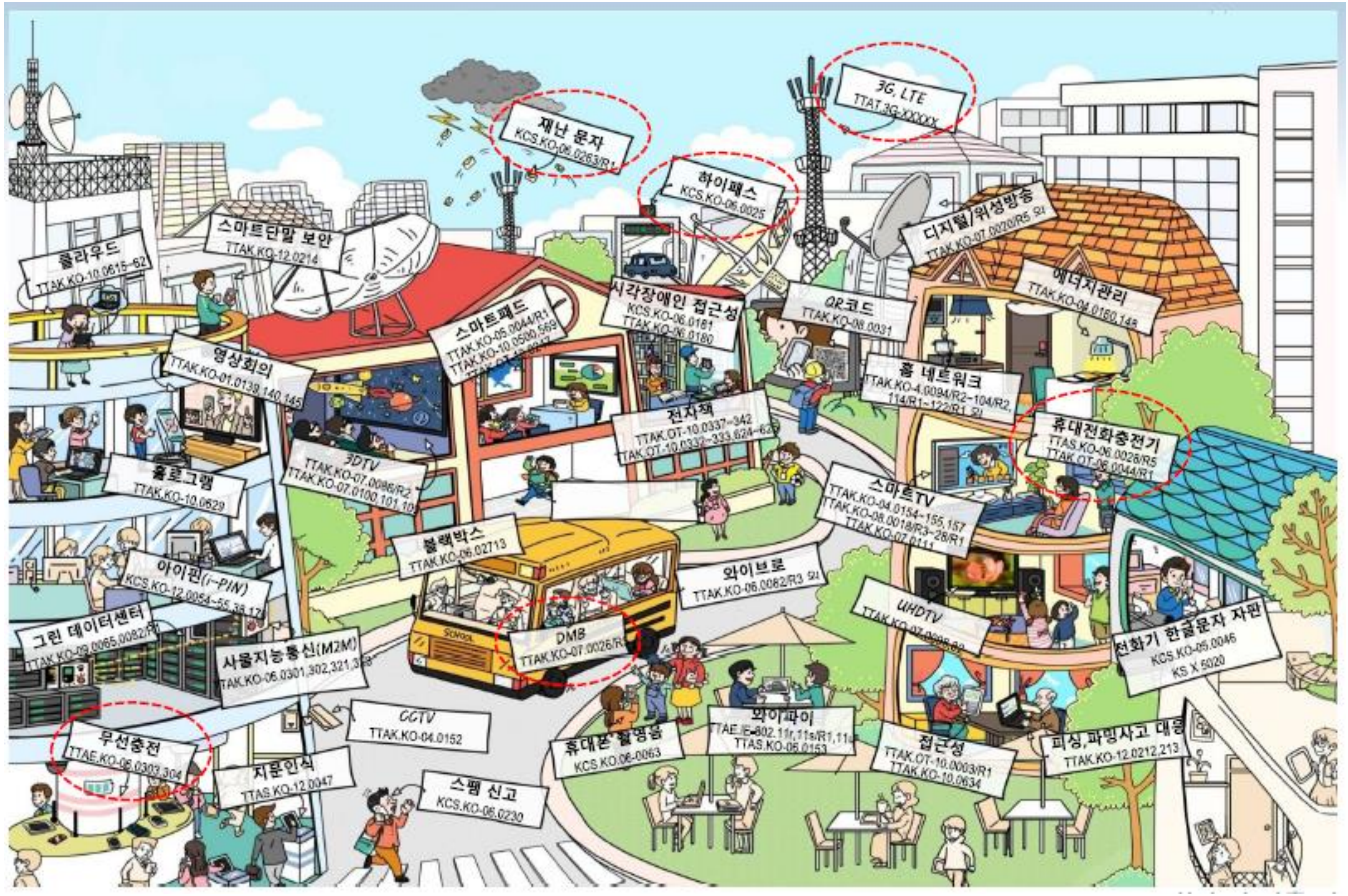
Video Coding Standard
& AI based Video Coding Trend

가천대학교 AI-Software 학부
최기호

Standard

Standards?



재난 문자
KCS.KO-06.0263/R1
3G, LTE
TTAT.3G-XXXXX
하이패스
KCS.KO-06.0025
클라우드
TTAK.KO-10.0615~62
스마트단말 보안
TTAK.KO-12.0214
디지털/위성방송
TTAK.KO-07.0020/R5 외
스마트패드
TTAK.KO-05.0044/R1
TTAK.KO-10.0500,569
시각장애인 접근성
KCS.KO-06.0181
TTAK.KO-06.0180
QR코드
TTAK.KO-08.0031
에너지관리
TTAK.KO-04.0160,148
영상회의
TTAK.KO-01.0139,140,145
홈 네트워크
TTAK.KO-4.0094/R2~104/R2,
114/R1~122/R1 외
전자책
TTAK.OT-10.0337~342
TTAK.OT-10.0332~333,624~62
휴대전화충전기
TTAS.KO-06.0026/R5
TTAK.OT-06.0044/R1
3DTV
TTAK.KO-07.0086/R2
TTAK.KO-07.0100,101,10
홀로그램
TTAK.KO-10.0629
스마트TV
TTAK.KO-04.0154~155,157
TTAK.KO-08.0018/R3~28/R1
TTAK.KO-07.0111
블랙박스
TTAK.KO-06.02713
아이핀(i-PIN)
KCS.KO-12.0054~55,38,17
와이브로
TTAK.KO-06.0082/R3 외
UHDTV
그린 데이터센터
TTAK.KO-09.0065,0082/R
DMB
TTAK.KO-07.0026/R
사물지능통신(M2M)
TTAK.KO-06.0301,302,321,3
전화기 한글문자 자판
KCS.KO-05.0046
KS X 5020
CCTV
TTAK.KO-04.0152
와이파이
TTAE.IE-802.11r,11s/R1,11
TTAS.KO-06.0153
접근성
TTAK.OT-10.0003/R1
TTAK.KO-10.0634
피싱,파밍사고 대응
TTAK.KO-12.0212,213
무선충전
TTAE.KO-06.0303,304
휴대폰 촬영음
KCS.KO.06-0063
지문인식
TTAS.KO-12.0047
스팸 신고
KCS.KO-06.0230

Standards?


가족과 TV시청
TV제품표준
전송방식 등 표준화
퇴근 후 저녁식사
식품안전 표준 등
업무과정
표준화된 사규 및 작업
표준 등에 따라 작업
일어나서 시계보기
표준 시간에 맞추어진
측정 표준의 대표사례
세수, 양치질
칫솔, 치약의 제조, 시험
방법도 표준화
출근과정
사무실 도착
엘리베이터는 승강기 안전
기준에 의한 검사 대상
1
2
3
4
5
6
7
9
낮
12

Standards?


AC CABLE FIT YOUR LOCAL STANDARD
EU-2P
US-2P
UK-2P
IT-2P/3P
AG-2P/3P
EU-3P
US-3P
UK-3P
CN-3P/2P
AU-3P/2P
kfdpower.en.alibaba.com

Standards?

- **International standards**
  - International standards are technical standards developed by international organizations (intergovernmental organizations), such as Codex Alimentarius in food, the World Health Organization Guidelines in health, or ITU Recommendations in ICT and being publicly funded, are freely available for consideration and use worldwide.

- **Standardization**
  - The implementation of standards in industry and commerce became highly important with the onset of the Industrial Revolution and the need for high-precision machine tools and interchangeable parts.
  - International standards are one way of overcoming technical barriers in international commerce caused by differences among technical regulations and standards developed independently and separately by each nation, national standards organization, or company.

Standard category

공식 표준 (De-Jure)

- 국제표준: ITU Recommendation, ISO/IEC의 IS 등
- 지역표준: 유럽의 EN 등
- 국가표준: 미국의 ANS, 영국의 BS, 우리나라의 KS 등
- 단체표준: ATIS(미국), TTC(일본), TTAK(한국) 표준 등

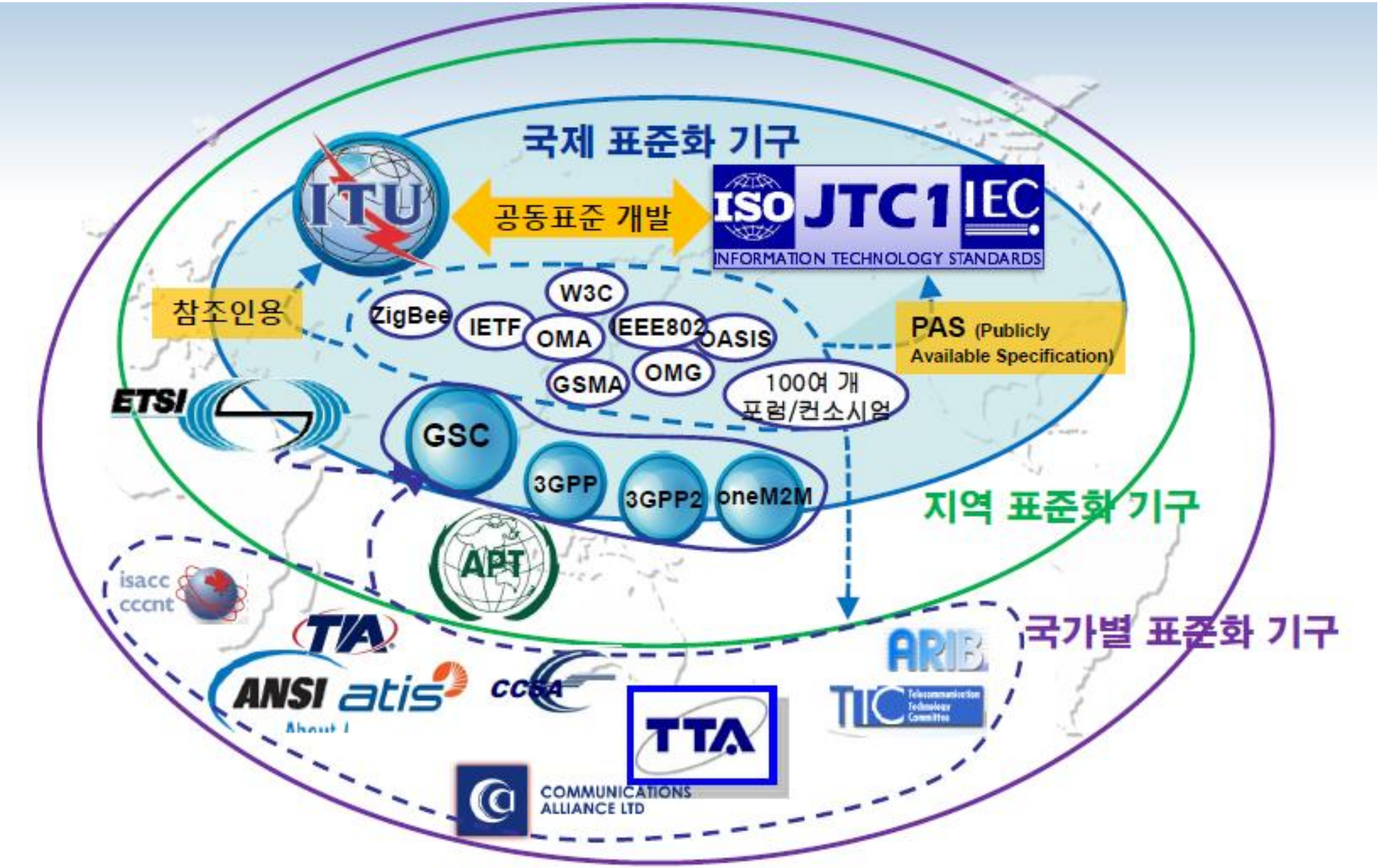
사실 표준 (De-facto)

- 포럼/컨소시엄 표준: IEEE, IETF, OMA, W3C 등 전세계 100여 개 사실 표준화기구에서 채택한 기술규격
- 시장표준: 시장 경쟁에서 승리한 기업의 규격
  예) 윈도우즈, 익스프로어, 애플의 iOS 등

Standard organization

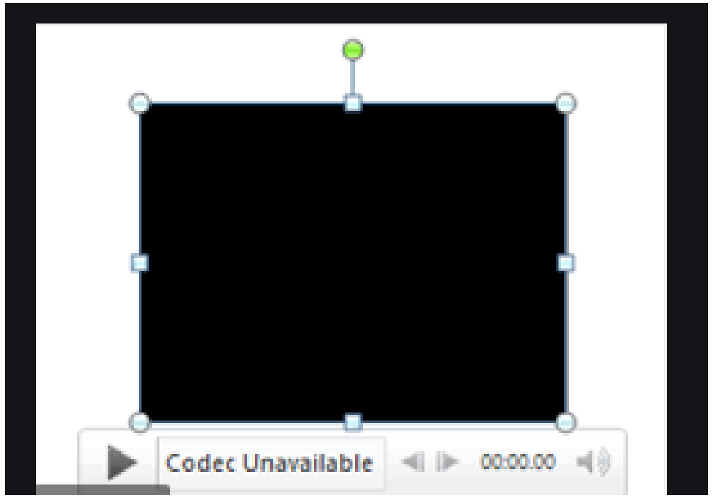
공식 표준화 기구	참여의 정도와 개발표준의 효력 범위에 따라 다음과 같이 구분		기술 분야 전반을 대상
	국제	국가간 조약기구인 ISO, 민간기구인 ITU 와 IEC 가 있음	상대적으로 충분한 협의를 거치는 표준 제정절차, 개방된 회원제도 등을 바탕으로 공신력을 인정받음
	지역	유럽의 ETSI, 아·태지역의 APT(ASTAP)	
	국가별	미국의 ATIS와 TIA, 일본의 TTC와 ARIB, 중국의 CCSA, 우리나라의 TTA 등	국가대표기관, 기업 등
			약 2년~4년 소요
사실 표준화 기구	**포럼/컨소시엄**	IEEE, IETF, OMA, W3C 등 전세계 100여 개가 넘음	특정 기술 분야에 이해관계가 있는 기업 및 개인이 시장의 필요 또는 연구 목적으로 연합하여 생성
			특정 기술 분야
			신속한 표준 제정 (약 12개월 이내)

