TEMPORAL CORELATION BASED HIERARCHICAL QUANTIZATION PARAMETER DETERMINATION FOR HEVC VIDEO CODING

Yimin Zhou⋆*, Hongyu Wang*⋆*, Ling Tian*⋆ *and Ce Zhu*†

⋆ School of Computer Science and Engineering, † School of Electronic Engineering,
University of Electronic Science and Technology of China, Chengdu, 611731

ABSTRACT

The quantization parameter (QP) value and Lagrangian multiplier (λ) are the key factors for an encoder to achieve the trade-off between visual quality and bit-rate in next generation multimedia communications. In this work, we propose a novel temporal redundancy ratio (TRR) model to determinate hierarchical QPs. Taking the temporal redundancy information, intensity and fluctuation into consideration, the TRR model is constructed as the ratio of mean and variance from the temporal redundancy, and then used to allocate a proper QP value for an individual picture to achieve bit-rate saving and improve the visual quality. We implement the TRR model based QP determination scheme into the reference software HM16.7. Under the common test condition (CTC), simulation results show that the proposed TRR model obtains 1.28% and 1.52% BD-Rate gain over HM16.7 for Low-Delay and Random-Access in 8-bit video coding respectively. Meanwhile, more BD-Rate gains are obtained in 10-bit coding and the screen coding class.

Index Terms— QP offset, Lagrange optimization, temporal redundancy

1. INTRODUCTION

As more and more video streams account for the ever-increasing Internet traffic, video coding technologies have drawn much attention from the research and industry communities. The most recent technology is high efficiency video coding (HEVC/H.265), which has been a state-of-the-art video compression standard since 2012. HEVC/H.265 provides improvements over traditional technologies in terms of sampled representations of pictures, coding tree units (CTUs), partitioning block and units by quadtree structures, slices, and tiles, intrapicture and interpicture predictions [1], range extensions [2], and scalable extensions [3]. Based on these technologies, HEVC outperforms previous coding standards in many aspects [4]. As a result, HEVC is becoming the de facto video coding technology.

One distinct feature of HEVC is the Hierarchical Coding Structure (HCS) [5], which divides pictures into different layers according to their temporal order. The layer index of a picture is correlated with the length of a group of pictures (GOP) and its present order count (POC). The reference picture management is organized to improve coding efficiency and reduce distortion. HCS can be used for bit allocation, quality control, and coding optimization. In HEVC, the HCS is also adopted to determine the quantization parameter (QP) value for each picture [6]. From a configured initial QP, the HCS increases the QP value with the layer index of the picture and directly uses the QP offset to describe the QP difference between two neighbouring layers. The QP value directly impacts the distortion and bit-rate of the coded stream. Hence, it is very important for an encoder to select proper QPs.

The traditional layer-based fixed-QP determination approach in HEVC possesses some limitations. First, since there is no response from the source content and no feedback from previous coding results, it lacks adaptation and scalability. Second, the fixed-QP allocation policy cannot simultaneously handle high and low bit-rate scenarios, which may result in visual quality degradation. Third, the fixed-QP allocation from the HEVC common test condition (CTC) is not an adaptive strategy for different content, resolutions, and frame rates. In the previous works, we proposed some Lagrangian multiplier adaptation algorithms for enhancing rate-distortion optimization (RDO) performance [7][8][9][10].

In general, a choppy inflection in the source contents indicates that successive pictures may be highly correlated. On the contrary, a smooth inflection means that neighbouring pictures are lowly correlated. Therefore, in HCS QP allocation, the appropriate strategy for QP offset assignment should concern the characteristics of the source contents to satisfy the global rate distortion optimization [11].

In this work, we propose a flexible QP value determination approach based on a novel temporal redundancy ratio (TRR) model, which can overcome the aforementioned shortcomings of the traditional approaches, to allocate the QP for a picture. The temporal redundancy value is calculated through the logarithm evaluation of the motion compensation from video sources. After the TRR model is constructed as the ratio of the mean and variance of the temporal redundancy, the current QP is adaptively calculated to achieve the BD-Rate [12] gain and better visual quality. The TRR model approach resolves three issues of traditional QP allocation. Furthermore, we exploit the motion characteristics of the source contents in the process of QP allocation.

