HEVC and Layered HEVC for UHD deployments

Ye-Kui Wang, Qualcomm

Workshop on Media Synchronization for Hybrid Delivery
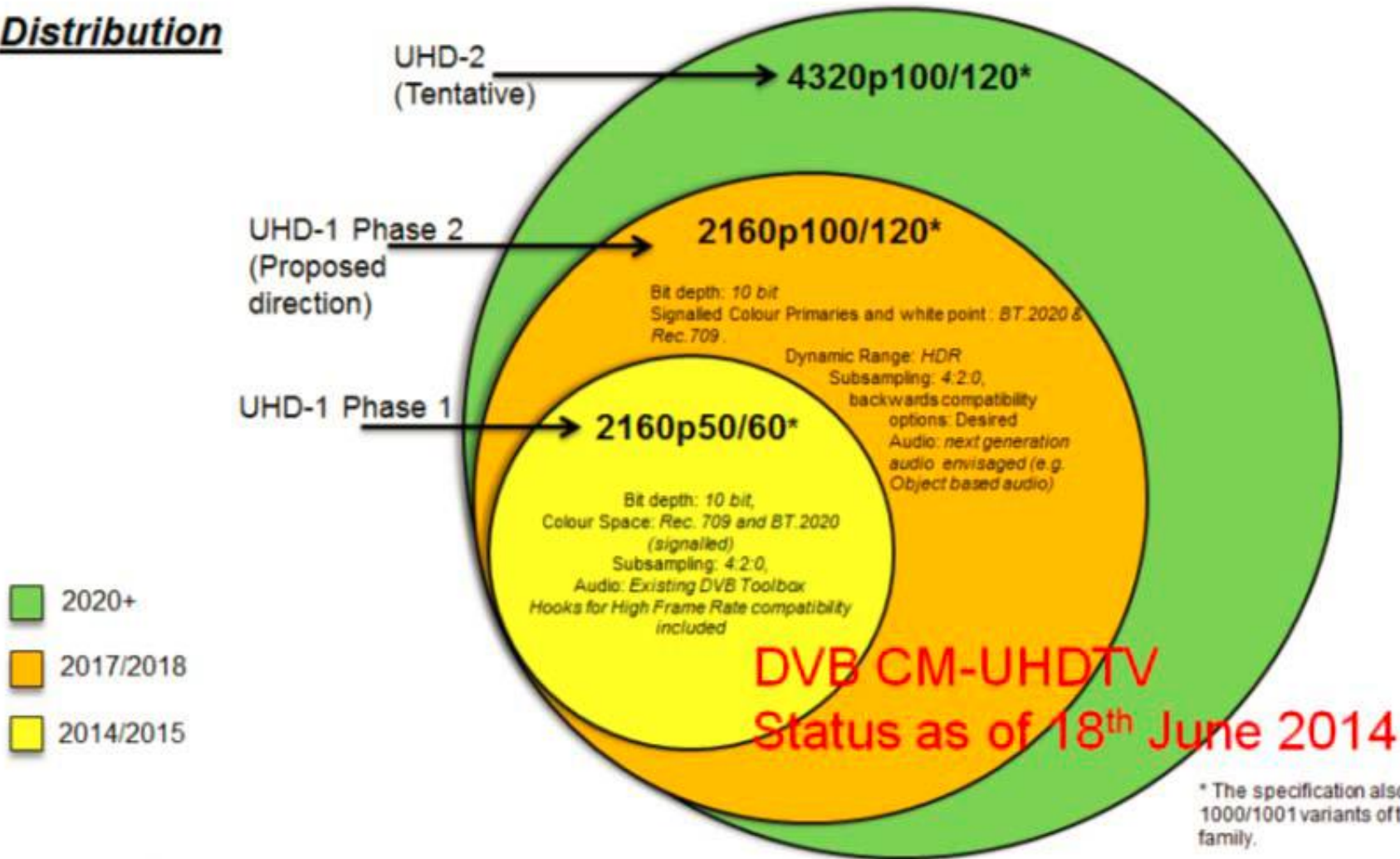
Oct. 22, 2014, Strasbourg, France

Outline

- Status and plan of DVB and ATSC
- Fundamental benefits and disadvantages of scalable video coding
- Comparison of SHVC vs SVC
- Use cases of SHVC in different applications
- SHVC coding performance
- Concluding remarks
- Acknowledgements

DVB status and plan

UHD-1 Phase 1 has been specified in ETSI TS 101 154 V1.12.1.

