Optimal Adaptive Quantization based on Temporal Distortion Propagation model for HEVC

Maxime Bichon, Julien Le Tanou, Michael Ropert, Wassim Hamidouche, Luce Morin

Optimal Adaptive Quantization based on Temporal Distortion Propagation model for HEVC

Maxime Bichon, *Student Member, IEEE*, Julien Le Tanou, Michael Ropert, Wassim Hamidouche, and Luce Morin

***Abstract*—Optimal adaptive quantization is one of the key points to optimize the coding efficiency of video encoders. Latest block-based video compression standards, such as *High Efficiency Video Coding* (HEVC), extensively use predictive coding techniques that create dependencies between blocks and increase the complexity of optimal block quantizers search. Specifically, the motion compensation is responsible for a dependency network connecting all blocks of the same GOP together. In this paper, this dependency network is estimated by a temporal distortion propagation model and an accurate estimation of Inter and Skip modes probabilities. Optimal quantizers are then designed per block in order to achieve the global optimization in terms of Rate-Distortion efficiency. By implementing the algorithm into the HEVC reference Model (HM), we report -16.51% PSNR-based and -26.26% SSIM-based average bitrate savings compared to no adaptive quantization. The proposed algorithm outperforms several related methods from state-of-the-art. Moreover, along with the demonstration of optimal quantizer solution, we propose an in-depth analysis of the algorithm behavior. This analysis includes, among others, the relative distribution of rates between frames and the control of quantizers dynamic range.**

***Index Terms*—Local Quantization, Rate-Distortion (R-D) Optimization, HEVC, Temporal Distortion Propagation, Skip probability.**

I. Introduction

THE total amount of video traffic over Global IP, cellular or satellite networks is in constant growth. The emerging video contents bring more immersive visual experience to the end users by developing new video technologies from acquisition to display such as Ultra High Definition (UHD), High Dynamic Range (HDR) and 360°. This large scale data brings new compression challenges to ensure efficient storage and transmission while preserving high quality for these contents. Moreover, the current diversity of video services offered by different platforms (social networks, on demand TV, 5G) increases the demand of end users for high quality, remotely and immediately accessible contents. This dynamic environment is the key motivation for achieving real-time and high efficiency coding.

The High Efficiency Video Coding (HEVC) [1] standard was released in 2013 by the Joint Collaborative Team on Video Coding (JCT-VC) established jointly by the ITU-T Video Coding Experts Group (VCEG) and the ISO Motion Picture Expert Group (MPEG). HEVC enables up to -50% bitrate savings [2], [3] for equal perceptual video quality compared to the previous Advanced Video Coding (AVC) [4] standard. This significant coding gain is achieved by the new coding tools introduced in HEVC. This standard relies on the common hybrid video coding scheme performing either Intra or Inter predictions at the block level to get benefits from spatial or temporal redundancies, respectively. HEVC encoders such as the reference software model (HM) [5] and the real time encoder x265 [6] aim to minimize the video distortion D subject to the total rate constraint $R \leq R_T$. The Rate Distortion Optimization (RDO) [7] usually minimizes the R-D cost function $J = D + \lambda R$ where λ is the Lagrange multiplier used to turn the constrained optimization problem into an unconstrained one [8]. λ controls the trade-off between distortion D and rate R.

There are mainly two approaches to enhance coding efficiency in the context of video standards. The first way consists in developing more efficient coding tools within the standardization process to build the next generation video standard. The Joint Video Exploration Team (JVET) has been recently investigating several new coding solutions [9] to show the evidence of developing a new standard with coding capability beyond HEVC. These new tools enable to increase the coding efficiency by up to 40% compared to HEVC [10]. However, this gain comes with a significant complexity increase of 10 times the HEVC complexity at both encoder and decoder sides [11], [12]. Moreover, this approach is a long term solution and usually requires a decade to release a new video standard.

