

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)

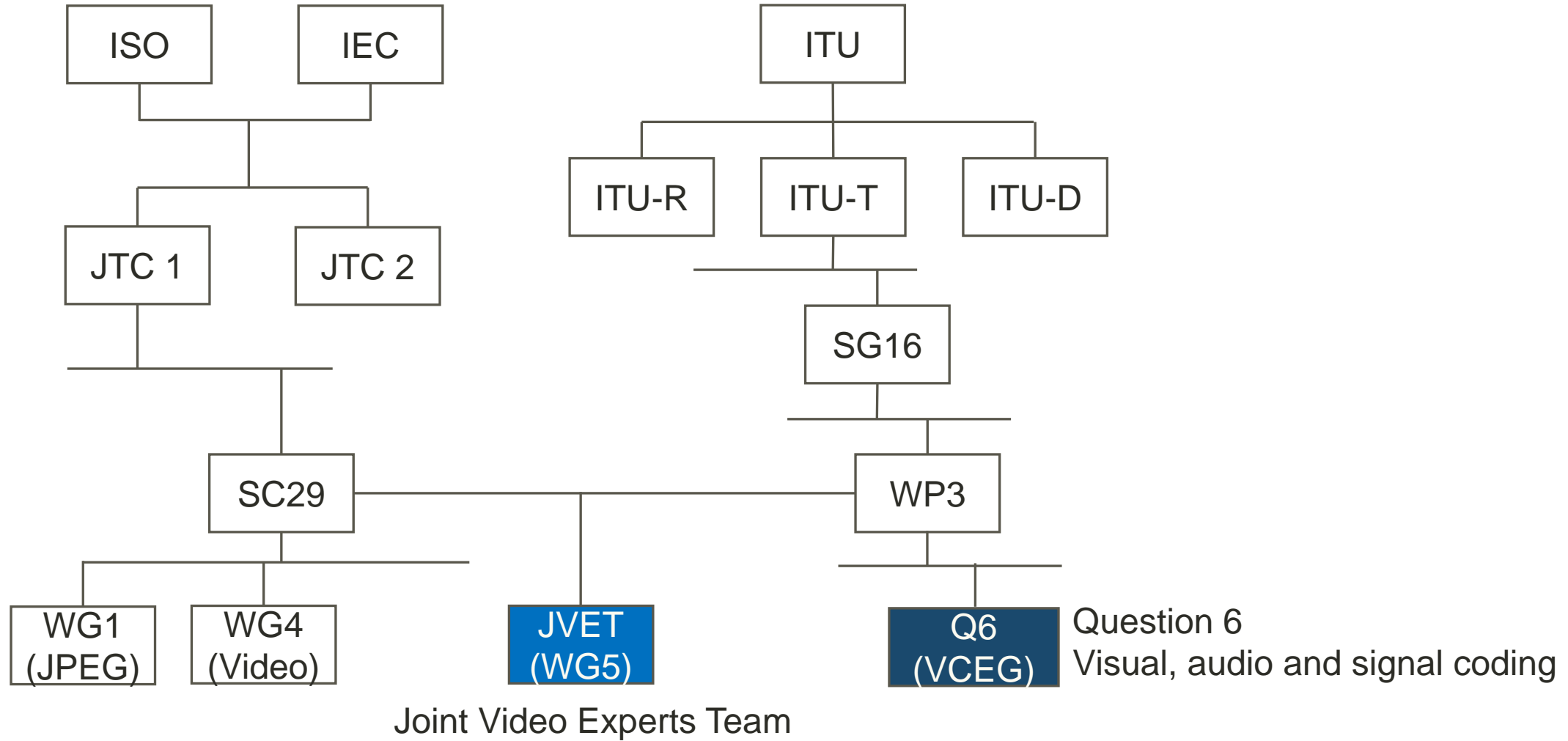
ITU-T SG16/Q6 及び JVETの活動

Sharp Corporation

范 哲銘

2023/2/3

画像符号化の国際標準化機関(ITU-T SG16/Q6 とJVET)



ISO: 国際標準化機構 (International Organization for Standardization)
IEC: 国際電気標準会議 (International Electrotechnical Commission)
ITU: 国際電気通信連合 (International Telecommunication Union)

- 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1回 ITU-T SG16会合 Geneva 2022/10/17-28
- 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26回 JVET会合(Z) Teleconference 2022/4/20-28
 - 第27回 JVET会合(AA) Teleconference 2022/7/7-16
 - 第28回 JVET会合(AB) Mainz, DE 2022/10/20-28
 - 第29回 JVET会合(AC) Teleconference 2023/1/11-20


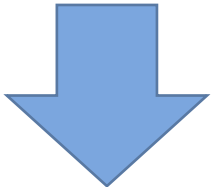
	参加人数	提案会社数	提案数	(シャープ提案数)
第26回JVET会合(Z) (2022/4)	360	約30社	120	(4)
第27回JVET会合(AA) (2022/7)	385	約30社	117	(6)
第28回JVET会合(AB) (2022/10)	387	約30社	156	(7)
第29回JVET会合(AC) (2023/1)	403	約30社	184	(7)

(中) Bytedance, Tencent, Alibaba, Kwai, OPPO, VIVO, Xiami, Dauha,

(日) Canon, KDDI, NHK, Panasonic, Sharp, Sony,

(米) Qualcomm, InterDigital, Apple, Dolby, Ofinno, (欧) Nokia, Ericsson, HHI, (韓) LGE, Samsung, ...