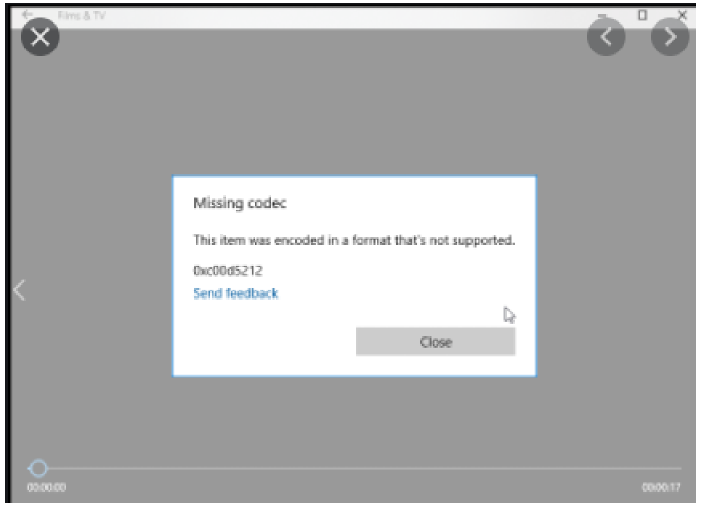
Standard organization

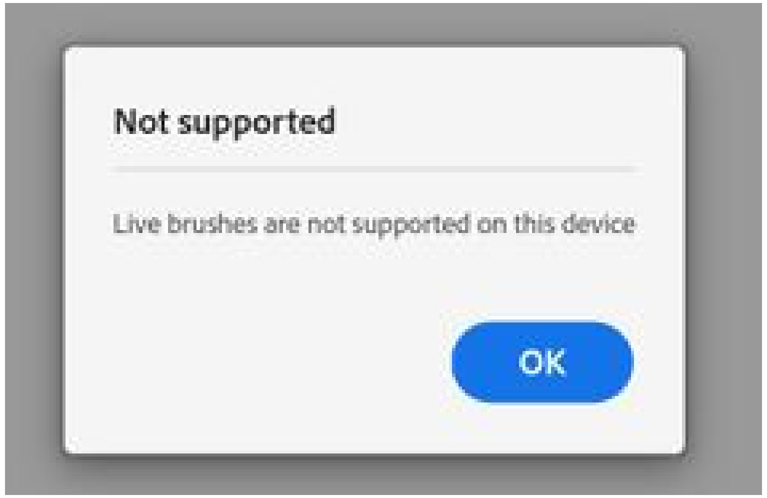


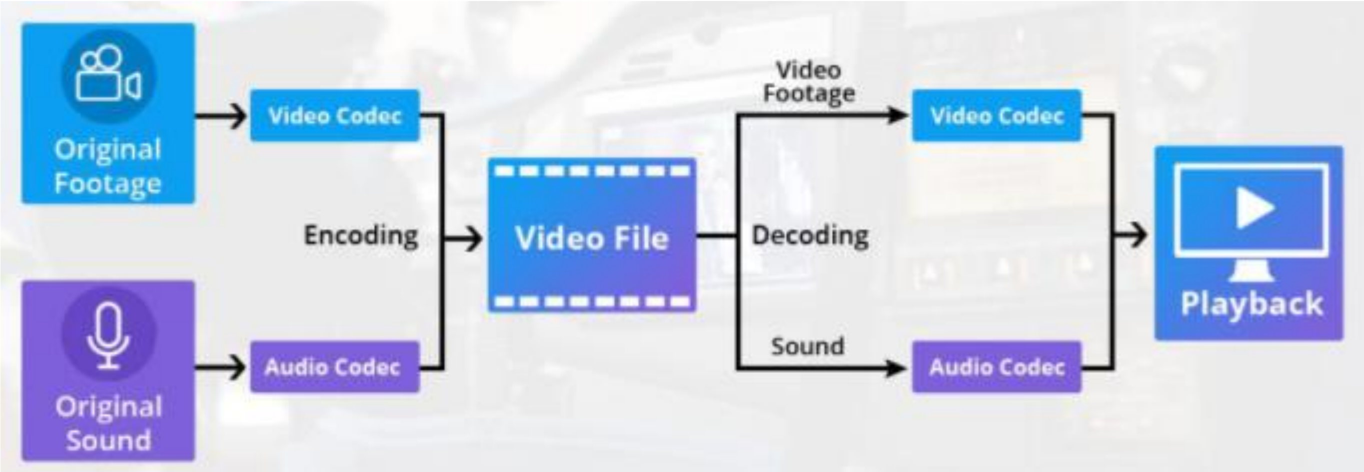
Video Coding Standards

Why we need video coding standard?



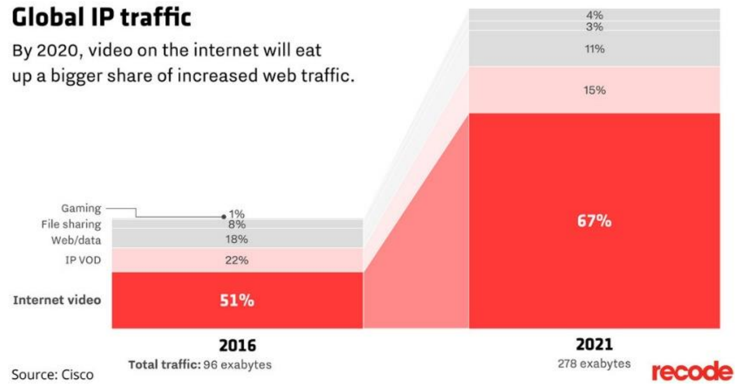




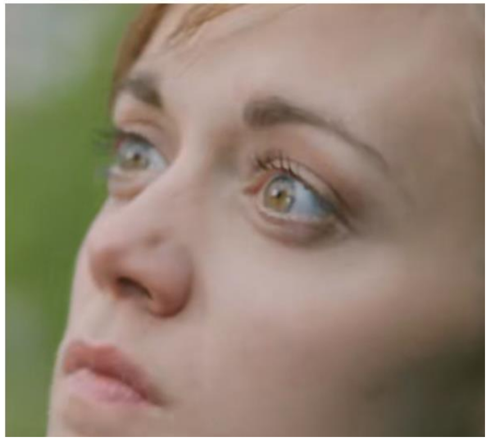


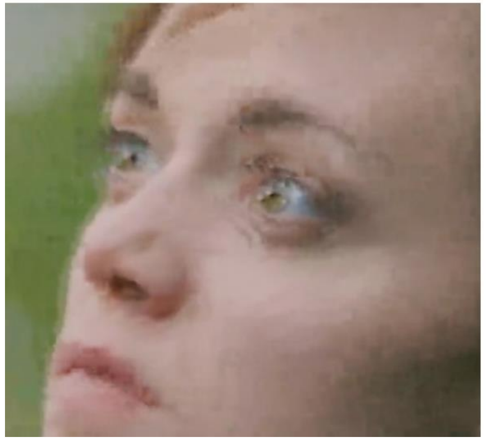
Why we need advanced video coding standard?

- **Background for a development of new video coding standards**
  - It's estimated that by 2022, 82 percent of the global internet traffic will come from video streaming and downloads (Cisco, 2019)



  - COVID-19 pandemic makes the following situation:
    - “Video use up 4x from six months ago” (RCRwireless, 29 June 2020, Limelight)
    - “YouTube is reducing the quality of videos for the next month, and it's because increased traffic amid the coronavirus outbreak is straining internet bandwidth” (Business Insider, 25 March 2020)
    - “Netflix to cut streaming quality in Europe for 30 days” (BBC, 19 March 2020)





Why we need advanced video coding standard?

- **Video codec example**
  - https://www.youtube.com/watch?v=AekAujM6ZcQ

Why we need advanced video coding standard?

- **Video codec example**

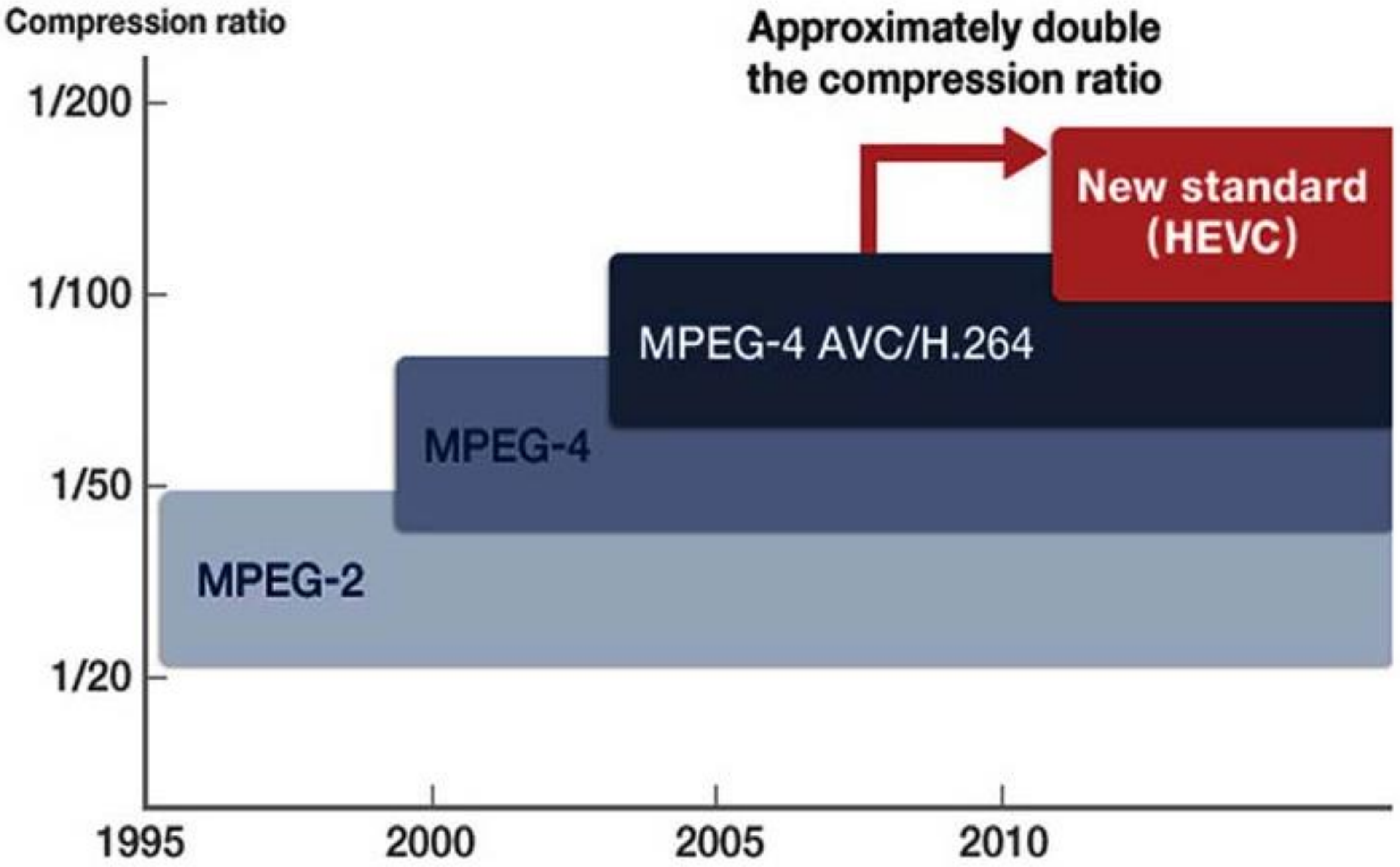
Why we need advanced video coding standard?

- **Video codec example**

Why we need advanced video coding standard?

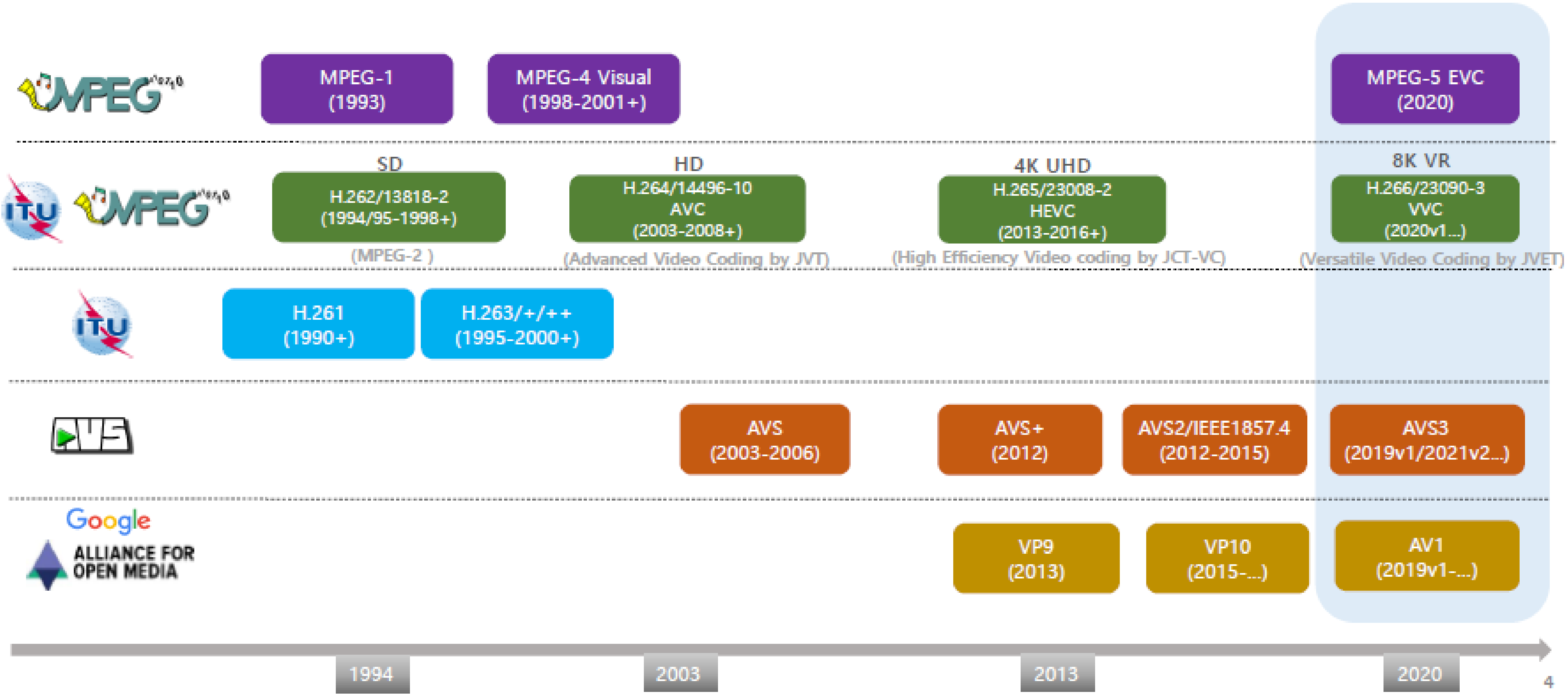
- **Video codec example**

Why we need advanced video coding standard?



History

- History of Video coding standard



Video coding standardization organizations

- Video coding standardization organizations
  - ISO/IEC MPEG = "Moving Picture Expert Group"
    - ISO/IEC JTC 1/SC29/WG 11 = International Standardization Organization and International Electrotechnical Commission, Joint Technical Committee 1, Subcommittee 29, Working Group 11
  - ITU-T VCEG = "Video Coding Expert Group"
    - ITU-T SG16/Q6 = International Telecommunications Union – Telecommunications Standardization Sector (ITU-T, a United Nations Organization, formerly CCITT), Study Group 16, Working Party 3, Question 6
  - JVT = "Joint Video Team"
    - Joint team of MPEG & VCEG, responsible for developing AVC
  - JCT-VC = "Joint Collaborative Team on Video Coding"
    - Joint team of MPEG & VCEG, responsible for developing HEVC (established Jan. 2010)
  - JVET = "Joint Video Expert Team"
    - Joint team of MPEG & VCEG, responsible for developing VVC (established Oct. 2015)
    - Previously called "Joint Video Exploration Team"
  - AVS = "Audio Video Standard"
    - Audio and Video coding standard workgroup of China (founded in June 2002)
  - AOMedia = "Alliance for Open Media"
    - Non-profit industry consortium (founded in Sept. 2015)

History of Video Coding Standards

- ITU-T H.120 [1] (not used much)
  - “Codecs for video-conferencing using primary digital group transmission”
  - First video conferencing specification
  - DPCM structure
  - Conditional replenishment

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

History of Video Coding Standards

- ITU-T H.261 [2]
  - "Video codec for audiovisual services at $p \times 64$ kbit/s"
  - Video conferencing (broad application)
  - First-time hybrid coding scheme
    Motion compensated prediction, transform
  - CIF resolution (Common Intermediate Format, 352×288 in Europe)

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

History of Video Coding Standards

- ISO 11172-2 MPEG-1 [3]
  - “Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 2: Video"
  - Video CD – first distribution of digital video format
  - Hybrid coding scheme with bi-prediction, “D-frames” for search
  - Audio layer-3: “mp3” format

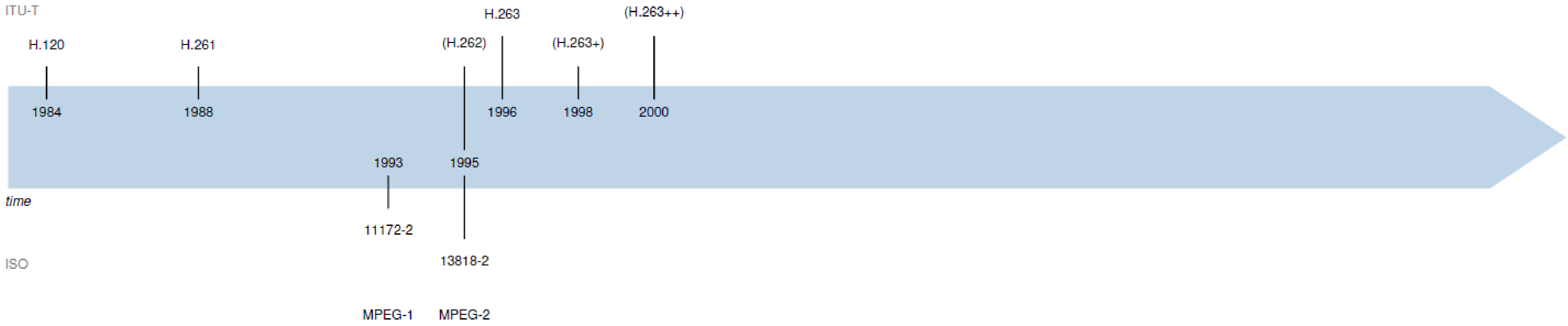
Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

