A Temporally Scalable Video Codec and its Applications to a Video Conferencing System with Dynamic Network Adaption for Mobiles

Hans L. Cycon, *Senior Member*, IEEE, Thomas C. Schmidt, *Member*, IEEE Matthias Wählisch, *Member*, IEEE Detlev Marpe, Senior *Member*, IEEE and Martin Winken

Abstract — *In this paper we present a multipoint video conferencing system that adapts to heterogeneous members including mobiles. The system is built upon a low complexity scalable extension of our H.264 codec DAVC, and a congestion-aware dynamic adaptation layer. We show that our temporally scaled video codec DSVC has the same RD performance as the non-scaled version with comparable configuration. We achieve this by QP cascading, i.e., assigning gradually refining quantization parameters to the declining temporal layers. The different quantization of frames does not lead to visually distinguishable quality fluctuations. We also present and analyze a mobile-compliant version of DSVC at reduced complexity that still admits comparable performance. Finally, we report on early work of dynamic layer tuning. Derived of delay variation measures, senders exploit scalable video layering to adapt the video transmission to varying network conditions. Initial results indicate that video performance remains close to optimal[1].*

Index Terms — **Mobile video conferencing, scalable video coding, heterogeneous conferencing environments, network-adaptive group communication.**

I. INTRODUCTION

Video applications in the Internet exhibit significant growth in several market segments. Conversational systems of video conferencing and immersive telepresence, video-assisted games, or high-definition IPTV broadcasting are more and more enabled by flexible and powerful nodes connected to the Internet at high speed, but also by recent advances in video coding and processing technologies. In addition, the number of devices capable of displaying moving images at reasonable quality is rapidly growing due to popular consumer devices such as smartphones and game boxes.

Today, a diverse variety of video applications are transmitted via networks with a wide range of capabilities, and displayed by largely heterogeneous receivers. Prior to transmission, video data need to be compressed and decoded with high-performance real-time video codecs that admit high flexibility in bit rate. The most efficient codec in sense of rate distortion (RD) performance has been defined in the H.264/AVC video coding standard [1], [2]. To work efficiently in such heterogeneous environments, suitable video codecs need to have some extended scalability properties in addition to high (RD) performance. Scalability in this context refers to the removal of parts of the video bit stream in order to adapt it to the varying terminal capabilities or network conditions [3].

Traditional video codecs already introduced some scalability features, but they came along with a significant loss of coding efficiency, as well as a large increase in decoder complexity as compared to the non scalable versions [19], [20], [21]. To overcome these problems, the scalable successor SVC of H.264/AVC video coding standard was defined in 2007 [4]. SVC enables the transmission and decoding of partial bit streams to generate video flows with temporal, spatial, and quality scalability. The fully implemented SVC, however, also comes with some increases of complexity and bit rate for the same fidelity as compared to single layer coding.

In this paper we present a real-world video conferencing system built upon a scalable extension of our H.264-implementation DAVC [22]. We can show that our temporally scaled video codec *DSVC* has the same RD performance and complexity as the non-scaled version with comparable configuration. We achieve this by QP cascading i.e. assigning to the declining temporal layers gradual refining quantization parameters. The different quantization of frames does not lead to visual distinguishable quality fluctuations. This codec is also part of our conferencing client for mobile devices, where it operates at reduced complexity with the specific need for dynamic scaling. Network adaptive dynamic scaling in our system is shown to solely work under the control of senders based on generally available delay variation measures. We discuss initial measurement results of this ongoing work on self-adaptive scaling.

The remainder of this paper is organized as follows. In

section II, we explore the problem space of scalable adaptive conferencing and review related work. Section III is dedicated to describing the design of the baseline codec, as well as its mobile companion. Section IV presents a detailed analysis of our scalable codec. Section V briefly exposes the conferencing application, and the network-adaptive video scaling. Finally, in Section VI we conclude with an outlook.

II. PROBLEM STATEMENT AND RELATED WORK

A. Heterogeneous Video Conferencing with Mobiles

Video conferencing is a highly conversational application and thus bound to rigid real-time constraints. Any component involved in preparing, transmitting, or displaying the audio-visual streams must be carefully controlled to not exceed performance limits. Such limits are threatened by resource exhaustion of processing capacities at end systems, as well as by network overloads. Conferencing demands are very different from those of streaming applications, even when transmitted in real-time, where many performance violations may be compensated by play-out buffers, elastic timing, etc.

In heterogeneous settings where capabilities are unevenly distributed among end-system and network connectivity is under-provisioned or temporally degraded, video conferencing performance at weak end points may easily drop down to an alienating experience for users. Audio-visual flows may stall or even come to a complete halt, whenever frames cannot be delivered in time for play-out. Conventional systems must adopt their resource demands (and thus quality) to comply with the lowest capacities available in a conference, or require a transcoding service, e.g., by an MCU, to reduce individual streams.

Challenges tighten when mobiles join a heterogeneous conference. For handhelds, bandwidth as well as processing and battery capacities commonly remain at least one order of magnitude below those of fixed systems. Using a highly optimized H.264 software codec, a mobile smartphone can reliably and simultaneously encode and decode a QCIF video at about 15 fps [5], resulting in data rates that comply to 3GPP offers. Thus, special treatment must be added to include mobile end systems into a conference of conventional quality at 30 fps of 384x288 (i.e. CIF) resolution.

B. Scalable Coding

Scalable coding of video flows is a promising approach of adapting data rates to capacities in heterogeneous and mobile environments [6]. In general, a video bit stream is called scalable when parts of it can be removed in a way that the resulting sub-stream forms another valid bit sequence for some target decoder. The sub-stream represents the source content with a reconstruction quality that is less than that of the complete original data. Bit streams that do not provide this property are referred to as single-layer. The usual modes of scalability are temporal, spatial, and quality scalability.

