Mobile Video Transmission Using Scalable Video Coding

Thomas Schierl, *Member, IEEE*, Thomas Stockhammer, *Member, IEEE*, and Thomas Wiegand, *Member, IEEE*

(Invited Paper)

***Abstract*—The Scalable Video Coding (SVC) standard as an extension of H.264/AVC allows efficient, standard-based temporal, spatial, and quality scalability of video bit streams. Scalability of a video bit stream allows for media bit rate as well as for device capability adaptation. Moreover, adaptation of the bit rate of a video signal is a desirable key feature, if limitation in network resources, mostly characterized by throughput variations, varying delay or transmission errors, need to be considered. Typically, in mobile networks the throughput, delay and errors of a connection (link) depend on the current reception conditions, which are largely influenced by a number of physical factors. In order to cope with the typically varying characteristics of mobile communication channels in unicast, multicast, or broadcast services, different methods for increasing robustness and achieving quality of service are desirable. We will give an overview of SVC and its relation to mobile delivery methods. Furthermore, innovative use cases are introduced which apply SVC in mobile networks.**

***Index Terms*—Content delivery, DVB-H, mobile, protocols, raptor codes, video, wireless, 3 GPP, H.264/AVC.**

I. Introduction

THE SCALABLE Video Coding (SVC) [1], [2] standard as an extension of H.264/AVC [3], [4] allows efficient, standard-based scalability of temporal, spatial, and quality resolution of a decoded video signal through adaptation of the bit stream. Scalability of a video bit stream allows for media bit rate as well as for device capability adaptation without the need of transcoding or re-encoding. The latter aspect is particularly relevant in emerging heterogeneous next generation networks. Herein, the capability of end-user devices motivates for scalability of the media, since terminals usually vary in display resolution and processing power capabilities according to their evolution state and category.

Initial mobile 3 G video services were largely based on H.263 and MPEG-4, but all recently introduced services are already almost exclusively based on H.264/AVC. The penetration of H.264/AVC capable terminals will increase over the next years such that mobile operators will exploit the increased efficiency of this codec. For example, 3 GPP recommends the use of H.264/AVC baseline profile for all services, including conversational services, packet-switched streaming services (PSS) [10], messaging services, and multimedia broadcast/multicast services (MBMS) [12]. Furthermore, mobile broadcast services such as DVB-H [13] and DAB [15] rely on H.264/AVC. The baseline H.264/AVC will be used to distribute Mobile TV services in 3 G, DVB-H and other mobile networks. However, the supported levels are currently quite restricted. It is obviously expected that future terminals will have significantly enhanced capabilities in terms of display, processing power, and access bit rates such that higher levels enabling significantly better quality will be supported. With the availability of SVC, the extension of H.264/AVC will not only be of interest to higher levels, but for certain services it is also very attractive to have a rate-scalable extension with a backward-compatible based layer.

Moreover, to support higher quality media, also the adaptation of the bit rate of a video signal is a desirable key feature. This provisions for cases, when limitations in network resources arise, that are mostly characterized by throughput variations, varying delays or transmission errors. Typically, in mobile networks, throughput, delay and transmission errors of a connection (link) depend on the actual quality of the reception conditions, which is influenced by transmission physics, but also by the availability of radio error control techniques. Thereby, unicast, multicast, and broadcast services provide different methods for increasing robustness or for achieving Quality of Service (QoS). For example, such methods can be categorized into methods for channels with and without feedback, or methods for applications with and without delay constraints, etc.

Rate and quality adaptation in a mobile network may happen in different network instances. Nowadays, classically rate-adaptation happens end-to-end, i.e., the multimedia server or the real-time encoder in a network selects the appropriate bit rate based on network information, and possibly also based on feedback from the receiver. A scalable bit stream extends these possibilities significantly: rate adaptation may be performed not only at the encoder/server, but in intermediate network nodes, or even only at the receiver. Rate adaptation may be applied at the streaming server, in intermediate network nodes for device adaptation, in radio link buffers for channel adaptation or only at the receiver to extract the appropriate resolution for the terminal display. Intelligent thinning of a scalable bit stream can be achieved without high costs in computational resources like

Manuscript received January 4, 2007; revised July 13, 2007. This paper was recommended by Guest Editor T. Wiegand.

T. Schierl and T. Wiegand are with the Fraunhofer Institute for Telecommunications—Heinrich-Hertz Institute, D-10587 Berlin, Germany (e-mail: thomas.schierl@hhi.fraunhofer.de; thomas.wiegand@hhi.fraunhofer.de)

T. Stockhammer is with Nomor Research GmbH, D-83346 Berlin, Germany (e-mail: stockhammer@nomor.de).

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

Digital Object Identifier 10.1109/TCSVT.2007.905528

required by transcoding methods [5]. To enable the adaptation not only at the media server, networks require extensions in the transport stream and stream signaling, for background we refer to [6].

Bit rate scalable media naturally combines with prioritization methods: It may be successfully combined with unequal error protection, selective retransmission, or hierarchical modulation schemes. The idea is to strongly protect the important part of the scalable media (the base layer) in order to overcome worst-case error scenarios and give less protection to the enhancement layer in order to overcome the most typical error situations. This approach results in graceful degradation of the play-able quality according to the channel condition. Such systems have been studies in great detail in many research publications and potentials of such technologies are well known [39], [40]. Most of these techniques have been limited by the non-availability of an efficient scalable video codec. With SVC in place, the impact of such cross-layer technologies will certainly grow over the next years.

A comprehensive treatment of wireless video transmission cannot be the objective. Therefore, this work attempts to provide some insight into potential use cases of SVC in wireless transmission networks. For this purpose, Section II briefly introduces to SVC including a discussion about the advantages of a scalable representation of a video bit stream. Topics such as coding structure, adaptation of the bit stream and network transport are highlighted. Section III discusses characteristics and features of mobile radio channels and delivery methods. We point out the challenges of reliable transmission of data in these types of networks and highlight the potential of SVC in this context. Moreover, we discuss state-of-the-art techniques for achieving error robustness in radio networks and further point out the relation of these techniques to SVC. To provide some further substance to the discussions in Section III, three specific example use cases for SVC in mobile networks in Section IV are provided.

II. Scalable Video Coding in Mobile Environments

The approach of SVC also known as layered video coding has already been included in different video coding standards in the past, like H.262 | MPEG-2 Video, H.263, and MPEG-4 Visual. But all these past standardization efforts produced results with inferior coding efficiency. Scalability has always been a desirable feature of a media bit stream for a wide range of services. This is especially the case for transport over best-effort networks that are not provisioned to provide suitable QoS and especially suffer from significantly varying throughput. Thus a real-time service needs to dynamically adapt to the varying transmission conditions: For example, it is expected that a video stream is capable to adapt its media rate to the transmission conditions to provide at least acceptable quality at the receivers, but also explores the full benefits of available higher system or device resources. Within multimedia sessions, typically the video consumes the major part of the total requested transmission rate compared to control and audio data. Therefore, an adaptation capability for the video bit rate is of primary interest in a multimedia session. A strong advantage of a video bit rate adaptation method relying on a scalable representation is the drastically reduced computational requirements in network elements compared to approaches that require video re-encoding or transcoding. With this motivation in mind, the H.264/AVC-based SVC is of major practical interest and is therefore briefly introduced in the following. For a more detailed description see other papers in this special issue in particular the overview in [2].

A. Scalable Video Coding Extensions of H.264/AVC

The SVC design [1], [2], which is an extension of the H.264/AVC [3], [4] video coding standard, can be classified as a layered video codec. SVC-based layered video coding is suitable for different use-cases like, e.g., supporting heterogeneous devices with a single, scalable bit stream. Such a stream allows for delivering a decode-able and presentable quality of the video depending on the device's capabilities. Here, presentable quality refers to resolution, frame rate and bit rate of a decoded operation point of the scalable video bit stream.

Another use-case, as mentioned before, is the adaptation to varying network conditions. Typically, end-to-end protocols cope with throughput variations by adjusting the transmission rate. If a real-time encoder per client is used or multiple streams for bit stream switching are available, adaptation would be applicable at the source for each client. But, if multiple clients should be served with the same video content and adaptation should be applied on the network, adaptability is required by the media itself. SVC explicitly provisions for removing packets from the bit stream, which implicitly results in bit rate and by that in presentation quality reduction of the video.

The coder structure and coding efficiency of SVC depend on the scalability features required by an application. Fig. 1 shows a typical coder structure with two quality layers for signal-to-noise ratio (SNR) fidelity scalability. When different resolutions shall be supported by a single bit stream, spatial scalability is used. Moreover, spatial and SNR fidelity scalability can be mixed. The enhancements are coded using the predictions from lower layers like the base layer. Temporal scalability is achieved by hierarchical B or P.

In SVC, the hybrid video coding approach of motion-compensated transform coding is extended in a way that a wide range of spatio-temporal and quality scalability is achieved. An SVC bit stream consists of a base layer and one or several enhancement layers. The removal of enhancement layers still leads to a reasonable quality of the decoded video at reduced temporal, SNR, and/or spatial resolution. The base layer is a bit stream conforming to H.264/AVC [3] ensuring backward-compatibility for existing receivers. The decoding process itself is still based on a single motion compensation loop keeping the processing overhead for the scalability small. The manageable complexity is one of the key features for a video codec being deployable in wireless transmission as receiving devices will always limited by processing power, memory and energy consumption.