映像符号化標準(H.266, H.274)の動向

- H.266 (V1) VVC (Versatile Video Coding)
 - 動画圧縮標準規格
 - H.265/HEVC に対して約50%の性能向上。スクリーンコンテンツ等の機能を網羅。
 - 2020年に承認
 - 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拡張情報規格
 - 2020年に承認
- 
- H.266 (V2) 12-bit、16-bit 画素ビット長拡張、高解像度のLevelの拡張(2022)
 - H.274 (V2) Alphaチャンネル/デプス/ViewIDなどの9つの補助拡張情報の追加(2022)
- 
- 今年度**
- H.266 (V3) 新しいレベルとシステム関連の補助拡張情報の追加(2023予定)
 - H.274 (V3) Neural network postfilterの補助拡張情報などの追加(2023予定)

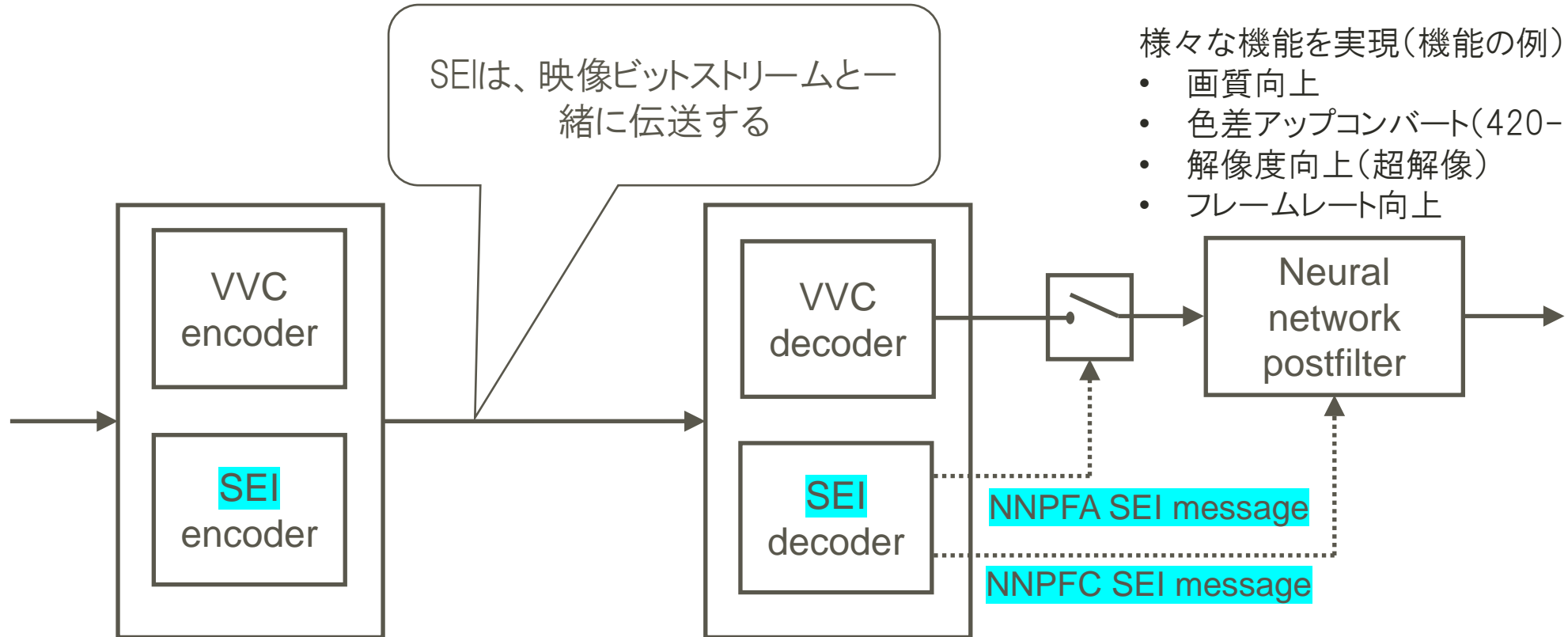
VSEI: 補助拡張情報(Versatile Supplemental Enhancement Information)

- ニューラルネットワークを用いたポストフィルタ処理を行う補助拡張情報(**SEI**)

2つのSEIを新しく定義

- Neural Network Post-Filter Activation (NNPFA) SEI
- Neural Network Post-Filter Characteristics (NNPFC) SEI

処理を適用するか否かを制御する情報
適用するネットワークの情報
(MPEG規格のNeural Network Codingを使って符号化)



- ニューラルネットワークを用いたポストフィルタ処理を行う補助拡張情報(SEI)
- H.274/VSEI (V3)に追加される。

SEI名	持続期間(適用フレーム)
Shutter interval information	The CVS containing the SEI message
Neural-network post-filter characteristics	The CVS containing the SEI message
Neural-network post-filter activation	Specified by the syntax of the SEI message
Phase indication	Specified by the semantics of the SEI message
Post-filter hint	Specified by the syntax of the SEI message

- SEIの処理の順序を決める補助拡張情報(SEI)
- H.266/VVC (V3)に追加されるSEIと前ページの追加SEIのインターフェース。

sei_processing_order(payloadSize) {	Descriptor
for(i = 0, b = 0; b < payloadSize; i++, b += 4) {	
po_sei_payload_type [i]	u(16)
po_sei_processing_order [i]	u(16)
}	
}	

