



L'impact de la sélection de transformées multiples (MTS) sur les performances de l'encodeur VVC

Sameh Samir, Taheni Damak, Matthieu Saumard, Mohamed Ali Ben Ayed,
Maher Jridi

► To cite this version:

Sameh Samir, Taheni Damak, Matthieu Saumard, Mohamed Ali Ben Ayed, Maher Jridi. L'impact de la sélection de transformées multiples (MTS) sur les performances de l'encodeur VVC. ORASIS 2023, Laboratoire LIS, UMR 7020, May 2023, Carqueiranne, France. hal-04219609

HAL Id: hal-04219609

<https://hal.science/hal-04219609>

Submitted on 27 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'impact de la sélection de transformées multiples (MTS) sur les performances de l'encodeur VVC

The Impact of Multiple Transform Selection (MTS) on the VVC Encoder performance

Sameh Samir^{1,2}, Taheni Damak², Matthieu Saumard³, Mohamed Ali Ben Ayed², Maher Jridi¹

¹ L@bISEN, Vision-AD, Yncréa Ouest, 33 Quater Chemin du Champ de Manœuvre
44470 Carquefou, France.

² Nouvelles Technologies et Systèmes des Télécommunications (NTS'Com), Sfax University, Tunisia.

³ L@bISEN, Vision-AD, Yncréa Ouest, 20 rue Cuirassé Bretagne
29200 Brest, France.

sameh.samir@isen-ouest.yncrea.fr

maher.jridi@isen-ouest.yncrea.fr

Résumé

Le développement des techniques de codage par le module transformée de l'encodeur vidéo s'est considérablement amélioré au cours des dix dernières années, et la norme de codage vidéo polyvalent (VVC) de nouvelle génération comprend désormais un certain nombre de nouveaux outils de transformées efficaces. Parmi ces outils, Multiple Transform Selection (MTS), une nouvelle approche pour l'unité de transformée. Cet article vise à fournir une étude statistique détaillée du nouvel outil de sélection de transformée multiple (MTS) dans le standard VVC. L'étude met également en évidence l'impact du MTS sur les performances du codeur à travers des résultats expérimentaux réalisés sur le logiciel de référence VVC VTM-14.

Mots Clef

Codage vidéo polyvalent (VVC), Encodeur vidéo, Codage par transformation, MTS, statistical analysis.

Abstract

The development of transform coding techniques has improved significantly over the past ten years, and the next generation Versatile Video Coding (VVC) standard now includes a number of efficient new transform tools. Among these tools, Multiple Transform Selection (MTS), a new approach for the transform unit. This paper aims at providing a detailed study of transform coding development with the new tool multiple transform selection (MTS) in VVC standard. The study also highlights the impact of the MTS on the encoder performance through experimental results made on VVC reference software VTM-14.

Keywords

Versatile video coding (VVC), video encoder, transform coding, MTS.

1 Introduction

With the rapid development of multimedia applications in recent years, such as the digital cinema, video surveillance, telemedicine, videoconferencing, live video streaming etc, the quantity of data exchanged through the internet continues to increase. According to Cisco Systems, the share of the video in global data traffic is already about 80% and is continuing to grow[1]. In addition, the data rates necessary to deliver this content are still height and this is due to steadily growing resolution of the 4k/8k ultra high definition (UHD) videos or 360° videos such as those for immersive and augmented reality. This poses major challenges with respect to broadband capacity and multimedia storage, which illustrates the need of efficient video signal compression that leaves the quality unchanged. In order to meet this demand, two major generations of video coding standards were finalized by the ITU-T Video Coding experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG)[2] including the standard Heigh Efficiency Video Coding (H265/HEVC). The H265/HEVC was developed in 2013 providing a 50% bitrate reduction compared with it's predecessor Advanced Video Coding Standard (H264/AVC)[3]. However, despite the major advances introduced by the H265/HEVC standard, it is no longer able to meet the potential needs of the industry, and future application scenarios require more efficient compression for the storage and transmission of data.