The temporal scaling functionality of SVC for configurations without low-delay constraints is typically based on a temporal decomposition using hierarchical B-pictures. Fig. 2 shows hierarchical B-pictures with two layers of SNR fidelity scalability: base layer and one fidelity enhancement layer. Pictures with labels T0 represent so-called *key pictures*. These pictures serve as

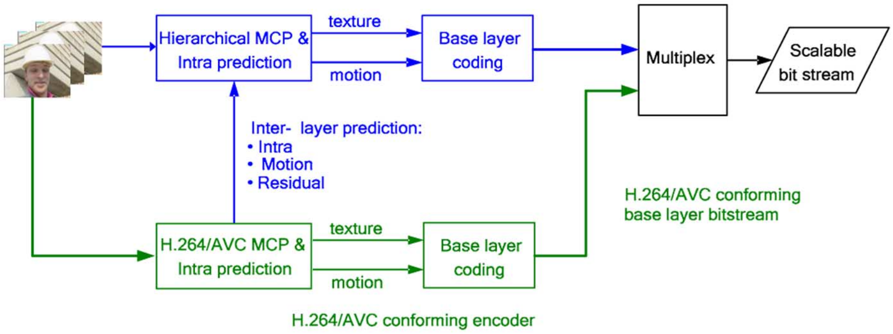

Fig. 1. Coder structure with two quality layers.

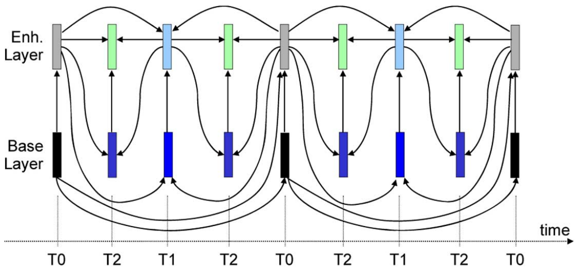

Fig. 2. SVC temporal prediction structure.

synchronization points between encoder and decoder. The encoder-decoder synchronization is achieved at the cost of coding efficiency since also the enhancement layer pictures for the latter two T0 pictures in Fig. 2 are inter-predicted from the preceding base layer pictures labeled with T0. The B-pictures between the key pictures are forming the temporal enhancement levels. Where pictures labeled T1 form the first temporal enhancement to the key pictures and pictures labeled T2 the second temporal enhancement. The base layer pictures labeled with either T1 or T2 are predicted from the highest available enhancement layer pictures. This approach, also known as medium granularity scalability (MGS) [1], provides high coding efficiency for the base layer in case the reference pictures are available and does not pose a significant problem when the reference pictures are not available; since only a few consecutive pictures depend on these through inter prediction. For more details on MGS, we refer to [2].

The SVC bit stream structure shown in Fig. 2 comprises a group of pictures (GOP) of size four. GOPs can be independently decoded, if the corresponding key picture has random access properties and the preceding reference is available.

For low-delay configurations, prediction dependencies can be selected in a way that no future dependencies are used. This allows for minimizing the structural decoding delay down to zero frames. Although this structure allows for the same temporal scalability functionality as those exploiting future dependencies, it reduces the coding efficiency at the price of low-delay.

In order to achieve multiple bit rate points in the enhancement layer rather than decoding or not decoding the whole enhancement layer, temporal scalability can be used within the MGS enhancement layer. That is, pictures are removed from the enhancement layer starting with the lowest temporal priority down to not decoding any of the enhancement layer pictures.

Another important scalability function is the spatial scalability, which, if carefully be used, can significantly reduce the bit rate required for serving heterogeneous receivers compared to simulcasting. The spatial scalability of SVC is achieved by different encoder loops with an over-sampled pyramid for each resolution (e.g., QCIF, CIF, and 4 CIF), including motion-compensated transform coding with independent prediction structures for each layer. In contrast to the encoder, the decoder can be operated in single loop, i.e., for decoding inter-layer dependencies it is not required to perform motion compensation in lower layers which a layer depends on. Note that resolution steps in the enhancement layers do not necessarily have to be of a factor of 2 of the aspect ratio.

In order to switch between different spatial layers random access points like H.264/AVC IDR pictures are required, i.e., the layer to be switched to must show an IDR picture a the time instance of switching. Within in SNR layers, switching at each picture is possible.

A combination of all three scalability functionalities within one bit stream is called combined scalability. Such a combined scalable bit stream allows for extraction of different operation points of the video, where each operation point is characterized by a certain level of SNR fidelity, temporal, and spatial resolution.

B. Network Transport and Adaptation of SVC

As mentioned before, one typical application for SVC is bit rate adaptation for transport over packet-switched networks, e.g., like IP-based radio networks. For this use-case, the media is typically delivered by either end-to-end protocols or broadcast mechanisms. Thus signaling within the bit stream is an important feature for allowing media-aware network elements (MANEs) [6] to apply bit stream adaptation or differentiation in protection of the layers according to their importance. Since an SVC bit stream can support up to three dimensions of scalability, a MANE needs detailed information about the resulting quality of the video, when reconstructed at the receiver. But a MANE can also rely on one absolute importance indicator for adaptation. Therefore, the encoder must already have selected the adaptation path through a global bit stream. More details about the transport interface of SVC can be found in [33].

For the aforementioned reasons, the identification of video data belonging to different layers is achieved by an extended approach of the network abstraction layer (NAL) concept of

H.264/AVC. The NAL hides the detailed bit stream structure of H.264/AVC and allows for high level (application layer) readability of the NAL packets. NAL packets typically represent a video frame, a part thereof, parameter sets (decoder initialization information), supplemental enhancement information (SEI)–supplying additional bit stream information like time stamps not required for decoding) and bit stream organizing information like end-of-stream indication. For SVC, the NAL header syntax has been extended for allowing identification of temporal, spatial and SNR scalability information per NAL packet. In order to give a pre-computed, single adaptation path through a bit stream, the NAL header also provides a one-dimensional priority indicator $(\mathrm{priority}_{\mathrm{id}})$. For more details on the NAL unit header and network adaptation, we refer to [33] and [6].

Besides knowing to which operation point a NAL packet belongs to, a MANE furthermore needs information about the characteristics of such points. For that, a Scalability Information SEI message provides detailed information about values like resolution, average bit rate and frame rate of operation points contained in the bit stream. This message is suitable for being transferred by in-band as well as out-of-band transport mechanisms.

III. Mobile Video Delivery: Networks, Services and Their Relation to SVC

A. Overview

The success of emerging mobile networks will, among other aspects, be determined by the extensive usage of the available bit rates. One way of achieving that is to offer attractive video services. Within this context, mobile users will expect to have access to similar video services as offered on their home appliances, including video-on-demand services, live mobile TV services, and clipcasting services. The distribution means can be quite different, e.g., real-time distribution through RTP/UDP, or non real-time distribution using classical Internet protocols such as HTTP/TCP, but also multicasting and broadcasting of streams and files is a hot topic in emerging system architectures. Finally, with the success of peer-to-peer networks, it can be expected that similar concepts will be deployed also for wireless networks.

In current service architectures, video is mainly included in three different types of services: 1) conversational services such as video telephony and video conferencing; 2) streaming and live TV services; as well as 3) services which download files to the end users' device before playout is actually started. All three service offerings have significantly different requirements especially in terms of delay and latencies requiring flexible and adaptive media coding algorithms.

In addition to these traditional stand-alone services, nowadays also mixtures of different delivery methods become popular, such as progressive download, where the user starts playing out the early parts of a file while still downloading, or services where the live-consumed stream may also be stored for later consumption. Furthermore, emerging network architectures will also combine different delivery networks and modes for most efficient service delivery, e.g., a combination of unicast transmission, multicast distribution and pure broadcasting may be envisaged. This mixture of delivery modes is quite crucial to allow the individual access of all the data in a unicast manner, but also to provide the option to distribute the popular content in an efficient manner by providing point-to-multipoint radio access bearers. Dynamic switching between distribution methods may apply, especially when handovers and roaming is supported.

Another trend on mobile communication are decentralized architectures. It can be foreseen that classical cellular and broadcast networks such as UMTS or DVB-H will be extended or replaced by alternative network designs reducing or completely dispensing with any centralized infrastructure. Relaying, femtocells, distributed content servers and distribution nodes, mobile ad-hoc and peer-to-peer networks, etc. have been researched extensively and first commercial systems making use of such technology are already deployed. Obviously, especially for mobile ad-hoc networks, the penetration of participating nodes significantly influences the capacity of the network and their dynamical behavior will result in heterogeneous and varying network throughput and available end-to-end bit rates.

Mobile and wireless distribution is also characterized by rapid development of improved receiving terminals and heterogeneous user preferences. Therefore, the mobile and portable device market offers many different capabilities of the receiving terminals in terms of bit rates, display sizes, memory, energy supply, and complexity for handheld devices, PDA-like devices, laptops or in-car receivers.

In this heterogeneous and dynamic environment it is obvious that one cannot expect that a single video bit rate stream can fulfill all the requirements without limiting the user experience of most other receivers. Rate-scalable media will be an important enabler to fully exploit the potentials of emerging delivery methods, improved receiver capabilities, and new service offerings to satisfy increasing user expectations in a resource and cost efficient manner. Therefore, it is appealing to identify and investigate use cases for flexible, adaptive and scalable video codecs in such environments. As nowadays, many mobile video services are already in operation using single layer codecs, particularly H.264/AVC, backward compatibility to such services is important, but also the provision of enhanced quality for better access networks, better network conditions and/or high-end receivers is desirable. A scalable extension to H.264/AVC is therefore highly preferable. Nevertheless, it is still not expected that in all mobile video service scenarios the addition a scalable video codec is necessary, but for at least a selected subset of applications significant benefits from a scalable solution can be expected.