Current standard H.264/AVC-decoders do not support spatial scalability, as this requires support of dedicated extension as defined in the Scalable Video Codec (SVC) amendment [4], while temporal and quality scaling can be achieved in compatibility to the widely deployed H.264/AVC [1] standard players.

Temporal scalability describes cases in which subsets of the bit stream represent the source content with a reduced frame rate (temporal resolution) [3]. A sequence of temporal layers consists of the base layer and temporal enhancement layers. Any bit stream obtained by a complete sequence of temporal layers starting from base layer to a suitable enhancement layer forms a valid input for the given decoder. Obviously, if the number of enhancement layers is enlarged, the bit rate and the frame rate of the video stream also increase.

Reference pictures form the basis for uni- or bidirectional predictions at the enhancement layer pictures. Conceptually, H.264/AVC allows for coding of picture sequences with arbitrary temporal dependencies. Following our real-time objective in conferencing, we consider only hierarchical prediction structures, here, where reference pictures are always temporally preceding the enhancement layer pictures (see Fig. 1). This implies to adhere to unidirectional predictions, which cause zero structural delay. In general, such low-delay structures decrease coding efficiency. For a general discussion we refer to [3].
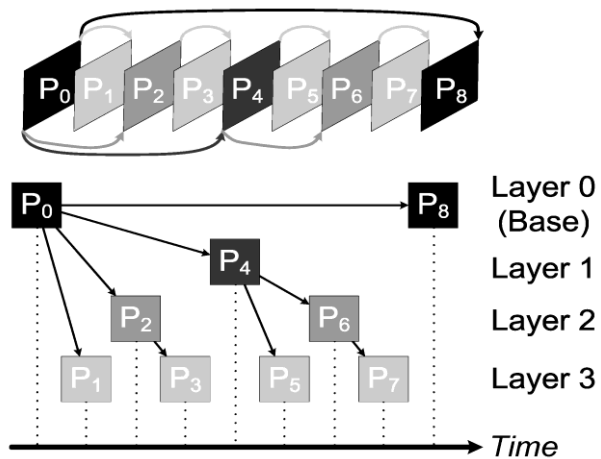

Fig. 1. Unidirectional dyadic hierarchical prediction structure with 4 temporal layers.

C. Adaptive Video Distribution

With the ability to scale video communication, a conferencing application may adapt streams to individual participants without the burden of transcoding. Layers can be selected or omitted statically according to initial media negotiations, or dynamically in reaction to network and runtime conditions. Dynamic adaptation faces the problems of reliably detecting network conditions and to enable a sender reaction in time so that media data neither accumulate in buffers or drop, nor under-utilize the available transmission resources. Finally, adaptation rates and algorithms need to remain stable and resistant against oscillating states even when network capacities fluctuate.

Previous work on adaptive layering has mainly focused on streaming scenarios. PALS [7] introduces a receiver-driven

approach to a layered peer-to-peer video concast. Receivers monitor sender performance and attempt to maximize throughput from multiple sources layer-wise. Another method of selecting peers from heterogeneous neighbors according to their support of quality layers is presented in [8]. In a progressive quality adaption, the authors dynamically select temporal layers according to a continuously measured network throughput at receivers. Baccichet et al. [9] distribute SVC streams via multiple overlay multicast trees with layer-awareness, while adapting to heterogeneous uplink capacities and network conditions. Kofler et al. [10] introduce an RTSP/RTP proxy at WiFi routers to facilitate scalable video distribution to mobiles, while an authentication scheme for SVC videos that enables verification of all possible substreams is presented in [11]. In an early study on multipoint video conferencing, Eleftheriadis et al. [12] compared the performance of a traditional MCU with a corresponding server system based on the SVC reference implementation and could identify a significant reduction in delay and complexity for the SVC. There are solutions, providing stream adaptation by application-layer routers from an infrastructure perspective. To the best of our knowledge however, there is no scalable adaptive solution for infrastructureless, peer-to-peer video conferencing systems.