History of Video Coding Standards

- ISO 13818-2 MPEG-2 [4]
  - "Information technology – Generic coding of moving pictures and associated audio information – Part 2: Video"
  - Digital TV, DVD, DVB-x – Widespread usage of digital video format
  - Support for interlaced video format (full SD resolution)
  - First standard supporting scalability features

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

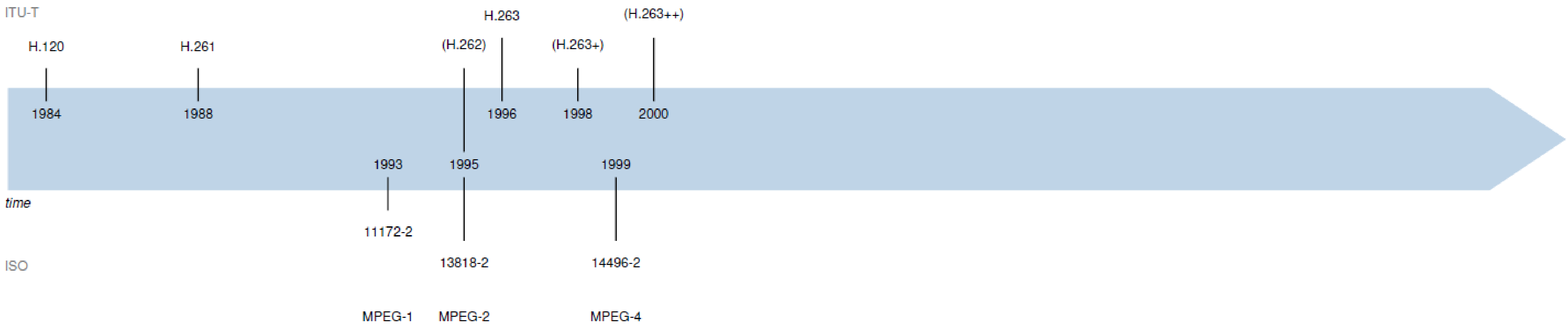
History of Video Coding Standards



- ITU-T H.263 [5]
  - "Video coding for low bit rate communication"
  - Video communication applications
  - Improved compression performance
  - Large set of extensions (18 Annexes)
    Organization into profiles

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

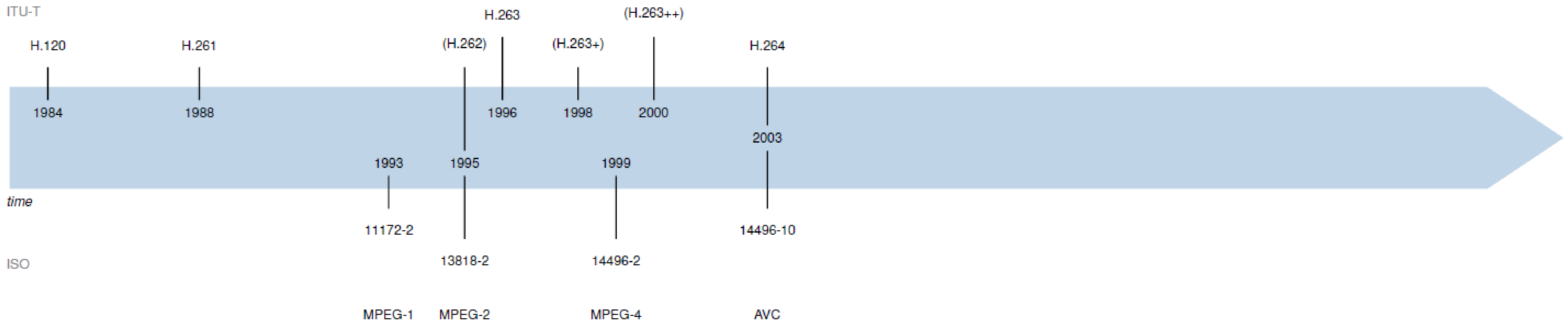
History of Video Coding Standards



- ISO 14496-2 MPEG-4 [6]
  - "Information technology – Coding of audio-visual objects – Part 2: Visual"
  - Multimedia object representation
  - Basically H.263 coding tools (plus quarter sample, global motion)
  - Arbitrarily shaped objects
    mostly used profile: Advanced simple profile

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

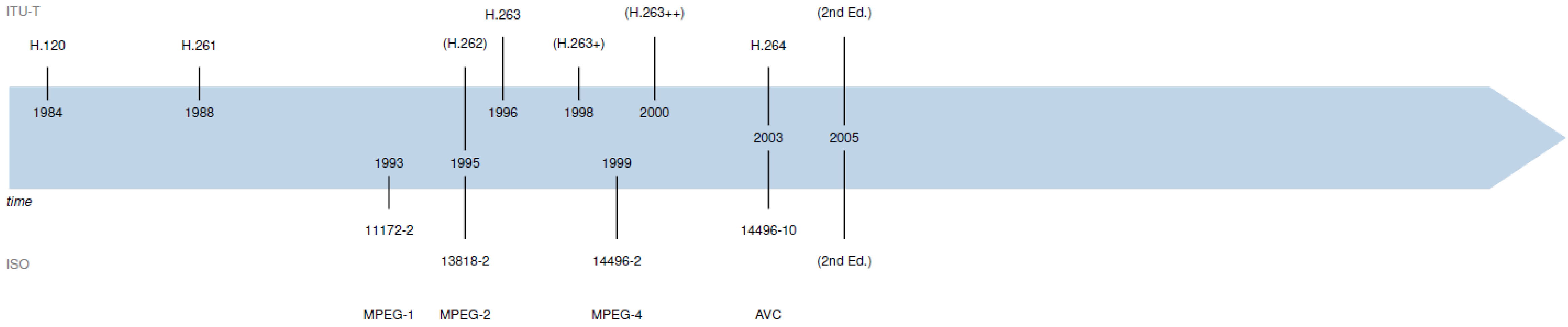
History of Video Coding Standards



- AVC (ITU-T H.264, ISO 14496-10) [7, 8]
  - Advanced Video Coding
  - HDTV distribution, YouTube, cameras (AVCHD), conferencing, . . .
  - Bit-exact specification, 4×4 integer transform in-loop filter, low-complex arithmetic coding
  - High compression performance

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

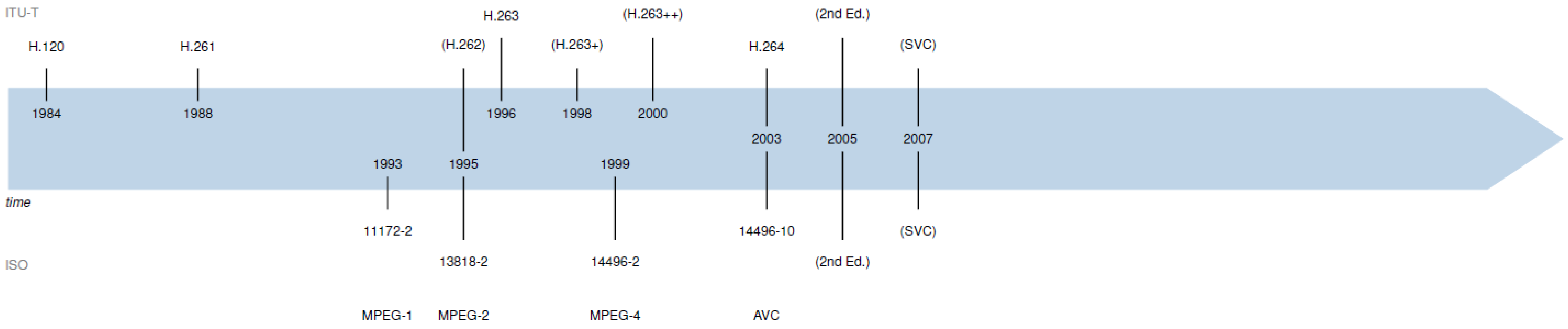
History of Video Coding Standards



- AVC (ITU-T H.264, ISO 14496-10) [7, 8]
  - Advanced Video Coding (High Profiles)
  - HDTV distribution, YouTube, cameras (AVCHD), conferencing, . . .
  - Improvements due to competitors (e. g. VC-1)
  - Additional 8×8 transform
  - Additional color spaces

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

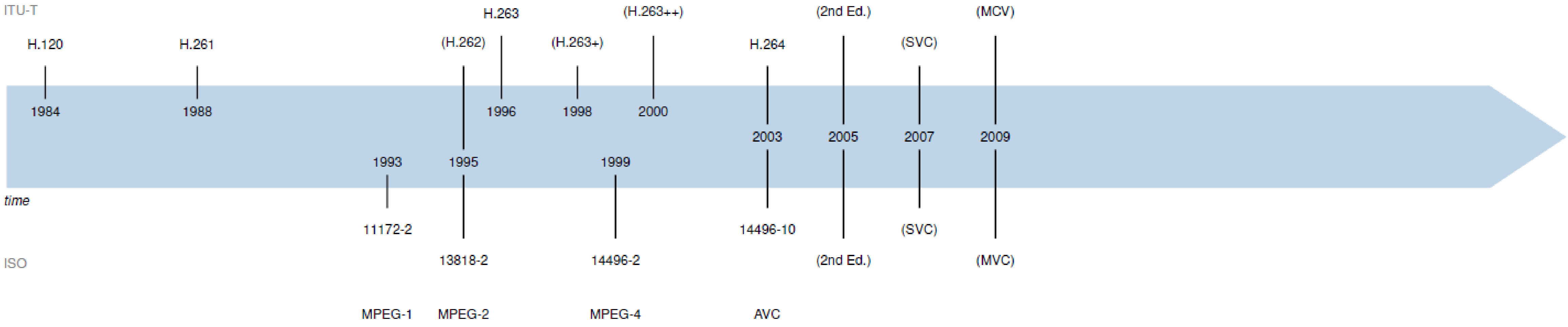
History of Video Coding Standards



- AVC (ITU-T H.264, ISO 14496-10) [7, 8]
  - Scalable Video Coding (extension)
  - General purpose, used in video conferencing applications (Vidyo)
  - Temporal, quality, spatial scalability
  - Single-loop decodability

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

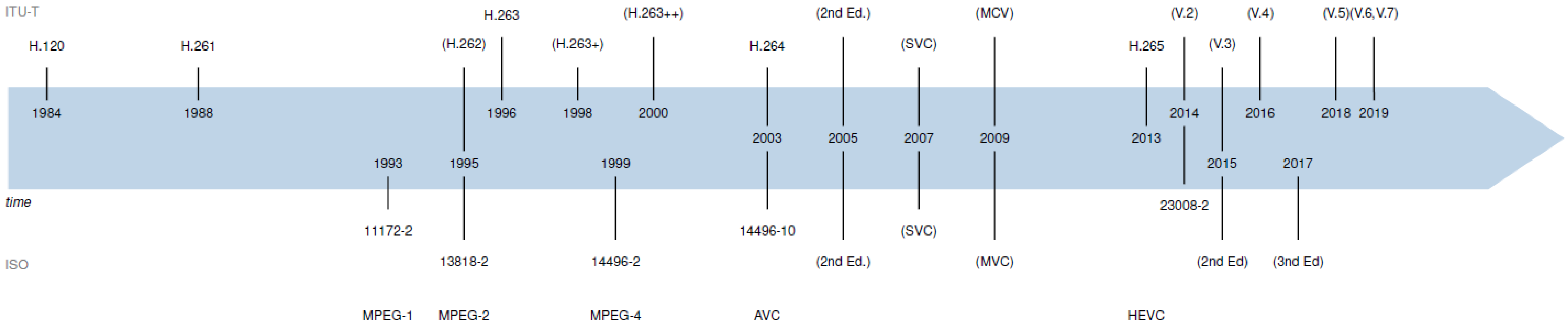
History of Video Coding Standards



- AVC (ITU-T H.264, ISO 14496-10) [7, 8]
  - Multiview video coding (extension)
  - Stereo and 3D video applications (Blu-Ray)
  - Two or more parallel views of the scene
  - No modifications on tool level

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

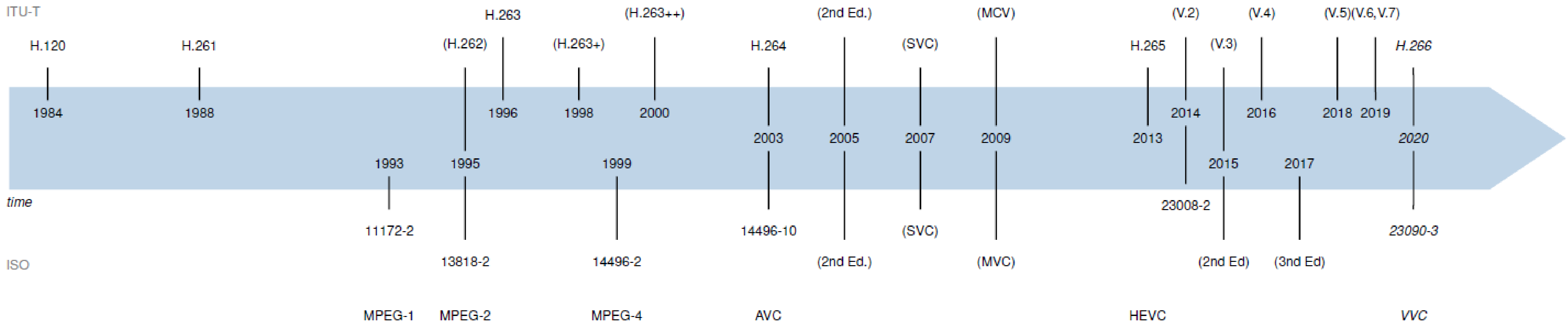
History of Video Coding Standards



- HEVC (ITU-T H.265, ISO 23008-2) [9, 10]
  - High efficiency video coding, general purpose, HD to UHD resolutions
  - Improved compression performance for application space
  - Edition 2 (10/2014): Range extensions, multiview, scalability; Edition 3 (04/2015): 3D video coding; Edition 4 (12/2016): Screen Content Coding; Edition 5 (02/2018): Omnidirectional Video SEI; Edition 6 (06/2019): SEI manifest; Edition 7 (11/2019): Fisheye and Annotated Regions SEIs

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

History of Video Coding Standards



- VVC (ITU-T H.266, ISO 23090-3) [11]
  - Versatile video coding, general purpose, UHD and larger resolutions
  - Extended application space ($360^\circ, \ldots$)
  - Improved compression performance for application space
  - Based on the hybrid coding scheme

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

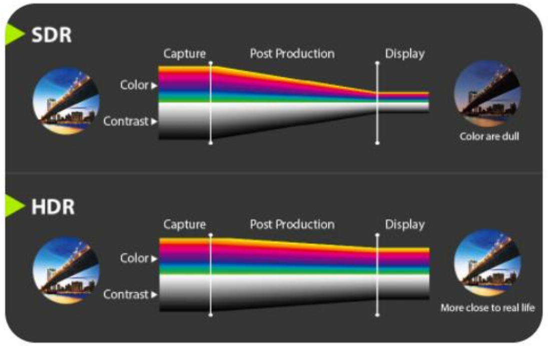
Why we need advanced video coding standard?