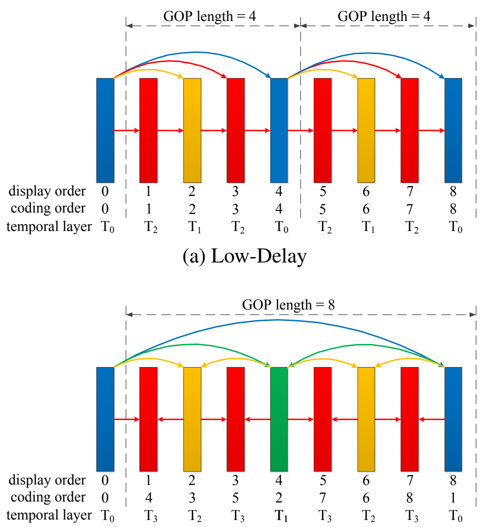
The remainder of this paper is organized as follows. The HCS and QP offset are described in Section 2. The TRR model is proposed in Section 3. The QP value and Lagrangian multiplier factor are calculated in Section 4. Section 5 presents the experimental results, and Section 6 concludes this work.

2. QP DETERMINATION WITH QP OFFSET

2.1. Hierarchical coding structures

The HCS virtually divides video pictures of temporal sequences into several layers for a GOP. The QP offset is a predefined value that depends on the index of the layer. The encoder system allocates the QP value based on the sum of the configured initial QP parameter and the QP offset during the encoding process.

Fig. 1 shows an example of the organization of the hierarchical structure with Low-Delay (LD) and Random-Access (RA) coding.



(a) Low-Delay

(b) Random-Access

Fig. 1. Hierarchical coding structure with temporal layers

The arrows between the pictures denote the reference direction. In Fig. 1(a), LD adopts the forward reference structure. There are eight pictures in the GOP, and the display order is the same as the coding order. However, as shown in Fig. 1(b), RA typically selects the binary pictures with a neighbouring temporal sequence as the reference lists. Hence, the display order and the coding order are different in RA.

2.2. Classical QP value for a picture

In general, the pictures in the lower layers of the HCS are more critical since they are referenced by pictures in higher layers. Thus, they should be allocated with smaller QP values. When the layer increases, a bigger QP value should be allocated. To simplify the QP allocation process, H.264/AVC, H.264/SVC, and HEVC/H.265 assume the constant value 1 as the increment for the QP offset between two neighbouring layers. Hence, the basic QP determination can be expressed by

$$QP_i = QP_{\text{ini}} + \Lambda + QP_{\text{offset}}, \tag{1}$$

where QP_i denotes the QP value for picture i, and QP_{ini} indicates the initial QP value from the coding configuration; the picture type modifier Λ is 0 for intra pictures and 1, otherwise. Moreover, QP_{offset} in Eq. (1) represents the offset value of the QP for picture i, which can be calculated as follows:

$$QP_{\text{offset}} = \text{layer}(L, i). \tag{2}$$

In Eq. (2), the function $\text{layer}(L, i)$ denotes the layer index based on the input of the GOP length L and POC i, which can be represented by the following formula:

$$\begin{aligned} &\text{layer}(L, i) = \min\{\log_2 L - k\} \\ &\text{s.t.} \;\; L - i \bmod L = (L - i \bmod L) >> k << k, \\ &\qquad k \in [0, \log_2 L], \end{aligned} \tag{3}$$

where k is the layer index, mod represents the modulo operation, and the symbols $<<$ and $>>$ denote the binary zero-fill left shift and right shift, respectively.

Combining Eqs. (1), (2), and (3), we obtain the final QP value for each picture. Since such an empirical QP selection approach lacks adaptation, X. Li et al. [13] introduced a QP adaptation approach with an offset value using the error propagation and distortion from the inter-frame dependency. In [14, 15], the authors propose an adaptive QP calculation method based on the temporal pumping artifact to take the temporal effect between pictures into consideration. However, the above works ignore the motion characteristics of the source and lacks effective adaptability in scenarios with different bit-rates, resolutions, or frame rates during the coding process. T.Zhao et al. [16] proposed a complex QP adaptation scheme which is deducted from traditional rate distortion model. Fiengo et al. [17] resolved the QP allocation problem using convex optimization.