III. CODEC IMPLEMENTATION

A. The Base Line Codec

DAVC, the core of the videoconferencing system, is a fast, highly optimized H.264/AVC implementation. It is based on the Constrained Baseline profile and is optimized for real-time encoding (as well as real-time decoding) by means of a fast motion-estimation strategy including integer-pel diamond search as well as a fast subpel refinement strategy up to ¼ -pel motion accuracy. Motion estimation includes the choice of several different macroblock (MB) partitions and multiple reference frames, as permitted by the H.264/AVC standard. For choosing between different MB partitions for motion-compensated (i.e., temporal) prediction and MB-based intra (i.e., spatial) prediction modes, a fast rate-distortion (RD) based mode decision algorithm with early termination conditions has been employed. The codec along with the H.264/AVC design also includes some suitable mechanisms to recover quickly from video packet loss [22].
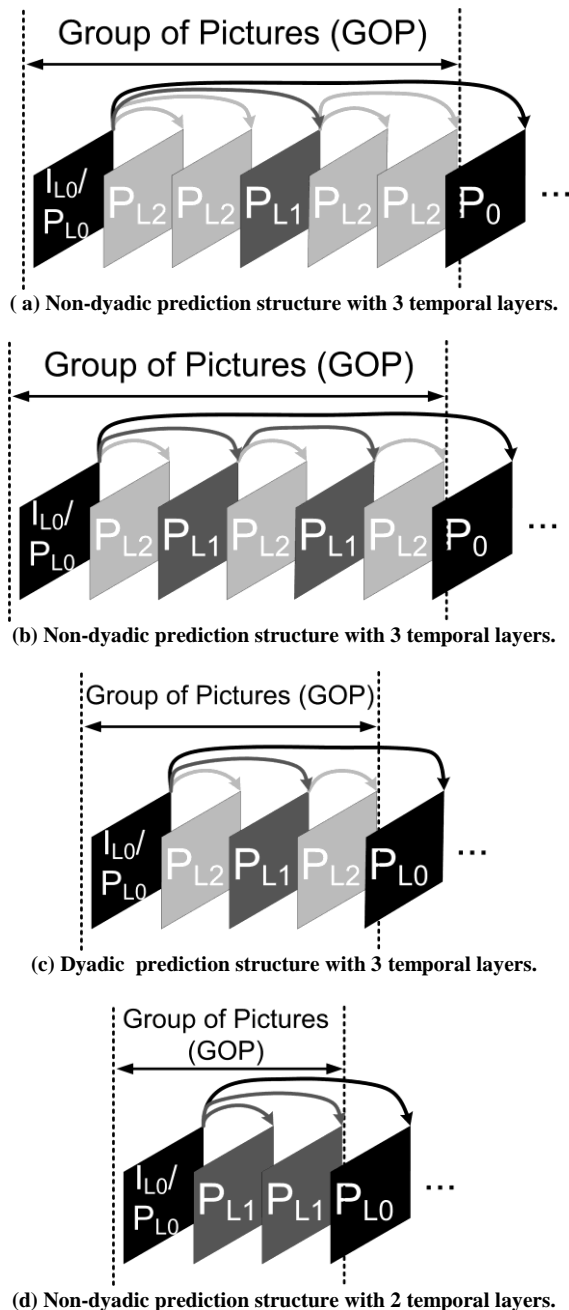
TABLE I

LAYER	BIT RATE			FRAME RATE			GOP (S.FIG.2)
	0	1	2	0	1	2	
3 TL	50%	15%	35%	1/6	1/6	2/3	(a)
	52%	27%	21%	1/6	1/3	1/2	(b)
	63%	17%	20%	1/4	1/4	1/2	(c)
2TL	70%	30%	--	1/3	2/3	--	(d)

Configurations of DSVC - relative bit rates and corresponding frame rates at different prediction structures presented in Fig. 2.

B. Temporal Scalability

Based on DAVC, we developed a temporal scalability structure [13]. The resulting codec, called *DSVC*, provides up to three temporal layers (TL). Depending on the frame partitioning, different bit rates per layer are attained. For the configuration of two enhancement layers, we implemented several decomposition strategies (cf., Fig. 2). To allow for a reduced weight of the base layer, we introduced a combination of both dyadic and non-dyadic layer decomposition. This reduces the data rate at the base layer down to 50 %. Strict dyadic prediction decomposition leads to a higher weight at the base layer, approx. 63 % of the overall bit rate in this case.


(a) **Non-dyadic prediction structure with 3 temporal layers.**

(b) **Non-dyadic prediction structure with 3 temporal layers.**

(c) **Dyadic prediction structure with 3 temporal layers.**

(d) **Non-dyadic prediction structure with 2 temporal layers.**
Fig. 2. Unidirectional hierarchical prediction structures. The symbols P_{Lk} specify pictures at temporal layers with layer identifier k.
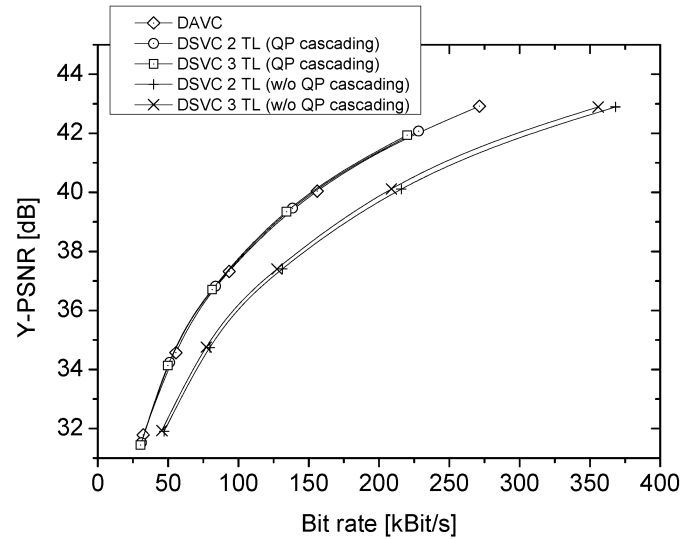
(a) Snapshot of the HHI video sequence "G4"
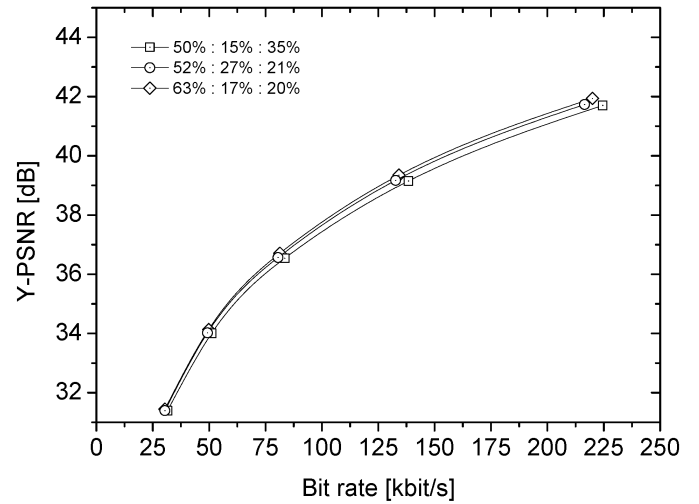
(b) Coding efficiency of DSVC codec (3TL) with different bit rate ratios at the Base layer: 1st enhancement layer: 2nd enhancement layer
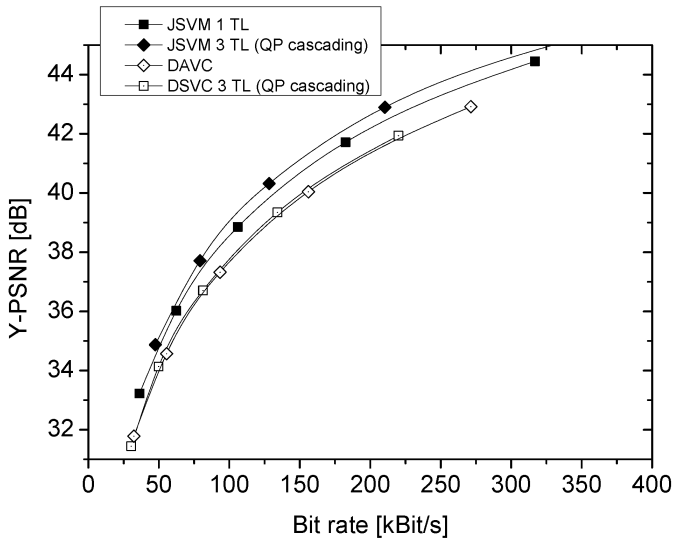
(c) Comparing Coding efficiency of reference codec JSVM with DAVC/DSVC