- **Background for a development of new video coding standards**
  - Video is continually increasing by resolution
    - HD existing, UHD (4Kx2K, 8Kx4K) appearing
    - Mobile services going towards HD/UHD
    - Stereo, multi-view, 360 video
  - Devices available to record and display ultra-high resolutions
    - Becoming affordable for home and mobile consumers
  - Video has multiple dimensions to grow the data rate
    - Frame resolution, Temporal resolution / color resolution, bit depth / Multi-view
    - Visible distortion still an issue with existing networks
  - Necessary video data rate grows faster than feasible network transport capacities
    - Better video compression (than current HEVC) needed in next decade, even after availability of 5G

Wien, Mathias, and Benjamin Bross. "Versatile Video Coding–Algorithms and Specification." *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020.

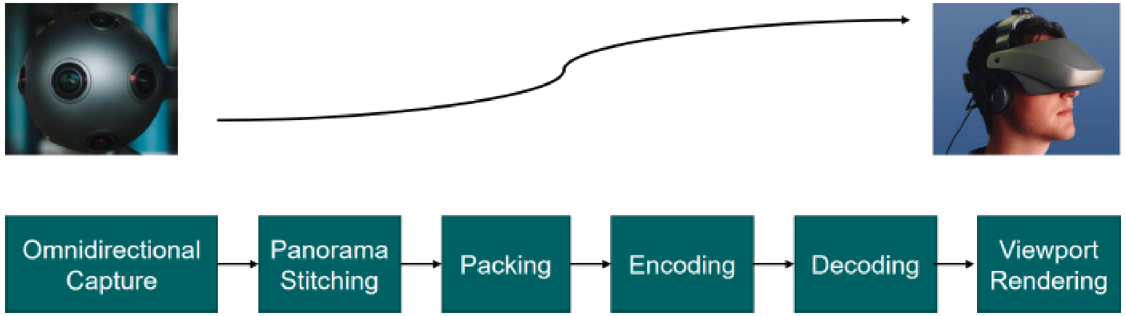
Why we need advanced video coding standard?

- **New applications**
  - Wide Color Gamut / High Dynamic Range coding



https://www.aten.com/eu/en/product-landing-page/true-4k-hdr/products/

  - VR / 360 Video Acquisition

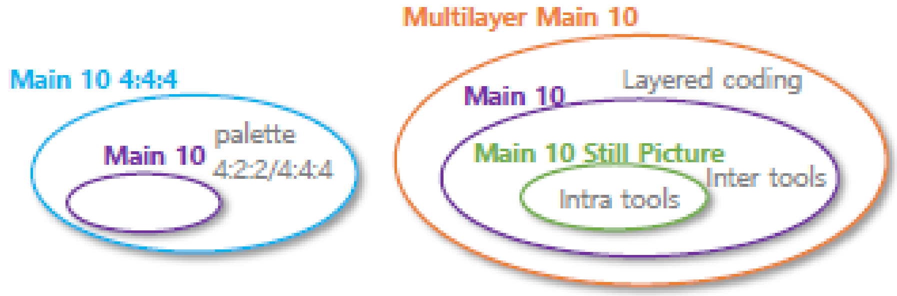


  - Screen contents / Game streaming

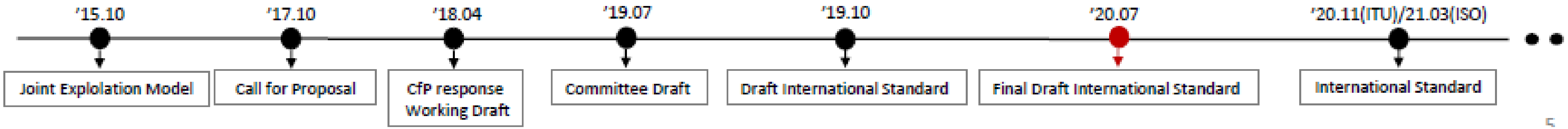
Photo: Valve

Overview - VVC

- **JVET: Versatile Video Coding (VVC)**
  - Goal of VVC standard
    - To provide significantly improved compression capability over existing video coding standards
    - To provide all functionality of video coding including UHD, HDR, 360°, screen contents, scalability etc.
    - VVC & VSEI (Versatile Supplemental Enhancement Information Messages for Coded Video Bitstream)
  - Profiles of Version 1
    - Main 10, Main 10 4:4:4
      - » support sub-picture, high tier will support higher frame rates up to 960
    - Multilayer Main 10, Multilayer Main 10 4:4:4
      - » support layered coding
    - Main 10 Still Picture, Main 10 4:4:4 Still Picture
      - » support a picture coding



  - Timeline



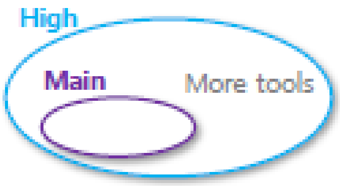
Overview - EVC

- **MPEG: Essential Video Coding (EVC)**
  - Goal of EVC standard
    - To provide a video compression solution which combines to Meet Business and Technical Requirements
    - Timely availability of published licensing terms to allow reliable business plans to be created
    - Coding efficiency at least as good as HEVC
    - Complexity suitable for real time encoding
    - The ability to address existing and emerging use cases, including offline encoding for streaming VOD, live OTT streaming
  - Profiles
    - Baseline Profile, Baseline Still picture profile
      - » Considered as royalty-free video codec, with tools made public more than 20 years
    - Main profile, Main Still picture profile
      - » High compression performance compared to HEVC

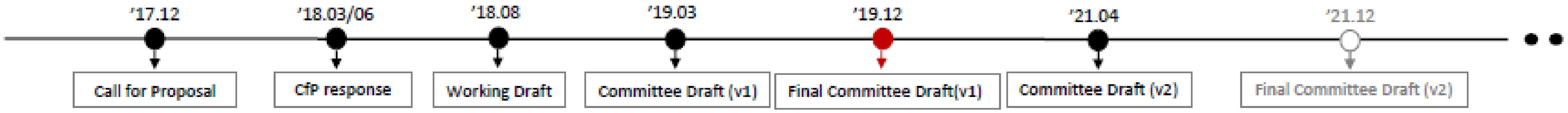
  - Timeline

Overview - AVS

- **AVS: Audio Video Standard 3 (AVS3)**
  - Goal of AVS workgroup
    - To develop video coding standards with low royalty rates (1RMB per hardware device for AVS1/AVS+/AVS2)
  - Goal of AVS3 standard
    - To provide significantly improved compression capability over existing video coding standards
    - Target applications Phase1: 8K, UHD TV, Broadcast, OTT
    - Target applications Phase2: Live video, VOD of VR, intelligent security
  - Profiles
    - Main Profile, Main 10bit Profile (Phase1)
      - » Outperform HEVC 27%
    - High 8bit Profile, High 10bit Profile (Phase2)
      - » Outperform HEVC 35%



  - Timeline



Video Coding Framework

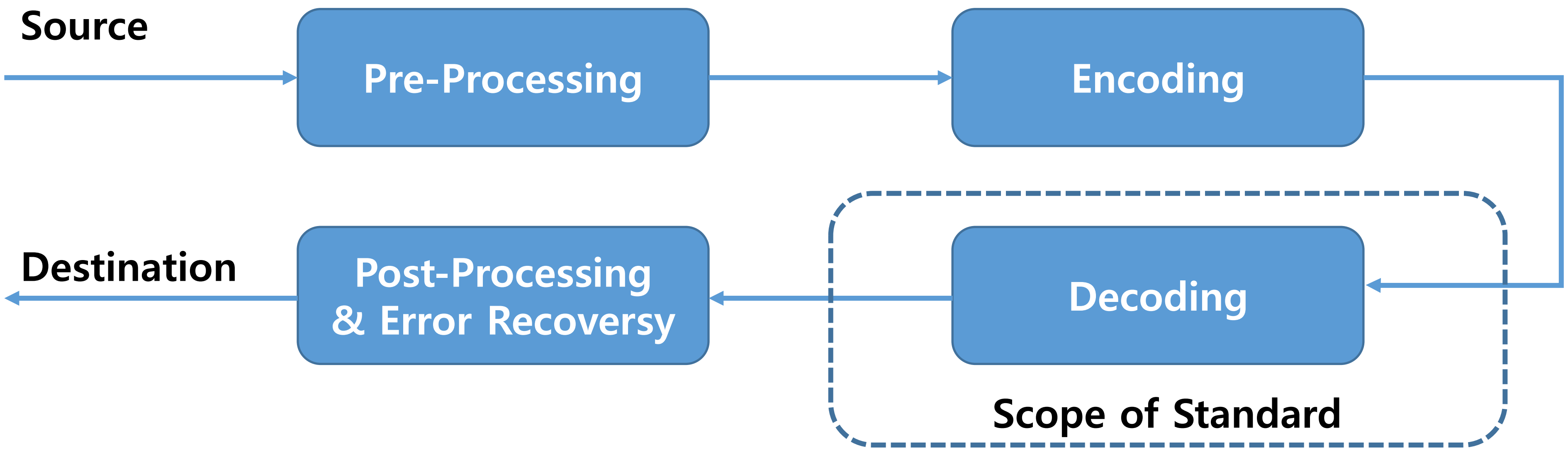
Video codec framework

- **Video codec framework**
  - Hybrid coding based on Prediction coding / Transform coding / Entropy coding
  - Prediction coding
    - Intra prediction
    - Inter prediction
  - Transform coding
    - DCT / DST based transform
    - Uniform quantization
  - Entropy coding
    - Binary arithmetic coding using context modeling
  - In-loop filtering
    - Deblocking filter / SAO / Etc.

Sze, Vivienne, Madhukar Budagavi, and Gary J. Sullivan. "High efficiency video coding (HEVC)." *Integrated circuit and systems, algorithms and architectures*. Vol. 39. Springer, 2014. 40.

Video codec framework

- **Scope of video coding standardization**



Video codec framework

- **Video codec framework**

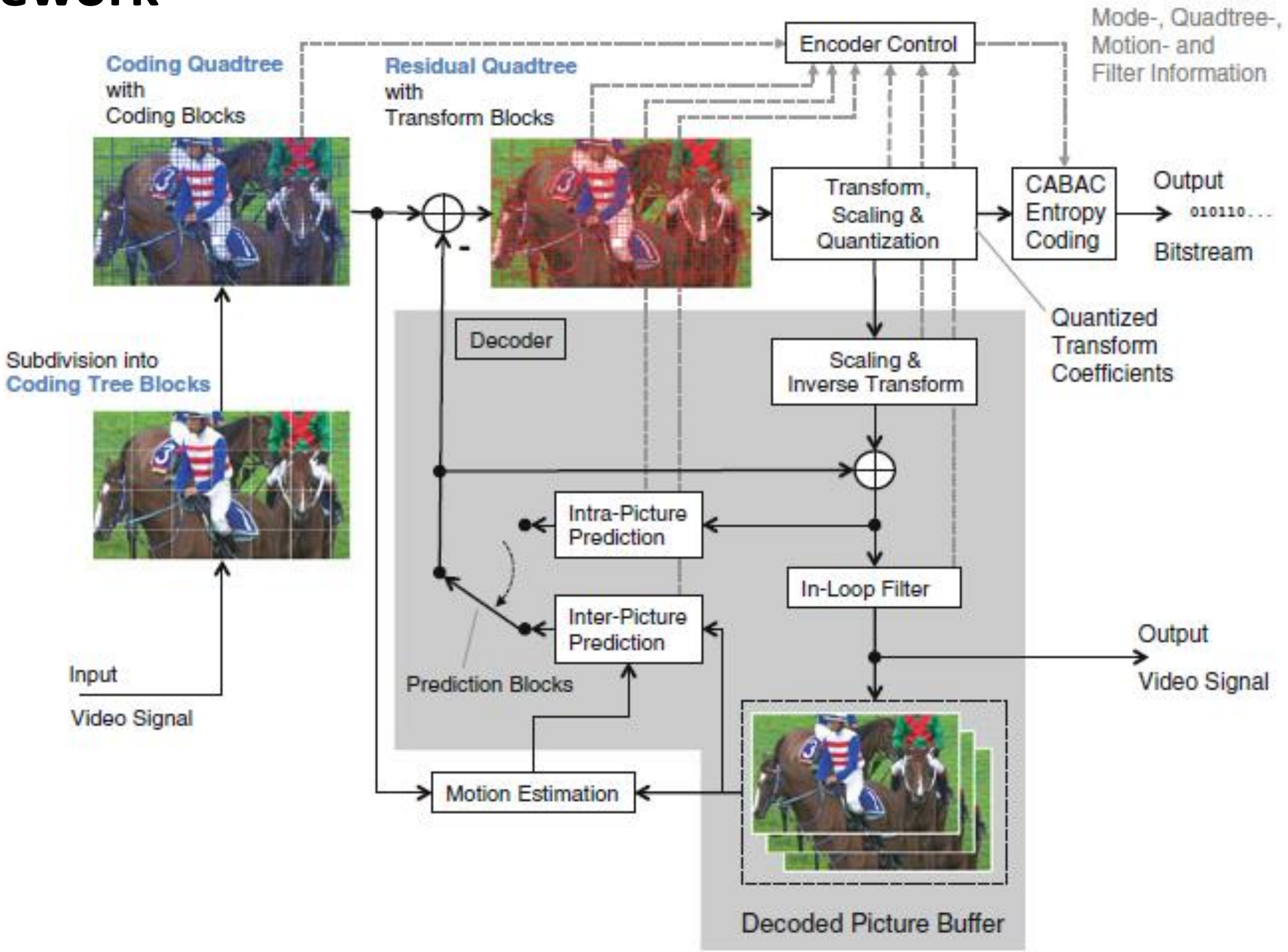


Fig. 3.1 Block diagram of an HEVC encoder with built-in decoder (*gray shaded*)

Sze, Vivienne, Madhukar Budagavi, and Gary J. Sullivan. "High efficiency video coding (HEVC)." *Integrated circuit and systems, algorithms and architectures*. Vol. 39. Springer, 2014. 40.

Video codec framework

- **Intra prediction**
  - The intra prediction framework of HEVC consists of three steps: reference sample array construction, sample prediction, and post-processing.