B. Radio Access Bearers and Mobile Distribution Networks

Mobile networks typically provide different modes to distribute packet data to service subscribers or also a combination of different access networks can provide different distribution means. In emerging systems, these radio access bearers are almost exclusively packet-based. We will provide an overview of different distribution means to highlight their properties, their benefits but also their deficiencies.

1) Unicast Streaming Bearers in Cellular Network: Deployed cellular networks such as UMTS provide packet-based

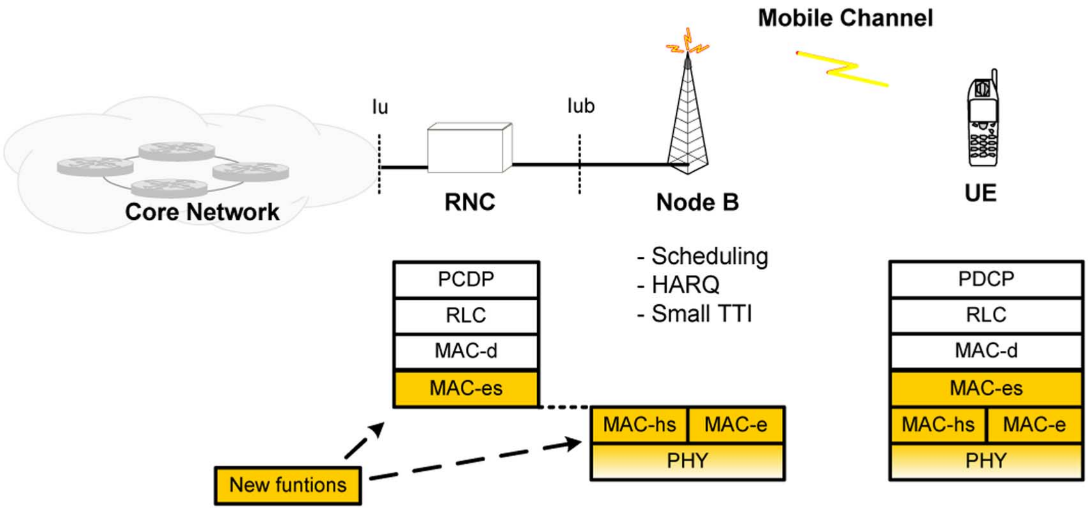

Fig. 3. HSDPA Architecture.

dedicated bearers with good streaming QoS, i.e., reasonable bit rate up to 128 or 256 kbps, high reliability, and reasonable but not excessive delay and jitter. This QoS is achieved by the use of physical layer forward error correction (FEC), fixed spreading codes, power control, and radio link layer retransmissions. Despite the latter technology is the source for some delay and jitter, these link layer retransmissions can solve the problem of occasional packet losses due FEC failure or link outages during handovers [11]. However, the provision of high-rate reliable radio bearers is resource expensive and will not scale well in case of increasing popularity of such a services and if higher bit rates are desired.

Due to this inefficiencies, a shift from traditional constant bit rate dedicated IP bearers to packet switched shared IP bearers is expected in the near future. Most recognized, high speed packet downlink access (HSDPA) has been introduced as a new technique in UMTS for downlink transmission. This technology provides significant enhancements in end-to-end service provisioning for IP-based services. The main paradigm shift is that instead of using dedicated resources for each user, the common resources are made available to all users and are shared dynamically based on different criteria. Additionally, HSDPA supports resource allocation with adaptive coding and modulation to exploit the varying radio channel and interference variations, fast hybrid ARQ to reduce retransmission round trip times, reduced transmission time interval (TTI) for latency reduction and to support fast scheduler decisions, and fast channel feedback. These added functionalities have been specified in the new MAC-hs sub layer and modifications of the physical layer as depicted in Fig. 3. The Node B needs to be aware of service or QoS parameters to employ appropriate scheduling algorithms. Nevertheless, the dynamic behaviour of a huge number of sharing users in a mobile system requires dynamic adaptations of the application and service data rates.

Other new data transmission modes in cellular systems will be based on very similar concepts: Among others, 3 GPP2's EDVO, IEEE's 802.16 family (also known as WiMAX), or 3 GPP's Long Term Evolution (LTE), operate along the same principles of dynamically and flexible sharing the available resources among users to optimize data throughput. Multimedia codecs which can cooperate with such modes and exploit their potentials are highly desirable.

2) Multicast/Broadcast Radio Access Bearer in Cellular Networks: Multicast IP transmission will be introduced in mobile cellular networks. The 3rd Generation Partnership Project (3 GPP) has taken the lead in this respect: Multimedia Multicast/Broadcast Service (MBMS) [12] extends the existing 3 GPP architecture by the introduction of an MBMS Bearer Service and MBMS User Services. The MBMS Bearer Service is provided by the packet-switched domain to deliver IP multicast datagrams to multiple receivers using minimum radio and network resources and provides an efficient and scalable means to distribute multimedia content to mobile phones. This is accomplished by point-to-multipoint (p-t-m) transmission. The architecture is complemented with two types of MBMS User Services. In streaming services, a continuous data flow of audio and/or video is delivered to the end user's handset. In download services, data for the file is delivered in a scheduled transmission timeslot.

The p-t-m MBMS Bearer Service does neither allow control, mode adaptation, nor retransmitting lost radio packets and hence, the QoS provided by the MBMS Bearer Service for the transport of multimedia applications is in general not sufficiently high to support a significant portion of the users for either download or streaming applications. As error resilience tools in multimedia codecs do neither provide sufficient efficiency nor quality in case of losses, 3 GPP included an application layer FEC based on Raptor codes [22], [23] for MBMS.

Other mobile cellular networks such as 3 GPP2 and WiMAX are likely to follow this direction and have decided to introduce similar multicast distribution means under the acronyms BroadCast/MultiCast Service (BCMCS) and Multicast Broadcast Service (MBS), respectively. In all cases the reception conditions of individual users might be quite different, for example depending on the position, the velocity, and the receiver capabilities of the handheld terminal. In addition, MBMS user services may be distributed also over p-t-p links if decided to be more efficient, they may be distributed in cell areas with different load or available technologies, or they may be delivered to terminals with different capabilities. Therefore, flexibility and rate adaptation is a desirable feature in cellular multicast distribution modes with different application scenarios.

3) Broadcast Bearer in Mobile Single Frequency Networks: Classical broadcast networks have been extended to provide IP-based distribution of multimedia and data services to handheld devices. Systems such as DVB-H [13], [14], T-DMB [15] and IP-based extensions of DAB like eDAB or DAB-IP target for mobile multimedia delivery in huge areas. Further extensions are currently considered by including satellite links as for example provided in DVB-SH (Satellite-to-Handhelds) and S-DMB.

On the radio layer DVB and DAB networks typically rely on the single frequency network (SFN) approach. In SFNs, all network cells are transmitting a particular channel on the same frequency. For coping with errors in case of mobility, DVB-H streaming services contain a Reed-Solomon FEC on MPEG-2 Transport Stream (TS) level. For broadcast file download services the Raptor FEC is applied on the application layer. Despite these advances, the coverage of such SFN systems is still limited, or the overall service quality in terms of bit rate is harmed by the worst-case receiver. Therefore, modes for the support of graceful degradation and different service quality support are desirable.

DVB-H provides the option of hierarchical modulation: In this case, receivers with a lower signal to noise ratio still receive the lower bit rate stream [high priority (HP) stream] while receivers with a high enough signal to noise ratio receive the higher bit rate stream [low priority (LP) stream]. Similarly, such priority mechanisms may also be supported by unequal error protection schemes such that mobile broadcast systems. This enables to map SVC bit stream layers according to their importance to different priority classes. For details on DVB-H modes we refer for example to [28] and [34].

4) Wireless Multihop or Mesh Networks: Wireless multihop networks [16] may use the ad hoc mode of the IEEE 802.11 Wireless Local Area Network (WLAN) specification [46], but also on the emerging IEEE 802.11s standard [47]. Mobile ad-hoc networks can be used for building short living network topologies for short term events or for setting up ad hoc topologies in areas where installation of fixed infrastructure is not possible or too costly. In the WiMAX IEEE 802.16 j standard [48], multihop relay extensions will be defined for enhancing coverage of WiMAX networks. In this case, infrastructure is partially used for serving the relay nodes, but the target node is served via relays only. Similar approaches are currently considered in beyond 3 G networks such as extensions to HSDPA or LTE using fixed relays to extend coverage, or to use micro base stations in home environments resulting in so called femtocells.

Especially in case of mobile relays, a major problem when utilizing these networks is the unreliable ad hoc network topology, which typically cannot provide a reliable infrastructure and QoS guarantees. This property implies unreliable paths between participants of a video transmission, thus route losses and unavailability of nodes through network partitioning are typical for such networks [17]. Despite specialized ad-hoc routing algorithms like Ad-Hoc On-Demand Distance Vector Routing (AODV) [43] and Optimized Link State Routing (OLSR) [44] are used, the network separation problem, as shown in Fig. 4, cannot be solved. In addition to loss of connectivity, unreliable radio link or congestion at intermediate hops may cause delays and losses. In [17] and [31] the suitability of such networks for video real-time streaming is shown (Section IV-B).

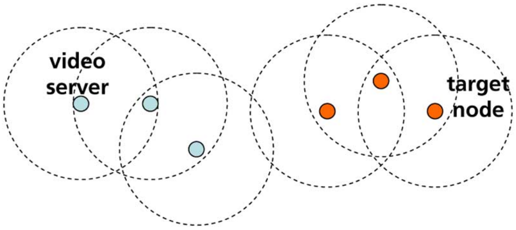

Fig. 4. Network separation in multihop topologies.