UHDTV currently only specified with HEVC.

ATSC status and plan

- ATSC 3.0 requirements have been stable for a while, technical specifications all being developed

- Fixed devices: Video resolutions up to Ultra High Definition 3840 × 2160 at 60 fps (or such higher frame rates and/or resolutions as may be determined to be desirable and practical)

  - System shall support aspects of UHD in addition to resolution and frame rates (e.g., higher dynamic range, greater color gamut, higher bit-depth, and denser chroma subsampling) as determined to be desirable and practical.

- Portable and mobile devices: Video resolution up to High Definition 1920 × 1080 at 60 fps

- HEVC is the only agreed video codec so far for ATSC 3.0, details of the support are not yet finalized

The fundamental benefits and disadvantages of scalable video coding (1/2)

- The benefits come from the compression advantage of layered coding utilizing inter layer prediction compared to "simulcast".

- Here, "simulcast" means that different versions of the same video content are encoded independently.

- The benefits would only apply in scenarios where different versions of the same video content are necessary.

  - They thus do not apply to one-to-one communications, such as video telephony, where only one version is needed for one video content

    - Even when adaptive bit rate is needed, rate control can be used, thus still just one version of the video content

    - There are corner cases where the network needs to do rate adaptation in order to react to immediate bitrate changes, for example if network buffers are overloaded. However, this requires deep integration of application layer technologies and network buffers and is therefore not a primary use case.

The fundamental benefits and disadvantages of scalable video coding (2/2)

- The benefits apply to one-to-many communications, including broadcast, streaming (live and VoD), and video conferencing where multiple receivers need different versions of the same video
  - Here concrete benefits include bandwidth saving in "backbone" network (between the origin server and caches/proxies), or more programs/channels, or higher qualities for the same programs --- good for service providers as well as reducing backbone network congestion
    - The fundamental disadvantages are higher bandwidth and higher decoding/decoder complexity for high-end users who only need the enhancement layer signal
    - Thus whether use of scalable coding makes sense to the extent for (wide) deployment depends on how big are the benefits vs the disadvantage, business model, and so on
    - One key aspect is whether the increased bandwidth and decoding/decoder complexity can be controlled to be non-significant
  - Another concrete benefit is reduced storage size, which in today's VoD environment can translate into significant cost reduction
  - Does not apply much for long-tail contents consumed by few users, which are similarly as one-to-one

SHVC vs SVC

Aspects		SVC	SHVC
Scalability features	Overall feature set	3 (temporal, spatial, SNR)	6 (temporal, spatial, SNR, bit depth, color gamut, hybrid codec)
	Spatial scalability	Limited to 2x and 1.5x in Baseline	Arbitrary ratio, field picture in BL, downsampling filter flexibility, etc
Backward compatibility		AVC coded base layer only	HEVC or non-HEVC coded base layer
Decoder design	BL decoder	Requires new API design in the BL	No change to the base layer decoder
	EL decoder	Cannot reuse AVC decoder	With no block level changes, easy to repurpose existing HEVC decoder products into SHVC decoder
Encoder design		Cross layer optimization complicates rate control	Easy to reuse existing HEVC encoder products
Splicing and random access		Decoders are required to be more intelligent than conforming decoders to utilize open-GOP RAPs and non-cross-layer aligned RAPs, and consequently IOP in this aspect is not guaranteed	Decoders are NOT required to be more intelligent than conforming decoders to utilize open-GOP RAPs and non-cross-layer aligned RAPs, and consequently IOP in this aspect is guaranteed

SHVC in video broadcast (1/2)