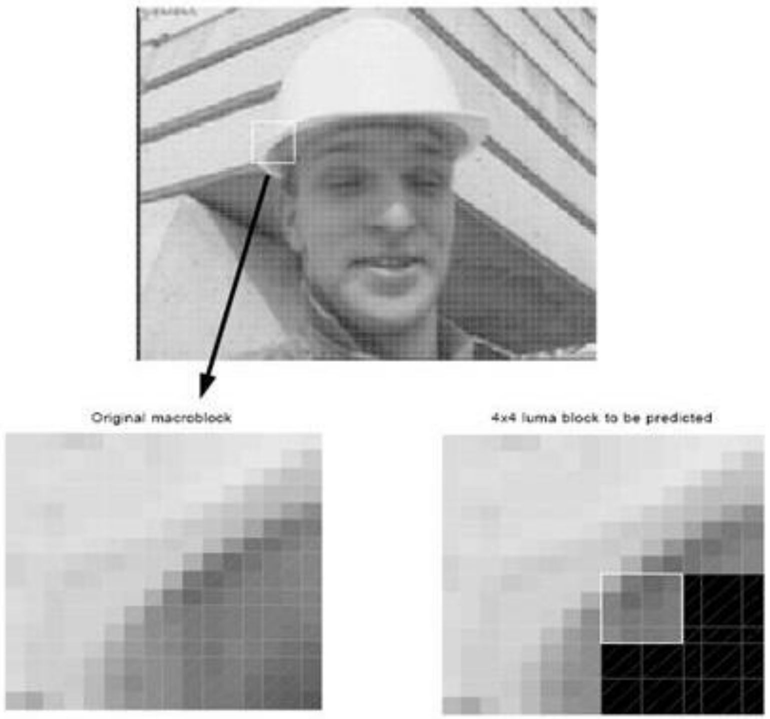


Table 4.1 Relationship between intra prediction mode number and associated names

Intra prediction mode number	Associated names
0	INTRA_PLANAR
1	INTRA_DC
2 . . . 34	INTRA_ANGULAR[i], i = 2 . . . 34

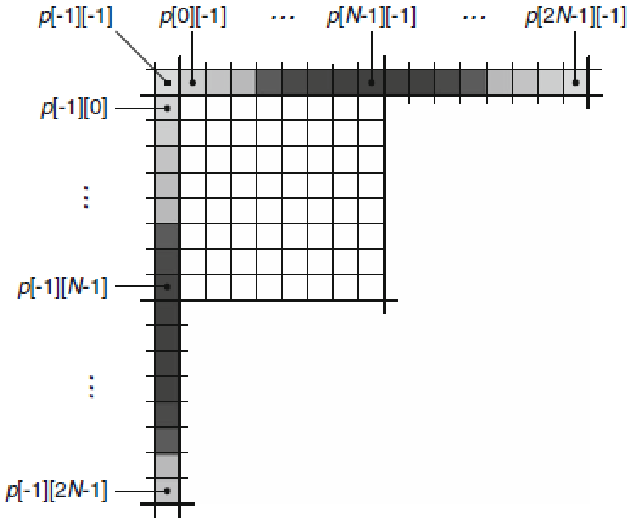


Fig. 4.1 An example of reference samples $p[x][-1]$, $p[-1][y]$ HEVC intra prediction uses for a block of size $N \times N$ samples

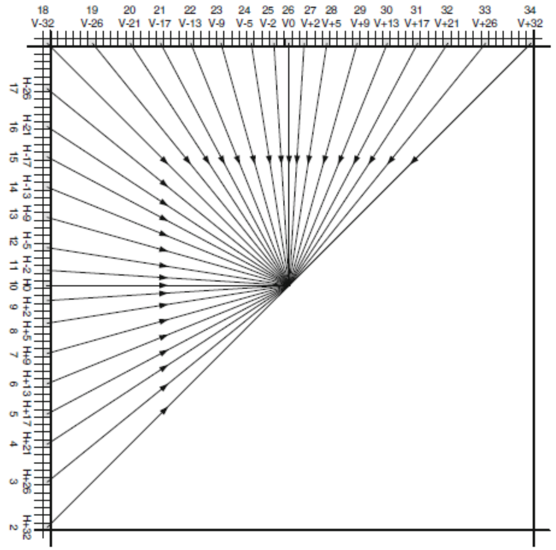
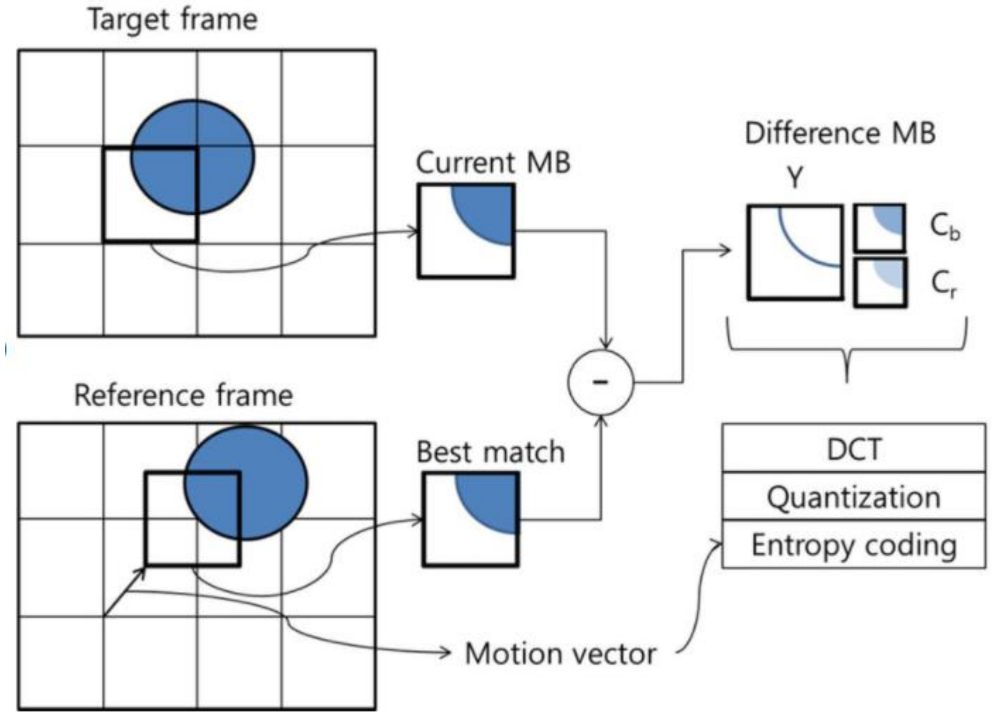


Fig. 4.5 Angle definitions of angular intra prediction in HEVC numbered from 2 to 34 and the associated displacement parameters. *H* and *V* are used to indicate the horizontal and vertical directionalities, respectively, while the numeric part of the identifier refers to the sample position displacements in 1/32 fractions of sample grid positions. Reproduced with permission from [8], © 2012 IEEE

Sze, Vivienne, Madhukar Budagavi, and Gary J. Sullivan. "High efficiency video coding (HEVC)." *Integrated circuit and systems, algorithms and architectures*. Vol. 39. Springer, 2014. 40.

Video codec framework

- **Inter prediction**
  - The inter-picture prediction makes use of the temporal correlation between pictures in order to derive a motion-compensated prediction (MCP) for a block of image samples.



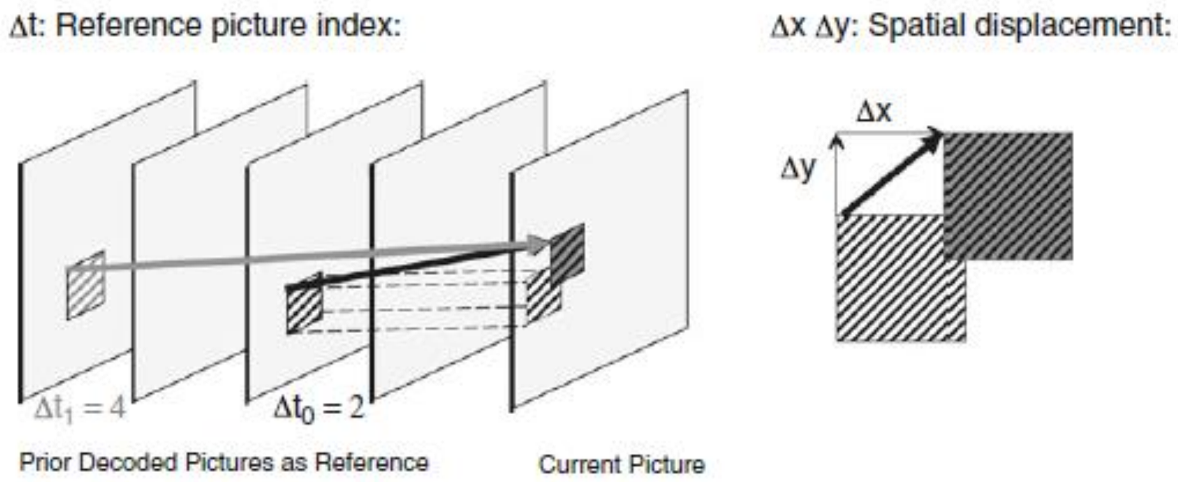
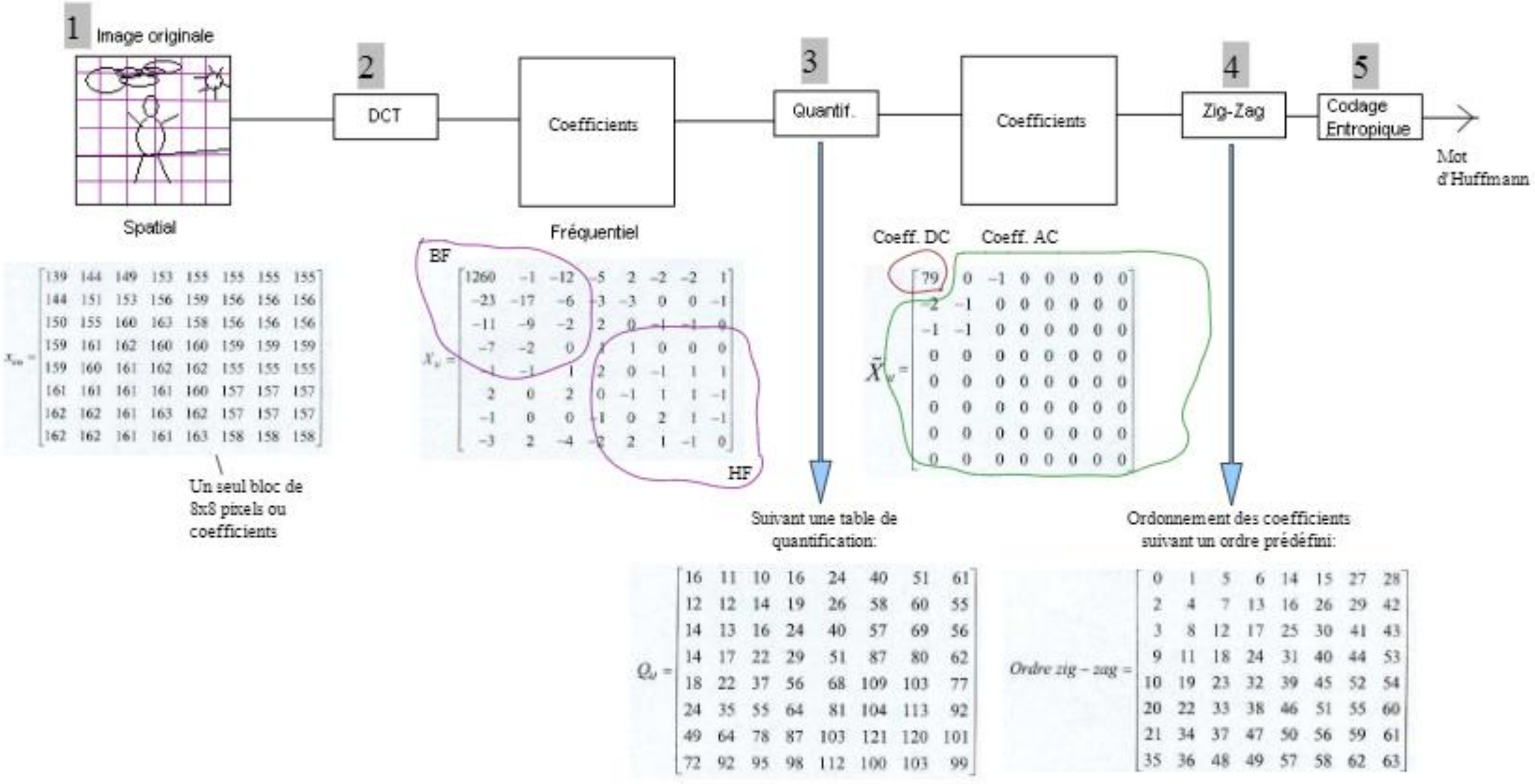


Fig. 5.1 Inter-picture prediction concept and parameters using a translational motion model

Sze, Vivienne, Madhukar Budagavi, and Gary J. Sullivan. "High efficiency video coding (HEVC)." *Integrated circuit and systems, algorithms and architectures*. Vol. 39. Springer, 2014. 40.

Video codec framework

- **Transform coding**
  - The transforms are applied to the residual signal resulting from inter- or intra-picture prediction.
  - The HEVC standard specifies core transform matrices of size 4x4, 8x8, 16x16 and 32x32 to be used for two-dimensional transforms in the context of block-based motion-compensated video compression.

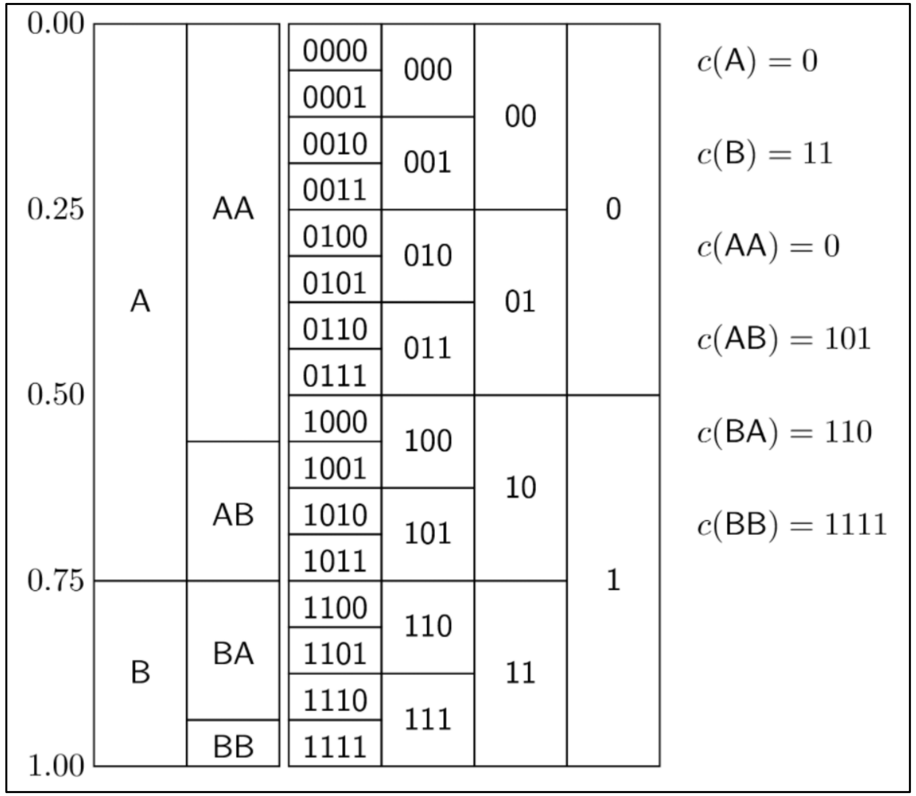


Video codec framework

- **Entropy coding**
  - Context-Based Adaptive Binary Arithmetic Coding (CABAC) is a method of entropy coding first introduced in H.264/AVC and now used in the latest High Efficiency Video Coding (HEVC) standard.

Symbol	Prob.	Codeword
a	0.05	0000
b	0.05	0001
c	0.1	001
d	0.2	01
e	0.3	10
f	0.2	110
g	0.1	111

Hoffman coding



Arithmetic coding

Video codec framework

- **In-loop filtering**
  - The in-loop filters are applied in the encoding and decoding loops, after the inverse quantization and before saving the picture in the decoded picture buffer.
  - The HEVC standard specifies two in-loop filters, a deblocking filter and a sample adaptive offset (SAO).

