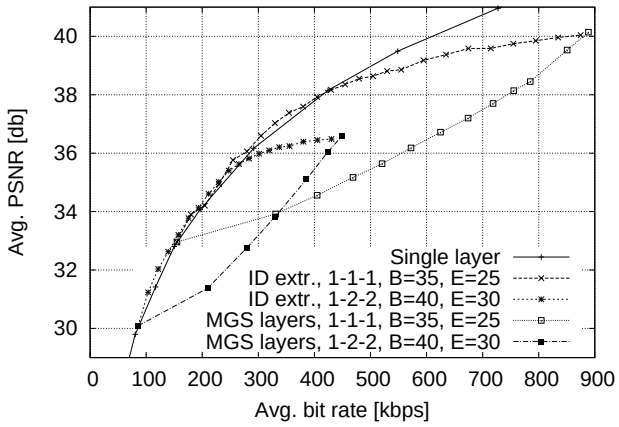



Fig. 7. RD curve for *Sony* sequence for different extraction methods ("1-2-2" denotes $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ MGS weights and "1-1-1" denotes $\mathbf{W} = [1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 3]$ MGS weights).

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.

### *F. Extraction Method*

In Figure 7 we compare the RD curves of MGS with the priority ID based extraction and the MGS layer based extraction. 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, the base layer with one MGS layer, 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 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).

Contrasting these results for the extraction methods with the results for the base and enhancement layer quantization parameters $B$ and $E$ in Sections IV-E, we observe that the extraction method has a relatively large impact on the RD performance. In particular, the extraction method strongly affect 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 IV-D, we observe the relatively minor impact of the MGS weights $\mathbf{W}$. Overall, further research on computationally efficient extraction mechanisms that maximize the RD performance, such as [55]–[57], is therefore important.

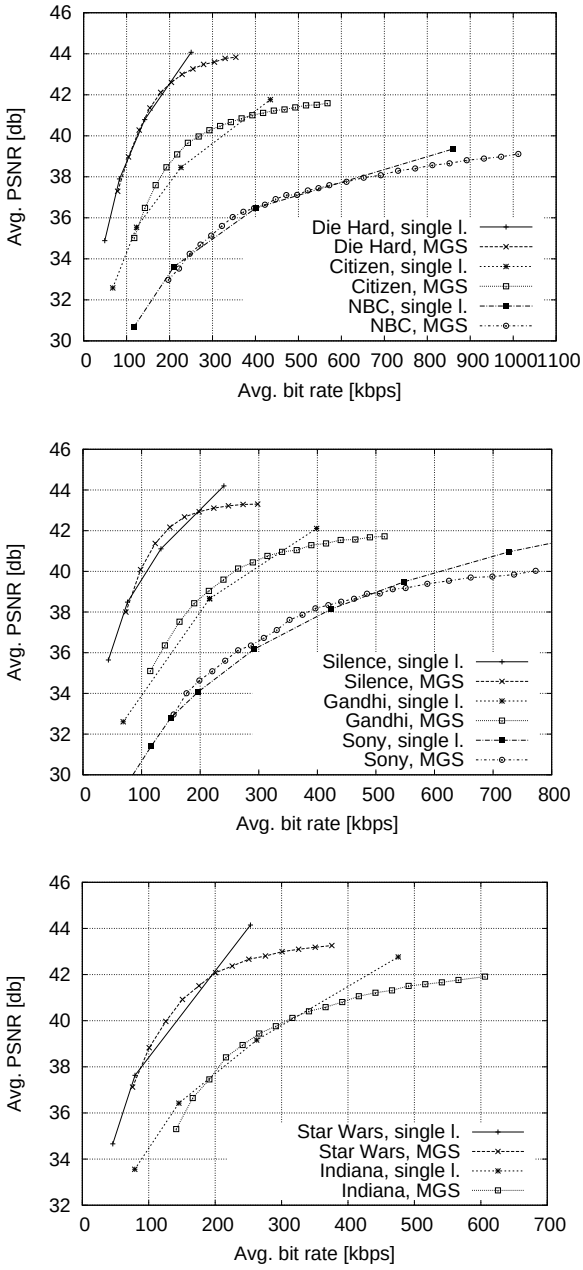
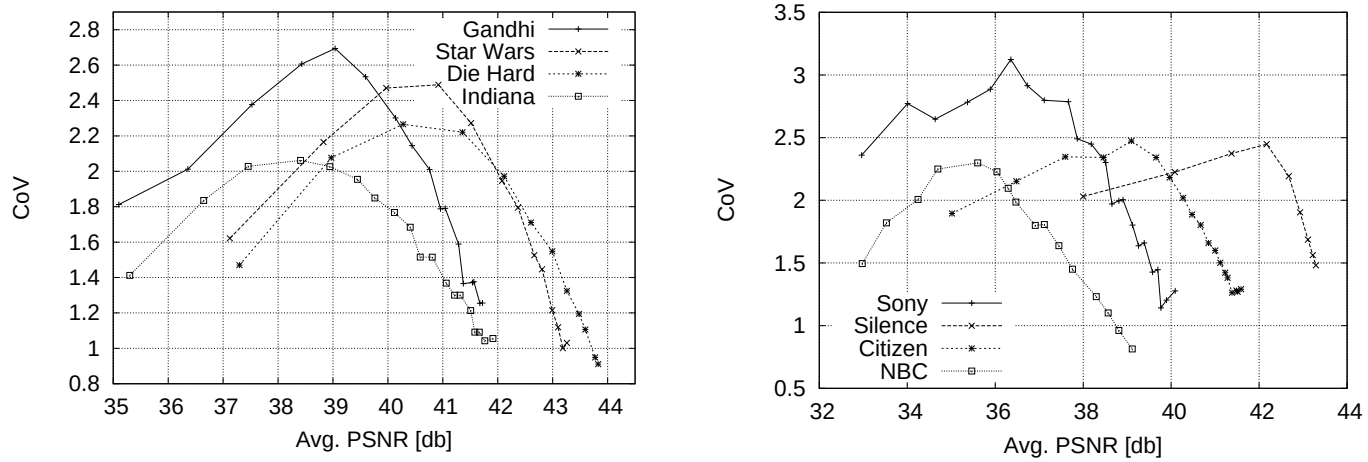