The second approach aims to enhance the coding efficiency of existing standard encoders without changing the syntax of the decoder. Common implementations of RDO consist in independent optimization of each block by testing all coding configurations and selecting the set of modes that minimizes the R-D cost. However, these consecutive local optimizations do not necessarily lead to the performance of global R-D cost minimization of the source signal [13]. Instead, the design of dependent models allows to reach higher coding efficiency with involving only encoder-side modifications. These solutions often use a look-ahead analysis of the video source, which introduces a manageable complexity increase. Indeed, the look-ahead usually runs in parallel with the encoder thanks to an efficient use of multi-threading on multi-core architecture . Apart from some delay, it significantly lowers the impact of look-ahead processing with respect to the encoding.

M. Bichon, J. Le Tanou and M. Ropert are with MediaKind, former Ericsson Media Solutions, Saint-Jacques-de-la-Lande 35136, France (e-mail: maxime.bichon@mediakind.com; julien.letanou@mediakind.com; michael.ropert@mediakind.com).

W. Hamidouche and L. Morin are with INSA Rennes, Institute of Electronics and Telecommunications of Rennes (IETR), CNRS - UMR 6164, VAADER team, 20 Avenue des Buttes de Coesmes, 35708 Rennes, France (e-mail: wassim.hamidouche@insa-rennes.fr, luce.morin@insa-rennes.fr)

Manuscript submitted on August, 2018.

Our previous works [14] and [15] present a global R-D optimization solution for the HEVC encoder. First, the temporal distortion propagation is modeled at the block level, called Coding Unit (CU) in HEVC. Propagation is introduced by temporal predictions between frames within the Group of Pictures (GOP) structure. Second, we accurately estimate the probability of a CU to be temporally predicted (i.e. Inter coded) and the probability of a CU to use the special coding mode named Skip based on a look-ahead analysis. The distortion propagation model and the probabilities enable to build an analytical solution deriving optimal local quantizers within a GOP at the CU granularity.

In this paper, we provide better insights about the analytical solution. Important assumptions are discussed and verified, such as the chosen Skip mode probability and rate independence hypothesis. We also prove that optimal delta quantizers are bounded and their dynamic is straightforward to control. The model is also extensively analyzed based on the rate distribution among several frames, the perceptual quality influence within the model and the Target Bitrate Deviation (TBD). Thanks to the development of a look-ahead within the HEVC reference Model (HM), new experimental results against close techniques from state-of-art are made possible. Subjective quality assessment of the model is performed showing consistent spatial and temporal quality improvements. Encoding runtimes are also provided for fairness. Optimal quantizers enhance the coding efficiency against no-Adaptive Quantization (AQ), in terms of bitrate savings, by -23.53% and -26.26% under the real time encoder x265 and the HM, respectively.

The rest of the paper is organized as follows. Section II gives an overview of the context and works related to global R-D optimization solutions followed by the motivations of this paper. The temporal propagation model considered in this paper is presented in Section III. Section IV investigates the proposed HEVC video optimization solution. Insights on the proposed solution and implementation details under two HEVC software encoders are provided in Section V. Section VI gives experimental results showing benefits of the proposed model within the two considered codecs. Finally, Section VII concludes this paper.

II. Context & Related Works

A. HEVC standard and its codecs

In the HEVC standard, frames are first equally divided into Coding Tree Unit (CTU), blocks of pixels processed in raster scan order. Each CTU can be further recursively split into multiple CU, based on a QuadTree decomposition. The coding of such units is achieved by three operations performed sequentially : prediction, transformation and quantization. One of the most important part in an HEVC encoder is the "decision core", which sets parameters of these processes in order to optimize the signal coding efficiency, under one or more constraints.

Prediction aims to estimate pixel values in a CU from previously coded data. In HEVC prediction is mainly based on three main modes: Intra, Inter and Skip. Intra mode makes use of spatial correlations in images to predict a CU by referencing the spatial neighboring pixels. Inter mode takes advantage of the temporal redundancy by referencing CUs from previously coded frames for motion prediction and compensation. Both Intra and Inter predictions are followed by transformation and quantization of residues (prediction errors). Finally, the Skip mode consists in Inter prediction, with no residue coding. This special mode leads to near-zero bitrate while setting distortion to its maximum, i.e. the prediction error energy. Note that, the HEVC Merge mode is assimilated to Inter mode in this paper.