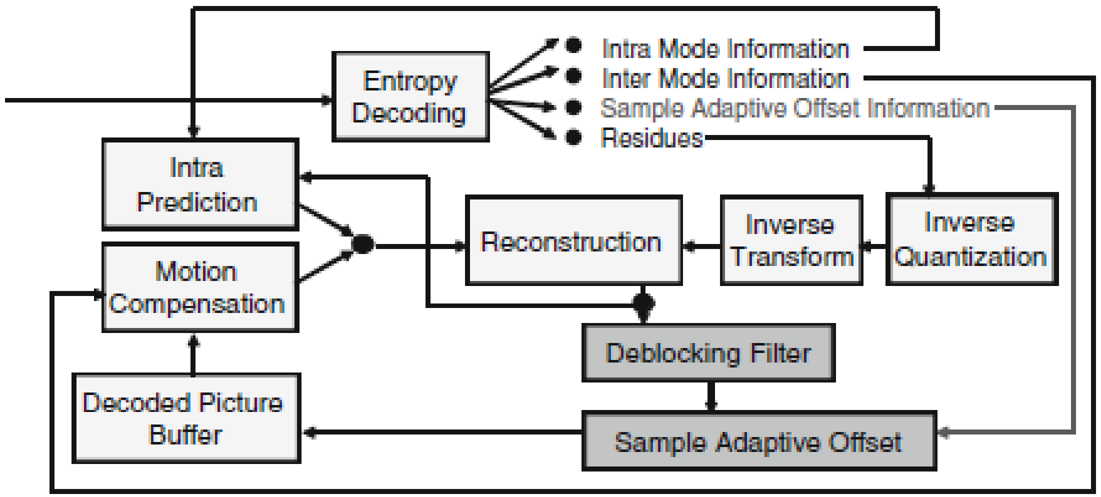


Fig. 7.1 Block diagram of HEVC decoder. Reproduced with permission from [13], © 2012 IEEE

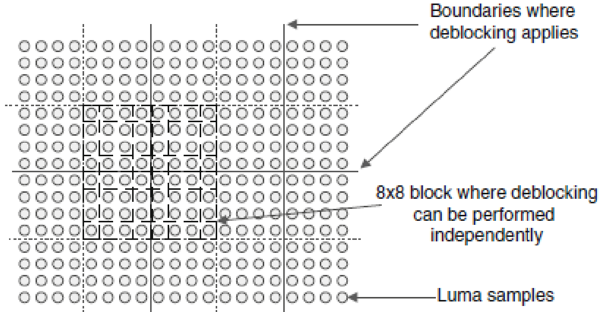


Fig. 7.20 Illustration of picture samples, horizontal and vertical block boundaries on the 8 × 8 grid, and those non-overlapping blocks of 8 × 8 samples (marked with *dotted lines*), which can be deblocked in parallel. The *dashed lines* mark samples used in deblocking decisions (vertical and horizontal)

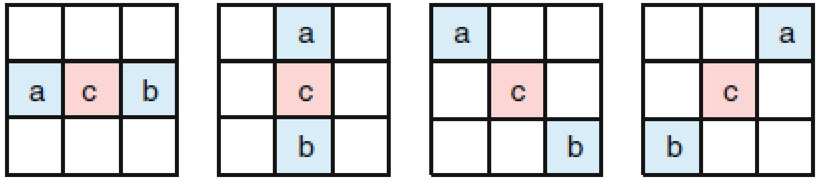


Fig. 7.14 Four 1-D directional patterns for EO sample classification: horizontal (EO class = 0), vertical (EO class = 1), 135° diagonal (EO class = 2), and 45° diagonal (EO class = 3). Reproduced with permission from [13], © 2012 IEEE

Sze, Vivienne, Madhukar Budagavi, and Gary J. Sullivan. "High efficiency video coding (HEVC)." *Integrated circuit and systems, algorithms and architectures*. Vol. 39. Springer, 2014. 40.

AI based Video Coding Development

AI based Video Coding Development

- **Examples of typical video applications**

Fig. 1. Examples of typical video applications.

Zhang, Yun, Sam Kwong, and Shiqi Wang. "Machine learning based video coding optimizations: A survey." *Information Sciences* 506 (2020): 395-423.

AI based Video Coding Development

- **Examples of typical video applications**

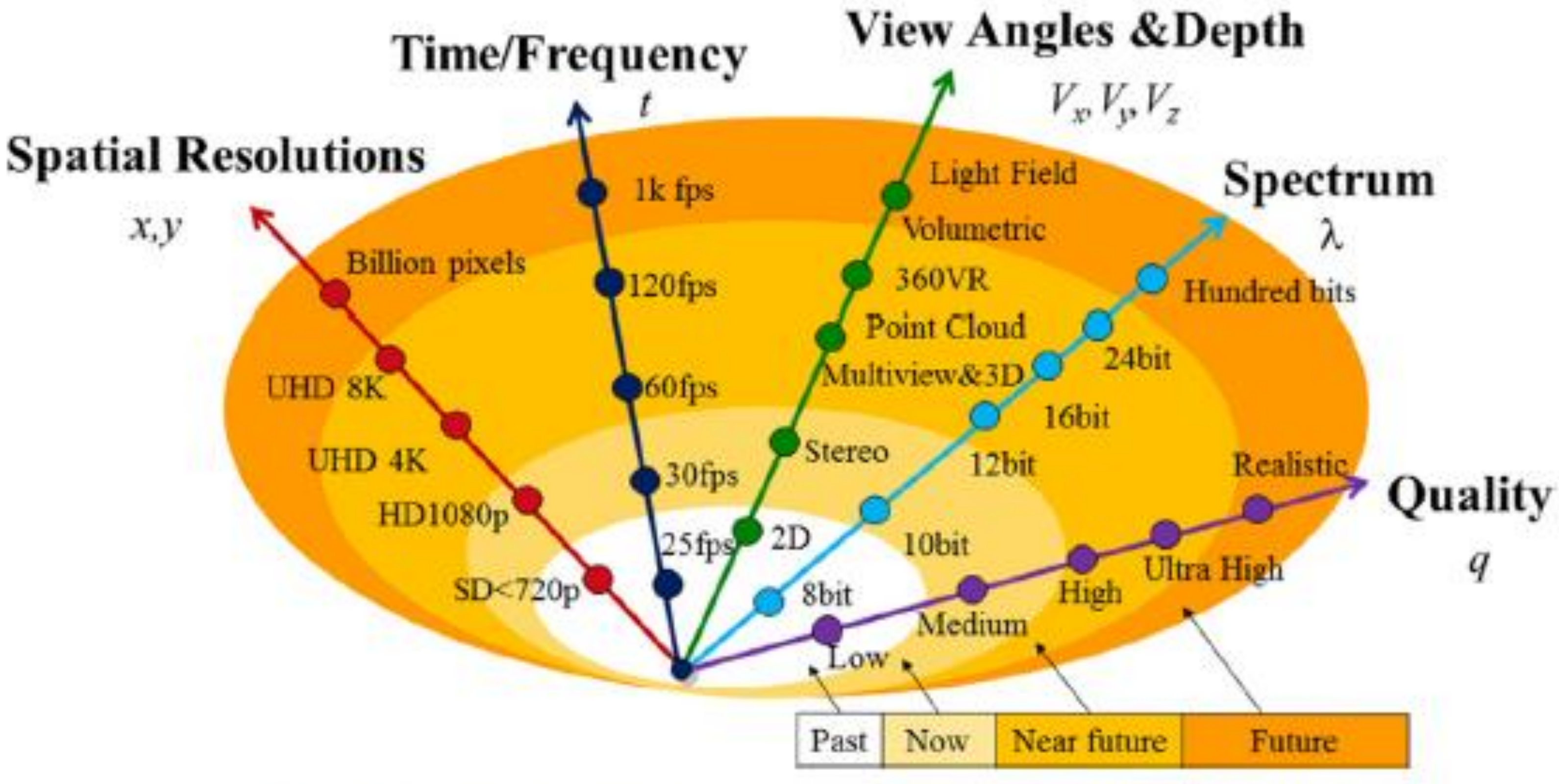


Fig. 2. Trends of video data representation from user end [16].

Zhang, Yun, Sam Kwong, and Shiqi Wang. "Machine learning based video coding optimizations: A survey." *Information Sciences* 506 (2020): 395-423.

AI based Video Coding Development

- **Examples of typical video applications**

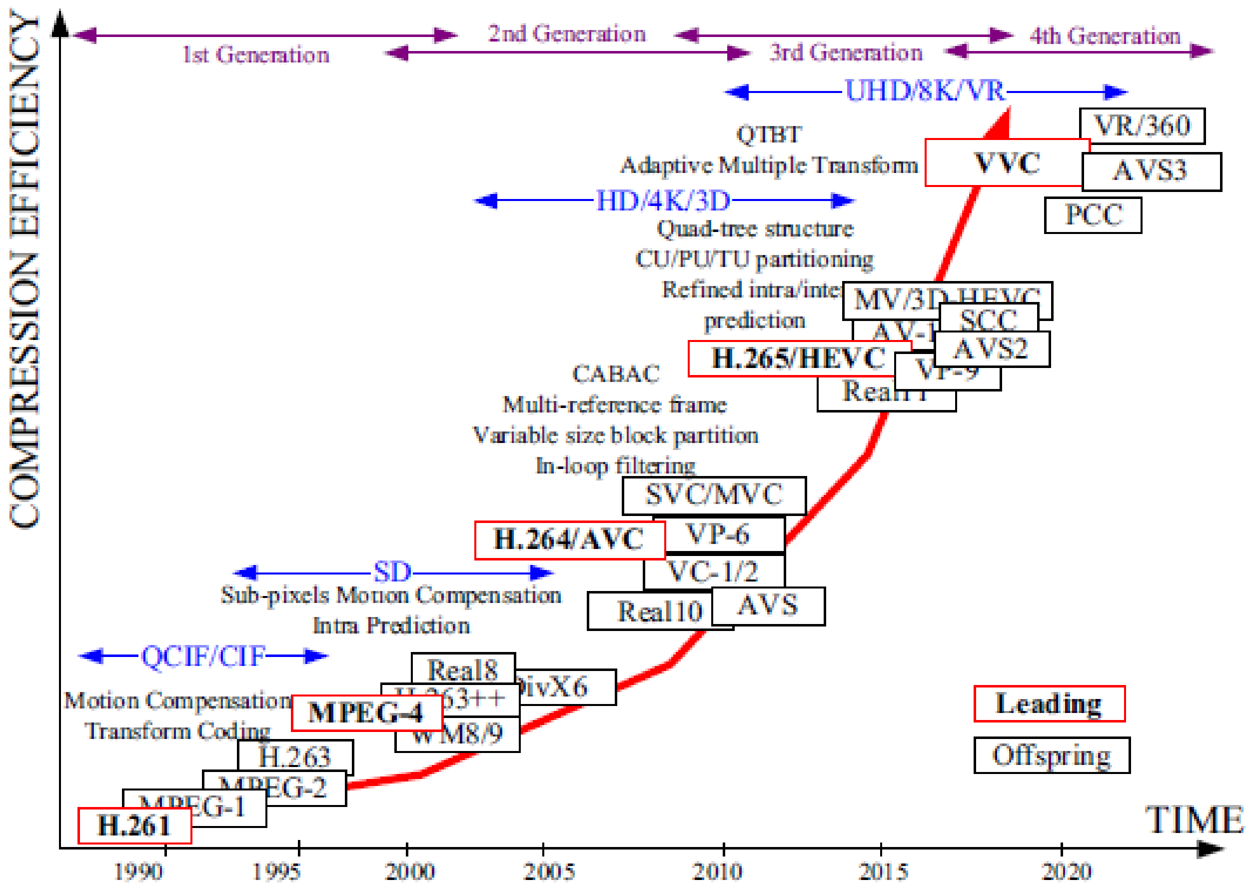


Fig. 5. Evolution of video coding standards.

Zhang, Yun, Sam Kwong, and Shiqi Wang. "Machine learning based video coding optimizations: A survey." *Information Sciences* 506 (2020): 395-423.

AI based Video Coding Development

- **History of AI based video codec development**

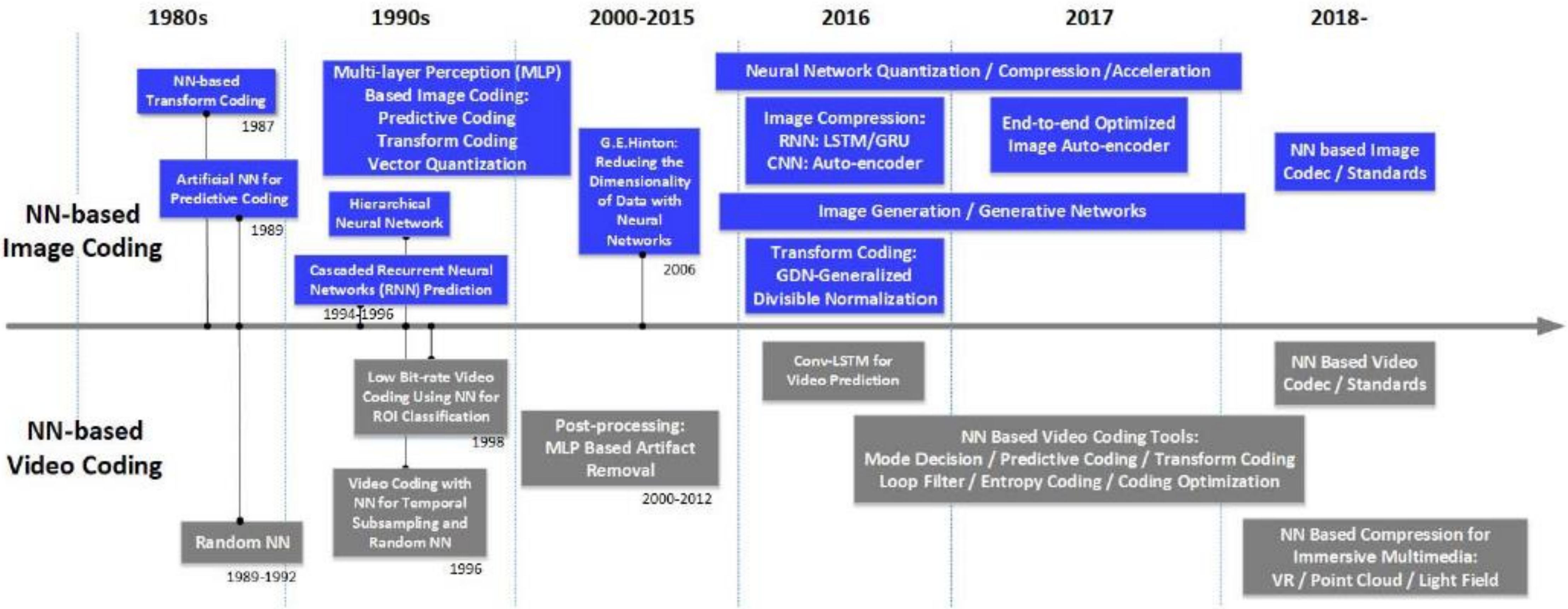


Fig. 4. The technical roadmap of neural network based compression algorithms.

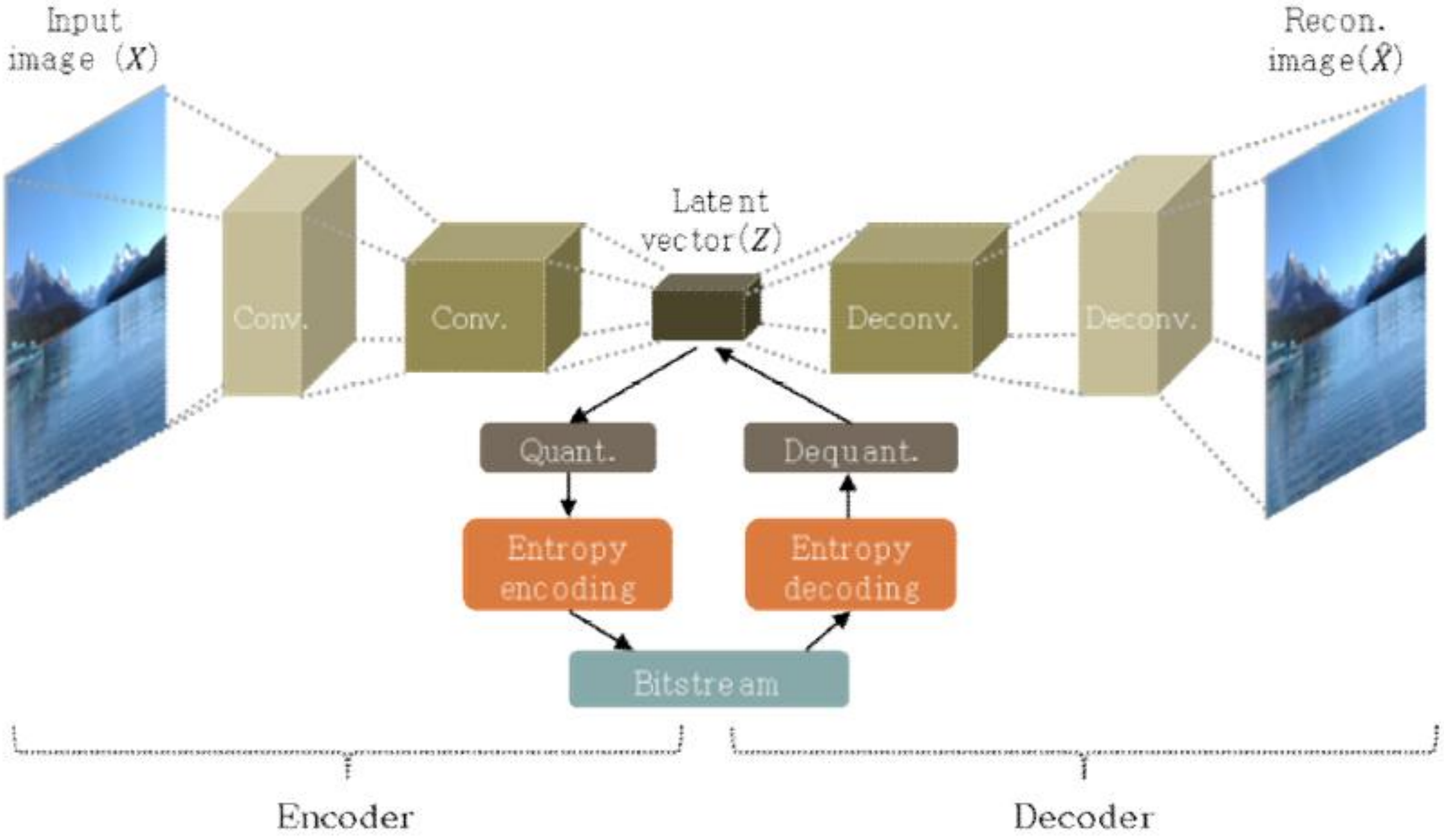
Ma, Siwei, et al. "Image and video compression with neural networks: A review." *IEEE Transactions on Circuits and Systems for Video Technology* 30.6 (2019): 1683-1698.

