

**SHARP**

Be Original.

ICTビジネス戦略オンラインセミナー

「デジュール及びフォーラム標準に関する国際標準化活動動向調査」(第一回)

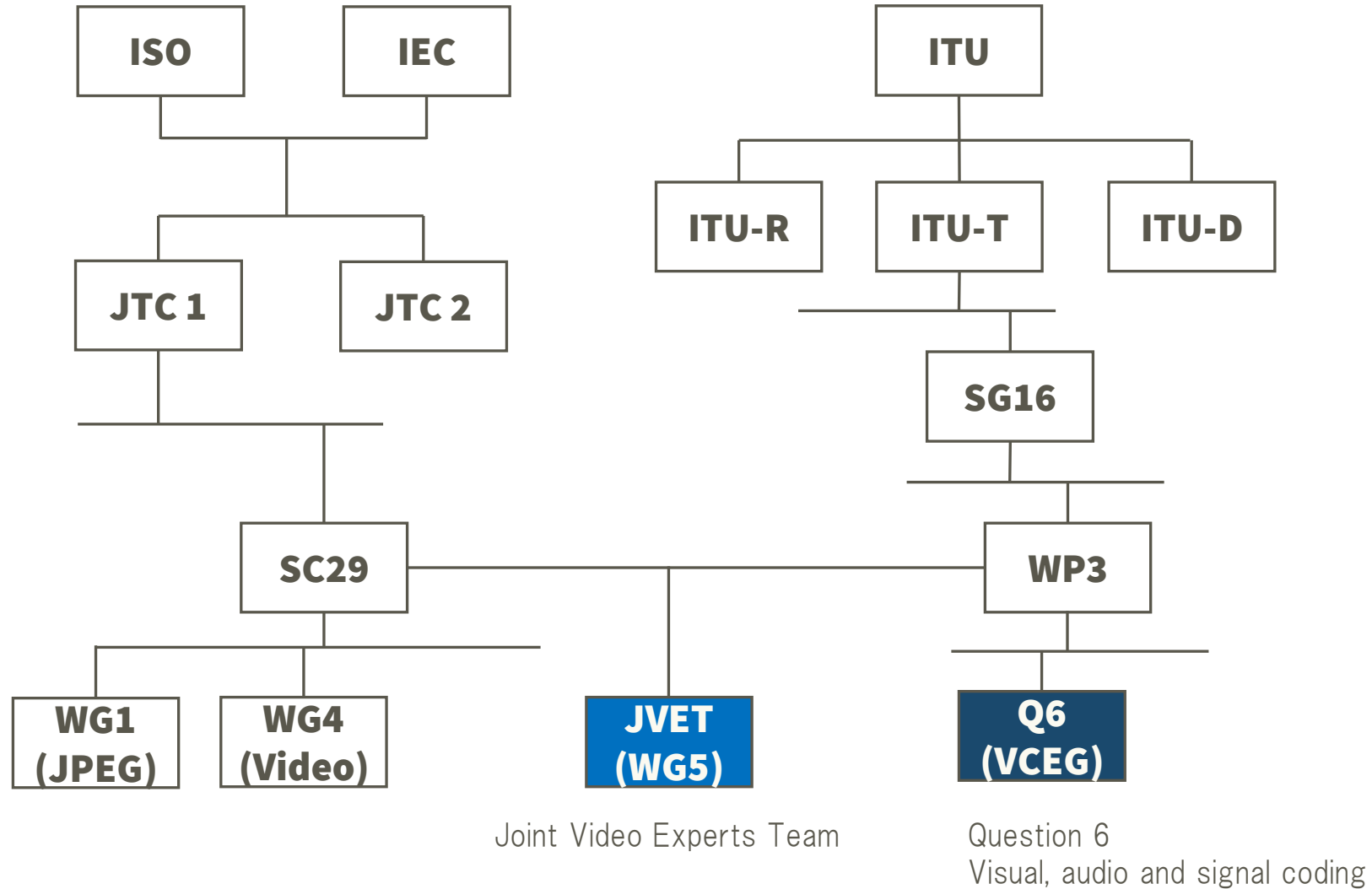
# ITU-T SG16/Q6 及び JVETの活動

Sharp Corporation

中條 健

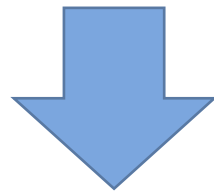
2022/1/28

• 組織図



- ITU-T SG16/Q6 Visual, audio and signal coding
  - H.シリーズ(H.266,H.265,H.264等)の動画像符号化とT.シリーズ(H.800,H.832等)の静止画符号化技術の標準化及び音声符号化関連の保守活動
  - 今年度の会合
    - 第7回 ITU-T SG16会合 Online 2021/4/19-30
    - 第8回 ITU-T SG16会合 Online 2022/1/17-28
- JVET(Joint Video Experts Team)
  - H.266/VVCとH.274/VSEIの拡張、H.265/HEVC、H.264/AVCの技術的メンテナンス
  - 今年度の会合
    - 第18回 JVET会合 Teleconference 2021/4/20-28
    - 第19回 JVET会合 Teleconference 2021/7/7-16
    - 第20回 JVET会合 Teleconference 2021/10/6-15
    - 第21回 JVET会合 Teleconference 2022/1/12-21

