STATISTICAL ANALYSIS OF INTER CODING IN VVC TEST MODEL (VTM)

Yiqun Liu[⋆†] *Mohsen Abdoli*[⋆] *Thomas Guionnet*[⋆] *Christine Guillemot*[†] *Aline Roumy*[†]
[⋆] ATEME, Rennes, France
[†] INRIA, Rennes, France

ABSTRACT

The promising compression efficiency improvement of Versatile Video Coding (VVC) compared to High Efficiency Video Coding (HEVC) [1] comes at the cost of a non-negligible encoder-side complexity. The largely increased complexity overhead is a possible obstacle towards its industrial implementation. Many papers have proposed acceleration methods for VVC. Still, a better understanding of VVC complexity, especially related to new partitions and coding tools, is desirable to help the design of new and better acceleration methods. For this purpose, statistical analyses have been conducted, with a focus on Coding Unit (CU) sizes and inter coding modes.

***Index Terms*—** Versatile Video Coding, Inter Coding, Rate Distortion Optimization, Complexity Analysis

1. INTRODUCTION

Standardization of VVC in 2020 has brought significant improvement to the capacity of video compression in terms of bitrate saving. Precisely, it offers 50% compression efficiency [2] compared to one of the most efficient video compression standards HEVC. This improvement by VVC is principally owed to newly adopted coding techniques. Most decoder devices could afford the additional complexity brought by these novel coding techniques taking into account current hardware capacities of these devices. Precisely, studies in [3] has shown that the relative decoder complexity of VVC is from 150% to 200% compared to HEVC in different configurations.

On the contrary, it is far from affordable for real time application for VVC encoder side since industrial encoding applications have strict limitations in terms of resources and execution time. Tests in [4] on VVC reference software VVC Test Model (VTM) 7 show that the encoding time of VVC is 5x, 7x and 37x times of HEVC encoding in Low-Delay (LD), Random-Access (RA) and All-Intra (AI) configurations respectively. Hence, it is vital to develop acceleration algorithms or methods to largely reduce the encoding complexity while preserving the majority of encoding efficiency. Papers of complexity analysis could help researchers to have beforehand a clear understanding of what is happening inside a VVC encoder (*e.g.* VTM), and what potentially interests them for their design process of acceleration method.

Various papers have contributed to the complexity analysis of VVC. In [5], a detailed complexity analysis based on VVC intra prediction tools has been performed. Pakdaman *et al.* in [3] have broke down the encoding process into encoding modules such as motion estimation, intra prediction, entropy coding etc. and then analysed the complexity partition of modules in multiple encoding configurations. [6] reviews complexity aspects of the different modules of the VVC standard and provide a complexity breakdown of these modules in a more precise way. In [4], VVC and HEVC are compared in terms of rate-distortion and complexity analysis. These aforementioned papers present complexity analysis at the level of encoding modules for inter coding. Our paper is the first to provide an analysis from CU sizes and coding modes perspective for inter coding in VVC.

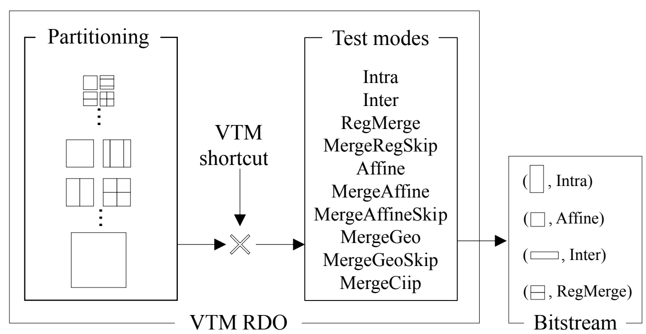

Fig. 1. High-level view of the RDO process involving partitioning, test modes and possible VTM shortcuts.

In this paper, a statistical analysis of the Rate-Distortion Optimization (RDO) process in inter coding of VTM-15.0 is presented. The main focus is put on the statistics of two factors: CU sizes and inter coding modes. The goal is to provide useful information for related works aiming at speeding up inter coding in VVC. The rest of this paper is organized as follows. Section 2 presents a summary of VVC specification in terms of available block sizes and coding modes. In Section 3, all statistical observations are presented, which are later analyzed and concluded in Section 4.