(d) Coding efficiency of DSVC for different temporal and QP cascading configuration

Fig.3. Encoding quality for the test sequence "G4" in 384x288 resolution at a frame rate of 30 Hz.

We also implemented a configuration of a single enhancement layer by a non-dyadic composition with two non-referenced P-frames at the topmost time level. The frame rate of this base layer as compared to the enhancement layer is 1/3.

The base layer carries approx. 70 % of the overall bit rate. For details of the temporal decomposition and resulting bit and frame rates, we refer to Table 1.

C. Quantization in Layers

The coding efficiency for hierarchical prediction structures are based on the amount of quantization per temporal layer This will be configured by the *quantization parameter QP*. Frames of the temporal base layer should be coded with highest fidelity, since they are used as references for all temporal enhancement layers. Consequently, a larger quantization parameter should be chosen for subsequent temporal layers as the quality of these frames influences fewer pictures [3]. A gradual quantization depending on the layer is called *QP Cascading*.

We have chosen the following strategy for QP cascading (cf., [15]): Based on a given quantization parameter QP_0 for pictures of the temporal base layer, the quantization parameter QP_T for pictures of a given temporal layer with an identifier $T > 0$ is determined by $QP_T = QP_0 + 3 + T$.

D. Real-time Complexity

Real-time compliance of the codec can be measured by evaluating the maximum number of frames that can be encoded per second on target hardware. The single layer

version (DAVC) of the DSVC codec achieves up to 284 frames per second on a standard desktop PC. It slightly outperforms comparable H.264 codecs [22]. Compared to a single layer encoded stream, the hierarchical prediction structure only changes the pointer to the reference frame. Parameters for the motion prediction (in particular the search range) remain unchanged. Temporal scalability thus does not introduce additional overhead and does not decrease the run time performance.

IV. EVALUATION OF ENCODING QUALITY

In this section, we analyze the encoding quality of our DSVC codec.

For reproducibility, we use as input data the HHI video test sequence "G4" (see snapshot Fig. 3(a)) in 384 x288 resolution at a frame rate of 30 Hz. Experiments have been conducted for other test sequences, which achieved similar results. Note, when configured with a single temporal layer (1TL), the DSVC corresponds to our DAVC codec.

A. The DSVC Codec

We evaluate the quality of our DSVC codec by measuring the peak signal-to-noise ratio (PSNR) depending on different bit rates. This quantifies the pure encoding quality, i.e., the distortion of the compressed stream in contrast to the original data without including network disturbances or layer adaptation. The rate distortion (RD) is analyzed for different layer configurations, which reflect the number of temporal layers (TL) used for encoding, effects of QP cascading, and variable bit rates per layer. We disable the intra refresh option.

DSVC is compared with the SVC reference software Joint Scalable Video Model (JSVM) version 9.16 [16]. It is worth noting that the JSVM encoder is designed for complete RD characteristics, but has no real-time abilities in contrast to the DSVC.

Table 2 shows similar results for different resolutions and additional sequences, i.e., there is no loss of RD performance in 3TL case if we use QP cascading in contrast to uniformly assigned QP parameters for the frame levels (i.e. no QP cascading). The test sequences refer to those used in [3].

TABLE 2

TEST SEQUENCE	3TL	3TL (NO QP CASCADING)
G4	-3.6 %	30.6%
KO	-2.2 %	31.5%
SM	-6.7%	27.8%
TC	-1.3%	30.0%
TW	-4.9%	20.7%
AVERAGE	-3.7%	28.1%

BD-Rate relative to single layer coding with resolution 768x576 and frame rate 30 Hz comparing results with and w/o QP cascading

The values give the so-called BD-Rate [26] results relative to single layer coding. So, for the 3TL case with QP cascading,

the bitrate is reduced on average by 3.7%, whereas bitrates increase by 28 % on average without QP cascading.