- H.266 (V1) VVC (Versatile Video Coding)
  - コンシューマ向けの機能は、スクリーンコンテンツを含め、ほぼ全て網羅
  - 4:4:4/4:2:2(H.265(V2)の機能)
  - マルチレイヤ (H.265(V2)の機能)
  - スクリーンコンテンツ (H.265(V4)の機能)
- H.274 (V1) VSEI (Versatile Supplemental Enhancement Information)
  - ビットストリームに入れることができる補助情報
  - H.264, H.265では、Annexとして記述されていたものを独立規格化



- H.266 (V2) 12bit-16bit ビットデプス拡張、高解像度のLevelの拡張
- H.274 (V2) Alphaチャンネル/デプス/ViewIDなどの9つの補助情報の追加

- プロファイル

- 9個の整理されたプロファイル(V2)
  - インター、イントラ、静止画を持つ
  - 8-bit、Monochrome、4:2:2プロファイルを作らない

- 5つの拡張ツールに整理

- ① 変換精度拡張  
sps\_extended\_precision\_flag
- ② ラスト符号化拡張  
sps\_reverse\_last\_sig\_coeff\_enabled\_flag
- ③ ライスパラメータ導出拡張  
sps\_rice\_extension\_flag
- ④ ライスパラメータ導出拡張  
sps\_persistent\_rice\_adaptation\_flag
- ⑤ 変換スキップのライスパラメータ導出  
sps\_ts\_residual\_coding\_rice\_present\_flag

**H.266 (V1)**

**Main 10**  
**Main 10 Still Picture**  
**Multilayer Main 10**

**Main 10 4:4:4**  
**Main 10 4:4:4 Still Picture**  
**Multilayer Main 10 4:4:4**

**H.266 (V2)**  
**Operation range extensions**

**Main 12**  
**Main 12 Intra**  
**Main 12 Still Picture**

**Main 12 4:4:4**  
**Main 12 4:4:4 Intra**  
**Main 12 4:4:4 Still Picture**

**Main 16 4:4:4**  
**Main 16 4:4:4 Intra**  
**Main 16 4:4:4 Still Picture**

ビットデプス拡張(その1)

① 変換精度拡張

変換の精度を  
**15-bit以上を増やす**

$\text{MaxTransformRange} = 15$   
 $\Rightarrow \text{Max}( 15, \text{Min}( 20, \text{BitDepth} + 6 ) )$

残差シンタックスの  
値域を増やす



$\text{Prefix} = 6 \Rightarrow 6$

$\text{maxPreExtLen} = 11 (=26 - 15) \Rightarrow 26 - \text{MaxTransformRange}$

$\text{truncSuffixLen} = 15 \Rightarrow \text{MaxTransformRange}$

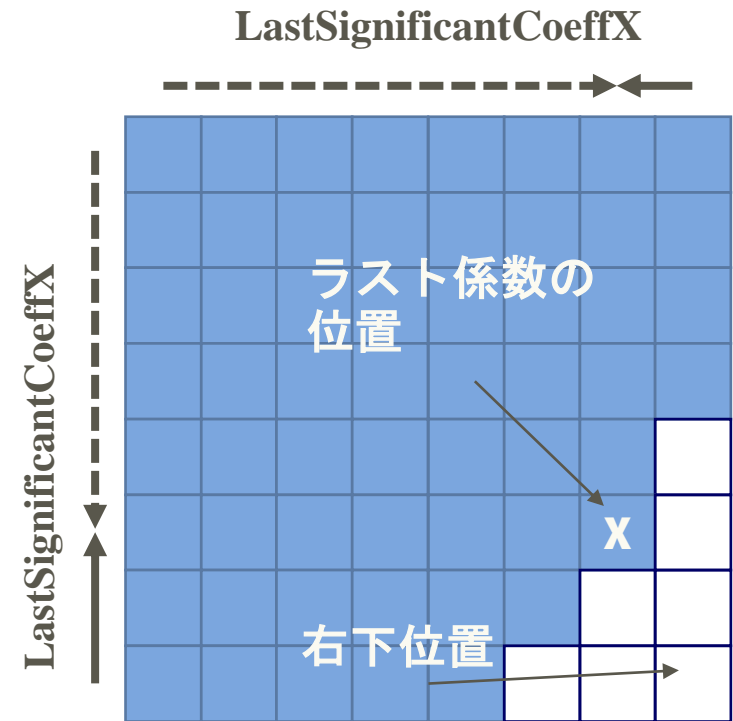
② ラスト符号化拡張

ラスト変数位置を右下位置からの距離で符号化する

sh\_reverse\_last\_sig\_coeff\_flag==1の場合、以下を適用

$\text{LastSignificantCoeffX} = ( 1 \ll \text{Log2ZoTbWidth} ) - 1 - \text{LastSignificantCoeffX}$