2. RDO OF INTER CODING IN VTM

To depict the RDO process in VTM, there exists numerous coding parameters such as Intra Prediction Mode (IPM) of intra coding, Motion Vector (MV) representation mode and choice of transform *etc*. Nevertheless, if we ignore these trivial parameters, the RDO process could be described as the search for the best trade-off between bit rate and distortion. More precisely, this search is executed on different coding modes of different CU sizes. Therefore, in this paper the statistics of block sizes and coding modes are jointly considered.

As presented in Fig.1, various CU sizes are the result of partitioning in the RDO process. The partitioning consists of splitting the CU of size 128x128 recursively by five split modes, namely Quaternary Tree (QT) split, Horizontal Binary Tree (HBT) split, Vertical Binary Tree (VBT) split, Horizontal Tenary Tree (HTT) split, Vertical Tenary Tree (VTT) split. Comparing to codec HEVC in which only QT is available for partitioning, the added directional splits gives rise to a larger variety of CU sizes. In VVC block size is authorised if its widths and heights are any power of two between 4 and 128, except for sizes 128×4, 128×8, 128×16 and 128×32. It is worth-noting that same CU size could be obtained by different series of split modes.

For each CU, eleven coding modes are available. Two of these modes, namely hash-based inter prediction and palette modes are not enabled in the Joint Video Exploration Team (JVET) Common Test Condition (CTC). Hence, they are not included in the analysis of this paper. Three principal modes coding modes are intra ("intra"), inter mode with new MV ("me") and inter mode with MV of multi resolutions ("aff"). The remaining modes are merged modes indicating the coded MV are shared among neighboring blocks. These modes could be then divided into two groups of skip modes and non-skip modes. The residual would not be encoded and transmitted under skip mode. The non-skip modes include regular merge ("mrgreg"), merge with geometric partitioning ("mrggeo"), merge with Combined Intra-Inter Prediction (CIIP) ("mrgciip") and merge with affine ("mrgaff"), while the merge modes consist of regular merge with skip("mrgregskp") merge with skipped geometric partitioning ("mrggeoskp") and finally merge with skipped affine MV ("mrgaffskp").

Though VVC is computationally expensive, JVET group has already adopted diverse shortcuts or conditional early exits as presented in [7] in VTM. We have deactivated existing shortcuts in VTM-15.0 and evaluated its performance. The assessment was that the complexity will increase by 138%. Moreover, the performance of the tested encoder (*i.e.* without shortcuts) is 0.76% better than the reference encoder (*i.e.* with shortcuts), in terms of BD-rate. This trade-off might be interpreted as an indicator that the shortcuts in VTM are efficient in terms of identifying useless tests and partitioning depths. Many of shortcuts are based on history of the tested split modes. Nonetheless, aspects of CU sizes and coding modes are overlooked.

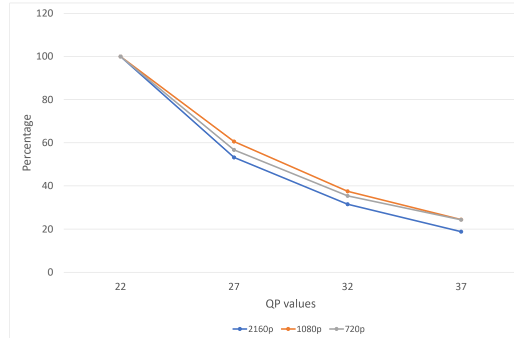

Fig. 2. Encoding time in different QPs comparing to QP 22

3. STATISTICS

Our major purpose in the following analysis is to find the CU sizes and/or coding modes with relatively high complexity occupation and low selection rate in RDO process. From the perspective of encoder acceleration, CU sizes or coding modes with higher complexity portion and significantly lower selection rate are more favorable to the design of block size/coding mode based acceleration rules. The size or mode with larger complexity portion has more potential in accelerating. Lower selection rate indicates it is less likely to make wrong decisions when skipping the RDO of current CU size or mode.