C. Sweet Spots for Scalable Video Coding

With the provided background on different video services, different receiver capabilities and user expectations, and heterogeneous transport and reception conditions, we identify some generic use cases for scalable media codecs, specifically SVC, in mobile environments. Scalable extensions ma not be necessary, or at least not be major importance for any mobile video service. Consider cases when rate adaptation on encoded content is of little interest, e.g., in messaging services, or end-to-end conversational services with online encoding, single layer codecs are usually sufficient. Also, for applications where the video needs to be encoded on handheld devices, the increased encoding complexity of SVC encoding may limit the applicability of SVC in such environments.

However, scalable video coding has its obvious merits if an encoded version of the video signal needs to be transmitted to receivers with different access bit rate or reception capabilities and the encoding cannot be done individually or is not economically viable for each and every receiver. Therefore, consider the following three different transport and service scenarios:

Scenario A: On-demand transmission of pre-encoded content to receivers with different and/or varying access bit rates. This scenario covers for example on-demand streaming services, for which the media server can host multiple bit rate and quality versions of the content. This allows efficient storage as well as smooth dynamic switching between those versions. This scenario, may also contain adaptation to channel conditions on the network path, e.g., in a MANE [6].

Scenario B: On-demand or live transmission of the same content in parallel to receivers with different and/or varying access bit rates and/or different reception capabilities. This scenario covers for example the distribution of different video resolutions over a mobile TV system. Whereas H.264/AVC legacy receivers with restricted profile/level specifications may for example only decode the 128 kbps stream, high performance SVC-capable receivers may add one or two additional layers for significantly better quality.

Scenario C: On-demand or live transmission of the same content to a receiver with unknown transmission conditions, e.g., due to missing or delayed feedback links. This scenario covers for example mobile TV broadcasting for which the reception quality determines the decoded SVC layer, for example by applying some unequal error protection.

D. Integration of SVC in Content Delivery Protocols

For the successful deployment of SVC in existing and emerging mobile systems it is essential that SVC is integrated in existing delivery protocols. For example in 3 GPP or DVB-H, especially RTP and 3 G/ISO file format are the most important means to deliver multimedia data in real-time streaming and download delivery services, respectively. In parallel to the standardization work of the SVC codec, it was also taken care that SVC is integrated into those network interfaces. Therefore, we will briefly summarize these efforts. Only if the video layering information can be exploited in the network in a simple manner, i.e., without full media-awareness, then SVC will be able to fully exploit its potentials.

The packetization of SVC data into **Real-time Transport Protocol (RTP)** network packets is described in [6] and it is outlined in detail how scalability information can be used on RTP level for network adaptation. The concept of directly encapsulating NAL units in RTP packets is maintained, but the idea of priority information as introduced in H.264/AVC with the four levels of nal_ref_idc is significantly extended to allow signaling of a linear priority as well as different layers within temporal, spatial and SNR scalability dimensions. The strongest advantage of a scalable representation of a video bit stream is obviously the ability of adaptation of the stream without need of re-encoding. By providing this signaling in the RTP payload header as well as in the SDP session signaling, adaptation can be applied in the network by nodes typically known as MANE [6]. Adaptation can be necessary or beneficial for different reasons, e.g., for bit rate or device capability adaptation.

Many live-media distribution protocols are based on the RTP including p-t-m transmission, e.g., in DVB-H [13], [14] or MBMS [12]. In this case, the provision of different layers on e.g., different multicast addresses, allows for applying different protection strength on different layers. Pioneer work in this area was introduced under acronym "priority encoding transmission (PET)" [21] as well as in [29], and many subsequent publications used similar concepts to show the benefits of multicasting scalable media coding based on priorities. However, almost all of these concepts were hindered from successful deployment due to the non-availability of efficient scalable video codecs. These concepts will likely be revisited with the availability of SVC.

A further motivation for adaptation in MANEs is the idea of rate/distortion (R/D) optimized decisions for rate allocation of different, competing video streams [8]. In this case rate allocation for the scalable video streams is applied in a way that the overall average decoded video quality at the connected receivers is optimized. This way of optimization requires further meta-information within the bit stream or as synchronized out-of-band information. In [9], an approach for calculation and assignment of quality information to an SVC bit stream is presented, but this approach lacks the missing relation between the quality values calculated for different streams. But just this would be the precondition for R/D optimized decisions in MANEs as discussed in Section II-B. For this reason, SVC allows for indicating such R/D values or at least R/D relations between bit streams in the linear priority identifier (PID) of the SVC NAL unit header. The method used for assigning the PID is described within the Scalability SEI by an URI (uniform resource identifier).

As a fallback for existing receivers, the RTP transmission of the AVC base layer of an SVC bit stream will be achieved using the native payload format of H.264/AVC [41]. This requires the separation of AVC and SVC at least into two different RTP sessions in p-t-m scenarios.

For non-realtime delivery, the integration of SVC in container formats is essential. The proposed file container format for SVC—the **AVC File Format** [7] is based on ISO base file format. 3 GPP makes use the 3 gp file format which is based on the ISO file format. Generally the media data (Access Units) of different media types in the same media file are stored in so-called "mdat" containers. Additional meta information about size, timing and location of the media data/Access Units (AUs) is stored separately in so-called "trak" containers, thus for each media type a "trak" container and data in a "mdat" container exists. Additionally, all AU of different media types can be also contained in an interleaved way within one "mdat" container for efficient file access.

For SVC, additional cases need to be considered: Access Units typically contain data of different scalability layers (different temporal, spatial or quality representations of the stream), which are stored in separate SVC NAL units. A SVC AU is composed by a set of SVC NAL units belonging to different scalability levels but to the same instance of time. Within the SVC file format, the NAL units of different layers can be stored in different "traks," where special NAL units (so called extractor NAL units, defined in the file format) within the "mdat" are used for referencing NAL units of other layers in other "traks."

The video layer-wise arrangement of SVC data within a file is a precondition for the file being used in some applications as for example in harmonic broadcasting (see Section V). Note, that especially for progressive download services the tracks can be interleaved to smaller movie fragments. With respect to the channel conditions and the required protection for successfully downloading the layers, the application may start with the base layer quality protected with highest rates, while SVC layers in different "trak" containers for higher quality are transferred with less protection rate, which results in longer download times for these parts.

Another important delivery protocol is the **MPEG-2 Transport Stream (TS)**, which is typically used for digital broadcast TV delivery over DVB-T, DVB-C, DVB-S or DVB-IP. Currently, there exists only a specification for embedding H.264/AVC according to [3] into MPEG-2 TS. Ongoing work [32] attempts to provide features for layered transmission of SVC in different MPEG-2 TS Elementary Streams. This will allow for layered multicast transmission on different broadcast channel.

E. Combining Mobile Transport Protocols and SVC

This section is dedicated to provide a high-level overview on radio protocols which in combination with SVC may provide new service opportunities, better user experience and/or advanced efficiency. Obviously, we can only provide a short overview on potential use cases. The integration of SVC and scalable media in general will be subject of study in upcoming standardization work.

Content delivery to mobile users is clearly dominated by some common trends, but also some diverging and competing technologies. IP-based packet delivery is common to most systems: However, whereas emerging 3 G systems such as HSDPA or LTE will continue to rely on QoS provision, less centralized architectures such as wireless ad-hoc or multihop networks, for example are based on IEEE 802.11 WLAN will have to exploit end-to-end adaptation mechanisms and application layer tools to support sufficiently good user experience.

Emerging cellular networks such as HSDPA, WiMAX, or LTE, make use of many physical layer and medium access control (MAC) layer features to support QoS and efficiency as mentioned in III-B.1). An important concept in all in these systems is the provision of fast and timely feedback, e.g., every 2 ms, on the physical and radio layer. This so-called channel quality indication (CQI) is be used by the centralized base station scheduler to dynamically select appropriate modulation and coding schemes, to adapt the transmit power, to select the user for the next transmission slot, or to use it for the selection of appropriate multiple antenna configurations. Furthermore, the application of fast and efficient ARQ methods allows QoS provision and can minimize the residual loss rates.

However, whereas these radio systems are highly sophisticated with respect to the transmission of arbitrary data flows, the differentiation of data is very coarse. In current system architectures, each flow is basically only differentiated among four QoS classes, namely conversational, streaming, interactive, and background. Individual packets within each flow are all treated the same. If for example packets would have to be dropped due to a temporary overload situation in the system, head-of-line or end-of-line dropping strategies are applied. Only just recently, it was recognized that such a coarse treatment of data flows might limit the efficiency and service quality of such systems. Investigations on "*per-packet QoS*" have been started. Thereby, one could rely on general packet marking strategies such as Differentiated Service [42]. For example, some mapping of SVC priority information to DSCP may be an option to introduce "*per-packet QoS*". Alternatively, the scheduler in such a radio system may itself be media-aware, e.g., by including some MANE-like functionality, and may therefore be able to use priority information in the SVC NAL unit header.

Despite such concepts are promising and may be of interest to further enhance mobile video services, still a significant amount of work needs to be done, for example on appropriate mappings, on appropriate packet dropping and delay strategies, on potential gains and also specifically on perceptual quality aspects. Some initial investigations into this direction have for example been presented by the authors in [18] and are discussed in more detail in Section IV-A.

Other combination of scalable video and emerging 3 G systems may be beneficial in intra- and intersystem handovers. Especially the latter case may result in significantly different bit rates, which need to be adapted quite fast. As new service architecture concepts will target the integration of seamless services over multiple radio access technologies, rate-adaptive applications will be essential. SVC may play an important role in such architectures, especially if the rate adaptation is simple enough such that network elements can easily be upgraded to perform this task.