$\text{LastSignificantCoeffY} = ( 1 \ll \text{Log2ZoTbHeight} ) - 1 - \text{LastSignificantCoeffY}$



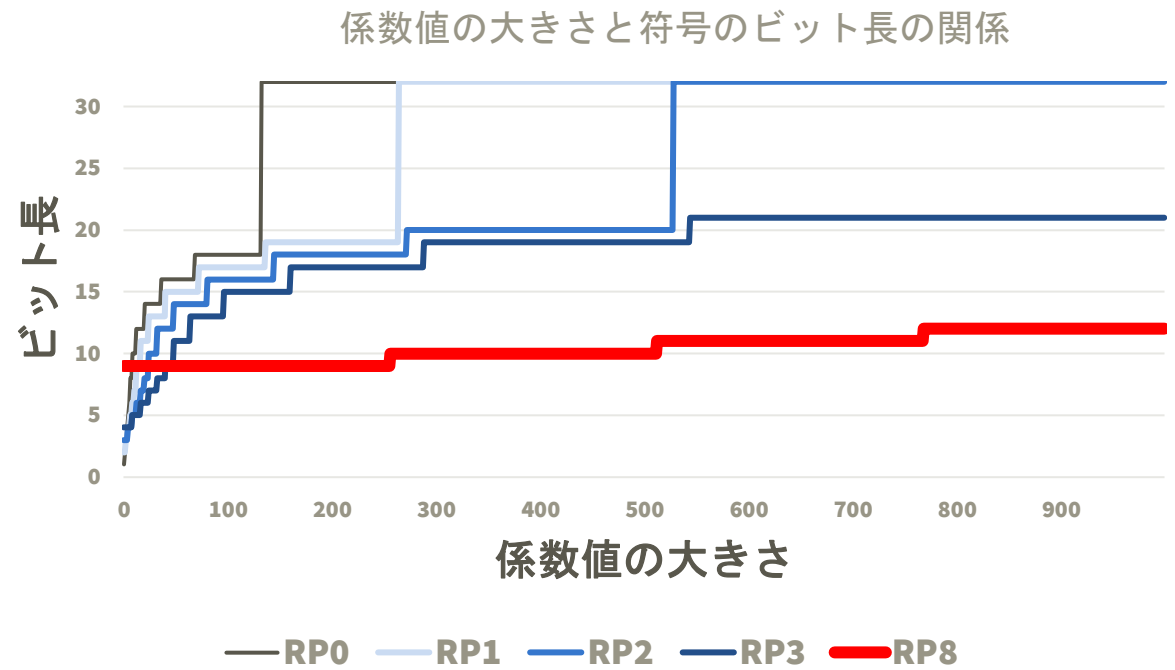
• 残差符号化の基本

- 小さい値⇒短く、大きな値⇒小さく。ライスパラメータ(RP)で変更
- V1では、RP=0~3を用いるが、**係数値が1000以上になるような高ビットでは足りない**

• エントロピー符号化拡張

- ③ライスパラメータ導出拡張
- ④ライスパラメータ導出拡張
  - 周囲の状態から大きなRPの値を予測可能に
- ⑤変換スキップのライスパラメータ導出
  - 変換スキップではRPの値を直接符号化

RP==0の符号例	RP=1の符号例	RP=2の符号例	RP=3の符号例
0 0	0 00	0 000	0 0000
1 10	1 01	1 001	1 0001
2 110	2 100	2 010	2 0010
3 1110	3 101	3 011	3 0011
4 11110	4 1100	4 1000	4 0100
5 111110	5 1101	5 1001	5 0101
6 11111100	6 11100	6 1010	6 0110
7 11111101	7 11101	7 1011	7 0111
8 1111111000	8 111100	8 11000	8 10000
9 1111111001	9 111101	9 11001	9 10001