All our experiments and analyses are conducted on the first 64 frames of CTC sequences in RandomAccess Group Of Picture 32 (RAGOP32) configuration in which intra frames are excluded. Exceptionally, Fig.2 is based on sequences in Class A, B, E of CTC.

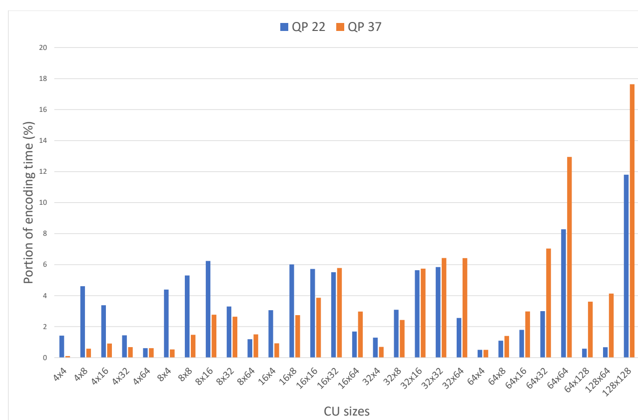

Fig. 3. Complexity distribution for block sizes in QP22 and QP37

The encoding complexity for Chrominance channel only accounts for a small part comparing to Luminance channel. Thus we focus on Luminance channel in the remaining of the paper. From a high-level view, the encoding complexity of VTM significantly depends on the selected Quantization Parameter (QP). Particularly, larger QP values tend to have faster encoding with VTM. Figure 2 is obtained by measuring encoding times of sequences of resolution 2160p, 1080p and 720p in QP 22, 27, 32 and 37. Then the average ratio between the encoding time of each QP and that of QP 22 is calculated. It shows that encoding time at QP 22 could be five times as much as QP 37.

Fig.3 shows the percentage distribution of encoding time spent on different block sizes in QP 22 and QP 37. Besides the fact that the overall encoding time is higher for QP 22, it can be observed that relatively higher portion of the time in QP 22 is passed on smaller block sizes. This could partly be explained by existing shortcuts in VTM disallowing excessively small blocks in QP 37. We could declare that larger block sizes are in general more crucial to speeding up the partitioning process, especially block 64x64 and 128x128 which take in total from 20% complexity in QP 22 to 30% in QP 37.

In another test, the selection percentage of different block sizes are computed. This metric is defined as the ratio between the total number of times it is selected and the total number of times a block size is tested. Fig. 4 shows the values of this metric in QP 22 and 37. As we can see from this figure, larger CU sizes correspond to larger selection rates compared with smaller blocks. Another worth-noting phenomenon is that the selection rate of 128x128 increase dramatically from 14% in QP 22 to 37% QP 37.

QP 22

H\W	4	8	16	32	64	128
4	0.25	0.4	0.49	0.4	0.27	NA
8	0.5	1	0.84	0.86	0.64	NA
16	0.77	1.15	1.64	1.58	2.21	NA
32	0.78	1.48	2.03	3.91	4.64	NA
64	0.48	0.9	1.71	3.89	11	5.61
128	NA	NA	NA	NA	5.87	15.73

QP 37

H\W	4	8	16	32	64	128
4	0.25	0.29	0.42	0.38	0.14	NA
8	0.35	0.61	0.7	0.8	1.35	NA
16	0.68	0.97	1.31	1.33	4.96	NA
32	0.73	1.47	1.67	3.11	4.24	NA
64	0.47	1.49	2.73	3.84	8.5	4.51
128	NA	NA	NA	NA	5.25	41.5

Fig. 4. Selection rate for different block sizes

Combining the above figures, it is observed that CU sizes such as 16x8, 8x16 and 16x16 are sizes with low selection rate and high complexity. For example, 16x16 blocks has the same level of complexity, while its selection rate is half of 32x32 blocks in QP 22.