We did observe in all measurements for the different frames in a sequence that in spite of the relatively large jumps in the quantization parameters (QP) and PSNR qualities, the reconstructed video appears temporally smooth and does not show visually distinguishable fluctuations in quality. Similar effects have been already reported in different configurations [3].

1) Varying Bit Rates per Layer

Fig. 3 (b) plots the coding quality for a scaling in three temporal layers and different bit rate ratios (as displayed in Table 1). All cases include QP cascading.

In general, coding efficiency improves with lower layers carrying a larger portion of the overall data stream. However, rate distortion decreases only slightly when the amount of data is significantly shifted to the highest temporal layer (cf., Fig. 3 (b)). This scaling characteristic is the basis to address bandwidth adaptation by omitting higher layers. Overall, the DSVC encoder produces similar quality results almost independent of the layer decomposition in use. Thus, we may limit our analysis to strict dyadic composition in the case of three temporal layers as displayed in Fig. 2 (c).

2) Comparison with Reference Encoder JSVM.

The coding efficiency of our real-time SVC implementation is compared to the reference encoder JSVM for one and three temporal layers in Fig. 3(c). The DSVC and DAVC encoder achieve a RD performance almost similar to the reference encoder, with decreases only about 1 - 1.5 dB. It should be recalled, though, that JSVM is far from operating in real-time.

3) Effects of Layering & QP Cascading.

Fig. 3 (d) shows the RD performance of the DSVC encoder for a varying number of available temporal layers and different QP options (i.e., with QP cascading and without QP cascading). A configuration with cascaded quantization parameter outperforms an equal quantization of frames in the following ways.

   i. The overall RD performance decreases for multi-layered streams without QP cascading.
   ii. Applying QP cascading yields an RD performance equal to the single layer stream.
   iii. Without QP cascading, the coding quality becomes dependent on the number of layers in use.

Thus we can observe that signaling overheads due to temporal layering are fully compensated by QP cascading in our DSVC implementation.

B. DSVC on Mobiles

The generic DSVC codec has been ported and tuned to the large OS platform. This included adaptation and optimization of the ANSI compliant C version to the wireless MMX instruction set for the mobile systems with target-specific code.

In order to enable real-time encoding performance, even

when appropriately reducing the resolution of the input video to a 240x144 pixel format, the DSVC codec has been restricted to perform motion estimation only for integer-pel displacements. At this coder configuration, our test system, a smartphone with 800MHz SC36410 Kernel, could reliably encode 8 fps while simultaneously decoding 15 fps without CPU exhaustion or packet drop. Corresponding values for another platform were slightly lower at about 6:12 respectively. It should be noted, though, that in the absence of an open API support for video capturing, we had to dedicate noticeable resources to platform-specific video image extraction of display buffers.
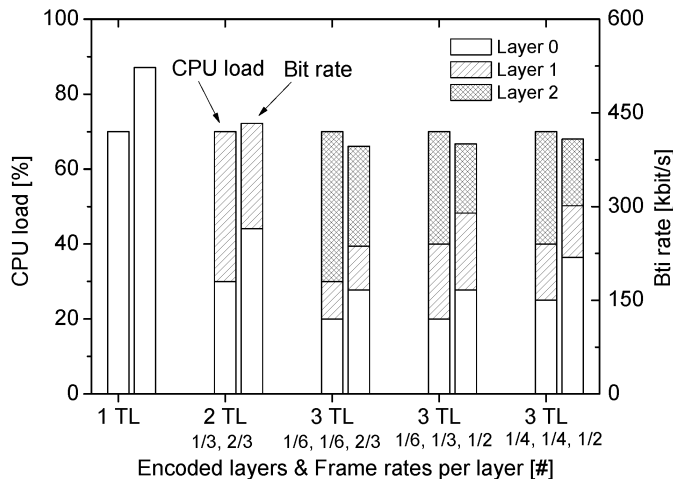

Fig. 4. CPU and bandwidth consumption at mobile device while receiving different layers of G4 384x288@15 Hz

To quantify resource consumption of layered video processing in a standardized, reproducible experiment, we sent the G4 test sequence (384x288@15 Hz) to the test mobile using different layers. Results obtained for the layer configurations in Table 1 are displayed in Fig. 4. The measurements revealed a large scaling factor of 3.5 for processing consumptions, while sustained bandwidth scales down to about 40 %.
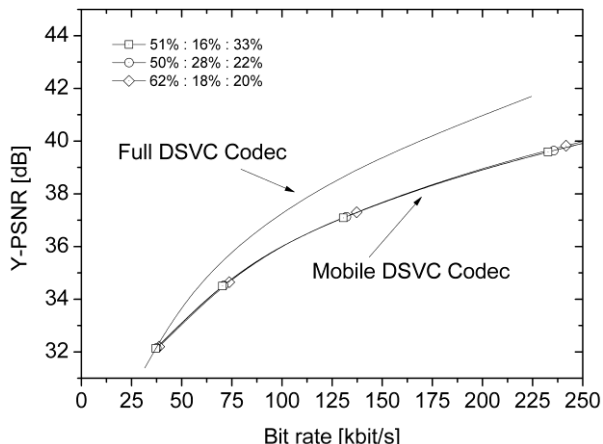

Fig. 5. Encoding quality for the Mobile-DSVC (3 TL) and G4 384x288@30 Hz at different bit ratios between layers compared to full DSVC

Fig. 5 presents RD diagrams for the scaled-down mobile codec and the G4 test sequence in 384x288@30 Hz resolution, compared to the full DSVC codec. Note redundancy with Fig. 3(c) is included for convenience of the reader. Frame rates have been chosen to distribute among layers as displayed in Table 1 for the full codec. Algorithmic down-scaling, however, leads to a slight modification in bandwidth ratios as visible in the legend. Results show a moderate loss of 0 - 2 dB in rate-distortion performance relative to our full DSVC encoder. Still, our mobile-based scalable video encoder produces acceptable video quality when conforming to the tight resource constraints of the mobility regime.