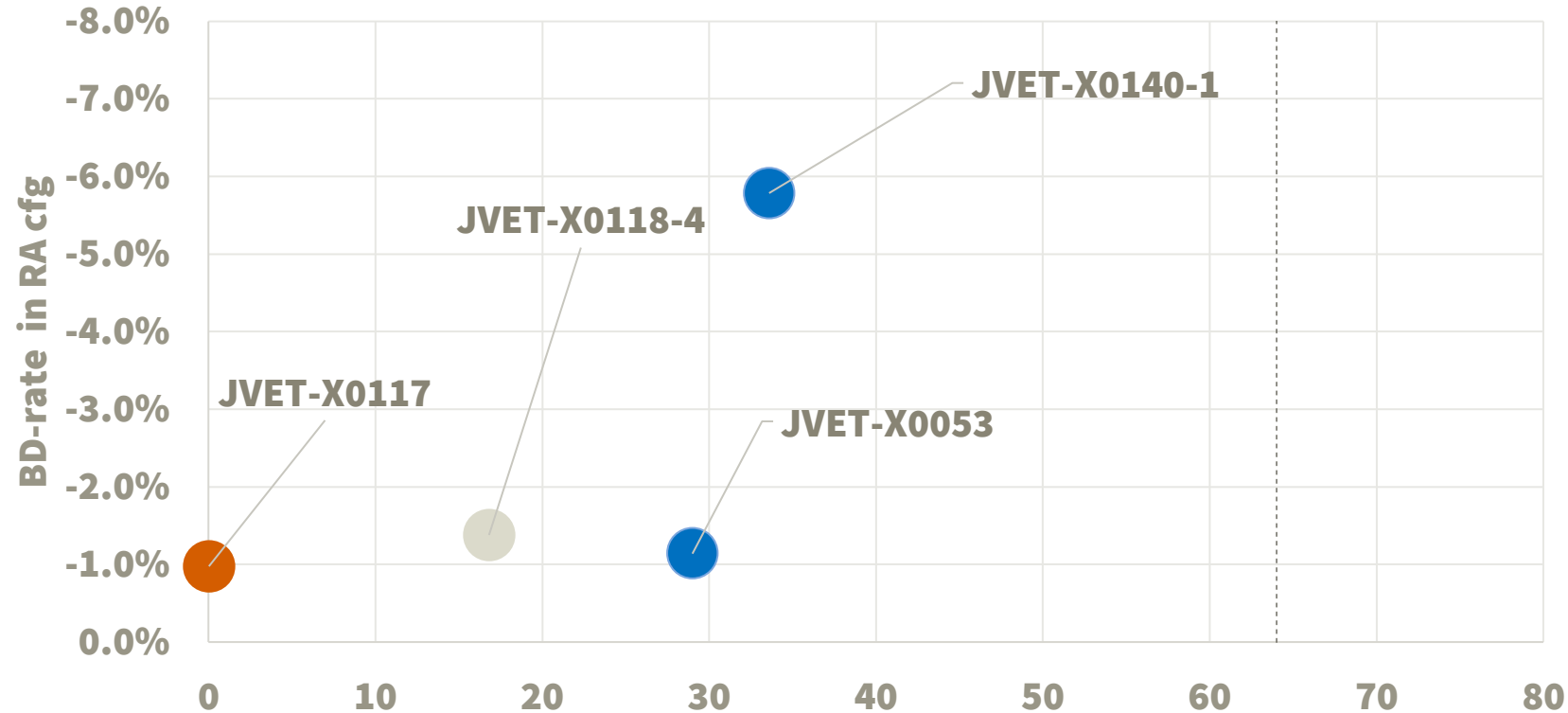


- H.266(V1)標準化あとの探索活動(EE: Exploration Experiment)をスタート
  - 2021年1月会合から
- EE1: Neural Network-based Video Coding
  - Neural Networkによる深層学習を用いた方式の検討
  - ループフィルタ、超解像、イントラ予測、インター予測
- EE2: Enhanced Compression beyond VVC capability
  - 従来の符号化ツールの拡張
  - ECM(Enhanced Compression Model)と呼ばれる参照ソフトウェアを作成
  - 最新のECM-3.1は、H.266の参照ソフトウェアVTMと比較して、Intraで約7%、Interで約16%の改善