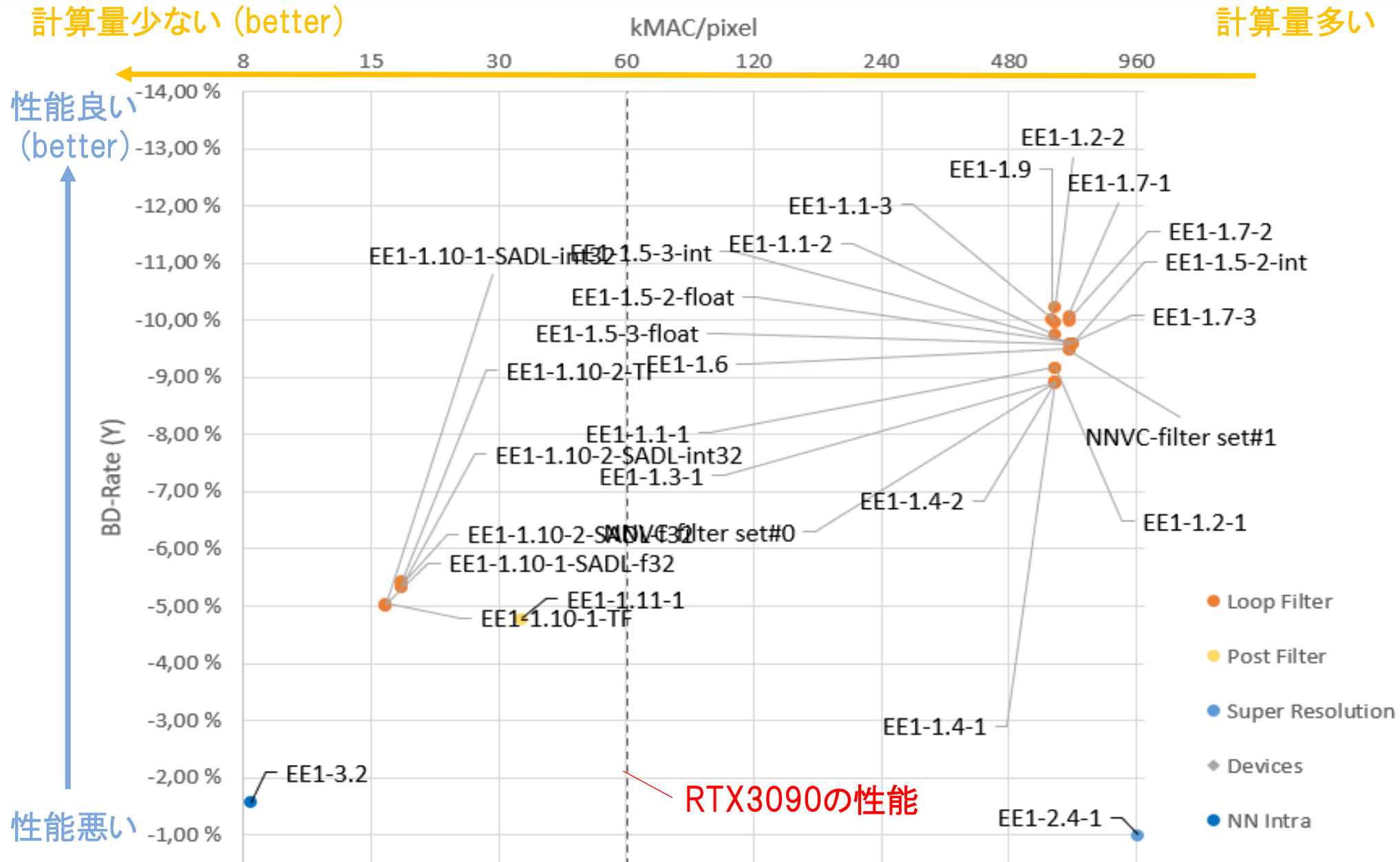
- H.266(V2)標準化あとの探索活動(EE: Exploration Experiment)
 - EE1: Neural Network-based Video Coding
 - Neural Networkによる深層学習を用いた方式の検討
 - EE2の活動と統合されて、VVCの新しい拡張profileもしくは次の符号化標準に繋がる技術探索。
 - ループフィルタ、超解像、イントラ予測、インター予測
 - EE2: Enhanced Compression beyond VVC capability
 - 従来の符号化ツールの拡張、VVCの新しい拡張profileもしくは次の符号化標準に繋がる技術探索。
 - ECM(Enhanced Compression Model)と呼ばれる参照ソフトウェアを作成
 - 最新のECM-7は、VVCの参照ソフトウェアVTMと比較して、Intraで9%、Interで約19%の改善

EE1: Neural Network-based Video Coding

EE1: Neural Network-based Video Coding

Year		2018				2019				2020				2021				2022				2023
MPEG	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141
JVET	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
VVC	CfP		WD1									v1 FDIS						v2 FDIS				
JVET NN活動		AHG9	AHG9	AHG9	AHG9	AHG9	AHG9 /CE13	AHG9 /CE10	AHG9			AHG1 1	AHG1 1	AHG1 1/EE	AHG1 1/EE1	AHG1 1/EE1	AHG1 1/EE1	AHG1 1/EE1	AHG1 1/EE1	AHG1 1/EE1	AHG11, 14/EE1	AHG11, 14/EE1
NN寄書				4	2	7	6	9	1				11	19	17	17	23	29	30	31	47	48
Loop-filter													6	11	9	8	9	9	17	14	17	18
Intra													1	1	1	2	2	1			1	1
Inter													1	1	2	1	2	2	2	1	2	2
E2E														2	1	0	1	2	1	2	1	1
Post-filter														1	2	1	3	1	7	9	15	22
Super-resolution													2	2	2	4	6	5	2	4	8	3
Other													2	1		1		1	1	1	3	1