In July 2020, the Versatile Video Coding Standard (H266/VVC)[4] was developed by the ITU-T and ISO/IEC under the joint video experts Team (JVET) aiming to double the coding efficiency compared to HEVC for the same visual quality. Versatile Video Coding (VVC), has enhanced coding complexity on both the encoder and decoder sides. This increase in complexity is the result of several

coding techniques that were suggested to promote coding effectiveness. The Multiple Transform Selection (MTS) concept, a novel method for the transform unit, is one of these tools[5]. Multiple Transform Selection (MTS) requires the best of various transforms to be obtained via the Rate-Distortion Optimization (RDO) process and involves three transform types : the DCT type-II, DST type-VII and DCT type-VIII which are used to both inter and intra picture coded blocks.

This paper provides an overall description of the MTS tool development in transform coding, illustrates how this tool is implemented in VVC standard and analyzes its contribution on the encoder performances through a statistical study on the reference software VTM-14.

The remainder of this paper is organized as follows: Section 2 presents the VVC transform unit background while going into more details concerning the multiple transform selection tool, Section 3 lays out the impact of MTS on the VVC encoder through a statistical analysis of different type of transform. Finally, this article is concluded with an outlook in section 4.

2 Multiple Transform Selection (MTS)

The VVC is a block-based hybrid video codec. A video signal is first separated into smaller blocks for the purpose of intra- or inter-picture prediction. The remaining residual signal is then transformed and quantized. Finally, the quantized coefficients, the parameters of partitioning and prediction are entropy coded.

At the transform step, the residual between the original image and the predicted image is decorrelated with the discrete cosine transform (DCT) or the discrete sine transform (DST). These tools are refined and shaped into their final design in VVC, and the most significant advancements in these approaches may be split into three categories:

1. Primary Transform: is specified as separable transform and include five different combinations of transform types including the multiple transform selection (MTS) that will be detailed in this paper.
2. Secondary Transform: is the Low-Frequency Non-separable Secondary Transform (LFNST)[6] is a non-separable transform that is applied to the top-left low-frequency region of primary transform coefficients and is used for intra coded blocks that employ DCT-2 as their primary transform.
3. Transform partitioning: is presented for inter-picture predicted coding units in VVC , only a subpart of the residual block is coded in this transform mode[7].

In the following, the primary transform, will be described with technical details with reference to the test model VTM-14.0 software for VVC.

In the transform step, in addition to transform skip which is another transform mode that bypasses the transform stage

entirely and directly quantise and codes the remaining samples in the spatial domain [8], multiple transform combinations are applied to residual blocks according to several factors such as the size of the block, in fact, for blocks sizes equal to 64, only the DCT-2 is used. The first 32 low frequency coefficients are retained and the others become zero. This technique is called the zeroing out and explained in figure 1.

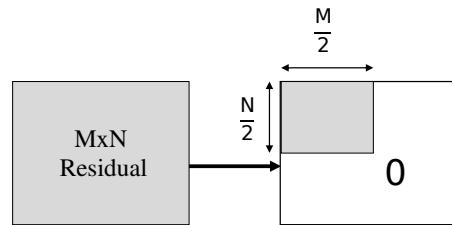


Figure 1: Zeroing out technique applied on a Residual Block of size $M \times N$.

For blocks size less or equal to 32, five combinations of transform types are applied, the conventional DCT2-DCT2 and four new MTS combinations which are DST7-DST7, DST7-DCT8, DCT8-DST7 and DCT8-DCT8. While doing the MTS transform combinations, zeroing out technique is applied where the 16 low frequency coefficients are kept and the high frequency coefficients are zeroed out. The primary transform presents two kinds :

- The Explicit MTS that can be applied to both intra or inter coded blocks and use the four previous MTS combinations.
- The Implicit MTS which is applied to only intra coded blocks and use in addition to the conventional DCT2-DCT2 and the DST7-DST7 combinations, the DCT2-DST7 and DST7-DCT2 combinations.

In figure 2, Examples of implicit MTS derivations are illustrated for various block sizes.

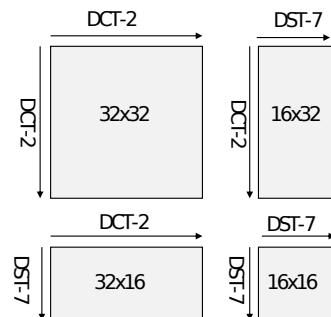


Figure 2: Examples of transform type selection for vertical and horizontal transforms based on the implicit MTS.