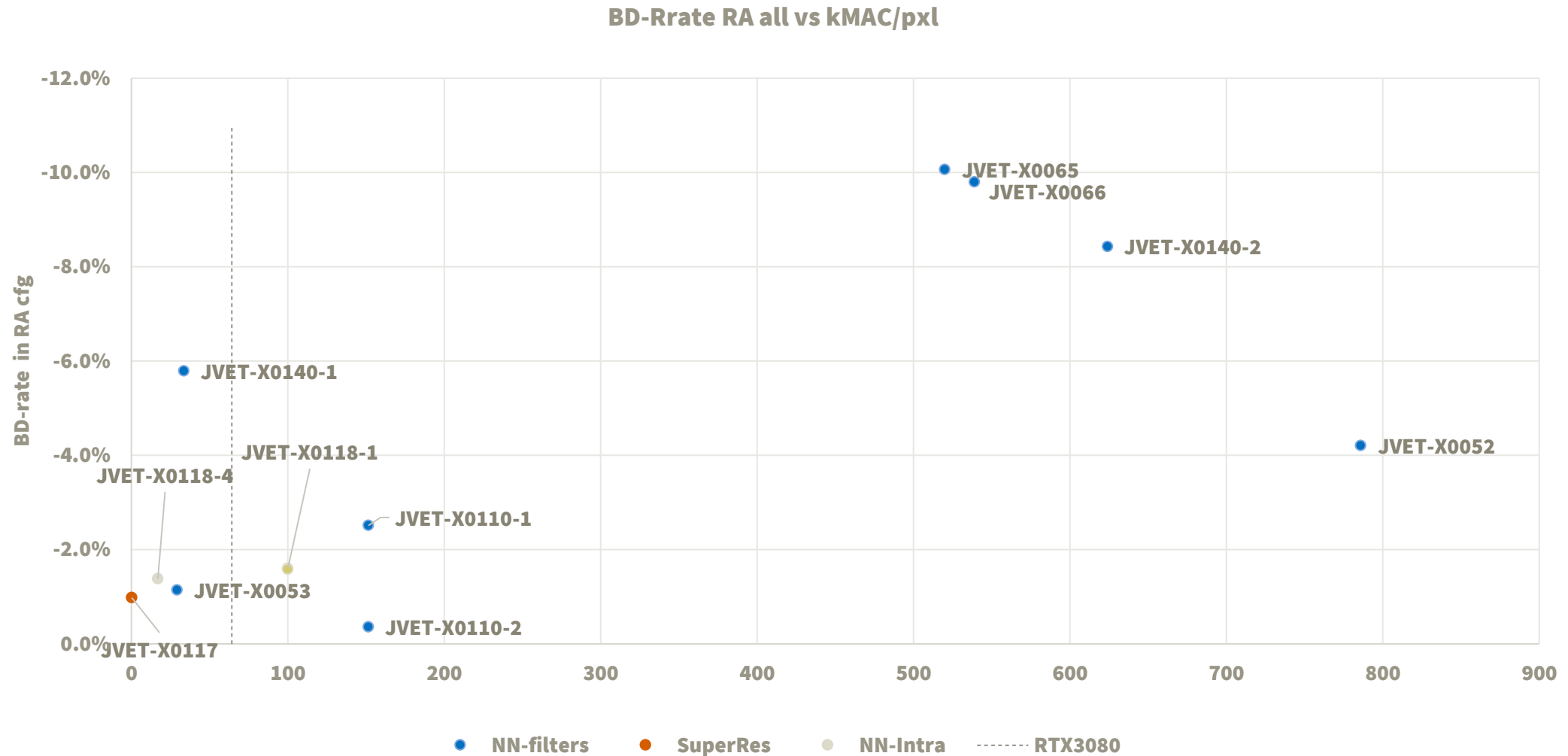
- **NN超解像:** パラメータ伝送でさらなる効果
- **NNループフィルタ:** 30k MAC(3万積/ピクセル)の計算量で6%近い性能向上
- **NNイントラ予測:** 1%以上の性能向上

**BD-Rrate RA all vs kMAC/pxl**

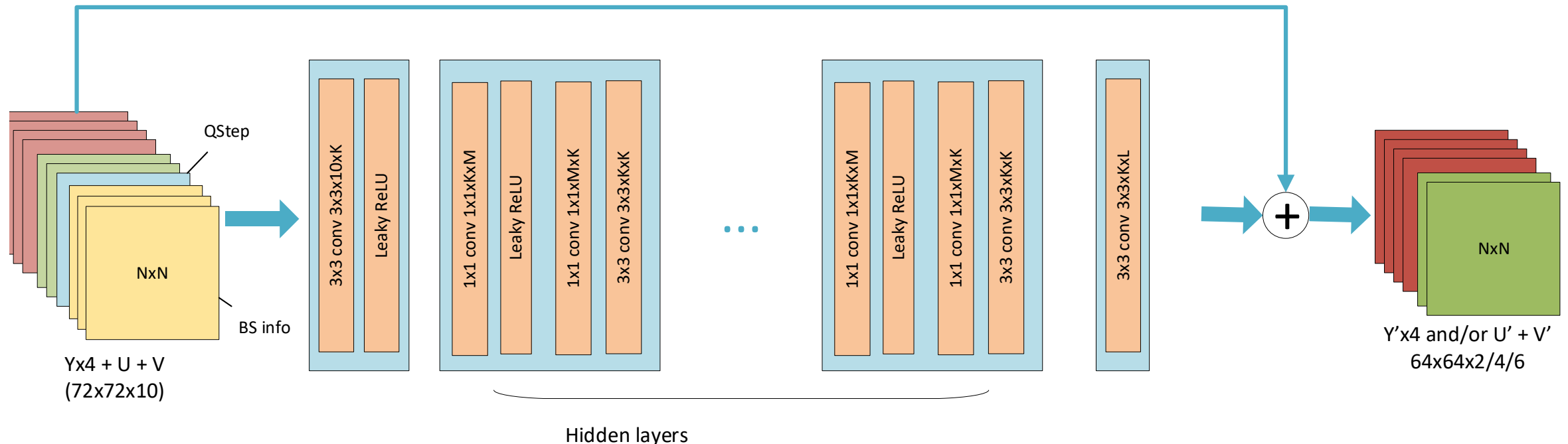


図は、JVET会合寄書の  
"EE1: Summary of Exploration  
Experiments on Neural  
Network-based Video Coding",  
JVET-X0023(2021/10)の  
スプレッドシートの図を利用  
(縮尺などを調整)