AI based Video Coding Development

- **Trend on AI based video codec development**
  - Auto encoder based video coding development
  - End to End based video coding development
  - Module (Tool) based video coding development
  - Down/up Sampling (SR) based video coding development

AI based Video Coding Development

- **Auto encoder based video coding development**



AI based Video Coding Development

- **End to End based video coding development**

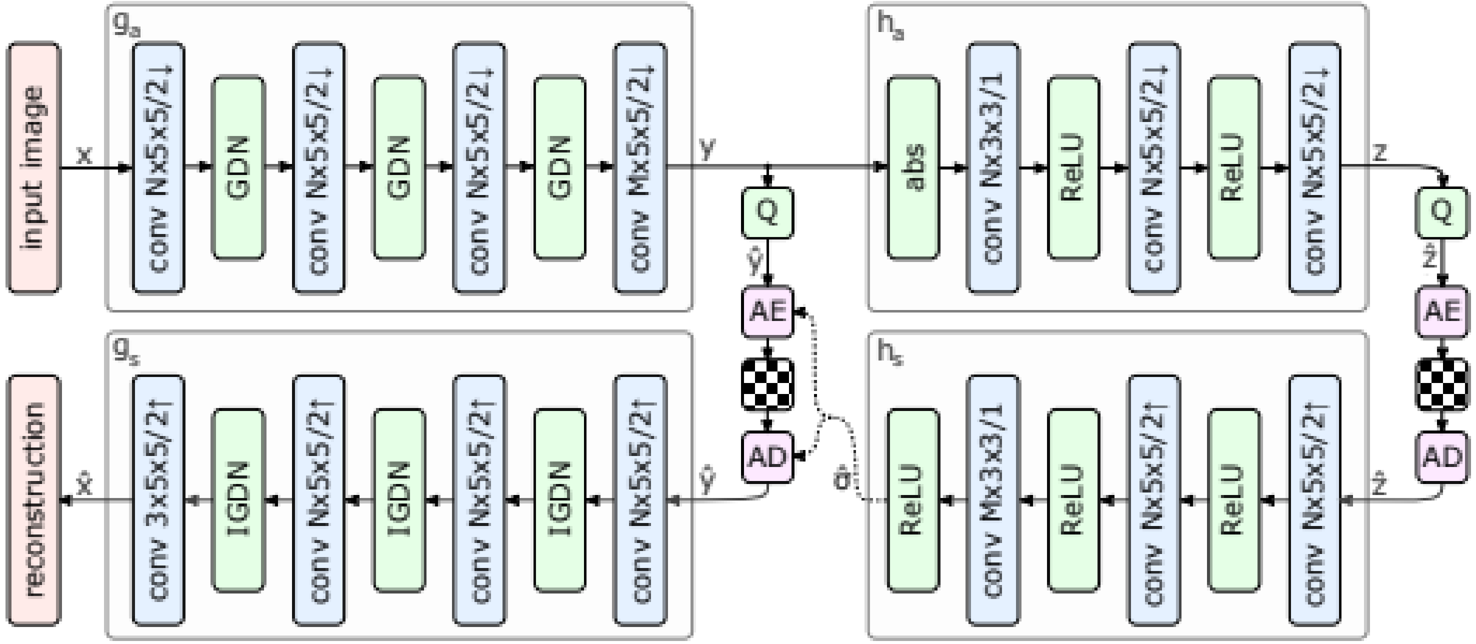


Figure 4: Network architecture of the hyperprior model. The left side shows an image autoencoder architecture, the right side corresponds to the autoencoder implementing the hyperprior. The factorized-prior model uses the identical architecture for the analysis and synthesis transforms g_a and g_s. Q represents quantization, and AE, AD represent arithmetic encoder and arithmetic decoder, respectively. Convolution parameters are denoted as: number of filters × kernel support height × kernel support width / down- or upsampling stride, where ↑ indicates upsampling and ↓ downsampling. N and M were chosen dependent on λ, with $N = 128$ and $M = 192$ for the 5 lower values, and $N = 192$ and $M = 320$ for the 3 higher values.

AI based Video Coding Development

- **AI based video codec tool development**
  - AI based tools have been studied to improve coding efficiency under the conventional codec framework that is hybrid codec structure.

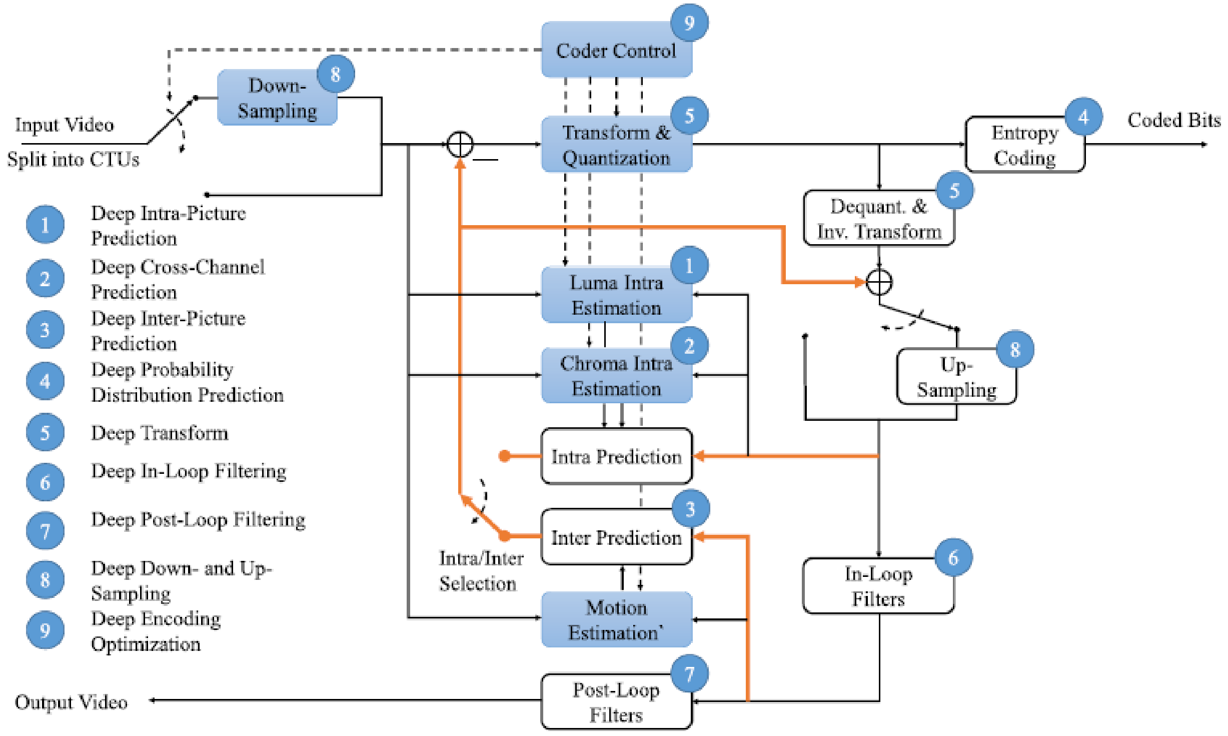
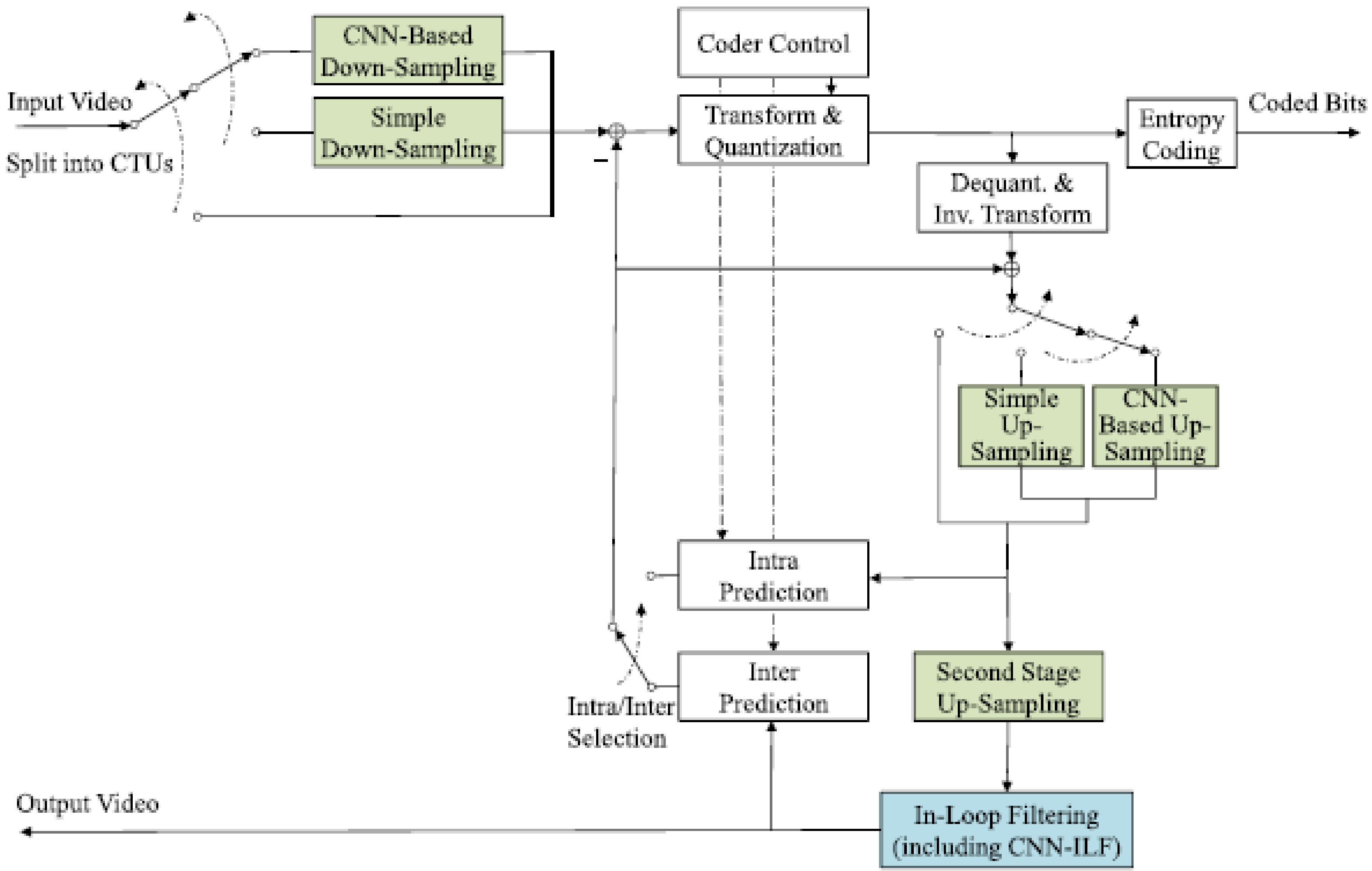


Fig. 3. Illustration of a traditional hybrid video coding scheme as well as the locations of deep tools inside the scheme. Note that the yellow lines indicate the flow of prediction, and the blue boxes indicate the tools that are used at the encoder side only.

Liu, Dong, et al. "Deep learning-based video coding: A review and a case study." *ACM Computing Surveys (CSUR)* 53.1 (2020): 1-35.

AI based Video Coding Development

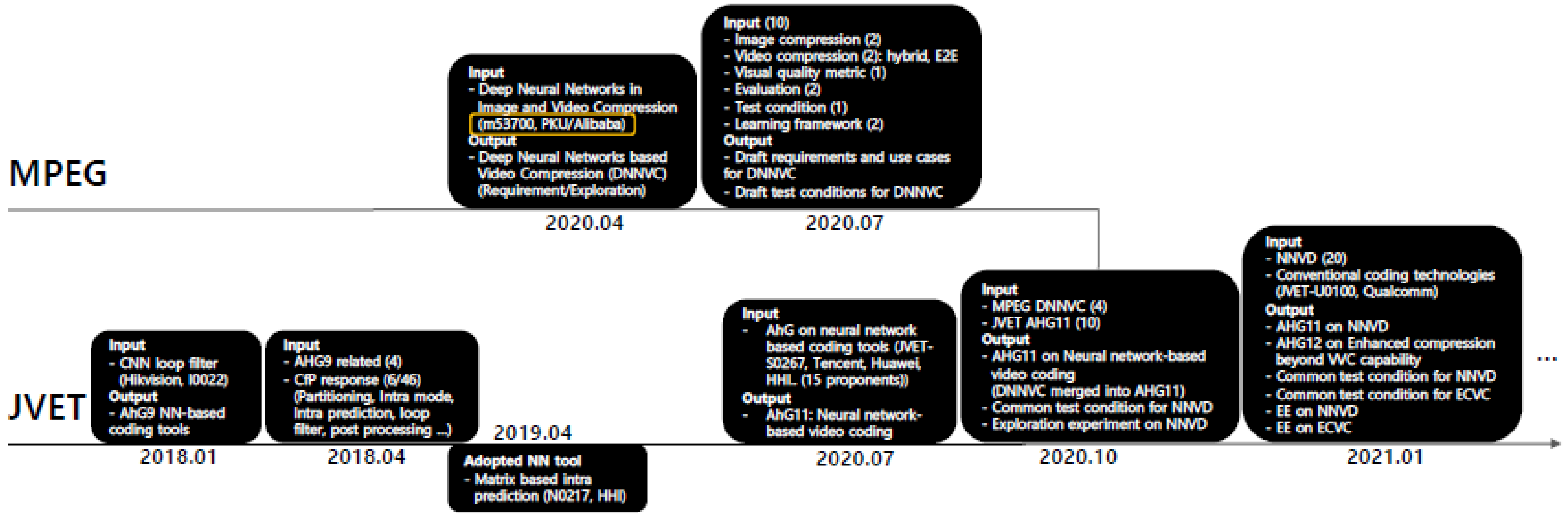
- **Down/up Sampling (SR) based video coding development**



AI related Standard activity

JVET activity on AI

- **History of AI based technology**



JVET activity on AI

- **Common Test Conditions (JVET U2016)**
  - Anchors: VTM12.0
  - Test sequences: A1|A2|B|C|D|E|F|H2 (4K|4K|1080p|WVGA|WQVGA | 720p | 720p | 4K)
    - ftp://jvet@ftp.hhi.fraunhofer.de/ctc/sdr or ftp://jvet@ftp.ient.rwth-aachen.de/ctc/sdr
  - Quantization parameters: fixed QP [22, 27, 32, 37, 42]
  - Evaluation metrics: PSNR, MS-SSIM (optional)
  - Configuration: All Intra(AI), Random access (RA), Low-delay B (LB), Low-delay P (LP)
  - Training sequences
    - JVET FTP dataset
      - https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/nnvc-ctc
      - 102 sequences identified from the JVET ftp site
    - BVI-DVC dataset (except source 9, 15)
      - University of Bristol: https://vilab.blogs.bristol.ac.uk/?p=2375
      - 764 sequences, 270p/540p/1080p/2160p, 10bit, 64frames
    - Tencent Video Dataset
      - Tencent: http://mmedia.oa.com/open/tvd
      - 113 sequences, 4K, 65frames
        - 89 sequences captured in R3D (10bit)
        - 24 sequences captured in BRAW(8bit)