In hybrid encoders, quantization is applied on transformed residue and it is controlled by the Quantization Parameter (QP) which lies in the range $[0..51]$ for HEVC. A low QP value leads to low distortion, while a high QP value tends to suppress the residue. Consequently, the Skip mode probability increases along with the QP value. We point out that the information removed by quantization in a CU may also be the redundant part of the signal used to predict subsequent CUs temporally or spatially. Therefore, ignoring Inter-CUs dependencies while setting locally the QP may be globally ineffective.

In this paper, two HEVC encoders have been considered for experiments: HM and x265. HM is the reference software used by JCT-VC for standard development. It implements all normative coding tools and it is based on a full RDO implementation, leading to high coding efficiency at the expense of high computational complexity. x265 is an open-source encoder developed by MulticoreWare to provide high coding efficiency under real-time constraint. Unlike HM and similarly to most of the real-time video encoders, x265 relies on a look-ahead process to efficiently drive the decision core. Commonly, a look-ahead mechanism consists in a video source analysis without encoding decisions. It differs from multi-pass approaches that perform multiple encodings of the video, refining the coding parameters at each pass. It is too computationally complex to suit real-time encoding.

B. Global RDO Solutions

Global RDO solutions in state of the art can be classified into two categories: exhaustive dependency exploration and dependency modeling.

1) Exhaustive dependency exploration: Methods that fall in this category proceed in an exhaustive exploration of the dependencies inherent to the coding scheme. It usually results in intractable computational complexity for real-time applications.

Ramchandran et al. [16] consider the frame bit allocation in video coding as a trellis problem solved with the Viterbi algorithm. Using a simple coding scheme and pruning rules, they succeed to achieve significant coding gain. Global optimization can also be opposed to local optimization when searching for optimal transformed coefficient levels. In [17], *Wen et al.* use the same trellis approach to jointly optimize all the transformed coefficient levels after quantization.

Fiengo et al. [18] express distortion as a convex function of all frames bitrate. Primal-Dual Proximal Algorithm is used to solve the convex optimization problem and achieve near optimal Rate-Control (RC). In [19] *Wiken et al.* measure the

dependencies between coefficients levels after Discrete Cosine Transform (DCT)/Discrete Sine Transform (DST), leading to an optimization problem solved by an iterative approach. *Bichon et al.* [20], [21] consider Intra prediction mode estimation as a joint optimization of spatially close Prediction Units (PUs).

Overall, approaches based on exhaustive dependency exploration show significant R-D efficiency improvement, but at the price of very high computational complexity. It is especially true when considering the large number of coding parameters combinations to explore in the HEVC standard.

2) Dependency modeling: The second category gathered methods that model the dependencies based on some estimators and theoretical assumptions. The optimization remains block-based, but new assumptions are used instead of the block independence hypothesis. These assumptions are usually established through a look-ahead, in order to apply the RDO locally, but consciously of the potential effects coding decisions may have on other CUs.

The temporal dependencies, that is the focus of this paper, can be modeled with various granularity. In [22], *Valenzise and Ortega* estimate a temporal dependency tree at the pixel level, further used to design an AQ method based on the tree depth. *Li et al.* [23] estimate distortion propagation frame by frame, in order to provide consistent video quality over an entire GOP.

This paper and following references model dependencies on a block basis, thus we define some notations to be comparable. We consider in this paper the CU with index i in the frame of index t, labeled i_t. A large number of studies modeling the dependencies between CUs use equation (1) to express the distortion propagation between CUs.

$$D_{i_t} = d_{i_t} + p_{i_t} D_{ref} \tag{1}$$

In equation (1), D_{i_t}, the distortion of CU i_t, is expressed as the sum of two terms: the local distortion d_{i_t}, depending only on the coding parameters set for CU i_t; and the reference distortion which depends on D_{ref}, the distortion of the reference samples CU used for prediction. p_{i_t} is a weighting factor that describes the relation between the distortion of a CU and its reference block's distortion.

The additive formulation in (1) is a simplification of a multiplicative formulation depicted in appendix A. By using the first order Taylor-Young expansion of D_{i_t} based on d_{i_t} and D_{ref} variables, local and reference distortions, the additive formulation can be obtained.