Choosing which MTS combination to apply for each implicit or explicit multiple transform selection depends on several factors such :

- Intra or Inter coded blocks,
- The sequence parameter Set's flags that controls the MTS,
- The MTS index defined in the reference software VTM-14 as « MTSIdx » and indicates the Horizontal and Vertical transform types represented respectively as « trTypeHor » and «trTypeVer ».

The transform type selection for vertical and horizontal transforms is resumed in Table 1.

Table 1: Identification of transform types depending on MTS index

MTSIdx	TrTypeHor	TrTypeVer
0	DCT2	DCT2
1	Transform Skip	
2	DST7	DST7
3	DCT8	DST7
4	DST7	DCT8
5	DCT8	DCT8

Table 2: Summary of the combination of different tools

MTS Flag	MTS Modes	Block intra	Block inter
0	Explicit MTS	N	N
	Implicit MTS	N	N
1	Explicit MTS	Y	N
	Implicit MTS	N	N
2	Explicit MTS	N	Y
	Implicit MTS	Y	N
3	Explicit MTS	Y	Y
	Implicit MTS	Y	N

From the configuration file, the MTS modes can be enabled or disabled. In fact, there are four flags ranges from zero to three controlling the MTS operation. When MTS flag is equal to zero, the MTS is disabled, in this case only the DCT2-DCT2 or transform Skip is enabled. When the MTS flag is set to one, the explicit intra MTS is enabled i.e. the explicit MTS is applied for intra coded blocks. When MTS flag is equal to two, whether the explicit inter MTS or the

implicit intra MTS is enabled, in this case, we apply respectively explicit MTS to inter coded blocks and implicit MTS to Intra coded blocks. The last flag is used to enable intra and inter MTS, in this case, only explicit MTS is applied for both intra and inter coded blocks. The mapping to the MTS modes are resumed in table 2 where Y/N means the associated MTS modes in the row can/cannot be applied with the type of the residual block in the column and depending on the initial MTS flag.

After applying the multiple transform selection, rate-distortion (RD) checking will be performed on combinations of DST-VII and DCT-VIII after applying DCT-II. With minimal RD costs, the VVC can determine the optimal transform during the coding unit (CU) partition and mode decision stages.

3 STATISTICAL ANALYSIS

In this section the impact of Multiple Transform Selection on compression performance will be studied from a statistical analysis and through these following two subsections.

3.1 VTM reference Software and Experimental conditions

VVC test model (VTM) is the name of the reference software for VVC[9]. The HEVC test model (HM) was the original source of its development[10]. The block level functionality was largely reworked to support the new coding tools in VVC. The software makes more frequent use of the C++ standard library and the more recent C++ version. The reference implementation serves as a framework for coding experiments and tool evaluation. To evaluate the proposals during VVC development, some common test conditions (CTCs) [11] were specified. Three mandatory test conditions are included in the CTC definition, which represent All-Intra, Random access, and Low-delay settings. Because of its much greater usage in applications, the random access (RA) situation is regarded as more essential than the others, that's why the experiments in this article will focus on random access configuration.

In the experiments, 6 video classes with 3 videos per class were employed, including classes A1 and A2 that contain 4K-UHD sequences (3840x2160), class B contains HD sequences (1920x1080), class C contains SD sequences (832x480), class D contains quarter-SD sequences and class E features video conferencing content (1280x720). These sequences use 8 or 10 input bit depth, 4:2:0 ratio of luminance to chrominance samples and 60 or 50 frame rate per second knowing that most sequences are 5 or 10 seconds long. For each sequence, 4 GOP of 16 frames were tested and the tested quantization parameter (QP) was adjusted at 32.