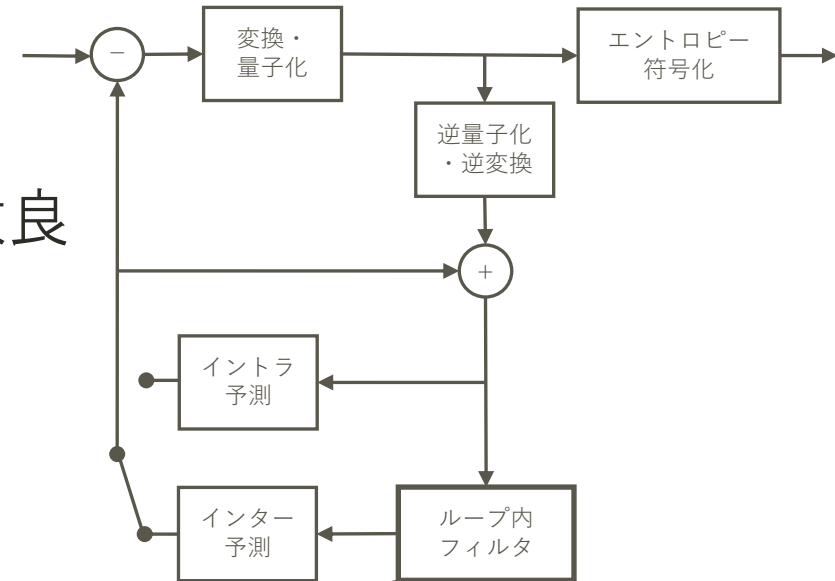
EE1: Neural Network-based Video Coding (2023年1月会合)



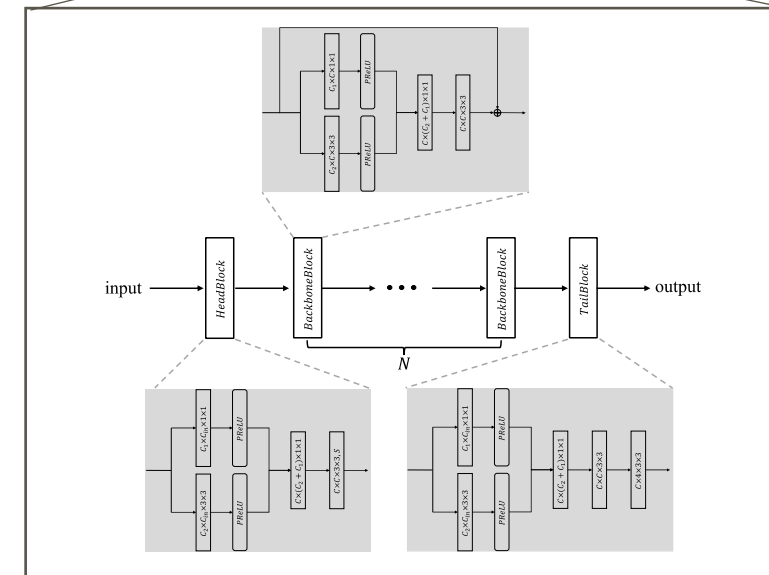
EE1: Neural Network-based Video Coding (Loop-filter)

• JVET-AB0179 (Bytedance)

- 2022/10会合で最も性能が高い方式の一つ
- NCS-1.0 set 1(後述)のResidual blocksの改良
- RA 11.45 % (性能向上約11%)
EncT 200%(約2倍),
DecT 56788% (約570倍)



BD-rate Over VTM-11.0_nvvc-2.0					
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT CPU
Class A1	-10.63%	-19.95%	-21.94%	214%	57805%
Class A2	-12.13%	-23.03%	-25.14%	204%	55842%
Class B	-10.93%	-26.84%	-26.05%	212%	60144%
Class C	-12.31%	-26.44%	-27.39%	173%	52818%
Class E					
Overall	-11.48%	-24.60%	-25.41%	200%	56788%
Class D	-13.96%	-27.50%	-29.42%	165%	49934%
Class F	-5.63%	-15.73%	-15.69%	280%	25531%



EE1: Neural Network-based Video Coding (Loop-filter)

NCS-1.0

VTM-11.0_nnvc-2.0ベースにC++で実装したNeural Network ベースの Loop-Filter 方法 (Set0とSet1の2種類)

Set0: JVET-AA0088 (Tencent)

YUVを同時に処理

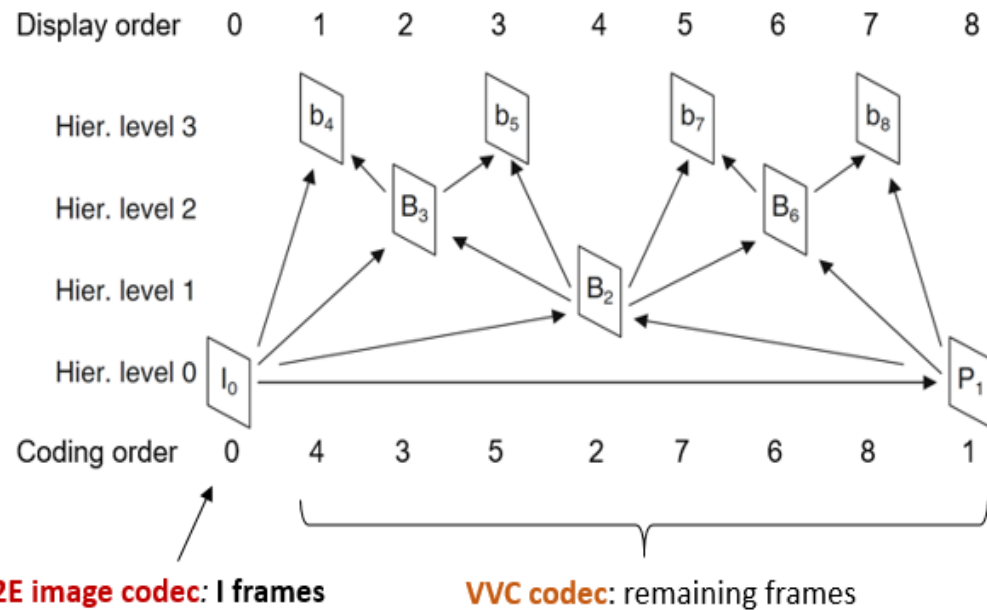
Set1: JVET-AA0111 (Bytedance, Qualcomm, Ericsson)