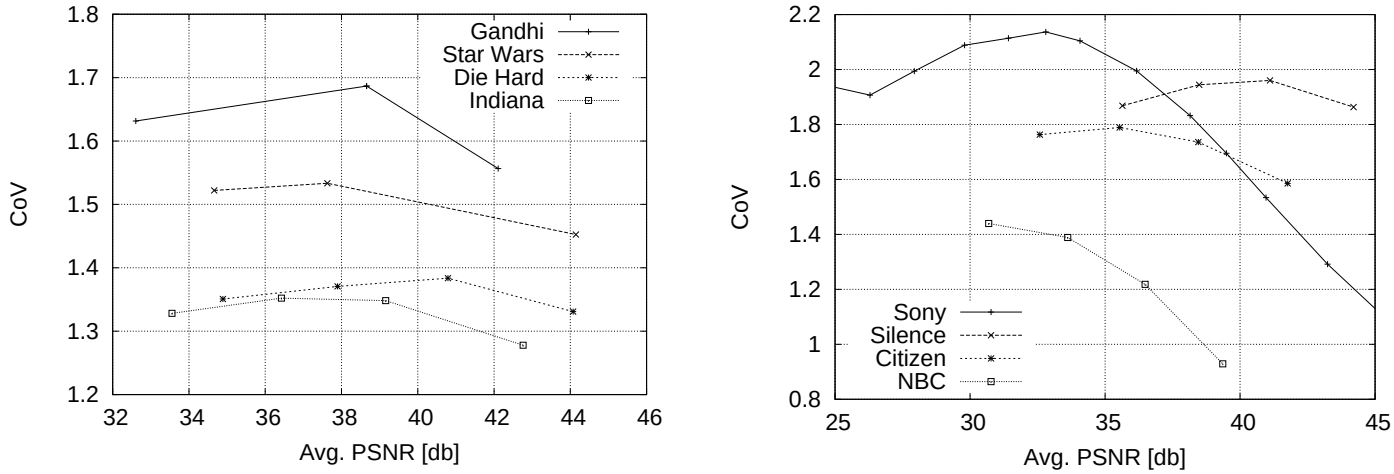

Fig. 8. RD curves for a wide range of test sequences with MGS weights $\mathbf{W} = [1, 2, 2, 3, 4, 4]$ and quantization parameters $B = 35,\ E = 25$ with priority ID based extraction.