3.2 Performance Analysis of Multiple Transform Selection

In this subsection, the coding performance of MTS is studied using VTM-14, in addition to the test conditions pre-

Table 3: Coding gain of bit-rate, Run-time and PSNR of MTS=3 comparing to MTS=0 for Random Access configuration and QP=32

Sequences Class	Sequences	ΔB	Average ΔB	ΔT	Average ΔT	$\Delta PSNR$	Average $\Delta PSNR$
Class A1	Tango2	1,47%	1.27%	-8.5%	-4.7%	-0.01%	-0.01%
	FoodMarket4	1,19%		-2.9%		-0.01%	
	Campfire	1,16%		-2.8%		-0.01%	
Class A2	DaylightRoad2	1,39%	1.05%	-5.1%	-8.24%	-0.01%	-0.04%
	CatRobot1	1,16%		-6.22%		-0.01%	
	ParkRunning3	0,60%		-13.4%		-0.10%	
Class B	BQTerrace	1,02%	0.93%	-9.7%	-8.66%	-0.04%	-0.05%
	BasketballDrive	0,65%		-10.7%		-0.04%	
	Cactus	1,14%		-5.6%		-0.07%	
Class C	BasketballDrill	0,12%	0.11%	-10.8%	-10.3%	-0.05%	-0.05%
	BQMall	0,10%		-9.8%		-0.05%	
	PartyScene	0,13%		-10.5%		-0.06%	
Class D	BasketballPass	0,93%	0.91%	-12.23%	-11.9%	-0.04%	-0.03%
	BQSquare	0,9%		-11.8%		-0.03%	
	RaceHorses	0,9%		-11.7%		-0.04%	
Class E	FouPeople	1,77%	1.77%	-6.04%	-6.3%	-0.02%	-0.02%
	Johnny	1,76%		-7.01%		-0.03%	
	KristenAndSara	1,8%		-6.01%		-0.02%	

sented in the previous section, LFNST parameter is initially turned off for this experiment. The coding performance is evaluated by comparing two MTS cases which are MTS = 3 means that the MTS is active and MTS = 0 means that the MTS is disabled, and to evaluate the video's quality, the peak signal-to-noise ratio (PSNR) weighted average[12] which provides information about the video's quality is used.

1. Effect of MTS on bitrates, PSNR and encoder run-time

To evaluate the effect of MTS on the bitrate, the run-time and the PSNR value, the following ΔX measurement is used:

$$\Delta X = \frac{X_{MTS0} - X_{MTS3}}{X_{MTS0}} \quad (1)$$

ΔX corresponds to the ΔT , ΔB or $\Delta PSNR$ in order to calculate respectively the bit-rate gain, the run-time gain or the PSNR gain when enabling/disabling MTS.

X_{MTS3} corresponds to the value of bit-rate, run time or PSNR when MTS is active (MTS = 3) and X_{MTS0} corresponds to the initial case where MTS is disabled (MTS = 0).

The bit-rate savings, the encoding software run-time and the PSNR are reported in table 3 for the case of active MTS compared to disabled MTS case.

For RA configuration, the average of bit-rate gain reaches 1.27% for classA1 when turning on the MTS, 1.05% bit-rate savings is noticed also for classA2 while class D represents an average of 0.91% bit-rate reduction. Thus, enabling MTS provides an encoding performance benefit through bit-rate savings.

However, table 3 lists different negative run-time averages. As an example, class A1 and class A2 represents -4.7% and -8.24% respectively. In addition, class B and C show that the encoder run-time is approximately increased by 8% and 10%. It is shown that the encoding time can increase by more than 12% for class D sequences. These results highlights the fact that enabling MTS increases the encoding run-time, which is expected since multiple transform selection tool introduce additional transform kernels of the various MTS combinations added in the transform module.

The encoding video quality variation is described by $\Delta PSNR$ in table 3, enabling MTS increases the video's quality to reach an average of 0.01% for class

A1 for example and 0.05% for class B. As a result we can say that MTS improves slightly the video's quality.

2. Statistical analysis of transform type of MTS

In order to emphasize the impact of MTS on transform module and more precisely on the choice of transform combinations, multiple statistics were made for 18 video sequences for the 6 classes (A1, A2, B, C, D and E) and have shown the number of transforms used by the VVC encoder including:

- Transform Skip,
- DCT2-DCT2,
- MTS explicit : DST7-DST7, DST7-DCT8, DCT8-DST7 and DCT8-DCT8,
- MTS implicit : DCT2-DST7,DST7-DCT2,
- Transform-ICT which is an integer transform used whether MTS is enabled or not and the case where the residual block is not transformed.