YとUVを別に処理

	Random access Main10				
	BD-rate Over VTM-11.0_nnvc-2.0				
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT CPU
Class A1	-8.92%	-15.94%	-18.51%	210%	52791%
Class A2	-10.15%	-20.20%	-17.29%	200%	50531%
Class B	-9.01%	-23.08%	-21.77%	209%	51646%
Class C	-9.82%	-21.78%	-22.31%	195%	50009%
Class E					
Overall	-9.44%	-20.73%	-20.37%	203%	51205%
Class D	-11.43%	-23.49%	-23.88%	196%	47037%
Class F	-4.69%	-11.70%	-10.66%	298%	25719%

• JVET-AA0063 (Huawei)

- IピクチャはEnd2End、Pピクチャ、BピクチャはVVCで、ピクチャ単位の切り替え
- End2Endに基づく画像コーデックは、Iフレーム上でVVCよりも優れた符号化性能を提供するため、Inter予測のPフレームとBフレームの符号化性能も改善される。
- RAで4.4%の性能向上



Random Access Test over VTM 14.0			
	Y	U	V
Class A1	-3.7%	-1.3%	-0.5%
Class A2	-5.6%	3.0%	9.5%
Class B	-3.4%	7.6%	4.3%
Class C	-1.5%	14.7%	9.8%
Class E	-9.4%	2.9%	-0.6%
Overall	-4.4%	6.1%	4.8%
Class D	-1.4%	7.1%	6.8%
Class F	-1.5%	13.0%	12.7%

- 前ページJVET-AB0149 (InterDigital)の性能

- 平均3.61%の改善。
- EncT163%、DecT820%(ニューラルネットワーク技術にしては計算量は少ないが従来型に比べるとまだ大きい)

	All Intra Main10							
	BD-rate Over VTM-11.0_nnvc-1.0							
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM	EncT	DecT CPU
Class A1	-4.31%	-3.67%	-4.03%	-4.54%	-3.52%	-3.86%	153%	636%
Class A2	-2.82%	-2.35%	-2.77%	-2.84%	-1.87%	-2.25%	169%	755%
Class B	-3.24%	-2.79%	-3.00%	-3.07%	-2.43%	-2.31%	168%	827%
Class C	-3.20%	-2.78%	-3.01%	-3.25%	-2.25%	-2.23%	165%	993%
Class E	-4.89%	-4.58%	-3.82%	-4.85%	-4.25%	-2.32%	157%	875%
Overall	-3.61%	-3.16%	-3.27%	-3.61%	-2.78%	-2.54%	163%	820%
Class D	-3.00%	-2.68%	-2.88%	-2.87%	-2.04%	-2.30%	165%	1020%
Class F	-2.21%	-1.77%	-1.88%	-2.18%	-1.46%	-1.50%	130%	498%

EE2: Enhanced Compression beyond VVC capability

EE2: Enhanced Compression beyond VVC capability

• ECM(Enhanced Compression Model)

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

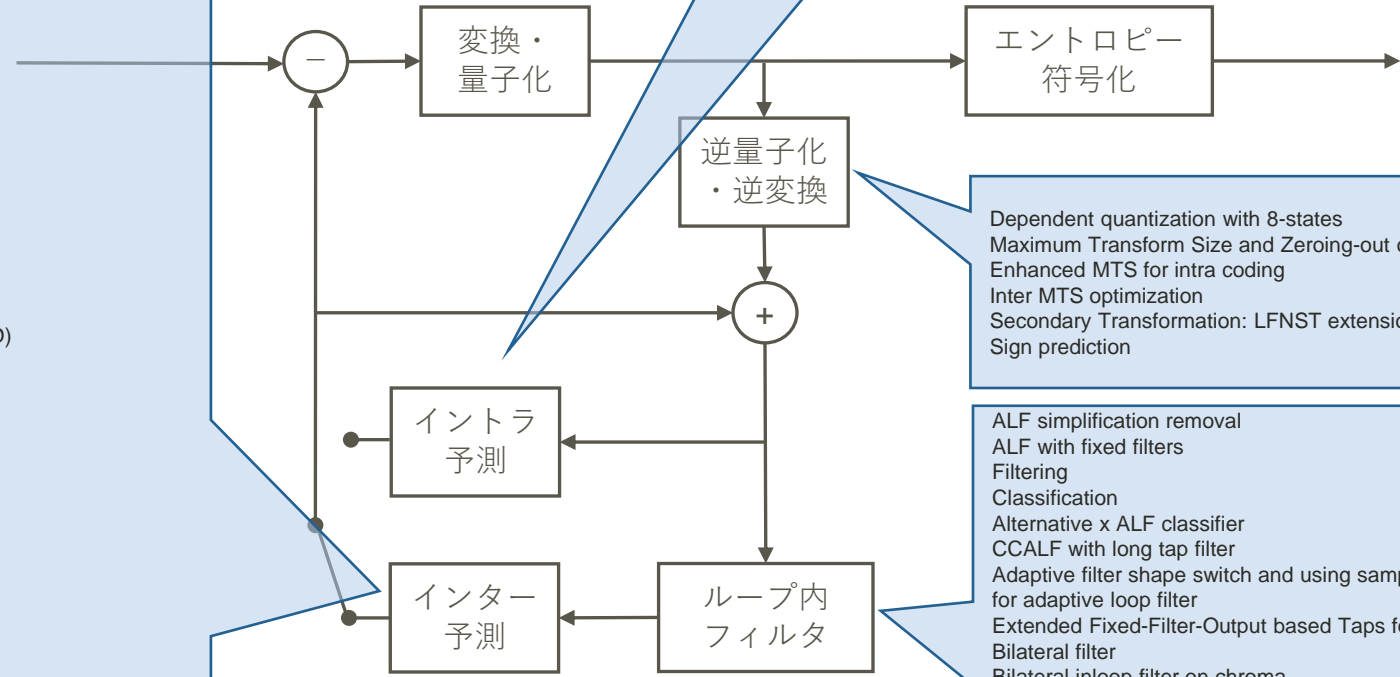
Multi-model LM (MMLM)
Gradient PDPC
Secondary MPM
Reference sample interpolation and smoothing for intra-prediction
Decoder side intra mode derivation (DIMD)
Fusion of chroma intra prediction modes
Intra template matching
Fusion for template-based intra mode derivation (TIMD)
Intra prediction fusion
Combination of CIIP with TIMD and TM merge
Extended multiple reference line (MRL) list
Template-based multiple reference line intra prediction
Convolutional cross-component intra prediction model
Spatial Geometric partitioning mode (SGPM)