Yang et al. present a Source Distortion Temporal Propagation (SDTP) model [24] that increases the coding efficiency by adaptively scaling the λ value for each CU. In this model, p_{i_t} is a function of the CU rate R_{i_t}. d_{i_t} is a function of R_{i_t} and the innovation $\sigma^2_{i_t}$ of CU i_t. $\sigma^2_{i_t}$ is defined here as the part of the signal which is unpredictable, i.e. the residue of prediction before quantization. Using equation (1), authors describe dependencies between CUs and they adaptively scale the λ value used for R-D cost computation, which leads to substantial coding efficiency improvement. The more the distortion of a CU impacts other CUs, the more the λ value decreases.

The model proposed in [24] has been further extended by *Xie et al.* in [25] for bit allocation strategy in the context of RC. Specific hierarchical coding schemes have also been investigated by *Gao et al.* for Low Delay (LD) [26] and Random Access (RA) [27] coding configurations. In the specific case of HM and RA configuration, coding efficiency increases by 2.2% and can be further improved to 5.2% when the method is coupled with the high-complexity Multi Quantization Parameter (MQP) optimization proposed by *Sullivan and Wiegand* [7].

Ropert et al. [14] propose the Rate Distortion Spatio-Temporal Quantization (RDSTQ), as a generalization of Macroblock-Tree framework designed for x264 open-source AVC encoder [28]. The main interests of this approach is to model the temporal distortion propagation between CUs from a R-D standpoint and to introduce a psycho-visual criteria to optimize perceived quality. In the case of RDSTQ, d_{i_t} and p_{i_t} are respectively presented as the local distortion and local probability of a CU to be Inter coded, i.e. predicted from previous frames. Using the high rate assumption and developing the total signal distortion as a weighted sum of local distortions, *Ropert et al.* provide an optimal quantizer for each CU. *Bichon et al.* [15] further introduce a more accurate Inter probability and the Skip probability consideration into a simplified RDSTQ solution. These improvements enable additional R-D gains and reduce the TBD compared to the initial algorithm.

C. Perceptually optimized AQ methods

Most of AQ algorithms aim to optimize a given perceptual quality metric other than Peak Signal to Noise Ratio (PSNR). As discussed later in the paper, the proposed model allows to minimize a perceptual distortion by smoothly considering any spatial psycho-visual factor. In particular, in our proposal we designed a variant of the model optimizing the Structural Similarity (SSIM) metric, proposed by *Wang et al.* [29]. It is then relevant to take a look at AQ methods from state of art that optimize [29]; despite, they do not necessarily target Global RDO.

Yeo et al. [30] optimize the SSIM by adaptively scaling the Lagrangian multiplier and local quantizer for each CU. Estimation of the coding parameters requires only local variance of pixels in this approach, which achieves -7.4% bitrate savings in HM8.0 with negligible computational complexity overhead. In [31], *Xiang et al.* argue that spatial AQ methods are more efficient when enhancing inter frame correlation. Consequently, authors compute CU delta quantizers based on the Sum of Absolute Transform Differences (SATD) cost of the Inter mode efficiency estimation, with a constraint of unchanged average delta quantizer for the entire frame. This solution enables -3.69% SSIM-based bitrate savings in average for RA coding configuration.

D. Objective and motivations

We focus on look-ahead-based global optimization approach and more specifically on the RDSTQ model. Despite the

high R-D efficiency of initial model [14] against no AQ ($-19,4\%$ SSIM-based bitrate savings in average for x265), the considered distortion propagation model was perfectible. *Bichon et al.* [15] improve the PSNR-based Bjøntegaard Delta Bit Rate (BD-BR) by -2% in average and reduce the TBD from 38.05% to 13.98%, in the context of x265 encoder. However, the performance evaluation in [15] is achieved in a restricted configuration (only x265 encoder and no perceptual criteria), the distribution of delta quantizers is not analyzed and the analytical solution is based on unverified assumptions.