### *G. MGS Rate-Distortion (RD) Performance*

In Figure 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




(a) MGS streams extracted with priority IDs from encoding with $B = 35,\ E = 25$ and $\mathbf{W} = [1, 2, 2, 3, 4, 4]$.




(b) Single-layer streams

Fig. 9. Comparison of frame time scale VD curves for MGS streams using priority ID based extraction with VD curves for single layer SVC streams.

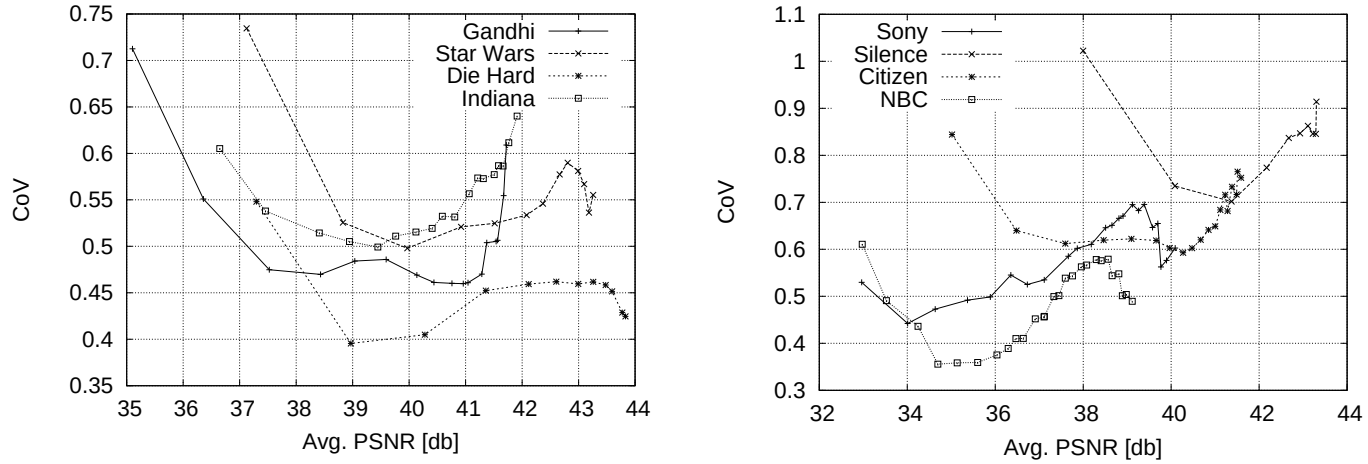
range of video content genres. In particular, we observe that for this wide set of test videos, the RD curve of MGS is very close or slightly exceeds the RD curve of the single layer encoding at low to moderate PSNR video qualities and corresponding bit rates. The improved RD performance with MGS is achieved through the RD optimized extraction which selectively discards MGS layer NALUs from selected frames if the NALUs provide relatively small PSNR improvements for their size (in bits).

At the upper end of the MGS RD curve 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.