Extended precision
Slice-type-based window size
Improved probability estimation for CABAC

Dependent quantization with 8-states
Maximum Transform Size and Zeroing-out of Transform Coefficients
Enhanced MTS for intra coding
Inter MTS optimization
Secondary Transformation: LFNST extension with large kernel
Sign prediction

ALF simplification removal
ALF with fixed filters
Filtering
Classification
Alternative x ALF classifier
CCALF with long tap filter
Adaptive filter shape switch and using samples before deblocking filter for adaptive loop filter
Extended Fixed-Filter-Output based Taps for ALF
Bilateral filter
Bilateral inloop filter on chroma
Cross-Component Sample Adaptive Offset (CCSAO)

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
Template matching based OBMC
History-parameter-based affine model inheritance and non-adjacent affine mode
Sample-based BDOF
Interpolation
Multi-hypothesis prediction (MHP)
Adaptive reordering of merge candidates with template matching (ARMC-TM)
MV candidate type based ARMC
TM based reordering for MMVD and affine MMVD
Regression based affine candidate derivation
Geometric partitioning mode (GPM) with merge motion vector differences (MMVD)
Geometric partitioning mode (GPM) with adaptive blending
Geometric partitioning mode (GPM) with template matching (TM)
GPM with inter and intra prediction
Template matching based reordering for GPM split modes
Bilateral matching AMVP-merge mode
IBC merge/AMVP list construction
IBC with Template Matching
IBC reference area
MVD sign prediction
Enhanced bi-directional motion compensation
Motion compensated picture boundary padding
Block level reference picture list reordering
Reference picture resampling (RPR)
Reconstruction-Reordered IBC (RR-IBC)
IBC merge mode with block vector differences (IBC-MBVD)
Template matching based BCW index derivation for merge mode



動画像符号化技術(1) Inter予測(フレーム間予測)

異なる時刻のフレームに基づいて予測画像を生成する技術

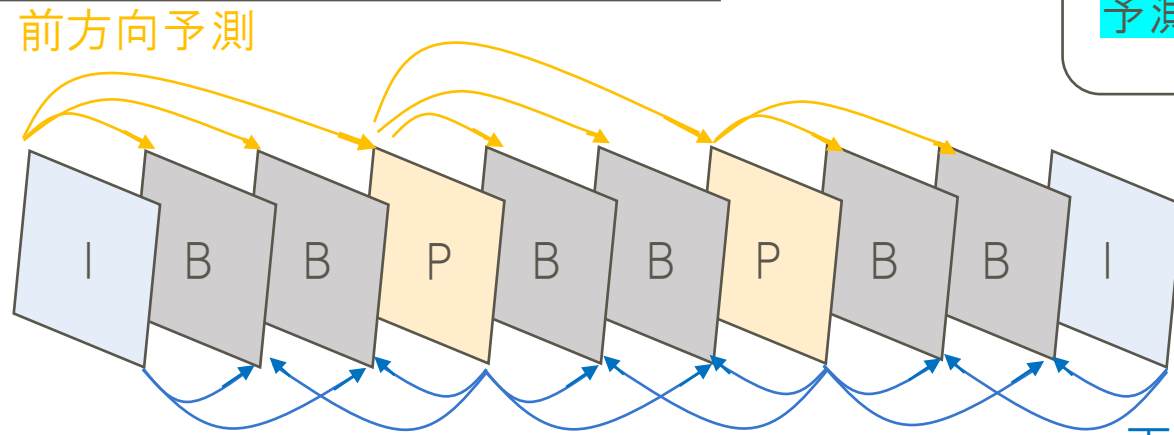
- Iフレーム (Intra-coded Frame) → 次ページ
フレーム間予測を用いずに符号化されるフレーム。イントラフレームやキーフレームとも呼ばれる。

- Pフレーム (Predicted Frame)
前方向予測のみを用いて符号化されるフレーム。
- Bフレーム (Bi-directional Predicted Frame)
前方向予測、後方向予測、両方向予測のうちいずれかを選択して符号化されるフレーム。

ECM(VVC以降)のインター予測技術の例:

- Local illumination compensation (LIC)
- Non-adjacent spatial candidate
- Template matching (TM)
- Multi-pass decoder-side motion vector refinement
- Adaptive decoder-side motion vector refinement
- Template matching based OBMC
-