In this paper, we aim to extensively detail the analytical solution of RDSTQ. We also aim to verify important assumptions, such as the hypothesis of rates independence and the Skip mode probability. We also tries to provide thorough observations on the model. One potential issue with AQ methods is the possible drift of delta quantizers, i.e. the lack of control on their dynamic range, that is proved fully controllable for the RDSTQ. Finally, we aim to demonstrate the efficiency and consistency of the model. The new results achieved in the reference model (HM) further allow a fair comparison with state-of-the-art methods. Encoding runtimes and subjective quality assessments, that are not presented in our previous papers, are also discussed. Finally, an analysis of the model behavior is proposed, including the rate distribution among frames within a GOP and the influence of a perceptual criteria in the spatial distribution of delta quantizers.

III. Temporal Distortion Propagation Model

The subscript i_t is used when referring to the CU with spatial index i in the frame with temporal index t. N denotes the number of CUs in a frame, and T denotes the GOP size. The video encoding process aims to find the optimal coding parameters $\vec{p}$ that minimize the total distortion D_{Tot} under the target rate R_{Tot} constraint, as expressed in (2).

$$\begin{aligned} \min_{\vec{p}} \quad & D_{Tot}(\vec{p}) \\ \text{s.t.} \quad & \textstyle\sum_{t=1}^{T}\sum_{i=1}^{N} R_{i_t}(\vec{p}) = R_{Tot} \end{aligned} \tag{2}$$

Video encoders aim to maximize the video quality perceived by the Human Visual System (HVS). To consider the HVS in the distortion model, a spatial psycho-visual weighting factor Ψ is introduced. This factor is applied on each CU to better reflect the quality perceived by the HVS and is discussed later in the paper. The D_{Tot} to minimize is then expressed by (3). In the particular case of $\Psi_{i_t} = 1\,, \forall i_t$, the minimized distortion is chosen to be the classical Mean Square Error (MSE).

$$D_{Tot}(\vec{p}) = \sum_{t=1}^{T}\sum_{i=1}^{N} \Psi_{i_t} D_{i_t}(\vec{p}) \tag{3}$$

The temporal distortion propagation model used in this paper defines the distortion D_{i_t} of a CU i_t as the weighted sum of its local distortion d_{i_t} and the distortion $D_{j_{t_{ref}}}$ propagated from its reference CU $j_{t_{ref}}$. Accounting the motion compensation, the propagation formula initially presented in [14] is given by

$$D_{i_t}(\vec{p}) = d_{i_t}(\vec{p_{i_t}}) + p_{i_t} \underbrace{\sum_{j_{t_{ref}} \in Ref(i_t)} r_{j_{t_{ref}},i_t} D_{j_{t_{ref}}}(\vec{p})}_{\eta_{i_t}}. \tag{4}$$

$Ref(i_t)$ is the set of reference CUs used for motion compensation, p_{i_t} is the probability of a CU to be Inter coded and $r_{j_{t_{ref}},i_t}$ the pixel surface ratio involved in the motion compensation to go from spatial position of $j_{t_{ref}}$ to spatial position of i_t. $d_{i_t}(\vec{p_{i_t}})$ is the local distortion, i.e. the distortion that only depends on $\vec{p_{i_t}}$, the coding parameters applied to encode the CU i_t. η_{i_t} is the amount of distortion from reference samples propagated into CU i_t after motion compensation. For writing simplification, distortion functions are expressed in the following without parameters, i.e. $d_{i_t}(\vec{p_{i_t}}) = d_{i_t}$, unless a particular coding parameter is necessary for understanding.

The main drawback of this model is to only consider Inter/Intra coding, i.e. modes involving the transmission of a residue, and to ignore the Skip coding mode where no residue is transmitted. To consider the Skip mode, *Bichon et al.* [15] introduce c_{i_t} as the probability of the CU i_t to be coded in Inter/Intra mode and $(1 - c_{i_t})$ as the probability of the CU to be coded in Skip mode.