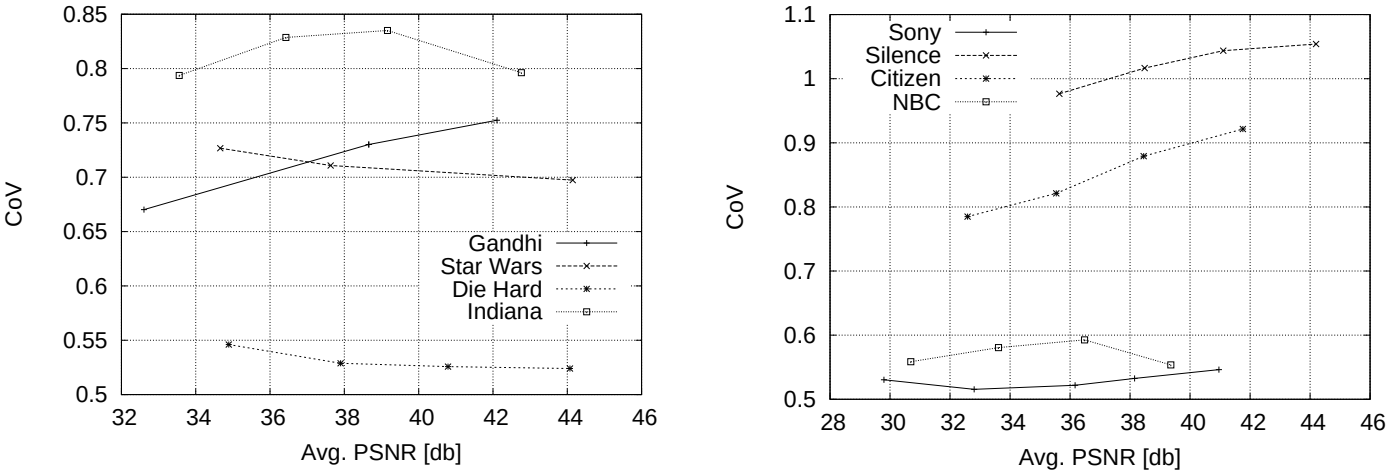
### *H. MGS Rate Variability-Distortion (VD) Performance*

In Figure 9, we compare the frame time scale rate variability-distortion (VD) curves of H.264 MGS and single-layer encodings for the wide set of test sequences. We plot the VD curves of the MGS streams extracted based on priority IDs in the top two plots in the figure, while the VD curves for the corresponding single-layer encodings are given in the bottom two plots. We also examined the VD curves for the MGS streams extracted based on MGS layers (not plotted here due to space constraints) 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 Figure 9 that, at the frame time scale, the MGS streams with priority ID based extraction have significantly higher bit rate variabilities than the corresponding single-layer encodings. We also observe that the relative order of the single-layer and MGS VD curves is typically the same, i.e., videos with relatively higher variability of the single-layer stream have relatively higher variability of the MGS layer streams. Videos with a high degree of heterogeneity in




(a) MGS streams extracted with priority IDs from encoding with $B = 35,\ E = 25$ and $\mathbf{W} = [1, 2, 2, 3, 4, 4]$.




(b) Single layer streams

Fig. 10. Comparison of GoP time scale VD curves for MGS streams using priority ID based extraction with VD curves for single layer SVC streams.

the levels of motion and texture complexity result in higher variability in the streamed frame sizes. For instance, *Sony* and *Gandhi*, 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 lower CoV values.

The MGS encoding and priority ID based extraction adds substantially to the frame time scale traffic variability. For instance, for *Die Hard*, the maximum CoV is increased from 1.38 for the single-layer stream to 2.27 for the MGS stream, while for *Sony*, the maximum CoV increases from 2.14 to 3.12. For 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.

We further observe from Figure 9 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. 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 IV-G. 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.

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 for both the MGS and single-layer streams compared to the frame time scale considered in Fig. 9. Moreover, the CoV values of the GoP sizes of the MGS streams are within similar ranges as for the single-layer streams. Thus, the added traffic variability of the MGS streams compared to the single-layer streams that was observed from Fig. 9 has effectively been eliminated. This implies that the added variability in Fig. 9 that was introduced by the selective inclusion of MGS layer NALUs for select frames has mainly added variability among the frames within a GoP.

Overall, these results indicate that, at the frame time scale, H.264 SVC MGS streams have significantly higher traffic variability than the corresponding H.264 SVC single-layer streams in the quality (bit rate) region where the RD efficiency of MGS is very close to or exceeds the RD efficiency of the

single-layer streams. 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. On the other hand, smoothing the video traffic to the GoP time scale effectively reduces the variability of MGS traffic to similar levels as experienced for single-layer video smoothed over GoPs.