2.3. Improvement approach for QP value determination

In hierarchical structure coding, higher layers contain more pictures, which deserve larger QP values to ensure better fundamental visual quality. Hence, we propose the following equation to determinate the QP offset value:

$$QP_{\text{offset}} = \log_2 L - \varphi(i) \cdot (\log_2 L - \text{layer}(L, i)), \tag{4}$$

where the function $\varphi(i)$ is the feedback of the temporal redundancy for picture i. Unlike the top-down QP allocation model in Eq. (1), Eq. (4) adopts a bottom-up QP allocation strategy, which can take better advantage of the rich picture contents from higher layers.

According to the HCS rule, half of all the pictures belong to the highest layer and are allocated with the same QP value in HEVC. This QP is the biggest QP in the coding process; equivalently, the highest layer has the biggest QP. Furthermore, the bigger the allocated QP value, the worse the quality distortion. Therefore, all of the pictures in the highest layer could potentially achieve the same distortion under the same QP value using Eq. (4).

More about the feedback function $\varphi(i)$ will be discussed in the next section. Next, we describe our proposed TRR model, which takes advantage of the motion characteristics and temporal redundancy information collected from the source rather than the inconsistent information gathered or predicted by the encoder to help dynamically allocate the QP values.

3. TRR MODEL

The source contents and motion variety have a significant impact on BD-Rate performance. This section proposes a novel TRR model to select suitable QP values based on the temporal redundancy and source contents.

3.1. Description of the distortion

The temporal redundancy has a great impact on picture distortion [18][19]. The motion compensated difference (MCD) is evaluated by the CTU level distortion. To depict MCD with the mean squared error (MSE), we use the following approximate formula:

$$C^{\mathrm{M}} \approx \alpha \cdot (C^{\mathrm{S}} + D). \tag{5}$$

In Eq. (5), C^{M} represents MCD for a CTU, C^{S} denotes the source motion compensated difference (SMCD) evaluated by the MSE for a CTU, D is the quantization distortion, and α is the estimation parameter.

Under a high rate assumption [20], the distortion model can be expressed by Eq. (6) since the source probability distribution can be approximated to be uniform within each quantization interval. The rate model for the uniform source can be obtained using the typical high-rate approximation curve for entropy-constrained scalar quantization.

$$D = \frac{q_{\mathrm{step}}^2}{12}, \lambda = c \cdot q_{\mathrm{step}}^2, \tag{6}$$

where q_{step} denotes the quantization step-size and λ is the Lagrangian multiplier in the rate-distortion optimize (RDO) method.

The RDO method has been shown to be practical and efficient and has been widely used in video codec implementations such as H.264/AVC as well as in the latest HEVC standards.

3.2. Linear distortion model

To investigate the relationship among C^{M}, C^{S} and λ on typical experimental data, Fig. 2 shows the scatter diagram from the typical *Johnny*, *BQMall*, and *BasketballDrill* sequences. The X, Y, and Z axes represent C^{S}, the Lagrangian multiplier λ, and C^{M}, respectively. Notice from Fig. 2 that the relation between C^{S}, λ, and C^{M} is approximately linear , and the distribution of the scatter dots approximates a plane. Accordingly, the linear model for estimating MCD in a CTU with distortions is

$$C^{\mathrm{M}} = p_1 \cdot C^{\mathrm{S}} + p_2 \cdot \lambda + p_3, \tag{7}$$

where λ is determined by its slice type and QP value, and p_1 , p_2, and p_3 are three model parameters. With Eq. (7), the MCD value C^{M} can be expressed by the SMCD value C^{S} and λ.