デコード側での
予測画像導出が増えている

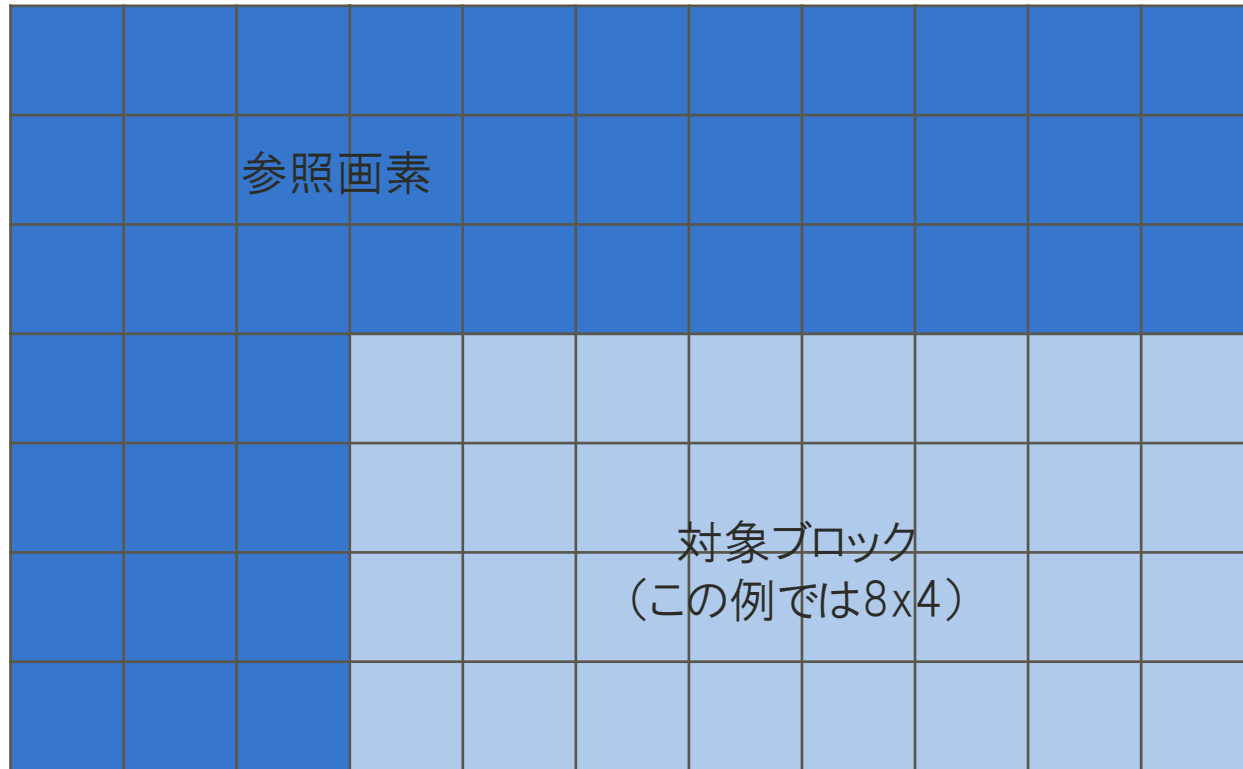


両方向予測

時刻方向

同一時刻のフレームに基づいて予測画像を生成する技術

復号済みの同一フレーム内の参照画素(下図青色)から、対象ブロック(水色)の予測画像を生成する



ECM(VVC以降)のイントラ予測技術の例:

:

Convolutional cross-component model (CCCM)

Multi-model LM (MMLM)

Gradient PDPC

Secondary MPM

Reference sample interpolation and smoothing for intra-prediction

Decoder side intra mode derivation (DIMD)

Template based intra mode derivation (TIMD)

.....

- ・ デコード側での予測画像導出が増えている
- ・ 参照領域が拡大している

EE2: Intra予測提案の例（2023年1月会合）

#	Test	AI					RA				
		Y	U	V	Enc	Dec	Y	U	V	Enc	Dec
Intra prediction											
1.1	Directional planar prediction	-0.04%	0.00%	-0.10%	103%	99%	-0.01%	-0.06%	-0.09%	100%	100%
1.2	Improvements on planar horizontal and planar vertical mode	-0.08%	-0.02%	-0.02%	102%	99%	-0.03%	-0.02%	-0.06%	100%	102%
1.3	Horizontal and vertical planar modes	-0.06%	0.00%	0.00%	107%	100%	incomplete				
1.4	Combination of directional planar prediction methods	-0.06%	-0.04%	-0.04%	104%	100%	-0.02%	-0.02%	-0.11%	99%	100%
1.6	Chroma fusion	-0.03%	-0.38%	-0.40%	101%	100%	0.00%	-0.20%	-0.36%	101%	100%
1.7	Adaptive reference region DIMD	-0.12%	0.03%	0.00%	115%	102%	-0.05%	-0.07%	-0.11%	104%	101%
1.8	Location-dependent DIMD	-0.05%	-0.08%	-0.10%	100%	100%	-0.01%	-0.04%	-0.08%	100%	99%
1.9	TIMD with directional blending	-0.12%	-0.03%	-0.05%	107%	102%	-0.07%	-0.14%	-0.12%	101%	97%
1.10	Optimizing the use of available decoded reference samples	-0.09%	-0.16%	-0.14%	101%	105%	-0.03%	-0.17%	-0.28%	101%	101%
1.11a	Test 1.8 + Test 1.9	-0.16%	-0.09%	-0.07%	106%	102%	-0.07%	-0.13%	-0.14%	100%	98%
1.11b	Test 1.8 + Test 1.9 + Test 1.10	-0.23%	-0.20%	-0.21%	110%	107%	-0.11%	-0.13%	-0.19%	102%	102%
1.11c	Test 1.7 + Test 1.8 + Test 1.9 + Test 1.10	-0.31%	-0.20%	-0.21%	129%	108%	-0.16%	-0.20%	-0.25%	108%	103%
1.12a	Gradient and location based CCCM, 7-tap	-0.04%	-0.56%	-0.59%	100%	100%	0.00%	-0.40%	-0.35%	100%	99%
1.12b	Gradient and location based CCCM, 5-tap	0.02%	-0.17%	-0.27%	99%	99%	0.03%	-0.07%	-0.15%	99%	100%
1.13	CCCM using non-downsampled luma samples with encoder optimization	-0.03%	-0.19%	-0.24%	100%	100%	-0.03%	-0.19%	-0.31%	100%	100%
1.14	CCCM using non-downsampled luma samples	-0.03%	0.01%	-0.04%	101%	100%	0.01%	-0.09%	-0.20%	101%	99%