• EE1結果全体(2021/10会合)



- JVET-X0140 EE1-1.4: Tests on Neural Network-based In-Loop Filter with Constrained Computational Complexity (Qualcomm)
  - 量子化情報とブロック情報を入力する
  - ブロック適応
    - CTU単位でNNフィルタのオンオフだけでなく、強さを制御するscale情報を伝送する
    - 出力画像 = scale \* NNフィルタ + (1 - scale) \* NNフィルタオフ
  - 1x1 Convを利用して計算量の割に層の数を増やす(3x3空間フィルタの1/9の処理量)



• **JVET**検討技術の特徴

- 計算量を大幅に削減して同程度の性能 (イントラで約**0.4%**の性能低下、インターでは**0.2%**の性能低下)
- イントラで約**3%**、インター(**RA構造**)でも**1.4%**以上の性能向上

計算量

29687 **MAC/pixel: 8 × 8 block**

All Intra Main10

BD-rate Over VTM-11.0 NNVC

	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM	EncT	DecT CPU	bit DIFF
Class A1	-3.93%	-3.10%	-3.37%	-4.14%	-2.65%	-3.10%	211%	1845%	3%
Class A2	-2.43%	-2.33%	-2.69%	-2.48%	-1.70%	-2.16%	217%	1805%	2%
Class B	-2.92%	-2.41%	-2.57%	-2.67%	-1.98%	-1.76%	219%	1903%	2%
Class C	-3.08%	-2.48%	-2.78%	-3.11%	-1.77%	-1.86%	217%	1927%	3%
Class E	-4.55%	-4.29%	-3.53%	-4.69%	-3.94%	-1.98%	222%	2198%	4%
<b>Overall</b>	<b>-3.31%</b>	<b>-2.84%</b>	<b>-2.93%</b>	<b>-3.32%</b>	<b>-2.33%</b>	<b>-2.11%</b>	<b>217%</b>	<b>1927%</b>	<b>3%</b>
Class D	-2.98%	-2.51%	-2.73%	-2.89%	-2.09%	-1.81%	208%	1926%	2%
Class F	-2.22%	-1.58%	-1.63%	-2.34%	-1.15%	-1.28%	139%	1134%	2%
Class H									

計算量

**2234 MAC/pixel: 8 × 8 block**

All Intra Main10

BD-rate Over VTM-11.0 NNVC

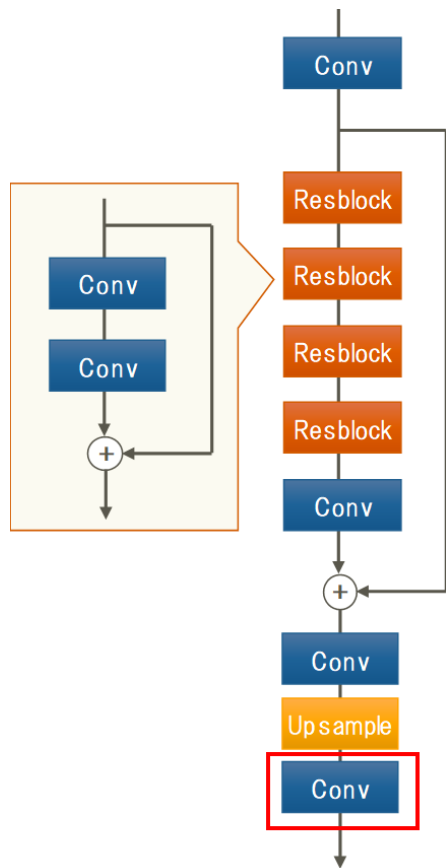
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM	EncT	DecT CPU	bit DIFF
Class A1	-3.76%	-2.87%	-3.21%	-4.02%	-2.38%	-3.02%	207%	1802%	3%
Class A2	-2.09%	-1.91%	-2.24%	-2.21%	-1.38%	-1.74%	222%	1767%	2%
Class B	-2.49%	-1.97%	-2.10%	-2.32%	-1.60%	-1.38%	221%	1735%	2%
Class C	-2.53%	-1.86%	-1.95%	-2.59%	-1.12%	-1.03%	218%	1790%	2%
Class E	-3.88%	-3.39%	-2.67%	-4.10%	-3.18%	-1.20%	222%	2023%	3%
<b>Overall</b>	<b>-2.87%</b>	<b>-2.32%</b>	<b>-2.37%</b>	<b>-2.94%</b>	<b>-1.85%</b>	<b>-1.60%</b>	<b>218%</b>	<b>1809%</b>	<b>2%</b>
Class D	-2.47%	-2.05%	-2.08%	-2.44%	-1.53%	-1.22%	211%	1749%	2%
Class F	-1.82%	-1.19%	-1.18%	-1.90%	-0.83%	-0.60%	137%	993%	2%
Class H									