3.3. TRR model

Without loss of generality, we use Eq. (8) to deduct the mean value of MCD for each CTU to represent the picture level MCD:

$$\mathbf{C}^{\mathrm{M}}(i) = \frac{1}{n} \cdot \sum_{b=1}^{n} C_b^{\mathrm{M}}. \tag{8}$$

In Eq. (8), $\mathbf{C}^{\mathrm{M}}(i)$ is the MCD for picture i, n denotes the number of CTUs in a picture, and C_b^{M} represents the MCD value for CTU b.

For picture i, we use a logarithm form to describe the picture level MCD intensity by Eq. (9).

$$m_i = \log_2 \mathbf{C}^{\mathrm{M}}(i), \tag{9}$$

where the MCD intensity m_i could partly reveal the feedback of temporal redundancy.

To estimate the temporal redundancy of picture i, the mean value of m can be obtained using pictures within a window as follows:

$$\mu_i = \frac{1}{\omega} \cdot \sum_{j=1}^{\omega} m_{i-j}, \tag{10}$$

where μ_i is the mean value of m, and ω is the window size.

Similarly, the variance of m_i can be calculated by

$$\sigma_i = \sqrt{\frac{1}{\omega} \cdot \sum_{j=1}^{\omega} (m_{i-j} - \mu_i)^2}. \tag{11}$$

Eventually, the proposed TRR model is as follow:

$$\varphi(i) = \kappa \cdot \log_2 \frac{\mu_i}{\sigma_i + \epsilon}. \tag{12}$$

where κ denotes the empirical parameter 1/6 and ϵ is a very tiny value 10^{-6} to prevent the zero divide error.

The TRR effectively takes the temporal characteristics including the relevance among pictures, intensity of images, and precision requirements of the encoding procedure into account.

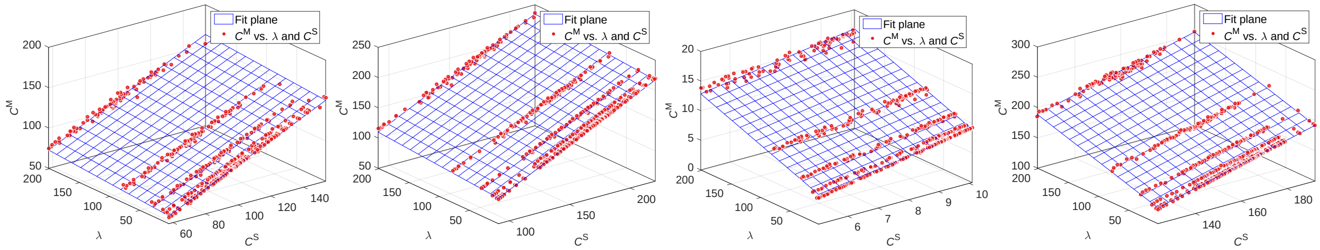
4. QP VALUE AND LAMBDA CALCULATION

Since the temporal redundancy is calculated based on previous pictures, the TRR model could be adaptively updated based on the fluctuation of picture contents. Finally, the QP value of the picture POC index i is calculated by

$$\begin{aligned} QP_i = {} & QP_{\text{ini}} + \Lambda + \log_2 L \\ & - \kappa \cdot \log_2 \frac{\mu_i}{\sigma_i} \cdot (\log_2 L - \mathrm{layer}(L, i)). \end{aligned} \tag{13}$$

The Lagrangian multiplier λ_i is recalculated as follows:

$$\begin{aligned} \lambda_i(QP_i) &= \lambda_i(QP_{\mathrm{sys}}) \cdot 2^{\frac{QP_i - QP_{\mathrm{sys}}}{6}} \\ &= \lambda_{\mathrm{sys}} \cdot 2^{\frac{QP_i - QP_{\mathrm{sys}}}{6}}, \end{aligned} \tag{14}$$



(a) *BasketballDrill* (b) *BQMall* (c) *Johnny* (d) *PartyScene*

Fig. 2. Scatter diagram samples for motion compensated difference (MCD) evaluation.

where λ_{sys} denotes the system Lagrangian multiplier value, which is calculated according to the QP value from Eq. (1).