シャープ
提案

最新探索モデル(ECM-7.0)の性能

- ECM-6.0との比較

- All Intra構造で約1%、Random Access構造で約0.9%の符号量削減

	All Intra Main 10				
	Over ECM-6.0				
	Y	U	V	EncT	DecT
Class A1	-0.57%	-1.41%	-1.14%	121%	104%
Class A2	-1.03%	-2.81%	-3.34%	119%	107%
Class B	-1.00%	-2.22%	-2.00%	120%	110%
Class C	-0.83%	-1.17%	-1.16%	123%	121%
Class E	-1.52%	-1.92%	-1.78%	125%	119%
Overall	-0.98%	-1.90%	-1.86%	122%	112%
Class D	-0.65%	-0.71%	-0.94%	115%	112%
Class F	-0.78%	-2.74%	-2.70%	105%	105%
Class TGM	-1.81%	-5.38%	-4.66%	102%	92%

	Random Access Main 10				
	Over ECM-6.0				
	Y	U	V	EncT	DecT
Class A1	-0.86%	-0.75%	-1.37%	104%	93%
Class A2	-1.08%	-1.93%	-2.16%	103%	98%
Class B	-0.93%	-1.55%	-1.60%	110%	96%
Class C	-0.62%	-0.84%	-0.71%	115%	107%
Class E					
Overall	-0.86%	-1.28%	-1.43%	109%	99%
Class D	-0.58%	-0.79%	-0.92%	115%	105%
Class F	-1.12%	-2.69%	-2.59%	109%	102%
Class TGM	-1.05%	-3.12%	-2.73%	108%	99%

ECM-7.0 の性能

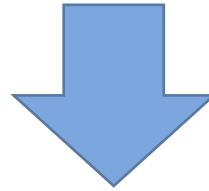
- VVCの参照ソフトウェアVTMとの比較
 - All Intra構造で約9%、Random Access構造で約19%の符号量削減

	All Intra Main 10				
	Over VTM-11.0				
	Y	U	V	EncT	DecT
Class A1	-8.67%	-17.52%	-24.11%	617%	348%
Class A2	-12.88%	-21.39%	-24.64%	576%	348%
Class B	-7.62%	-20.59%	-18.79%	630%	392%
Class C	-8.12%	-12.50%	-12.91%	635%	411%
Class E	-9.72%	-17.97%	-16.91%	619%	435%
Overall	-9.13%	-17.98%	-19.03%	618%	387%
Class D	-6.68%	-10.00%	-10.01%	624%	416%
Class F	-18.45%	-26.68%	-26.24%	444%	397%
Class TGM	-31.47%	-38.80%	-37.65%	386%	367%

	Random Access Main 10				
	Over VTM-11.0				
	Y	U	V	EncT	DecT
Class A1	-19.46%	-21.91%	-29.05%	583%	646%
Class A2	-21.45%	-28.23%	-31.25%	586%	894%
Class B	-17.65%	-28.04%	-25.93%	614%	830%
Class C	-19.26%	-22.56%	-22.39%	605%	892%
Class E					
Overall	-19.20%	-25.39%	-26.67%	600%	817%
Class D	-20.25%	-23.80%	-24.06%	605%	959%
Class F	-22.05%	-29.05%	-29.34%	520%	556%
Class TGM	-30.00%	-35.63%	-35.75%	556%	438%

- JVET と MPEG WG2, WG4 の連携

- VCM (Video Coding for Machines) について連携して活動
 - VCM は画像認識のための映像符号化
 - もともと MPEG WG2 (Requirement) で議論して探索活動を行い、提案募集 (CfP: Call for Proposal) を実施
 - 2022/10 会合で提案募集 17 提案の評価が行われ、VVC アンカーに対して Object tracking で最大 57 %、Instance segmentation で最大 45%、Object detection で最大 39% の同一性能に対する符号化効率改善が確認



- JVET では Video Codec の規格外の技術をまとめた Technical Report (H.Supp.MACVC) 作成する
- MPEG WG4 (Video) で、VCM の標準化に向けた技術検討を行う

- **ITU-T SG16/Q6 及び JVETの活動**

- **標準化活動**

- **H.266/VVC (V3)**
- **H.274/VSEI (V3)**

2023年中に標準
化完了予定

- **技術探索活動 (JVET Exploration Experiment)**

- **EE1: Neural Network Video Coding**
- **EE2: ECM (Enhanced Compression Model)**

- **JVET, WG2, WG4の連携**

- **VCM (Video Coding for Machines)**

SHARP

Be Original.