- Advantages:
  - The fundamental benefits (for all contents: less bandwidth in the backbone network, more programs, or higher quality; less backbone network congestion; reduced storage size)
  - Reuse of two HEVC 4K decoders in set-top boxes (which was said to be a standard feature for pic-in-pic function) to do 1080p-to-4K scalability almost for free (only a re-sampler and some firmware changes needed)
  - Use of multi-layer SHVC to add more temporal sub-layers (i.e. to enhance frame rate) or layers in the future – a use case discussed in DVB
  - Use of SHVC in order to add advanced video experience (HDR, BT.2020 color gamut)
  - Support enhancement to legacy AVC devices as well (with AVC as the base layer, enhancement layers coded using SHVC)
  - Other benefits, such as those documented for eMBMS in 3GPP TR 26.904 for SVC but applicable to SHVC as well
    - Unequal error protection, layer-aware transmission for coverage and capacity improvements, graceful degradation, fast stream switching

SHVC in video broadcast (2/2)

- Disadvantages:
  - The fundamental disadvantages (higher bandwidth and decoder/decoding complexities for high-end users who only need the enhancement layer signal

- Ways to alleviate the disadvantages
  - SHVC general encoding optimization
  - More frequent IRAPs for lower layers, could potentially let high-end receivers have lower bandwidth than simulcast
  - Apply UEP to the layered bitstream, where less channel coding redundancy is added to the enhancement layer, partially or fully offsetting the increased bandwidth of source coding
  - Etc.

SHVC in video streaming (1/2)

- Advantages:

  - The fundamental benefits (for popular contents: less bandwidth in the backbone network, more programs, or higher quality, less backbone network congestion; reduced storage size)

  - Efficient delay optimization in live streaming based on DASH

    - A currently popular low-latency live streaming solution is to simulcast multiple rates/qualities, and the receiver uses whatever that have been received and playable at a time based on the low-latency requirement and discard other data. Layered coding can potentially provide very good benefit compared to simulcast in this scenario.

  - Improved cache hit ratio for adaptive streaming (DASH)

    - Some detailed results can be seen from Section 6.1.3.7 of 3GPP TR 26.904

SHVC in video streaming (2/2)

- Disadvantages:
  - The fundamental disadvantages (higher bandwidth and decoder/decoding complexities for end users)

- Ways to alleviate the fundamental disadvantages
  - Higher bandwidth for end users:
    - SHVC general encoding optimization
    - Different IRAP frequencies for different layers
    - Duration between IRAPs larger than desired end-to-end latency
  - Decoder and decoding complexities: more severe than other services due to the higher number of layers needed
    - Decoder complexity: Less of an issue for SHVC compared to SVC because of the high-level syntax only design of SHVC
    - Decoding complexity: Control of the inter-layer prediction chain length to trade-off between decoding complexity and compression efficiency

SHVC in video conferencing (1/2)

- Advantages:
  - The fundamental benefits (less bandwidth in the backbone network, more participants, or higher quality; less backbone network congestion)
  - Reduced end-to-end delay (improved user experience for a given end-to-end delay)
  - Using 3 layers and unequal error protection, e.g.,
    - Client 1 encodes at 1080p with 3 layers (1080p, 540p, 270p).
    - Client 2 is on a high quality network with a powerful machine, and decodes all 3 layers, with a target output layer of 1080p
    - Client 3 is on a lower bitrate network and/or lower CPU machine, and decodes 2 layers (540p and 270p).
    - The base 270p layer is used for error resiliency and network adaptation when the 540p layer is occasionally lost or intentionally dropped.

SHVC in video conferencing (2/2)

- Disadvantages:
  - The fundamental disadvantages (higher bandwidth and decoder/decoding complexities for end users)
- Ways to alleviate the fundamental disadvantages
  - SHVC general encoding optimization
  - More frequent IRAPs for lower layers, could potentially let high-end receivers have lower bandwidth than simulcast

SHVC common test conditions

- Types of scalability tests include:
  - Spatial scalability of 2x and 1.5x
  - SNR scalability (1x spatial ratio, low quality to high quality)
  - Combined scalability (1080p 8-bit BT.709 to 4K 10-bit BT.2020)

- Test sequences:
  - 7 pairs of sequences for spatial and SNR scalability
  - 5 pairs of sequences for combined scalability

- Coding configurations:
  - All Intra, random access, low delay
  - Random access is the most relevant for broadcast video applications and this set of results is provided

Spatial scalability (HEVC coded base layer)