## V. 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.

We found that the RD curves of H.264 SVC MGS encodings 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. We found that 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; smoothing over each GoP effectively removes this added variability. We also 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.

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. Another direction is to examine how transport mechanisms for both wired and wireless networks can efficiently transport the highly variable MGS streaming traffic.

## Acknowledgment

We are grateful to Dr. Heiko 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. of IEEE Int. Symp. on 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 Transactions on Broadcasting*, 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. of Advanced Satellite Multimedia Systems Conference (ASMA) and Signal Processing for Space Communications 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 Systems Journal*, 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. of IEEE Vehicular Technology Conference (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 Transactions on Broadcasting*, 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. of IEEE Symp. on World of Wireless Mobile and Multimedia Networks (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. of Packet Video*, 2007, pp. 26–35.

[10] D. Hu and S. Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Transactions on Wireless Communications*, 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 Transactions on Mobile Computing*, 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 Transactions on 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 Transactions on Mobile Computing*, 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 Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[15] G. V. der Auwera, P. T. David, M. Reisslein, and L. Karam, "Traffic and quality characterization of the H.264/AVC Scalable Video Coding extension," *Advances in 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 Transactions on Broadcasting*, 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 Transactions on Broadcasting*, vol. 48, no. 4, pp. 337–347, Dec. 2002.

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

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

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

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

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

[23] G. V. 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 Transactions on Broadcasting*, vol. 54, no. 3, pp. 698–718, Sep. 2008.

[24] 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.

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

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

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

[28] E. Gurses and O. Akan, "Multimedia communication in wireless sensor networks," *Annals of Telecommunications*, vol. 60, no. 7–8, pp. 799–827, July/August 2005.

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

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

[31] 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. of IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2009, pp. 1–6.

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

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

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

[35] J. Xu, R. Hormis, and X. Wang, "MIMO video broadcast via transmit-precoding and SNR-scalable video coding," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 456–466, Apr. 2010.

[36] 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. of Int. Symp. on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2009, pp. 114–121.

[37] D. Fiems, B. Steyaert, and H. Bruneel, "A genetic approach to Markovian characterisation of H.264/SVC scalable video," *Multimedia Tools and Applications, in print*, 2011.

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

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

[40] C. Mazataud and B. Bing, "A practical survey of H.264 capabilities," in *Proc. of Communication Networks and Services Research Conference (CNSR)*, 2009, pp. 25–32.

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

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

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

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

[45] 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 Transactions on Multimedia*, vol. 11, no. 7, pp. 1209–1224, Nov. 2009.

[46] 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 Transactions on Broadcasting, in print*, 2011.

[47] 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," in *Picture Coding Symposium (PCS)*, Lisbon, Portugal, Nov. 2007.

[48] D. Marpe, H. Schwarz, and T. Wiegand, "Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 620–636, Jul. 2003.

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

[50] 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 Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1186–1193, Sep. 2007.

[51] E. Maani and A. K. Katsaggelos, "Optimized bit extraction using distortion modeling in the scalable extension of H.264/AVC," *IEEE Transaction on Image Processing*, vol. 18, no. 9, pp. 2022–2029, Sep. 2009.

[52] J. Reichel, H. Schwarz, and M. Wien, "Joint scalable video model 11 (JSVM 11)," Joint Video Team, Doc. JVT-X202, Jul. 2007.

[53] 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

[54] 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. of IEEE Consumer Communications and Networking Conference (CCNC)*, Jan. 2010.

[55] H. Lee, Y. Lee, D. Lee, J. Lee, and H. Shin, "Implementing rate allocation and control for real-time H.264/SVC encoding," in *Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, Jan. 2010, pp. 269–270.

[56] R. Li, J. Sun, and W. Gao, "Fast weighted algorithms for bitstream extraction of SVC Medium-Grain scalable video coding," in *Proc. of IEEE Int. Conf. on Multimedia and Expo (ICME)*, 2010, pp. 249–254.

[57] L. Wei, G. Sen, C. Xu, and Z. Jihong, "SVC bitstream extraction based on the importance of MGS slice," in *2nd International Conference on Industrial and Information Systems (IIS)*, vol. 1, Jul. 2010, pp. 148–151.