Similar to radio protocol retransmissions, 3 GPP streaming services delivered via User Datagram Protocol (UDP) also allow for application layer retransmissions to combat residual radio losses. In [11] application layer retransmission was investigated for targeting packet loss outside the QoS-controlled 3 G network. With the use of selective retransmissions, retransmissions in bandwidth constrained networks can be solved. Furthermore, in [19] it is shown how to prioritize retransmission data in the 3 GPP PSS context by differentiating audio and video flows. It is expected that similar approaches combined with SVC will provide benefits for such environments as has been shown by many generic research contributions, e.g., [35].

3 G-based networks tend to offer more and more personalized video services such as video-on-demand or interactive TV. However, also classical p-t-m video broadcasting systems are also getting significant attention, for example through DVB-H, DMB, and MediaFlo deployments. In these systems, the reception conditions for different users are quite heterogeneous. Therefore, these systems include modes which allow for differentiating the received bit rate based on the terminal's location and the resulting receiving conditions. As already mentioned, some sort of hierarchical modulation is used, for example in DVB-H [28] and MediaFlo [36]. Such modes may be successfully combined with SVC to support graceful degradation. In a similar manner, in [37], the integration of several multiresolution broadcast systems for wideband code division multiple access (WCDMA) cellular mobile networks, specifically into the p-t-m mode of MBMS, has been investigated. These features will allow for graceful degradation. However, despite the concepts are well-known, the combination with SVC as well as an optimized system design for each of these networks will be a challenging task requiring input from research and from initial deployments using single layer codecs. Similar concepts may also be applied not on the physical layer, but on the application layer using unequal erasure protection schemes. An example for scalable multimedia file delivery is discussed in Section IV-C. The proposed scheme not only provides graceful degradation, but extends the user experience by a second dimension, namely the startup delay.

The discussed use cases in this section show the potential usage of SVC in different mobile radio environments. However, even if the architecture and protocol integration is completed, such systems still leave a significant amount of freedom to exploit the usage of SVC. In this case it essential to understand the cross-layer effects of different system parameters [20], and to optimize the system parameters to maximize user experience and system resources. Research has already provided many ideas and concepts about this, and many of these ideas

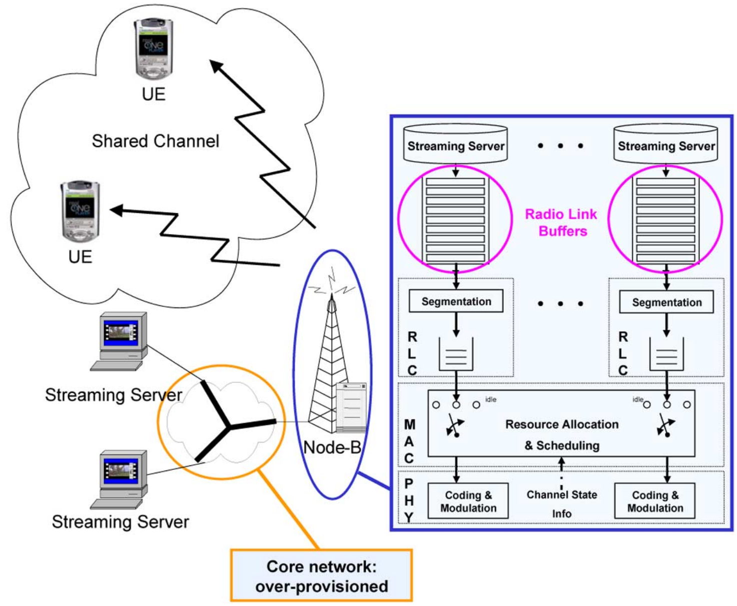

Fig. 5. Media-aware multi user HSDPA scheduling.

will have to be revisited and refined for successful deployments of the SVC technology in different environments.

IV. Example Applications—Media Delivery in Mobile Networks Using SVC

In this section we discuss three mobile video transmission systems using SVC. One approach discusses the use of SVC for streaming in wireless multiuser network environments and the other two approaches use SVC in combination with a rateless FEC code [22] and show the benefit of combining these two basic techniques.

A. *Wireless Multiuser Video Streaming Using SVC*

The first example presents a dynamic sharing of radio resources in a wireless multiuser system by combining SVC with appropriate radio link buffer management for multiuser streaming services [18].

Let us consider a wireless multiuser streaming environment similar to a system such as HSDPA as shown in Fig. 5. Assume that in total M users in the coverage area of a base station receive streaming multimedia data from a server. We assume that the core network is overprovisioned such that neither congestion nor losses between server and base station are an issue. The streaming server forwards the NAL units encapsulated in RTP packets directly into the *radio link buffers* at the base station, where they are stored until transmission to the media clients over the shared wireless link is scheduled. The Node-B in this case acts as a MANE. For each radio access slot a scheduler decides which users can access the wireless system resources, and a resource allocation unit assigns them appropriately. If the radio link buffers are not served fast enough because of bad scheduler decisions or too many streams are competing for the common resources, the system is in overload and typical congestion problems arise. In previous works [30], it has been shown that for real-time applications it is beneficial to operate with finite radio link buffer sizes and to drop data units already at the radio link buffer to reduce the excess load and avoid late-loss at the media client.

For H.264/AVC the 2-bit NRI header field and the NAL unit type differentiation between single slice and IDR (to determine the GOP structure) have been shown to be sufficient for an efficient drop strategy. For SVC, the improved layering combined extended priority labeling can be used to modify the radio link buffer management strategy such that higher priority data is kept in the buffer and low priority data is dropped earlier. For the details of this approach we refer to [18] and we will present selected simulation results.

For those, a looped Foreman sequence of 300 pictures (10 s) has been encoded with both, H.264/AVC and SVC at CIF resolution at 30 Hz. For both streams we apply a GOP size of 16 pictures and an intra frame distance of 2 s. The SVC stream has an H.264/AVC base layer at 160 kbps and two SNR refinement layers with an overall bit rate of about 390 kbps, which is the same bit rate as the H.264/AVC anchor. The encoder Y-PSNR is 36.4 dB for SVC and 37.0 dB for H.264/AVC. The wireless multiuser scenario contains a model of a HSDPA system (including fast fading and shadowing on the mobile radio channel), for details we refer to [18] and references therein. The scheduling strategy applied at the air interface is *maximum throughput*, i.e., the user which allows for highest data rate during the next 2 ms, is scheduled for transmission. The size of the radio link buffer is restricted to 110 KBytes.

In the experiment, $M = 4$ streaming users are connected to the base station. Table I shows selected simulation results: The average channel quality [signal-to-noise interference ratio (SNIR)] of each user is given in the first row. The overall playable picture rate of the medium (channel) quality user 1 is significantly increased in case of SVC when compared to H.264/AVC. This is due to the fact that the buffer load of this user can be reduced to the most important base layer fragments to achieve continuous playout. Furthermore, priority-based dropping of SNR refinement layers results in a smoother

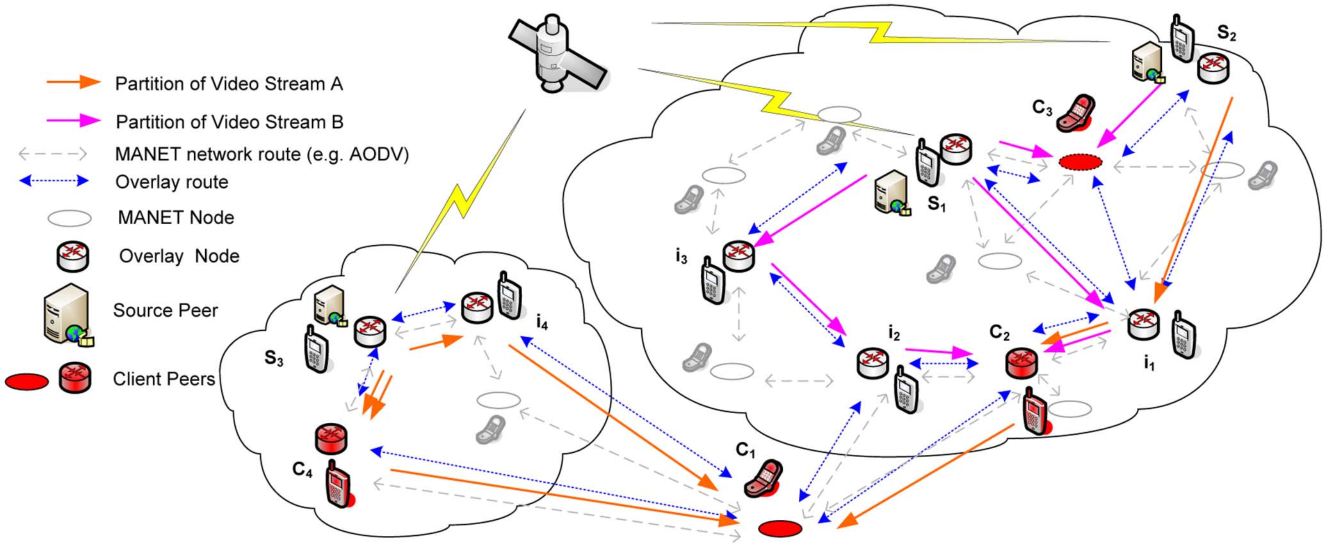
Partition of Video Stream A
Partition of Video Stream B
MANET network route (e.g. AODV)
Overlay route
MANET Node
Overlay Node
Source Peer
Client Peers
S2
C3
S1
i3
i2
C2
i1
S3
i4
C1
C4

Fig. 6. Overlay on top of a MANET for distributed and R/D optimized delivery of SVC.