V. NETWORK-ADAPTIVE CONFERENCING

A. The Video Conferencing System

Our work is implemented in a digital audio-visual conferencing system, which is realized as a lightweight multipoint videoconferencing software. It has been designed in a peer-to-peer model as an Internet conferencing tool.
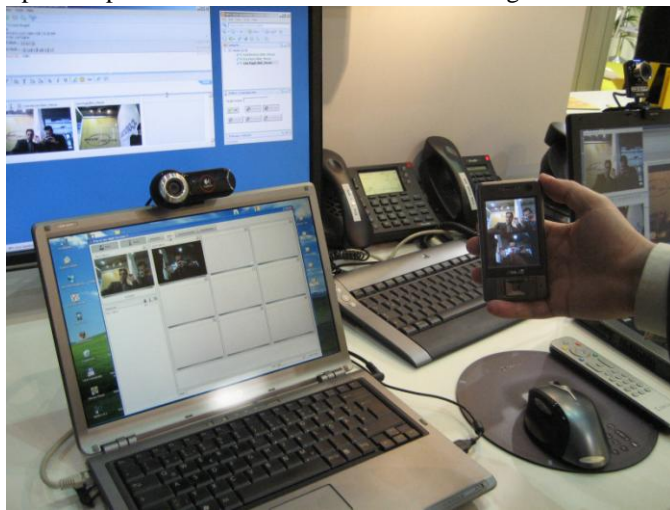

Fig. 6. Video conference integrating heterogeneous members and a mobile at CeBIT. 09.

The Internet conferencing tool works without MCU server on a hybrid P2P network structure [25], and seamlessly complies with different Internet Protocol versions, as well as the conference management signaling of SIP [24] and H.323 legacy MCUs [23]. It is designed for heterogeneous network conditions and components, where the scalability of the codec enables dynamic adaption of data streams to the available capacity at network and the receiving side.

The conferencing system works on desktop computers running common operating systems, as well as on common handhelds [5]. Further porting to popular platforms is ongoing.

B. The Adaptation Layer

Heterogeneous, and in particular mobile clients can signal their system capacities within initial SDP negotiations, while network conditions may change at runtime. The objective of the adaptation layer is to achieve a dynamic scaling of the

video transmission appropriate to network resource changes that are visible at the sender, without explicitly involving receivers. Network congestions or overloads should be quickly detected and immediately answered by a reduction in layers. After a reduced network load has been observed, the activation of layers should act tardy to avoid oscillating transmission rates and to maximize the user experience of continuous, uninterrupted play out.

The temporal scaling in our conferencing system is adjusted according to changes in the available bandwidth between sender and receivers [14]. Traditional bandwidth estimation follows the general observation that the delay continuously increases when links start to become congested [17]. Identifying the transition from an almost constant delay to a continuous growth can approximate the available bandwidth. Consequently, control measurements are taken on delay changes, as was outlined in the seminal work of Van Jacobson [18].
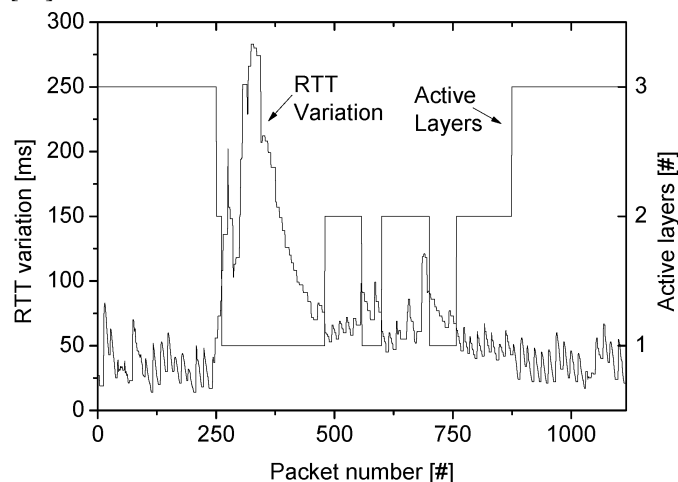

Fig. 7. Video layer adaptation stirred by changing jitter values

Instead of static bandwidth estimation, the layer adaptation of DSVC requires a measurement of change which correlates to second order delay changes, or the variation of jitter. With respect to a minimal deployment support and lightweight mobility regimes, we impose the following constraints. An autonomous estimation of bandwidth-changes at senders should be implemented without additional active measurements in the restricted wireless regimes. The DSVC layer adaption is thus based on sensing changes of the inter-arrival jitter. The inter-arrival jitter is commonly available, e.g., by RTCP. Its variation can likewise serve as an indicator for accumulating queuing delays at routers and access gateways. DSVC senses the transmission of transport stacks and thus does not introduce additional packet overhead or requires functional updates at the receiver.

In detail, we observe the inter-arrival jitter packet wise. Whenever the jitter increases by at least a threshold during a short sequence of packets (e.g., 10), the adaptation reduces the video layers. Layers are added again to the transmission after a long sequence of packets (e.g., 50) is observed that continuously decreases delay variation. Parameters in this

ongoing work are subject to current optimizations. To evaluate our adaption scheme, we set up a test network. The source transmits the "G4" test sequence to a receiver, while the available bandwidth between source and receiver is manually varied. Our results are displayed in Fig. 7 which shows a smooth layer adaptation to jitter conditions in the test network. Video quality remains fluent in our test environment and does not stall even in rapidly changing network conditions.