Different pie charts were created to quantify the percentages of each transform type. For this study, we will present an example for 2 videos of class A2 and B, in random access configuration and with a quantification parameter (QP) equal to 32.

According to figure 3, we notice that when MTS is disabled, a significant percentage of the combination DCT2-DCT2 appears and reaches 35% for day-lightRoad2 video of the class A2 while the absence of the other combinations of the MTS . On the other hand, when MTS is active, a remarkable percentage of 12% which appears for the DST7-DST7 combination as shown in the figure in addition of the DCT8-DST7 combination which reaches 1%.

Figure 4 shows another example that details the percentages of MTS transform combinations for Cactus video from class B. According to figure 4, Cactus video presents also an important percentage of DCT2-DCT2 combination appears (39%) in addition to the case where there is no transform (58%). However, if MTS is enabled, DCT2-DCT2 percentage is decreased, and new MTS transform combinations appears like 15% for DST7-DST7 and 3% for DCT8-DST7, we notice also 1% for an implicit MTS combination (DST7-DCT2).

We can deduce from the results of the used transforms for the test videos that enabling MTS causes the appearance of different transform combinations, which are presented with variable percentages, these percentages are generally low for the combinations: DST7-DCT8, DCT8-DCT8, DCT2-DST7, DST7-DCT2, transform Skip, and ICT transform and which fluctuates between 0% and 5% for the 18 videos. On the other hand, the percentages of DCT2-DCT2,

DST7-DST7, and the no transform case are significant and range from 10% to 65% for the 18 videos.

4 CONCLUSIONS

The new tool of The VVC encoder's transform module, Multiple transform selection has been described in this paper through the reference VTM-14 software. After a statistical analysis done on several videos, the multiple transform selection tool was more detailed by presenting, first, its impact on the encoder performance, through the bitrate, the PSNR and the encoder run-time and then, this impact was more thoroughly examined on transform types and sizes. Through the testing results of The MTS tool, this statistical study proved one of the sources of the new VVC encoder's complexity in comparison to its predecessors.

Acknowledgment

This work is funded by the PHC Maghreb ECO VVC program, and within a co-supervised thesis between Higher Institute for Electronics and Numerique of Nantes (ISEN) France, and New technologies and telecommunication systems Laboratory (NTS'COM) of Sfax, Tunisia.

References

- [1] V. Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," *White paper*, vol. 1, no. 1, 2018.
- [2] H. E. V. Coding, "Recommendation itu-t h. 265," *International Standard ISO/IEC*, pp. 23008–2, 2013.
- [3] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (hevc)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, 2012.
- [4] B. Bross, J. Chen, S. Liu, and Y.-K. Wang, "Versatile video coding (draft 10)," *ITU-T and ISO/IEC JVET-S2001*, 2020.
- [5] B. Bross, J. Chen, J.-R. Ohm, G. J. Sullivan, and Y.-K. Wang, "Developments in international video coding standardization after avc, with an overview of versatile video coding (vvc)," *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463–1493, 2021.
- [6] M. Koo, M. Salehifar, J. Lim, and S.-H. Kim, "Low frequency non-separable transform (lfnst)," in *2019 Picture Coding Symposium (PCS)*, pp. 1–5, IEEE, 2019.
- [7] X. Zhao, S.-H. Kim, Y. Zhao, H. E. Egilmez, M. Koo, S. Liu, J. Lainema, and M. Karczewicz, "Transform coding in the vvc standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3878–3890, 2021.