TABLE I
SELECTED SIMULATION RESULTS FOR WIRELESS MULTIUSER STREAMING WITH H.264/AVC UND SVC

	User 1	User 2	User 3	User 4
Channel SNIR	12.9 dB	18.0 dB	17.9 dB	9.7 dB
H.264/AVC Y-PSNR	30.1 dB	37.0 dB	36.7 dB	15.4 dB
SVC Y-PSNR	34.9 dB	36.4 dB	36.1 dB	24.2 dB
H.264/AVC frame rate	19.7 Hz	30.0 Hz	29.5 Hz	0.8 Hz
SVC frame rate	30.0 Hz	30.0 Hz	29.5 Hz	13.5 Hz

variation in the PSNR over the duration of the stream. The performance of the worst quality user 4 is also increased in case of SVC. However, this user is still not able to continuously receive the full base layer, and temporal scalability is required to perceive at least a "slide show" at the media client. The improvements of both user 1 and 4 are not achieved at the expense of a bit rate reduction for users 2 and 3 when changing from H.264/AVC to SVC bit streams, but the slight decrease of at 0.6 dB in peak SNR (PSNR) in both cases is due to the loss in coding efficiency due to scalability. For more details on coding efficiency for SVC, see [45].

B. Distributed Video Streaming in Mobile Multi-Hop Networks Using SVC

The second approach [25] is related to real-time streaming in mobile ad hoc networks (MANETs) where the main problem solution is the increase of robustness in frequent playout for real-time media services in MANETs and is related to the approach presented by the authors in [17]. Mobile multihop networks or MANETs as discussed in Section III-B.4) have gained interest for delivery of multimedia content and other mobile services. However, if real-time delivery is an essential service requirement in a MANET service and streaming delivery needs to be used due to the associated delay constraints, reliability of such services is hard to achieve in MANETs. With common point-to-point transmission techniques such as link layer forward error correction or retransmission protocols, sufficiently good service quality in MANETs is often not possible, because of entire link outages and disconnections.

Therefore, the approach presented in [25] combines the benefits of cooperative interaction of a client (c_x) with multiple intermediate nodes (i_x)/source nodes (s_x) in an overlay network on top of a MANET for enhancing reliability in connectivity by source and path diversity. Fig. 6 gives an overview of such a topology. For suitable application layer QoS, the approach in [25] relies on two technologies: SVC and application layer FEC (AL-FEC).

The Raptor code [22] as an AL-FEC is an erasure correction code mainly used in environments with packet losses. Further, the rateless property of the Raptor code allows that a virtually infinite amount of independent encoding (output) symbols can be generated from a limited number of source (input) symbols. For a multiple source scenario, a randomization mechanism has been proposed in [17] for making the different Raptor encodings from different sources linear independent without the need for coordination among those sources. Because of this property, a Raptor decoder at a receiver is not concerned with which source a symbol originates from, but only the amount of received symbols. For successfully reconstructing source symbols, a number of encoding symbols only slightly higher than the number of source symbols has to be received. With this approach, the layers of an SVC bit stream are separately encoded at each of the multiple source nodes. The different layer encodings are then transferred via the overlay in different network streams including rate-distortion information of the whole SVC bit stream. This allows for rate-distortion optimization also at intermediate nodes.

Consider the situation where multiple clients are requesting partitions of different video streams from different sources. In this case, the transmission bit rate at intermediate nodes may be limited, which does not allow serving each rate request for a particular layer and bit stream for the connected clients. In this case the intermediate nodes apply an R/D optimization for the competing video streams. As input for this distributed optimization, the client feeds back the received rate from all sources, it is connected to. The client acts as the central instance for coordinating the R/D optimization by the different connected sources. The optimization has as result a new rate allocated for each connected client and is propagated to the clients. Based on these messages from the intermediate nodes, clients decide which actual rates for each SVC layer should be requested from the connected intermediate/source nodes, i.e., the client is partitioning its overall allocated rate (sum of rates allocated for a client by all connected sources) to individual subscription rates for each SVC layer. The described optimization procedure fulfills the important aspects of distributed processing. Each participating node carries out its own optimization and propagates the decisions which other nodes use for their own optimization.

We encoded three different ITU-T video sequences with an H.264/AVC base layer and two SVC fidelity enhancement layers with MGS, a GOP size of 16 and 1 IDR per GOP for random access. The streams are SVC encoded at rates from 40 kbps (base layer) to 150 kbps (highest layer), QCIF, 15 fps using SNR scalability. Five different rate points are achieved by removing NAL units of the enhancement layer from the bit stream starting with the lowest temporal priority. We simulated 40 scenarios with 30 nodes and 3 available servers over 60 min simulation time. We conducted experiments with different number of clients. Increasing the number of clients results in network saturation due to limited transmission capacity at overlay nodes. The nodes are moving with random patterns at random speed.

Fig. 7 shows avg. received video quality averaged over all clients in terms of PSNR for different methods. DRD denotes the distributed RD-optimized method proposed, PET refers to the Priority Encoding Transmission method of [17] and SINGLE refers to a state-of-the art single server system with rate adaptation. The results show that the RD-optimized approach performs consistently better than the other two (between 1–4 dB better than the single-server approach and approx. 2 dB better than PET). With the number of clients, the degrees of freedom increase for applying the RD-optimization. Therefore, the performance gain experienced by the DRD approach increases with the number of clients. Due to the connectivity-preserving property of PET, it performs better than the single server, but performs worse with a low number of clients due to the PET rate overhead. The DRD approach gives an average performance gain over the other two systems, since the connectivity of clients and the RD-information about the video streams is taken into account.

C. Scalable On-Demand Service Over Broadcast Channels

Real-time services over broadcast channels are especially attractive for live events as they allow many receivers to view the same content at the same time with highest radio efficiency. However, mobile users are more heterogeneous and diverse than users enjoying TV programs in front of their home TV equipment. They tune into a service, when they have spare time, but generally do not schedule their agenda to the offered TV program. Preferably, such users tune into an ongoing program quite arbitrarily which results in the unfortunate case that often the start of movie or an episode is missed. Therefore, for some services, it is highly desirable that content can be accessed at arbitrary time, i.e., *on-demand*. However, unidirectional mobile broadcast systems such as DVB-H do not offer this feature, nor is it bandwidth efficient and economically viable if popular content needs to be distributed to several or possibly many users in parallel.

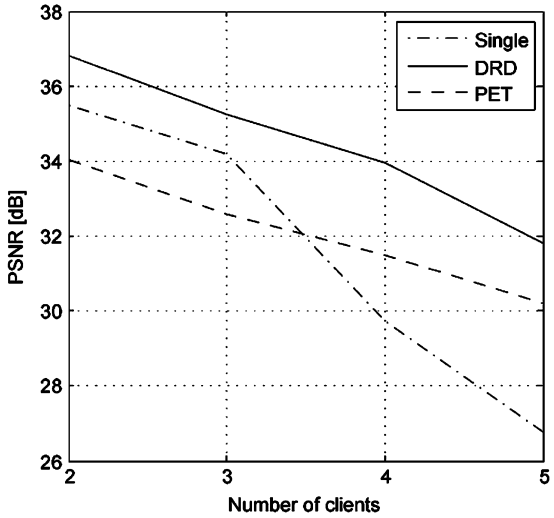

Fig. 7. Avg. PSNR as function of number of clients for different transport methods.

In this case, so called *clipcasting* is a far more attractive way to broadcast or multicast video or music clips. Users subscribe in order to “download” the collection of desired audio and video files. Once the receiver is in operation, it can tune to the service and collect data. The IETF protocol File Delivery over Unidirectional Transport (FLUTE) [38] (being part for example of MBMS and DVB-H IPDC [28]) provides means to deliver files over a unidirectional network. Most suitably, FLUTE is used with Raptor codes as proposed in MBMS and DVB-H [22].

Assume now the case of distributing multimedia files using FLUTE and SVC. Media files are encapsulated in 3 G file format, and for the case of SVC files, the encapsulation rules according to Section III-C have been used. If the sender exploits the fountain property of the Raptor code, receivers can arbitrarily tune to the service and receive the broadcast file. Thereby, losses do not affect the reliability of the reception, but only the time it takes a receiver to acquire the file. Once the amount of received Raptor symbols is just slightly larger as the included file, reconstruction of the broadcast file is possible and it is accessible for playout.

However, the video content of the file may not have the appropriate quality or resolution for the all receivers. If different target receivers are in the field, then simulcasting of the different quality versions may be an approach. In this case, three separate files, each with different size may have to be distributed. Due to the different size of the individual quality versions, the reception duration of one or the other may vary, always depending on the receiving conditions and the transmission rate. In this case it is obviously advantageous to use SVC and to distribute different layers in different FLUTE sessions with Raptor fountains.

The concept is shown in Fig. 8, whereby the different layers are broadcast individually, and receivers listen to only those streams/fountains, which match their presentation capability.

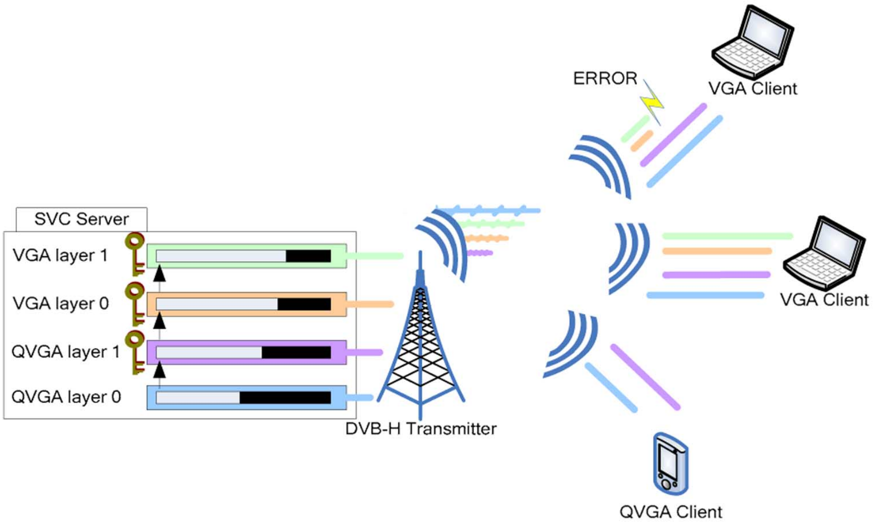

Fig. 8. Clipcasting SVC files in DVB-H.