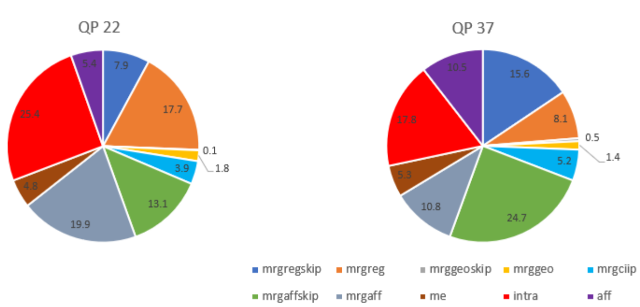

Fig. 5. Pie chart of complexities of inter coding modes

To take one step further in the statistical analysis of inter coding, we present how different inter coding modes are involved in RDO search. The first experiment in Fig.5 present the distribution of encoding time at inter coding mode level in QP 22 and QP 37.

In general, intra, mrgaff, mrgaffskip mrgreg and mrgregskip are main contributors to encoding time. Fig.6 provide selection rates as the ratio between number of selected inter coding modes and number of tested modes. We could remark that mrgaff, mrgaffskip and mrgreg these three modes have minor selection rates though they are accounting for nearly half of the complexity.

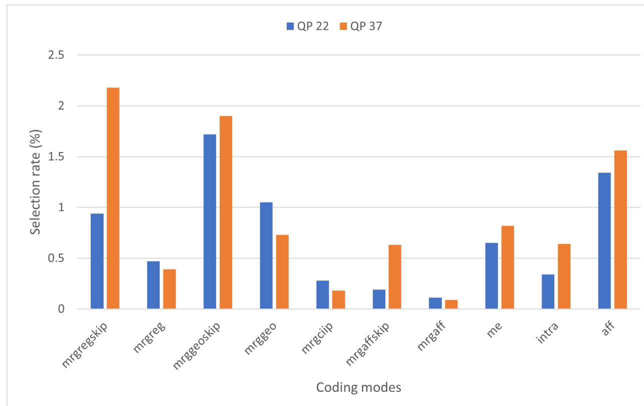

Fig. 6. Selection rate of inter coding modes

Fig.7 shows the distribution of inter modes for encoded CUs of different sizes. The fact that the aforementioned three coding modes are less chosen could also be proved by this figure. We could find out that the number of mrgregskip is dominant for most CU sizes and that the number of these three modes are relatively small, which is coherent with Fig.6. From Fig.7 we could also observe that smaller blocks tend to be encoded with intra mode. Another remark is that the skip modes are more frequently selected for larger CUs. It is probably because the residual of larger CUs is more expensive to be encoded. In addition, mrgreg mode have higher chance to be selected in smaller QP which is in contrast to mrgaffskip and mrggeoskip. For some CU sizes, we could make shortcuts or condition for early termination for RDO of inter modes which are rarely selected to speed up the encoding process, such as mrgaff mode with merely 1.9% selected for bloc 128x128.

4. ANALYSIS AND CONCLUSION

In this study, complexity analysis of CU sizes and inter coding modes has been combined with selection rate analysis. From CU size perspective, block sizes with high complexity generally correspond to high selection rate. Block sizes 128x128 and 64x64 are responsible for one third of the complexity.