EE1-3.1: BD-rate gains vs complexity of NN-based intra prediction, JVET-X0118

色々な解像度や種類のテスト  
画像のクラスでテスト

• シャープ提案(JVET-V0063)

- ユースケース例: VVC4Kを伝送し、AI超解像により8Kにする
- シーケンスで最適モデルが異なるので、NNモデルの情報をメタデータとして伝送
- メタデータサイズを最小化するため最終層だけを伝送



補間フィルタで拡大した場合と超解像+メタデータの比較

		Random access Main10								
		BD-rate Over VTM-11.0_NNVC RPR with half resolution QP=22, ... ,42								
		BD-rate5			BD-rate4 q2345 (piecewise cubic)			EncT	DecT	
		Y	U	V	Y	U	V			
Class A1 4K	Tango2	-0.36%	-3.97%	-2.32%	-0.40%	-4.92%	-2.47%	100%	795%	
	FoodMarket4	-1.27%	1.50%	-0.83%	-1.05%	-0.25%	-1.82%			
	Campfire	-3.71%	-0.88%	-8.43%	-2.48%	-0.81%	-7.84%			
Class A2 4K	CatRobot1	-6.69%	-2.47%	-6.11%	-4.81%	-3.85%	-5.89%	100%	733%	
	DaylightRoad2	-8.29%	-1.98%	-4.51%	-5.82%	-2.28%	-3.58%			
	ParkRunning3	-3.60%	-2.81%	3.52%	-2.43%	-4.57%	1.55%			

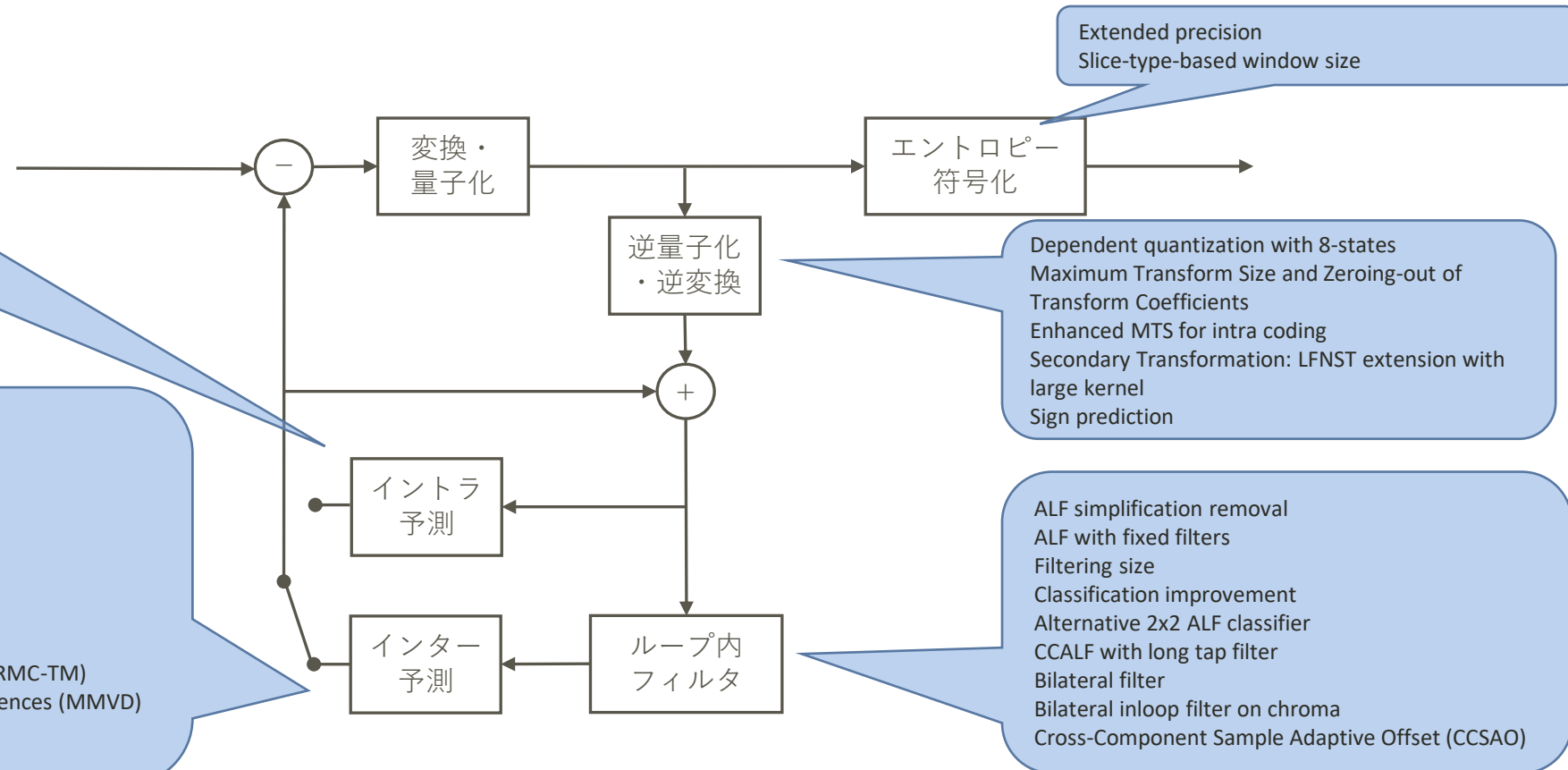
## EE2: Enhanced Compression beyond VVC capability