A large residue should lead to a high probability for Intra/Inter mode, while a large quantization step Δ should lead to a high probability for Skip mode. Hence, c_{i_t} is proposed to be defined as:

$$c_{i_t} = \frac{12\sigma^2_{src_{i_t}}}{12\sigma^2_{src_{i_t}} + \Delta^2_{i_t}} \tag{5}$$

$$c_{i_t} \to \begin{cases} 1 & \text{if } 12\sigma^2_{src_{i_t}} \gg \Delta^2_{i_t} \\ \frac{12\sigma^2_{src_{i_t}}}{\Delta^2_{i_t}} & \text{if } 12\sigma^2_{src_{i_t}} \ll \Delta^2_{i_t} \end{cases} \tag{6}$$

where $\sigma^2_{src_{i_t}}$ is the variance of predicted residue obtained by motion compensation between source samples, and Δ_{i_t} is the quantization step used to code the CU i_t. The behavior analysis of (5) is given in (6). If $12\ \sigma^2_{src_{i_t}} \ll \Delta^2_{i_t}$, c_{i_t} tends toward $12\sigma^2_{src_{i_t}}/\Delta^2_{i_t}$, which can be approximated as 0. It is intuitively adequate that a large residue leads to a high probability for Intra/Inter mode, while a large quantization step leads to a high probability for Skip mode. The choice of Skip probability model is also a mathematical workaround to draw a more robust solution, as described in Section V-C.

In order to include c_{i_t} in the propagation model in (4), we first define $D^C_{i_t}$ and $D^S_{i_t}$ in equation (7) as the distortion of a CU i_t to be coded in Inter/Intra and Skip, respectively.

$$D^C_{i_t} = d_{i_t} + p_{i_t}\eta_{i_t}, \quad D^S_{i_t} = \sigma^2_{src_{i_t}} + p_{i_t}\eta_{i_t} \tag{7}$$

As stated in Section II-A, the Skip mode introduces a distortion, i.e. d_{i_t}, that is equal to the prediction error, i.e. $\sigma^2_{src_{i_t}}$. It is our justification for the second equation in (7). According to (7), the propagation model in (4) is turned into (8).

$$\begin{aligned} D_{i_t} &= c_{i_t} D^C_{i_t} + (1 - c_{i_t}) D^S_{i_t} \\ &= c_{i_t} d_{i_t} + (1 - c_{i_t})\, \sigma^2_{src_{i_t}} + p_{i_t}\eta_{i_t} \end{aligned} \tag{8}$$

By developing the total distortion D_{Tot} from (3) and using the temporal propagation defined in (8), we can express the total distortion as a weighted sum of local distortions (9). The details of the calculation are explained in appendix C, as well as how indexes should be interpreted.

$$D_{Tot} = \sum_{t=1}^{T}\sum_{i=1}^{N}\left(c_{i_t}d_{i_t} + (1-c_{i_t})\,\sigma^2_{src_{i_t}}\right)U_{i_t}, \tag{9}$$

where U_{i_t} is the accumulation factor recursively defined by

$$\begin{cases} U_{i_T} &= \Psi_{i_T} \\ U_{i_t} &= \Psi_{i_t} + \sum_{i_{t+1}} p_{i_{t+1}} r_{i_t,i_{t+1}} U_{i_{t+1}}. \end{cases} \tag{10}$$

The different steps to establish the recursion are given in appendix D. U_{i_t} can be semantically interpreted as the *proportion* of the local distortion d_{i_t} that impacts the total distortion D_{Tot}.

The main interest of the formulation in (9) is to isolate local distortions d_{i_t} that depend only on local coding parameters $\vec{p_{i_t}}$. Consequently, the problem stated in (3) can be solved by locally setting coding parameters that optimize the overall R-D efficiency of a GOP. The application case of adaptive local quantization and its related analytical solution are both described in the next section.

IV. RDSTQ Algorithm

In this Section the RDSTQ algorithm initially proposed by *Ropert et al.* [14] and improved by *Bichon et al.* [15] is detailed. First, the local quantization problem is introduced and the total distortion D_{Tot} derivatives are simplified to lead to a simple mathematical formulation. Then, an analytical solution is provided, based on R-D Shannon bound, independence of rates and high rate assumptions, which leads to optimal encoding from an R-D standpoint.