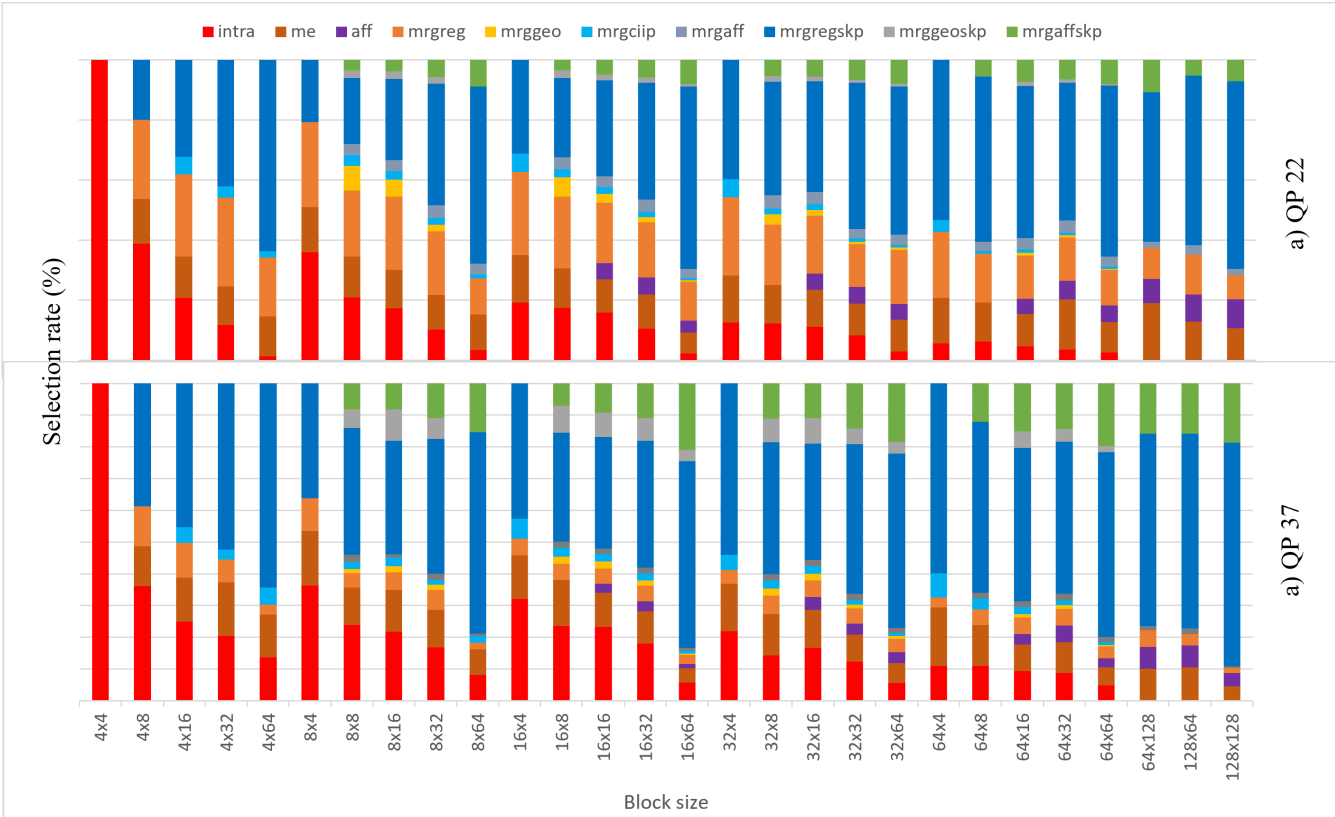

Fig. 7. Stacked chart of selected inter modes in different block sizes

In addition, block sizes like 16x8, 8x16, 16x16 exhibit relatively low selection rate while requiring a significant share of the overall complexity. Therefore, these block sizes are relevant targets for acceleration algorithms. From coding mode perspective, mrgaff, mrgaffskip and mrgreg tend to be less likely to be selected. Thus, shortcuts dedicated to skipping adaptively these coding modes might be promising. Shortcut on coding modes and partitioning acceleration method are in different scopes. The former focus on reducing number of CU for RDO. The latter speeds up the RDO for CU of certain sizes. The combination of these two could lead to a larger speed-up of encoding.

5. REFERENCES

[1] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] B. Bross et al., "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, 2021.

[3] Farhad Pakdaman, Mohammad Ali Adelimanesh, Moncef Gabbouj, and Mahmoud Reza Hashemi, "Complexity analysis of next-generation vvc encoding and decoding," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3134–3138.

[4] Alexandre Mercat, Arttu Mäkinen, Joose Sainio, Ari Lemmetti, Marko Viitanen, and Jarno Vanne, "Comparative rate-distortion-complexity analysis of vvc and hevc video codecs," *IEEE Access*, vol. 9, pp. 67813–67828, 2021.

[5] Mário Saldanha, Gustavo Sanchez, César Marcon, and Luciano Agostini, "Complexity analysis of vvc intra coding," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3119–3123.

[6] Frank Bossen, Karsten Sühring, Adam Wieckowski, and Shan Liu, "Vvc complexity and software implementation analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3765–3778, 2021.

[7] Adam Wieckowski, Jackie Ma, Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, "Fast partitioning decision strategies for the upcoming versatile video coding (vvc) standard," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 4130–4134.