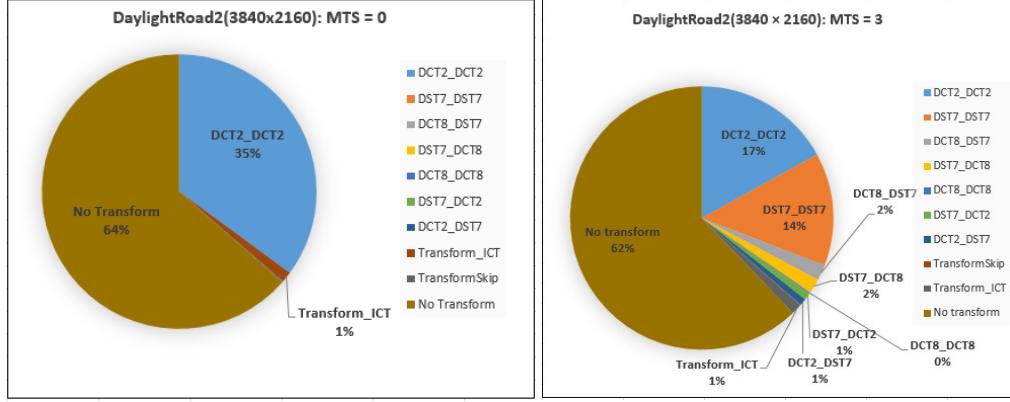


Figure 3: Pie Charts representing the percentages of transform types for DaylightRoad2 video for MTS = 0 and MTS = 3 and QP=32.

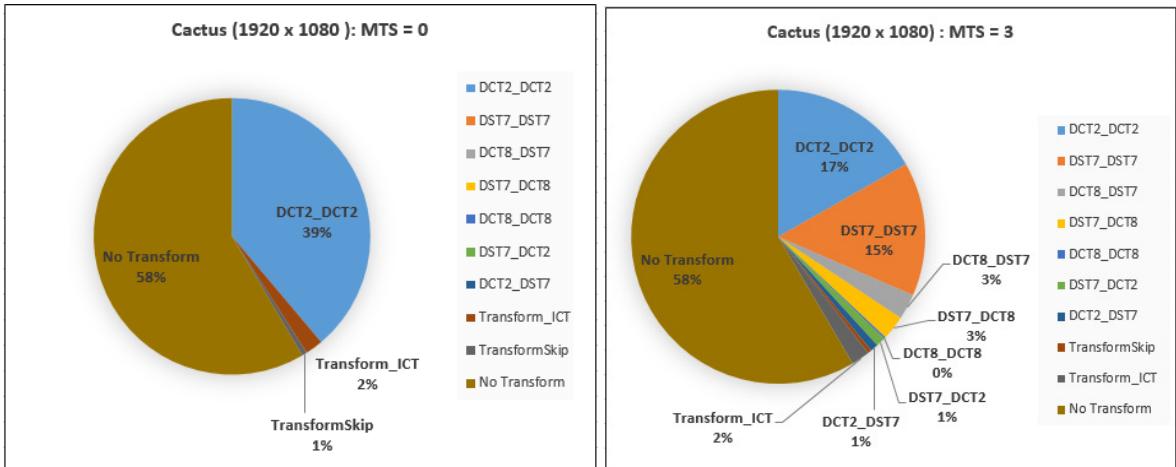


Figure 4: Pie Charts representing the percentages of transform types for Cactus video for MTS = 0 and MTS = 3 and QP=32.

- [8] B. Bross, J. Chen, J.-R. Ohm, G. J. Sullivan, and Y.-K. Wang, “Developments in international video coding standardization after avc, with an overview of versatile video coding (vvc),” *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463–1493, 2021.
- [9] “Vtm vvc reference software.. accessed:[online] available:<https://vcgit.hhi.fraunhofer.de/jvet/vvcsoftware-vtm>,” Mar. 31,2021.
- [10] “Hm hevc reference software. accessed:[online]. available: <https://vcgit.hhi.fraunhofer.de/jvet/hm>,” Mar. 24,2021.
- [11] F. Bossen, J. Boyce, X. Li, V. Seregin, and K. Sühring, “Jvet common test conditions and software reference configurations for sdr video,” *Joint Video Experts Team (JVET) of ITU-T SG*, vol. 16, pp. 19–27, 2019.
- [12] W.-J. Chien, J. Boyce, W. Chen, R. Chernyak, K. Choi, R. Hashimoto, Y. Huang, H. Jang, S. Liu,

and D. Luo, “Jvet ahg report: Tool reporting procedure (ahg13),” *JVET document N*, vol. 13, 2020.