Spatial ratio	2x		1.5x	
Anchors	BL+EL	EL only	BL+EL	EL only
Traffic	-14.2%	-22.8%		
PeopleOnStreet	-22.7%	-37.1%		
Kimono	-23.7%	-38.8%	-35.0%	-64.5%
ParkScene	-15.2%	-23.3%	-27.0%	-48.2%
Cactus	-14.2%	-22.6%	-26.1%	-47.5%
BasketballDrive	-17.8%	-26.9%	-29.7%	-50.9%
BQTerrace	-7.7%	-10.7%	-17.2%	-26.9%
Average	-16.5%	-26.0%	-27.0%	-47.6%

* In this and the next 3 slides: performance is measured against simulcast for both "BL+EL" and "EL only"

Spatial scalability (AVC coded base layer)

Spatial ratio	2x		1.5x	
Anchors	BL+EL	EL only	BL+EL	EL only
Traffic	-11.3%	-19.7%		
PeopleOnStreet	-17.6%	-31.1%		
Kimono	-15.8%	-29.1%	-22.4%	-49.1%
ParkScene	-12.7%	-20.4%	-21.8%	-41.7%
Cactus	-10.9%	-18.5%	-18.6%	-37.9%
BasketballDrive	-12.1%	-19.9%	-20.4%	-39.8%
BQTerrace	-6.2%	-9.2%	-13.2%	-22.2%
Average	-12.4%	-21.1%	-19.3%	-38.1%

SNR scalability (HEVC coded base layer)

Spatial ratio	1x (SNR)	
Anchors	BL+EL	EL only
Traffic	-18.5%	-29.3%
PeopleOnStreet	-25.2%	-38.6%
Kimono	-23.4%	-35.8%
ParkScene	-20.6%	-31.7%
Cactus	-19.3%	-29.6%
BasketballDrive	-22.9%	-33.4%
BQTerrace	-16.8%	-21.8%
Average	-21.0%	-31.5%

Combined scalability (HEVC coded base layer)

Scalability type	Spatial, bit depth, and color gamut scalability		Color gamut scalability only	
Anchors	BL+EL	EL only	BL+EL	EL only
Birthday	-17.2%	-30.0%	-31.2%	-60.1%
BirthdayFlash1	-19.3%	-32.2%	-29.2%	-55.9%
BirthdayFlash2	-18.1%	-30.7%	-28.5%	-55.3%
Parakeets	-23.6%	-38.9%	-33.8%	-64.1%
Tablecar	-15.2%	-22.8%	-26.2%	-50.2%
Average	-18.7%	-30.9%	-29.8%	-57.1%

Compared to single-layer coding - Spatial and SNR scalability

Spatial ratio	2x	1.5x	SNR (1x)
Traffic	24.0%		20.1%
PeopleOnStreet	18.8%		13.8%
Kimono	13.9%	9.2%	14.1%
ParkScene	18.5%	17.8%	15.9%
Cactus	22.6%	20.4%	16.8%
BasketballDrive	17.6%	13.9%	12.2%
BQTerrace	17.2%	18.6%	7.3%
Average	18.9%	16.0%	14.3%

* In this and the next slide: performance is measured against single-layer coding of the higher-quality signal.

Compared to single-layer coding - Combined scalability

Scalability type	Spatial, bit depth, and color gamut scalability	Color gamut scalability only
Birthday	28.2%	23.3%
BirthdayFlash1	21.5%	27.5%
BirthdayFlash2	23.5%	29.2%
Parakeets	15.6%	18.5%
Tablecar	15.8%	38.9%
Average	**20.9%**	**27.5%**

Concluding remarks

- HEVC will definitely play a key role in UHD deployments
- Layered HEVC extensions can be utilized in UHD deployments with backward compatibility in an efficient manner

Acknowledgements

- The slides on SHVC coding performance and SHVC vs SVC comparison were mainly from Yan Ye of InterDigital

- Thomas Stockhammer and Michael Luby of Qualcomm and Yan Ye of InterDigital reviewed much of the content and contributed to various aspects

- The author heard some of the mentioned use cases firstly from Minhua Zhou of Broadcom, Jill Boyce of Vidyo, and Ajay Luthra and Sam Narasimhan of Arris