VI. CONCLUSION & OUTLOOK

Video conferencing in real-world heterogeneous environments needs significant scaling abilities to flexibly adapt to various conditions. Even though the principle methods and tools are around, its realization in a ready-to-use system are still hard to achieve.

In this paper we presented such a conferencing application built upon a fast and efficient temporally scalable video codec, its mobile-based variant, as well as an adaptation layer that dynamically selects appropriate frame rates. In an extensive analysis we could demonstrate the strength of our video solution, but also identify further needs for optimization.

Future work will proceed in two different directions. At first, we will extend experimental analysis and optimizations of the network adaptation to maximize video performance. Second we will extend the scalability of our codec in particular by including spatial scaling options as offered by the SVC standard.

REFERENCES.

[1] "Advanced Video Coding for Generic Audiovisual Services," ITU-T, Tech. Rep. Recommendation H.264 & ISO/IEC 14496-10 AVC, v3, 2005.

[2] J. Ostermann, J .Bormans, P. List, D. Marpe, N. 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, April 2004.

[3] 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, September 2007.

[4] "Advanced Video Coding for Generic Audiovisual Services, Annex G," ITU-T, Tech. Rep. Recommendation H.264 & ISO/IEC 14496-10 AVC/Amd.3 Scalable Video Coding, November 2007.

[5] H. L. Cycon, T. C. Schmidt, G. Hege, M. Wählisch, D. Marpe, and M. Palkow, "Peer-to-Peer Videoconferencing with H.264 Software Codec for Mobiles," in *WoWMoM08 - WS on Mobile Video Delivery (MoViD)*, IEEE Press, June 2008, pp. 1-6.

[6] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile Video Transmission Using Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1204-1217, Sept. 2007.

[7] R. Rejaie and A. Ortega, "PALS: Peer-to-Peer Adaptive Layered Streaming," in *Proc. 13th Intern. WS on Network and Operating Systems Support for Dig. Audio and Video*. ACM, 2003, pp. 153-161.

[8] O. Abboud, K. Pussep, A. Kovacevic, and R. Steinmetz, "Quality Adaptive Peer-to-Peer Streaming Using Scalable Video Coding," in *Proc. 12th IFIP/IEEE (MMMS 2009),* ser. LNCS, no. 5842. Springer Verlag, 2009, pp. 41-54.

[9] P. Baccichet, T. Schierl, T. Wiegand, and B. Girod, "Low-Delay Peer-to-Peer Streaming using Scalable Video Coding," in *Proc. Intern. Packet Video Workshop (PV2007)*. IEEE, 2007, pp. 173 -181.

[10] I. Kofler, M. Prangl, R. Kuschnig, and H. Hellwagner, "An H.264/SVC-based adaptation proxy on a WiFi router," in *Proc. 18th Intern. Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'08)*. ACM, 2008, pp. 63-68.

[11] K. Mokhtarian and M. Hefeeda, "End-to-End Secure Delivery of Scalable Video Streams," in *Proc. 19th Intern. Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'09)*. New York, USA: ACM, 2009, pp. 79-84.

[12] A. Eleftheriadis, R. Civanlar, and O. Shapiro, "Multipoint videoconferencing with scalable video coding," *J. Zhejiang Univ. SCIENCE A*, vol. 7, no. 5, pp. 696-705, 2006.

[13] H. L. Cycon, T. C. Schmidt, M. Wählisch, M. Winken, and D. Marpe : "Optimized Temporal Scalability for H.264-based Codecs and its Application Video Conferencing," *Proc. 15th IEEE International Symposium on Consumer Electronics (ISCE 2010),* Braunschweig, Germany, June 7-10, 2010.

[14] H. L. Cycon, G. Hege, D. Marpe, M. Palkow, T. C. Schmidt, and M. Wählisch: "Adaptive Temporal Scalability of H.264-compliant Video Conferencing in Heterogeneous Mobile Environments," to be presented at *IEEE Global Communications Conference (IEEE Globecom 2010),* Miami, FL, USA, Dec 6-10, 2010.

[15] J. Reichel, H. Schwarz, and M. Wien (Eds.), "Joint Scalable Video Model 11 (JSVM 11)," ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-X202, July 2007.

[16] Joint Video Team (JVT), "SVC Reference Software (JSVM Software)", Heinrich-Hertz-Institut, 2011, http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm.

[17] R. Prasad, C. Dovrolis, M. Murray, and K. C. Claffy, "Bandwidth Estimation: Metrics, Measurement Techniques, and Tools," *IEEE Network*, vol. 17, no. 6, pp. 27-35, November--December 2003.

[18] V. Jacobson. "Congestion avoidance and control." In SIGCOMM '88: Symposium on Communications architectures and protocols, pages 314-329, NewYork, NY, USA, 1988. ACM.

[19] "Generic Coding of Moving Pictures and Associated Audio Information -Part 2: Videos," ITU-T, Tech. Rep.ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2 Video), ITU-T and ISO/IEC JTC 1, 1994.

[20] "Video Coding for Low Bit Rate communication," ITU-T, Tech. Rep. ITU-T Rec. H.263, ITU-T, Version 1, 2, 3, 1995 (1), 1998 (2), 2000 (3).

[21] "Coding of audio-visual objects - Part 2: Visual," ISO/IEC, Tech. Rep. ISO/IEC 14492-2 (MPEG-4 Visual), ISO/IEC JTC 1, Version 1, 2, 3}, 1999 (1), 2000 (2), 2004 (3).