A. Local Quantization Problem

The coding parameters of interest in this paper are the local quantization parameters, noted q_{i_t} for CU i_t. For ease of reading, the set of local quantizers for all CUs in a GOP is noted $\{q\}$, with $\{q\} = \{q_{i_t}\}_{i=1..N,t=1..T}$. The overall constrained minimization problem is then:

$$\begin{aligned} \{q^*\} &= \arg\min_{\{q\}} \sum_{t=1}^{T}\sum_{i=1}^{N} \Psi_{i_t} D_{i_t}(\{q\}) \\ \text{s.t.} \quad & \sum_{t=1}^{T}\sum_{i=1}^{N} R_{i_t}(\{q\}) = R_{Tot}. \end{aligned} \tag{11}$$

A simplification is made for solving the problem and achieve an analytical solution. This simplification is to consider the Inter probability p_{i_t} and the references distortion η_{i_t} that affects i_t independent of q_{i_t}. According to (10), U_{i_t} is then also independent of q_{i_t}. Intuitively, the local quantizer should affect the Inter probability. However, its influence is negligible in most cases, i.e. Intra and Inter modes efficiencies are far from each other.

The non-Skip probability c_{i_t} and the local distortion d_{i_t} both depend on the local quantization parameter q_{i_t}. The necessary condition to find the minimum of D_{tot} is determined by the condition of all the derivatives equal to zero $\forall i \in \{1,...,N\}, \forall t \in \{1,...,T\}$:

$$\frac{\partial D_{Tot}}{\partial \Delta_{i_t}} = \left(\frac{\partial d_{i_t}}{\partial \Delta_{i_t}}c_{i_t} + \frac{\partial c_{i_t}}{\partial \Delta_{i_t}}d_{i_t} - \frac{\partial c_{i_t}}{\partial \Delta_{i_t}}\sigma^2_{src_{i_t}}\right)U_{i_t} \tag{12}$$

(12) can be simplified into (14) as described below:

$$\frac{\partial c_{i_t}}{\partial \Delta_{i_t}} = \frac{-24\sigma^2_{src_{i_t}}\Delta_{i_t}}{144\sigma^4_{src_{i_t}} + \Delta^4_{i_t} + 24\sigma^2_{src_{i_t}}\Delta^2_{i_t}} \approx 0 \tag{13}$$

We justify this approximation by observing that whatever the values of $\sigma_{src_{i_t}}$ and Δ_{i_t}, denominator should always be much larger than numerator. This simplification leads to

$$\frac{\partial D_{Tot}}{\partial \Delta_{i_t}} \approx \frac{\partial d_{i_t}}{\partial \Delta_{i_t}}c_{i_t}U_{i_t} \tag{14}$$

B. Analytical Solution

In this subsection, we depict the analytical solution which makes use of (14) to solve the constrained problem described in (11). The analytical solution results in obtaining the optimal delta quantizers dQP for all CU, that maintain the GOP total rate identical. The problem in (11) is modeled thanks to the Lagrangian multiplier method with λ the Lagrangian multiplier. The new function to minimize is the total R-D cost J_{tot} defined in (15).

$$J_{Tot} = D_{Tot} + \lambda\left(\sum_{t=1}^{T}\sum_{i=1}^{N} R_{i_t} - R_{Tot}\right) \tag{15}$$

The necessary condition to find the minimum of J_{tot} is that all partial derivatives with respect to quantization parameters are equal to zero $\forall i \in \{1,...,N\}, \forall t \in \{1,...,T\}$:

$$\frac{\partial J_{Tot}}{\partial \Delta_{i_t}} = \frac{\partial D_{Tot}}{\partial \Delta_{i_t}} + \lambda\frac{\partial}{\partial \Delta_{i_t}}\sum_{t=1}^{T}\sum_{i=1}^{N} R_{i_t} = 0 \tag{16}$$

We express the rate R_{i_t} of a CU i_t as a function of $R^C_{i_t}$ and $R^S_{i_t}$ as the rates of a CU i_t to be coded in Inter/Intra and Skip, respectively. However, rate of skipped CUs is theoretically equal to zero. Thus, we have

$$R_{i_t} = c_{i_t}R^C_{i_t} + (1-c_{i_t})\underbrace{R^S_{i_t}}_{\approx 0} = c_{i_t}R^C_{i_t}. \tag{17}$$