Furthermore, as likely smaller files are received earlier, high end terminals may still decide to playout lower resolution files early, if the user wants to get access to the content as fast as possible. This flexible clipcasting approach allows for reliably serving receivers with different reception conditions, different display capabilities and different urgency to access the content.

Such a scheme has been investigated in more detail in [24] along with discussions on optimizations, etc. We present selected simulation results that highlight the mentioned benefits. For the assessment we assume that the system provides three quality layers, Q1 (QCIF, 15 fps, 37.1 dB), Q2 (CIF, 15 fps, 34.7 dB), and Q3 (CIF, 30 fps, 37.7 dB). We encode single layer H.264/AVC and SVC to achieve these values, whereby the quality is the same, only the bit rates are different. The clip has duration of 5 min. Then for single layer transmisson ($M = 1$) each of the clips is fountain encoded and transmitted with a rate of total 256 kbps, but from three surrounding transmission sites. Each transmission site supplies the network with an independent fountain such that in case of multiple site reception, the Raptor symbols can be combined before decoding. In case of simulcast, each of the three quality versions are encoded and transmit, whereby the rates among the different layers are split according to some optimization [24]. In the same way for the SVC encoded files, each layer is transmitted in a separate fountain according to the same optimization criteria. For different reception conditions [in total 6 cases are evaluated from 3 BS6 (worst case) to 3 BS1 (best case)], it is measured, how long it takes on average until a certain quality layer can be recovered. Fig. 9 shows the average delay, normalized by the length of the clip, over the receiver quality.

Obviously, in all cases with better reception quality, the access delay decreases as the Raptor decoding process has access to a sufficient symbol set earlier in time. The single layer case with Q1 provides best quality, but then the system will only provide the lowest quality. If single layer Q2 or Q3 are provided only, then receivers with lower capabilities will be excluded. Obviously the time until the playout lasts longer for higher qualities, as the size of the files are larger. To support the feature of multiple file reception with a single layer video codec, simulcast (SimC) may be applied. In this case, the rate must be shared by the quality versions, which obviously prolongs the time to reconstruct the file for each quality. If we apply SVC instead of simulcast, then this time can be reduced significantly for quality layers Q2 (by a factor of 1.5) and Q3 (by a factor of 2). In addition, receivers with better display capabilities may still decide to decode the lower resolution earlier. Therefore, the entire user experience flexibility is shifted from the sender to the receiver.

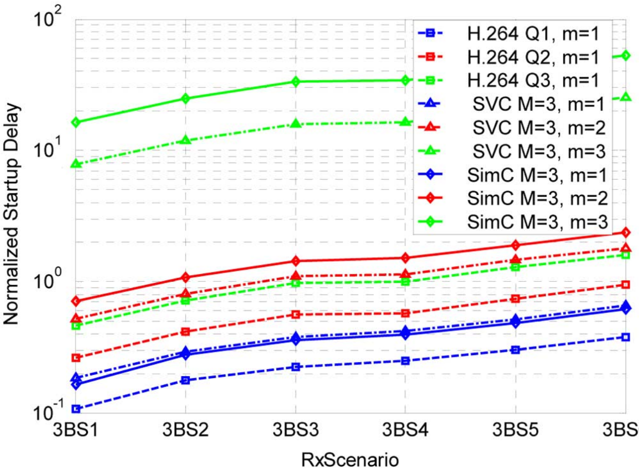

Fig. 9. Delay over channel quality (improving from right to left) for clipcasting of H.264/AVC with quality Q1, Q2, and Q3, SVC coding, and simulcast.

The clipcasting as presented has one drawback compared to live streaming as one has to wait quite long until the playout of the clip will start. For this purpose, virtual on-demand broadcasting schemes exist which allow this property: The most popular schemes are known asg *Harmonic Broadcasting* (HB) [26] and *Pyramid Broadcasting* (PB) [27]. The basic idea is to provide different segments of the content on different bearers. Once being tuned to the desired content, and after having received the first segment, the playout of this segmented is started while the remaining segments are received. The same may be applied to the other segments as well. Whereas HB divides the content in equally sized segments but distributes the segments at different rates (decreasing with segment number). For PB, the distribution rates are equal, but the segment sizes are different (increasing with segment numbers).

In [24], HB has been combined with the layered fountain approach referred to as Layered HB (LHB). Then, instead of providing a fountain just for each layer, for each segment in each layer a fountain is provided. The segmentation is still harmonic but the number of segments may be different on each layer. Among other aspects, it is shown, that if users want to access a low quality stream quite fast, then only a very low start-up delay may be necessary. If however, one waits longer or only accesses the clip at a later stage, the full resolution of the video is decodable. LHB is characterized that a significant amount of quality control is shifted to receivers, the service includes inherent scalability in terms of quality and playout delay, and that due to the fountain approach basically full reliability can be achieved. The approach may be integrated into an existing DVB-H IPDC CDP with only very minor changes, see [24].

This LHB concept may be further extended, by providing a mixture of low-quality real-time streaming with a base layer (possibly backward compatible to existing systems) and a download delivery based on the clipcasting approach, such that the live stream may be recorded, and enhanced by using additional quality layers and FEC symbols. Services like this may also be attractive for conditional access.

V. Conclusion

In this work, we describe the potential use of SVC in mobile networks. Further we outline use cases of mobile media delivery, which can benefit from using SVC. We give examples showing the impact of SVC on existing media delivery services and techniques. In general, it is obvious from all the results that the flexibility provided by SVC provides significant opportunities for network integration. Nevertheless, it is important that SVC is integrated in existing and emerging networks in established environments, e.g., by the use of the 3 G file format and RTP. Then, a smooth extension of emerging H.264/AVC-based services will provide new potentials for network operators and end users.

Acknowledgment

The authors would like to thank all co-authors of cited papers, and in particular C. Hellge, G. Liebl, T. Gasiba, W. Abdel-Samad, and S. Johansen. They also appreciate the valuable comments of the reviewers.

References

[1] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC, v3: 2005, Amendment 3: Scalable Video Coding.

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

[3] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, v3: 2005.

[4] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[5] M. Kalman, P. van Beek, and B. Girod, "Optimized transcoding rate selection and packet scheduling for transmitting multiple video streams over a shared channel," in *Proc. IEEE Int. Conf. Image Process.*, Genoa, Italy, Sep. 2005, pp. 165–168.

[6] S. Wenger, Y.-K. Wang, and T. Schierl, "Transport and signaling of SVC in IP networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1164–1173, Sep. 2007.

[7] P. Amon, T. Rathgen, and D. Singer, "File format for scalable video coding (SVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1174–1185, Sep. 2007.

[8] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," Microsoft Research, Tech. Rep. MSR-TR-2001-35, Feb. 2001.

[9] H. Schwarz, D. Marpe, and T. Wiegand, *Closed Loop Coding With Quality Layers* Joint Video Team (JVT) of ISO-IEC MPEG & ITUT VCEG, Nice, JVT-Q030, Oct. 2005.

[10] *Transparent End-To-End Packet-Switched Streaming Service (PSS); Protocols and Codecs (Release 6)*, 3 GPP TS 26.234., Apr. 2006.

[11] T. Schierl, M. Kampmann, and T. Wiegand, "3 GPP compliant adaptive wireless video streaming using H.264/AVC," in *Proc. IEEE Int. Conf. Image Process.*, Genoa, Italy, Sep. 2005, pp. 696–699.

[12] *Multimedia Multicast and Broadcast Service (MBMS); Protocols and Codecs (Rel. 6)*, 3 GPP TS 26.346., Sep. 2005.

[13] *Digital Video Broadcasting (DVB);*, IP Datacast over DVB-H: Content Delivery Protocols, ETSI TS 102 472 (V1.1.1), Jun. 2006.

[14] *Digital Video Broadcasting (DVB);*, Specification for the use of Video and Audio Coding in DVB services delivered directly over IP Protocols, ETSI TS 102 005 (V1.2.1), Apr. 2006.

[15] *Digital Audio Broadcasting (DAB);*, DMB video service; User Application Specification, ETSI TS 102 428 V1.1.1, Jun. 2005.

[16] IETF Mobile Ad-Hoc Network (MANET) Working Group [Online]. Available: http://www.ietf.org/html.charters/manet-charter.html

[17] T. Schierl, K. Gänger, T. Stockhammer, and T. Wiegand, "SVC-based multi source streaming for robust video transmission in mobile ad-hoc networks," *IEEE Wireless Commun. Mag.*, vol. 13, no. 5, pp. 96–103, Oct. 2006.

[18] G. Liebl, T. Schierl, T. Wiegand, and T. Stockhammer, "Advanced wireless multiuser video streaming using the scalable video coding extension of H.264/AVC," in *IEEE Int. Conf. Multimedia Expo (ICME'06)*, Toronto, ON, Canada, Jul. 2006, pp. 625–628.

[19] T. Schierl, M. Kampmann, and T. Wiegand, "H.264/AVC interleaving for 3 G wireless video streaming," in *Int. IEEE Conf. Multimedia Expo (ICME)*, Amsterdam, The Netherlands, Jul. 2005, pp. 868–871.