The computation complexity of the proposed QP allocation mainly comes from the ratio function calculation in the TRR model. The ratio is derived from Eqs. (8), (9), (10), and (11), whose time complexities are $O(n)$, $O(1)$, $O(\omega)$, and $O(\omega^2)$, respectively. In fact, the CTU number n in a picture normally falls in the range of dozens to thousands. It is smaller than the picture resolution, and the window size ω is generally set to the same value as the GOP length L. Hence, the computation complexity in Eqs.(8)–(11) is negligible, particularly when comparing it with the computation load during the whole coding process.

5. NUMERICAL RESULTS AND ANALYSIS

To evaluate the performance of the proposed TRR model, we integrated it with HEVC reference software HM16.7 [21], and conducted the experiments with the LD and RA coding structures. The tested sequences included Class A (2K), B (1080P), C (WVGA), D (WQVGA), E (720P), and resolution sequence Class F (Screen) [22]. The tests adopted the CTC coding results as the baseline.

Table 1. BD-Rate for Low-Delay (LD) coding.

	LDB Main			**LDB HE10**		
	Y	U	V	Y	U	V
Class B	-1.02%	0.72%	-0.04%	-1.01%	0.54%	-0.22%
Class C	-0.90%	-3.34%	-3.38%	-0.98%	-3.52%	-3.69%
Class D	-0.79%	-5.25%	-4.72%	-0.70%	-5.28%	-5.80%
Class E	-2.88%	-8.01%	-7.32%	-2.82%	-8.21%	-7.72%
Overall	-1.28%	-3.43%	-3.41%	-1.26%	-3.57%	-3.89%
Class F	-2.66%	-4.70%	-3.88%	-2.73%	-4.75%	-5.58%
Enc Time[%]	101%			102%		
Dec Time[%]	101%			101%		

Table I shows the BD-Rate for Classes B to F on LD structures. The Y, U, and V columns represent one luma (Y) and two chrominance (U and V) components for a sequence. The BD-Rate was calculated by choosing piece-wise cubic interpolation in HEVC. Compared with the benchmark, the negative and positive BD-Rate values in the table denote bit-rate savings and waste, respectively. As shown in Table I, almost all the tested results are negative, which indicates the proposed TRR model achieves bit-rate savings. In particular, the sequences of Class E (720p) have an average BD-Rate gain of 2.88%. The average BD-Rate gain of the other common sequences is 1.28% for both 8-bit and 10-bit coding. This is because the TRR model adopts the inflection of the source contents while considering the correlation between successive pictures. Furthermore, the TRR value feedback is directly employed in the QP determination, which enables the encoder to achieve obvious coding gains.

Table 2. BD-Rate for Random-Access (RA) coding.

	RA Main			**RA HE10**		
	Y	U	V	Y	U	V
Class A	-1.77%	-6.66%	-8.53%	-1.77%	-6.80%	-8.86%
Class B	-1.02%	-8.81%	-11.27%	-1.08%	-8.96%	-11.55%
Class C	-2.16%	-6.31%	-6.59%	-2.17%	-6.79%	-6.75%
Class D	-1.40%	-7.82%	-7.51%	-1.35%	-8.13%	-7.96%
Overall	-1.52%	-7.59%	-8.65%	-1.54%	-7.87%	-8.95%
Class F	-2.42%	-5.52%	-4.68%	-2.33%	-5.58%	-5.01%
Enc Time[%]	102%			103%		
Dec Time[%]	101%			102%		

Similarly, Table II shows the superiority of the proposed approach over the traditional HM16.7 baseline method when RA coding is applied. Based on the observations, the TRR model achieves gains under all resolutions, which indicates that the proposed model achieves bit savings. This benefits video transmission, distribution, and storage.

6. CONCLUSION

A novel TRR model was proposed and applied for QP calculation during the HEVC hierarchical structure video coding process. The TRR model takes into account the distortion of SMCD and facilitates the allocation of a proper and adaptive QP for picture coding. Unlike the traditional QP offset setting, the proposed TRR model adopts a bottom-up QP allocation strategy for better BD-Rate gain. Our extensive experiments demonstrate that the proposed approach can achieve significant bit-rate savings for both LD and RA coding processes. In the future work, we will investigate the empirical parameters κ in the QP calculation to improve model precision and promote BD-Rate performance.