	Request
Platform	
GPU Type (including Memory Size)	M
Framework	M
Number of GPUs per task	M
Sum of Peak TeraFLOP	O
Inference and Training	M
Training	
Epoch	M
Batch size	M
Number Iterations	O
Training Time	M
Patch size	O
Learning rate	O
Optimizer	O
Loss function	O
Preprocessing (Y/N)	O
Training sets	M
Training configuration per R-D point. This could include which QP points where used for training each R-D point.	M
Learning Curve (Training and Validation)	M
Inference	
Architecture / Network Summary / Graphical	M
Total Conv. Layers	
Total FC Layers	
Framework	N/A
Parameters – Number (Per Model and/or Total)	M
Parameter – Precision	M
Parameter – Memory (Per Model and/or Total)	M (Auto)
Total Memory	
Total MACs (or per pixel)	M
Any changes to network configuration or weights required to ge nerate rate points	
Peak Memory Usage	
Patch and Batch size	
Run Time: Proponent may choose to report one of the below: - CPU run-time - GPU run-time - CPU+GPU run-time Participants should provide a description of what run-time is rep orted and the environment used for computing the run-time, in cluding the patch size, batch size and number of GPUs	M

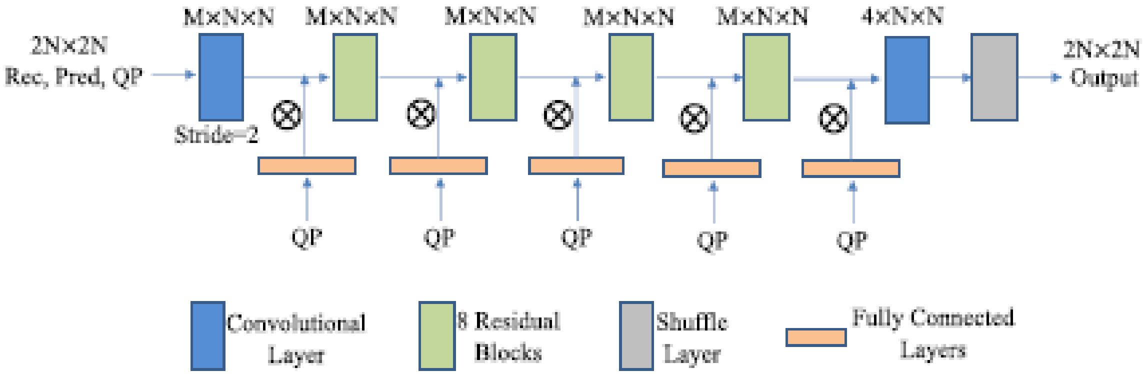
JVET activity on AI

- Exploration experiment (EE)
  - NN In-loop filter
  - NN Super resolution
- Non-EE contributions
  - NN In-loop filter
  - NN Post-filter
  - NN SAO
  - NN Inter prediction
  - NN Intra prediction
  - E2E Intra coding

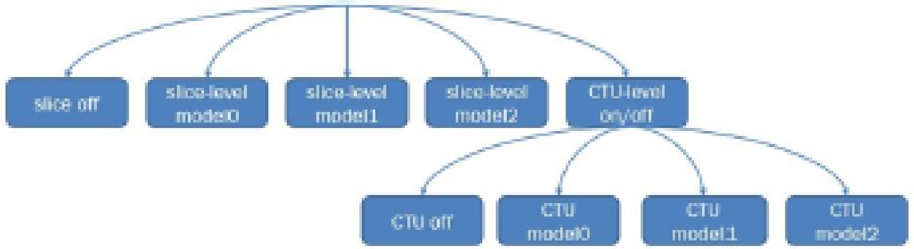
Doc.#	Proponent	Category	Performance	EncT/DecT	#params	kMAC /pxl
JVET-V0114	Qualcomm	NN In-loop filter	-7.5%/-14%/-14%@RA	x1.5/x146	4M	319
JVET-V0115	Qualcomm	NN In-loop filter	-8.5%/-22%/-22%@RA	x2.1/x344	8M	624
JVET-V0137	Tencent	NN In-loop filter	-1.1%/-4.9%/-3.8%@RA	x1.0/x30	23K	23
JVET-V0073	Sharp	Super Resolution	-6.5%/10.8/11.5@RA(4K)	x1.0/x1.5	483.6K	122
JVET-V0096	Qualcomm	Super Resolution	-8.1%/26.5%/15.5%@RA(4K)	x1.2/x0.3	2.7M	344
JVET-V0149	NJU/OPPO	Super Resolution	-9.6%/8.5%/-1.7%@RA(4K)		42.5M	
JVET-V0074	Alibaba	NN In-loop filter	-5.7%/-15.3%/-15.6%@RA	x1.1/x1.0 (16 core)	1 M	
JVET-V0090	Tencent	NN In-loop filter	-5.x%/-6.x%/-6.x%@RA(BCD)	x30/x40000	100M	2000
JVET-V0100	Bytedance	NN In-loop filter	-11.8%/-27.6%/-27.3%@RA	x2.6/x324	528M	1400
JVET-V0101	Bytedance	NN In-loop filter	-11.4%/-22.7%/-22.6%@RA	x2.6/x302	90M	1400
JVET-V0075	Nokia	NN post-filter	-2.5%/-5%/-5%@RA(CD)	1.2/x464	0.15M (323 var)	
JVET-V0092	InterDigital	NN SAO	-3.5%/-7.2%/-6.3%@RA	1.8/x125	2.5M	2898
JVET-V0076	InterDigital	NN Inter	-1.0%/-0.3%/-0.3%@RA	x7.6/x11	8K	16.6
JVET-V0105	InterDigital	NN Intra	-1.69%/-3.48%/@RA/AI (TF) -1.69%/-3.46%/@RA/AI (16bit)	x 2 / x 8 x 1.7 / x 5	16.1M	1600~11500
JVET-V0055	Alibaba	E2E Intra	-1.44%/24.74%/11.91%@AI	x 37		

JVET activity on AI

- **JVET-V0101: Conditional In-loop filter with parameter selection (Bytedance)**
  - Training
    - Network
      - Kernal size: 3x3/feature map M=96/#Resblock=32/patch=128x128
      - Total Conv.Layers=69 / FC Layers=10
    - Input to network
      - Recon + Prediction + QP
    - Total 4 models, ~22.6M/model
      - Different group models for I/B slice, Y/CbCr
    - Training data: DIV2K, BVI-DVC

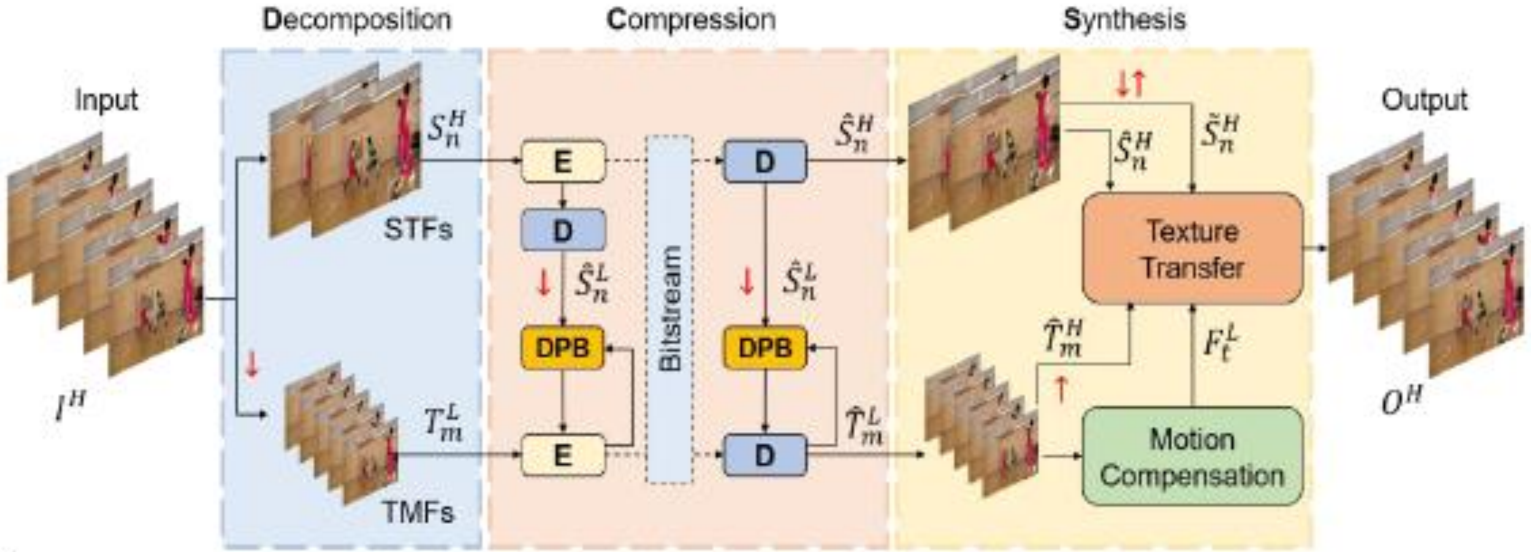


  - Apply to VTM
    - Slice-level multi-models, CTU-level multi-models
    - Adaptive granularity for filter determination and model selection
      - For higher resolution and a larger QP, use coarser granularity
      - Patch size: 32x32~256x256
    - Candidate model list: {QP, QP-5, QP-10}
    - Candidate model list is different for different temporal layers
      - Lower temporal layer: {QP, QP-5, QP-10} or {QP, QP-5, QP(I)}
        - if ratio of intra-coded samples >=threshold, replace 3rd model
      - Higher temporal layer:, {QP, QP-5, QP+5}
    - Performance
      - RA: -11.4%/-22.7%/-22.6%, ×2.6/x302 encT/decT
      - AI: -8.7%/-21.3%/-22.1% , x1.7/x288 encT/decT



JVET activity on AI

- **JVET-V0149: Decomposition, Compression, Synthesis (DCS)-based Technology (NJU/OPPO)** [2]
  - Motion compensation network
  - Texture Transfer Network
  - TMF, STF using VTM10.0
  - Down-sampling filter
    - Bicubic in FFmpeg
  - Total 10 models, 4.25M/model
    - Y/UV, 5 QP values
  - Performance
    - RA: -9.6%/4.3%/-4.5% in class A, overall -4.6%



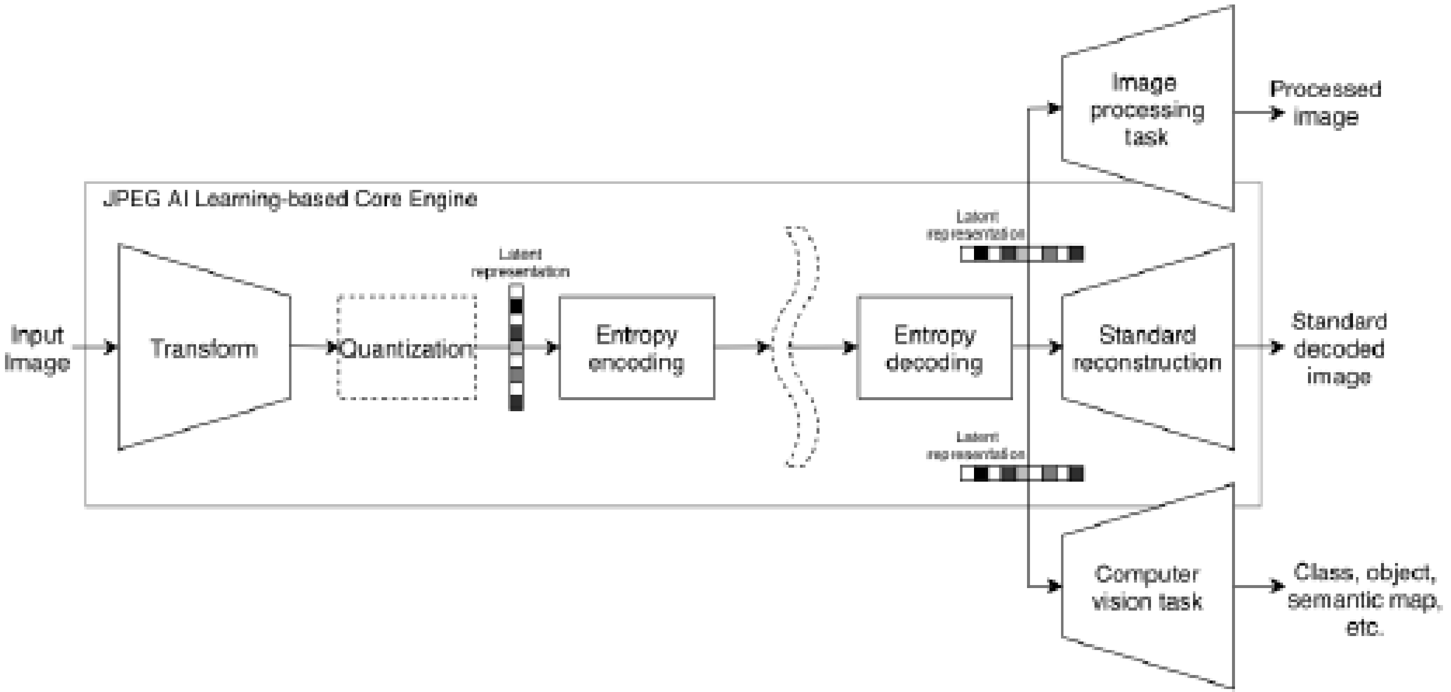
↓(in red) is for down-sampling, and ↑ (in red) is for up-sampling; **E**, **D** and **DPB** represent video encoder, decod er and decoding picture buffer, respectively.

JPEG activity on AI

- **JPEG AI history**
  - Exploration within JPEG standardization committee to assess performance of learning-based image compression in view of standardization since 2019
  - Call for Evidence combined with MMSP Workshop Grand Challenge
    - 6 submissions: 5 deep learning(DL) based image codecs, 1 hybrid VVC+DL based codec
  - New Work Item approved as ISO/IEC 6048
  - 2021.04: Established Joint AHG of JPEG & Q5 (Intelligent multimedia) & Q6(VCEG)
  - 2021.06: Call for Proposal
  - 2021.12: CfP response (2022.01: review result)
- **JPEG-AI Scope (ISO/IEC 6048)**
  - Creation of a learning-based image coding standard offering a **single-stream, compact** compressed domain representation, targeting both **human visualization**, with significant compression efficiency improvement over image coding standards in common use at equivalent subjective quality, and effective performance for **image processing and computer vision tasks**, with the goal of supporting a **royalty-free baseline**.

JPEG activity on AI

- **JPEG AI Framework: a Triple-Purpose Approach**
  - Standard JPEG AI decoding
  - Image processing tasks
    - Super-resolution
    - Denoising
    - Low-light enhancement
    - Color correction
    - Exposure compensation
    - Inpaiting
  - Computer vision tasks
    - Image retrieval and classification
    - Object detection, recognition and identification
    - Semantic segmentation (generate classes, labels, regions, etc.)
    - Event detection and action recognition
    - Face detection and recognition



Conclusion

Conclusion

- **Conclusion and outlook**
  - AI based video coding tool developments have been started recently.
  - Recently, Standard groups (e.g., JVET, MPEG, MPAI) have intensively studied on the AI based video coding tool development.
  - From the analysis of the recent studies on the video coding tools using AI network, we can find some commonalities as follows:
    - Focusing on how to use NN with available contexts more rather than network design
    - Focusing on how to design training data rather than network design
    - Separating models and using dedicated models depending on categories (e.g., QPs, Frame types, etc)
    - Increasing models and the number of network depths
  - It should be studied on memory and computation efficient design for practical image and video codec.
  - Considering the fact that the requirement about video coding technology is still increasing, the AI based video coding development seems promising.

Thank you!