[20] T. Stockhammer, "Robust system and cross-layer design for H.264/AVC-based wireless video applications," *EURASIP J. Appl. Signal Process., Special Issue on Video Analysis and Coding for Robust Transmission*, vol. 2006, Mar. 2006, 89371.

[21] A. Albanese, J. Bloemer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.

[22] M. Luby, T. Gasiba, T. Stockhammer, and M. Watson, "Reliable multimedia download delivery in cellular broadcast networks," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pt. 2, pp. 235–246, Mar. 2007.

[23] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.

[24] T. Stockhammer, T. Gasiba, W. A. Samad, T. Schierl, H. Jenkac, T. Wiegand, and W. Xu, "Nested harmonic broadcasting for scalable video over mobile datacast channels," *J. Wireless Commun. Mobile Compu., Special Issue on Video Communications for 4 G Wireless Systems*, vol. 7, no. 2, pp. 235–256, Feb. 2007.

[25] T. Schierl, S. Johansen, C. Hellge, T. Stockhammer, and T. Wiegand, "Distributed rate-distortion optimization for rateless coded scalable video in mobile ad-hoc networks," in *ICIP 2007*, San Antonio, TX, Sep. 2007, to appear.

[26] L. Engebretsen and M. Sudan, "Harmonic broadcasting is bandwidth-optimal assuming constant bit rate," in *Proc. Annual ACM-SIAM Symp. Discrete Algorithms 2002*, San Francisco, CA, Jan. 2002, pp. 431–432.

[27] H. Jenkac and T. Stockhammer, "Asynchronous media streaming over wireless broadcast channels," in *Proc. Int. Conf. Multimedia Expo (ICME)*, Amsterdam, The Netherlands, Jul. 2005, pp. 1318–1321.

[28] M. Kornfeld and G. May, "DVB-H and IP datacast-broadcast to handheld devices," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pt. 2, pp. 161–170, Mar. 2007.

[29] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," in *Proc. ACM SIGCOMM*, Aug. 2006, vol. 26, no. 4, pp. 117–130.

[30] G. Liebl, H. Jenkac, T. Stockhammer, and C. Buchner, "Radio link buffer management and scheduling for wireless video streaming," *Telecommun. Syst.*, vol. 30, no. 1–3, pp. 255–277, Nov. 2005.

[31] S. Adlakha, X. Zhu, B. Girod, and A. Goldsmith, "Joint capacity, flow and rate allocation for multiuser video streaming over wireless ad-hoc networks," in *Proc. IEEE Int. Conf. Commun.*, Glasgow, Scotland, Jun. 2007, pp. 1747–1753.

[32] *Transport of SVC video over ISO/IEC 13818-1 streams*, ITU-T Rec. H.222 & ISO/IEC 13818-1:2006/PDAM 3.

[33] Y.-K. Wang, M. M. Hannuksela, S. Pateux, and A. Eleftheriadis, "System and transport interface to SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1149–1163, Sep. 2007.

[34] G. Faria, J. A. Henriksson, E. Stare, and P. Talmola, "DVB-H: digital broadcast services to handheld devices," *Proc. IEEE*, vol. 94, no. 1, pp. 194–209, Jan. 2006.

[35] M. Podolsky, S. McCanne, and M. Vetterli, "Soft ARQ for layered streaming media," Univ. California, Computer Science Division, Berkeley, CA, Tech. Rep. UCB/CSD-98-1024, Nov. 1998.

[36] M. R. Chari, F. Ling, A. Mantravadi, R. Krishnamoorthi, R. Vijayan, G. K. Walker, and R. Chandhok, "FLO physical layer: An Overview," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pt. 2, pp. 145–160, Mar. 2007.

[37] A. M. C. Correia, J. C. M. Silva, N. M. B. Souto, L. A. C. Silva, A. B. Boal, and A. B. Soares, "Multi-resolution broadcast/multicast systems for MBMS," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pt. 2, pp. 224–234, Mar. 2007.

[38] T. Paila, M. Luby, R. Lehtonen, V. Roca, and R. Walsh, *FLUTE—File Delivery Over Unidirectional Transport* Oct. 2004, IETF RFC3926.

[39] U. Horn, K. W. Stuhlmüller, M. Link, and B. Girod, "Robust internet video transmission based on scalable coding and unequal error protection.," *Image Commun., Special Issue on Real-Time Video over the Internet*, vol. 15, no. (1-2), pp. 77–94, Sep. 1999, .

[40] H. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 53–68, Mar. 2001.

[41] S. Wenger, M. M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, *RTP payload format for H.264 video* IETF RFC 3984, Feb. 2005.

[42] K. Nichols, S. Blake, F. Baker, and D. Black, *Definition of the differentiated services field (DS Field) in the IPv4 and IPv6 headers* IETF RFC 2474, Dec. 1998.

[43] C. Perkins, E. Belding-Royer, and S. Das, *Ad hoc On-demand distance vector (AODV) routing* IETF RFC 3561, Jul.4 2003.

[44] T. Clausen and P. Jacquet, Optimized Link State Routing Protocol (OLSR)," IETF RFC 3626, Oct. 2003.

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

[46] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, ISO/IEC 8802-11: 2005, Aug. 2005.

[47] *IEEE P802.11 ESS mesh working group*, IEEE 802.11s draft D1.0, May 2007.

[48] *Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Multihop Relay Specification*, IEEE 802.16 j, IEEE P802.16 relay task group, July 2007.

Thomas Schierl (M'06) received the Dipl.-Ing. degree in computer engineering from the Berlin University of Technology, Berlin, Germany in December 2003.

He has been with Fraunhofer Institute for Telecommunications, Heinrich-Hertz Institute (HHI), Berlin, Germany, since 2003. As Project Manager, he is responsible for various scientific and industry research projects. He conducted different research works on reliable real-time transmission of H.264/MPEG-4 AVC and scalable video coding (SVC) in mobile point-to-point, point-to-multipoint, and broadcast environments like used by 3 GPP and DVB-H. He submitted various inputs on real-time streaming to standardization committees like 3 GPP, JVT/MPEG; ISMA, and IETF. He is one of the authors of the IETF RTP SVC payload format and author of the related SDP draft on signaling layered codecs. Furthermore, he is one of the authors of the SVC Amendment for MPEG-2 TS. In 2007, he was visiting the group of Prof. B. Girod at Stanford University, Stanford, CA, for different research activities. His current research work mainly focuses on developing new real-time streaming techniques for video delivery in mobile ad hoc networks (MANETs). Further research interests are in reliable transmission of real-time media in mobile networks and joint source channel coding, as well as the deployment of SVC in mobile networks.

Thomas Stockhammer (M'99) has been working at the Munich University of Technology, Munich, Germany, and was a Visiting Researcher at Rensselear Polytechnic Institute (RPI), Troy, NY, and at the University of San Diego, California (UCSD).

He has published more than 100 conference and journal papers, is member of different program committees and holds several patents. He regularly participates and contributes to different standardization activities, e.g., JVT, ITU-T, IETF, 3 GPP, and DVB and has co-authored more than 150 technical contributions. He is acting Chairman of the video adhoc group of 3 GPP SA4. He is also co-founder and CEO of Novel Mobile Radio (NoMoR) Research, a company developing simulation and emulation of future mobile networks such as HSxPA, WiMaX, MBMS, and LTE. The company also provides consulting services in the respective areas. Between 2004 and June 2006, he was working as a research and development consultant for Siemens Mobile Devices, later BenQ mobile in Munich, Germany. Now he is consulting for Digital Fountain, Inc., in research and standardization matters for CDPs, IPTV, and mobile multimedia transmission. His research interests include video transmission, cross-layer and system design, forward error correction, content delivery protocols, rate-distortion optimization, information theory, and mobile communications.

Thomas Wiegand (M'05) received the Dipl.-Ing. degree in electrical engineering from the Technical University of Hamburg-Harburg, Germany, in 1995 and the Dr.-Ing. degree from the University of Erlangen-Nuremberg, Germany, in 2000.

He is the Head of the Image Communication Group, Image Processing Department, Fraunhofer Institute for Telecommunications—Heinrich Hertz Institute Berlin, Germany. From 1993 to 1994, he was a Visiting Researcher at Kobe University, Kobe, Japan. In 1995, he was a Visiting Scholar at the University of California at Santa Barbara. From 1997 to 1998, he was a Visiting Researcher at Stanford University, Stanford, CA, and served as a consultant to 8 × 8, Inc., Santa Clara, CA. He is currently a member of the technical advisory boards of the two start-up companies: Layered Media, Inc., Rochelle Park, NJ, and Stream Processors, Inc., Sunnyvale, CA. Since 1995, he has been an active participant in standardization for multimedia, with successful submissions to ITU-T VCEG, ISO/IEC MPEG, 3GPP, DVB, and IETF. In October 2000, he was appointed as the Associated Rapporteur of ITU-T VCEG. In December 2001, he was appointed as the Associated Rapporteur/Co-Chair of the JVT. In February 2002, he was appointed as the Editor of the H.264/AVC video coding standard and its extensions (FRExt and SVC). In January 2005, he was appointed as Associated Chair of MPEG Video. His research interests include video processing and coding, multimedia transmission, semantic image representation, and computer vision and graphics.

Dr. Wiegand received the SPIE VCIP Best Student Paper Award in 1998. In 2004, he received the Fraunhofer Award for Outstanding Scientific Achievements and the ITG Award of the German Society for Information Technology. Since January 2006, he has been an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT). In 2003, he was a Guest Editor of IEEE TCSVT Special Issue on the H.264/AVC Video Coding Standard in July 2003.