### • ECM(Enhanced Compression Model)

- 従来型の符号化効率改善
- H.266で未導入の符号化ツールや、ツールの改善で構成

Multi-model LM (MMLM)  
Gradient PDPC  
Secondary MPM  
Reference sample interpolation and smoothing for intra-prediction  
Decoder side intra mode derivation (DIMD)  
Intra template matching  
Fusion for template-based intra mode derivation (TIMD)  
Combination of CIIP with TIMD and TM merge

Local illumination compensation (LIC)  
Non-adjacent spatial candidate  
Template matching (TM)  
Multi-pass decoder-side motion vector refinement  
Adaptive decoder-side motion vector refinement  
OBMC  
Sample-based BDOF  
12-tap Interpolation filter  
Multi-hypothesis prediction (MHP)  
Adaptive reordering of merge candidates with template matching (ARMC-TM)  
Geometric partitioning mode (GPM) with merge motion vector differences (MMVD)  
Geometric partitioning mode (GPM) with template matching (TM)  
Bilateral matching AMVP-merge mode



- H.266の参照ソフトウェアVTMとの比較

- All Intraで約6%、Random Access構造で約15%の符号量削減

	All Intra Main 10				
	Over VTM-11.0 + V0056				
	Y	U	V	EncT	DecT
Class A1	-6.98%	-12.44%	-16.63%	311%	241%
Class A2	-6.44%	-13.13%	-11.99%	298%	232%
Class B	-6.05%	-13.75%	-14.21%	338%	257%
Class C	-7.06%	-10.21%	-10.67%	322%	252%
Class E	-7.58%	-11.55%	-12.41%	322%	294%
<b>Overall</b>	<b>-6.75%</b>	<b>-12.28%</b>	<b>-13.16%</b>	<b>320%</b>	<b>255%</b>
Class D	-5.86%	-8.45%	-8.23%	316%	265%
Class F	-10.80%	-15.69%	-15.66%	244%	298%
Class TGM	-15.62%	-18.73%	-18.25%	232%	293%

	Random Access Main 10				
	Over VTM-11.0 + V0056				
	Y	U	V	EncT	DecT
Class A1	-14.95%	-16.42%	-21.85%	367%	466%
Class A2	-15.84%	-20.38%	-20.42%	364%	511%
Class B	-13.57%	-19.94%	-19.32%	384%	493%
Class C	-15.35%	-17.27%	-16.98%	387%	427%
Class E					
<b>Overall</b>	<b>-14.77%</b>	<b>-18.61%</b>	<b>-19.42%</b>	<b>377%</b>	<b>473%</b>
Class D	-16.10%	-17.65%	-16.99%	378%	459%
Class F	-13.54%	-18.02%	-17.76%	345%	407%
Class TGM	-14.65%	-18.96%	-19.20%	348%	304%

- H.265の参照ソフトウェアHMとVTMの比較

- All Intraで約25%、Random Access構造で約38%の符号量削減

- JPEGとの連携

- Q5、Q6とISO/IEC JTC 1/SC29/WG1(JPEG)とAIについて協力関係を議論

- JPEG AI

- 学習ベースの静止画符号化規格、ISO/IEC 6048 JPEG AI learning-based image coding systemとして標準化予定
  - 2022/1会合で提案募集(CfP: Call for Proposal)
  - CfP 2022/01, WD 2022/10, CD 2023, FDIS 2024
- Q6のNew Work ItemとしてITU-Tでも規格化



- ITU-T SG16/Q6 及び JVETの活動
  - H.266(V2)
  - 技術探索活動(JVET EE)
    - EE1: Neural Network Video Coding
    - EE2: ECM(Enhanced Compression Model)
  - JPEGとの連携(Q6)
    - JPEG AIをQ6のNew Work Item

# SHARP

Be Original.