[22] H. L. Cycon, T. C. Schmidt, G. Hege, M. Wählisch, and M. Palkow, "An optimized H.264-based Video Conferencing Software for Mobile Devices," in *ISCE 2008 - The 12th IEEE International Symposium on Consumer Electronics*, IEEE Press, April 2008, pp. 1-4.

[23] H. L. Cycon, G. Hege, D. Marpe, M. Palkow, T. C. Schmidt, and M. Wählisch, "Connecting the Worlds: Multipoint Videoconferencing Integrating H.323 and IPv4, SIP and IPv6 with Autonomous Sender Authentication," in *13th International Symposium on Consumer Electronics*. Piscataway, NJ, USA: IEEE Press, May 2009, pp. 890-893.

[24] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol," IETF, RFC 3261, June 2002.

[25] T. C. Schmidt, G. Hege, M. Wählisch, H. L. Cycon, M. Palkow, and D. Marpe, T. C. Schmidt, G. Hege, M. Wählisch, H. L. Cycon, M. Palkow, and D. Marpe "Authenticated Sources and its Application to an H.264 Videoconferencing Software for Mobiles," *Multimedia Tools and Applications, 2011*, in print.

[26] Gisle Bjøntegaard, "Calculation of average PSNR differences between RD curves," Doc. VCEG-M33, ITU-T VCEG, 13th Meeting, Austin, TX, USA, Apr. 2001.

BIOGRAPHIES

Hans L. Cycon (SM'10) is Associate Director of Daviko GmbH Berlin, Germany and Professor emeritus from HTW Berlin, University of Applied Sciences Berlin. He received his Diploma in Physics in 1975, his PhD in Mathematics in 1979 and his Habilitation in 1984 from the Technical University Berlin, Germany. Research activities led him to the Courant Institute of Mathematical Sciences (NYU) New York; CALTEC in Pasadena; Old Dominion University in Virginia and ICSI at Berkeley University in California. His publication fields are mathematical physics and signal processing i.e., image coding. Recent publications deal with e- and m-Learning using video-based techniques. Hans L. Cycon has been leading several projects in developing wavelet-based still image and video compression codecs. He is member of the IEEE and of the German delegation on the ITU/ISO standardization committee for JPEG 2000 still image standard.

Thomas C. Schmidt (M'05) is professor of Computer Networks & Internet Technologies at Hamburg University of Applied Sciences (HAW) and leads the Internet Technologies research group (INET) there. Prior to moving to Hamburg, he headed the computer centre of FHTW Berlin for many years, and continued work as an independent project manager later. He studied mathematics and physics at Freie Universität Berlin and University of Maryland. He has continuously conducted numerous national and international projects. His current interests lie in next generation Internet (IPv6 and beyond), mobile multicast and multimedia networking, as well as XML-based hypermedia information processing. He serves as co-editor and technical expert in many occasions, is actively working in IETF and co-chair of the IRTF research group on scalable adaptive multicast (SAMRG).
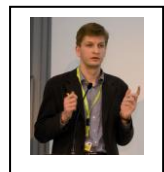
Matthias Wählisch (M'05) studied computer science and contemporary German literature at Freie Universität Berlin, where he completed his diploma thesis on structured hybrid multicast routing. He continues his research at the Computer Systems & Telematics group there, and is also with the INET research team at HAW Hamburg. He started professional activities at the networking group of the computer centre of FHTW Berlin while at high school. He is the cofounder of link-lab, a start-up company in the field of next generation networking. His major fields of interest lie in the design and analysis of networking protocols, with a special focus on mobility and group communication in underlay and overlay, where he looks back on ten years of professional experience in project work and publication.

Detlev Marpe (SM'00) received the Dr.-Ing. Degree from the University of Rostock, Rostock, Germany, in 2005 and the Dipl.-Math. degree (with highest honors) from the Technical University Berlin (TUB), Berlin, Germany, in 1990. From 1991 to 1993, he was a Research and Teaching Assistant in the Department of Mathematics, TUB. Since 1994, he has been involved in several industrial and research projects in the area of still image coding, image processing, video coding, and video streaming. In 1999, he joined the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI), Berlin, Germany, where as a Project Manager in the Image Processing Department, he is currently responsible for the development of advanced video coding and video transmission technologies. Since 1997, he has been an active contributor to the standardization activities of ITU-T VCEG, ISO/IEC JPEG and ISO/IEC MPEG for still image and video coding. From 2001 to 2003, as an Ad hoc Group Chairman in the Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, he was responsible for the development of the CABAC entropy coding scheme within the H.264/AVC standardization project. He also served as the Co-Editor of the H.264/AVC Fidelity Range Extensions. He has published numerous papers on a variety of topics in the

area of image and video processing and holds various international patents in this field. His research interests include still image and video coding, image and video communication as well as computer vision, and information theory. Dr. Marpe received the Fraunhofer Prize for outstanding scientific achievements in solving application related problems and the ITG Award of the German Society for Information Technology in 2004. As a Co-Founder of the Berlin based daViKo Gesellschaft für audiovisuelle Kommunikation mbH, he was recipient of the Prime Prize of the 2001 Multimedia Start-Up Competition of the German Federal Ministry of Economics and Technology.
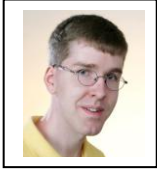
Martin Winken is research engineer at Fraunhofer HHI Berlin, Germany. He received his Diploma in computer engineering from the Technical University Berlin in 2006. Currently, he is pursuing his Ph. D. on video compression technology. He has published conference contributions and contributed to standardization activity in this area.