References

[1] G. J. Sullivan, J. R. Ohm, W.-J. Han, and T. 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, Dec. 2012.

[2] D. Flynn, D. Marpe, M. Naccari, T. Nguyen, C. Rosewarne, K. Sharman, J. Sole, and J. Xu, "Overview of the range extensions for the HEVC standard: Tools, profiles, and performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 4–19, Jan. 2016.

[3] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of shvc: Scalable extensions of the high efficiency video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, Jan. 2016.

[4] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, vp9, and H.264/MPEG-AVC encoders," in *Picture Coding Symposium (PCS), 2013*, Dec. 2013, pp. 394–397.

[5] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[6] H. L. Cycon, T. C. Schmidt, M. Wahlisch, D. Marpe, and M. Winken, "A temporally scalable video codec and its applications to a video conferencing system with dynamic network adaption for mobiles," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1408–1415, August 2011.

[7] B. Xiong, X. Fan, C. Zhu, X. Jing, and Q. Peng, "Face region based conversational video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 917–931, July 2011.

[8] S. Li, C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, and M. T. Sun, "Lagrangian multiplier adaptation for rate-distortion optimization with inter-frame dependency," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 117–129, Jan 2016.

[9] Y. G. S. L. Q. P. L. Y. C. Zhu, Y.M. Zhou, "Temporal dependent rate-distortion optimization in internet video coding," in *ISO/IEC JTC1/SC29/WG11 MPEG2014/M35891*, Feb. 2015.

[10] Y. Gao, C. Zhu, S. Li, and T. Yang, "Temporal dependent rate-distortion optimization for low-delay hierarchical video coding," *IEEE Transactions on Image Processing*. (Accepted in May 2017, in press)

[11] M. Winken, D. Marpe, and T. Wiegand, "Global and local rate-distortion optimization for lapped biorthogonal transform coding," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept 2010, pp. 173–176.

[12] G. Bjøntegaard, "Calculation of average psnr differences between rd-curves," in *Document VCEG-M33*, Apr. 2001.

[13] X. Li, P. Amon, A. Hutter, and A. Kaup, "Model based analysis for quantization parameter cascading in hierarchical video coding," in *16th IEEE International Conference on Image Processing (ICIP)*, Nov. 2009, pp. 3765–3768.

[14] S. Wan, Y. Gong, and F. Yang, "Perception of temporal pumping artifact in video coding with the hierarchical prediction structure," in *IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2012, pp. 503–508.

[15] Y. Gong, S. Wan, F. Yang, H. R. Wu, and B. Li, "A frame level metric for just noticeable temporal pumping artifact in videos encoded with the hierarchical prediction structure," in *IEEE International Conference on Image Processing (ICIP)*, Sep. 2015, pp. 3034–3038.

[16] T. Zhao, Z. Wang, and C. W. Chen, "Adaptive quantization parameter cascading in hevc hierarchical coding," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 2997–3009, July 2016.

[17] A. Fiengo, G. Chierchia, M. Cagnazzo, and B. Pesquet-Popescu, "Rate allocation in predictive video coding using a convex optimization framework," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 479–489, Jan. 2017.

[18] T. Yang, C. Zhu, X. Fan, and Q. Peng, "Source distortion temporal propagation model for motion compensated video coding optimization," in *IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2012, pp. 85–90.

[19] S. Li, C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, and M.-T. Sun, "Inter-frame dependent rate-distortion optimization using lagrangian multiplier adaption," in *IEEE International Conference on Multimedia and Expo (ICME)*, Jun. 2015, pp. 1–6.

[20] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. 14, no. 5, pp. 676–683, Sep 1968.

[21] HEVC reference software (HM16.7). https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.7/.

[22] F. Bossen, "Common test conditions and software reference configurations," in *Document JCTVC-L1100*, Jan. 2013.