In order to keep formula easy to read, in the following we simply write $c_{i_t}R^C_{i_t} = c_{i_t}R_{i_t}$. If we suppose the independence of rates, which is discussed in appendix B and experimentally validated in Section V-A, (16) is simplified into (18).

$$\frac{\partial J_{Tot}}{\partial \Delta_{i_t}} = \frac{\partial d_{i_t}}{\partial \Delta_{i_t}}c_{i_t}U_{i_t} + \lambda\frac{\partial R_{i_t}}{\partial \Delta_{i_t}}c_{i_t} = 0 \tag{18}$$

The Shannon bound is a R-D model commonly used in video coding for its high mathematical tractability. Further details about this model can be found in [32]. The R-D

Shannon bound is injected into (18) to obtain the optimal λ as (19). Developments are detailed in appendix E.

$$\lambda = 2\ \ln(2) U_{i_t} D_{i_t} \tag{19}$$

To simplify writing, we define $\lambda^{'}$ as

$$\lambda^{'} = \frac{\lambda}{2\ \ln(2)} \tag{20}$$

We then have $\forall i_t$

$$\log_2\left(\lambda^{'}\right) = \log_2\left(U_{i_t} D_{i_t}\right). \tag{21}$$

The RDSTQ aims to keep the average bitrate R_{Tot} of the GOP unchanged. It is achieved if the total rate obtained through RDSTQ is equal to the total rate obtained with a unique quantization step applied to all CUs in the GOP. In next developments from (22) to (31), we exhibit the total GOP rate and further apply the rate constraint.

By summing the log values weighted according to non-Skip probability c_{i_t} on both sides of (21) over all CUs of the GOP, we have

$$\log_2(\lambda') \underbrace{\sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t}}_{=N_{Tot}} = \sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} \log_2\left(U_{i_t} D_{i_t}\right), \tag{22}$$

$$\log_2(\lambda') = \frac{1}{N_{Tot}} \sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} \log_2\left(U_{i_t} D_{i_t}\right). \tag{23}$$

We consider a given CU k_τ and mix (21) with (23):

$$\frac{1}{N_{Tot}} \sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} \log_2\left(U_{i_t} D_{i_t}\right) = \log_2\left(U_{k_\tau} D_{k_\tau}\right) \tag{24}$$

We introduce the R-D Shannon bound as

$$R_{i_t} = -\frac{1}{2} \log_2\left(\frac{D_{i_t}}{\alpha\, \sigma_{i_t}^2}\right), \tag{25}$$

with α the parameter that model the source distribution and $\sigma_{i_t}^2$ the variance of the residue. In order to remove the cumbersome sum of all local distortion logarithms, we compute the $\frac{2R_{Tot}}{N_{Tot}}$ using the R-D Shannon bound.

$$\frac{2R_{Tot}}{N_{Tot}} = \frac{2}{N_{Tot}} \sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} R_{i_t} \tag{26}$$

$$\frac{2R_{Tot}}{N_{Tot}} = \frac{-1}{N_{Tot}} \sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} \left(\log_2\left(D_{i_t}\right) - \log_2\left(\alpha\, \sigma_{i_t}^2\right)\right) \tag{27}$$

The term depending on all local distortions can be eliminated by using (24) and (27) in order to obtain (28).

$$\begin{aligned}\frac{2R_{Tot}}{N_{Tot}} = &-\log_2\left(U_{k_\tau}\right) - \log_2\left(D_{k_\tau}\right) \\ &+ \frac{\sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} \log_2\left(\alpha\, \sigma_{i_t}^2\, U_{i_t}\right)}{N_{Tot}}\end{aligned} \tag{28}$$

This result is necessary for applying the rate constraint. The high bitrate approximation (29) is injected into (28) in order to make appear the quantization parameter QP_{k_τ} as follow:

$$D_{k_\tau} = \frac{\Delta_{k_\tau}^2}{12} = \frac{2^{\frac{QP_{k_\tau}-4}{3}}}{12} \tag{29}$$

$$\log_2\left(D_{k_\tau}\right) = \frac{QP_{k_\tau}-4}{3} - \log_2(12) \tag{30}$$

$$\begin{aligned}\frac{2R_{Tot}}{N_{Tot}} = &-\left(\frac{QP_{k_\tau}-4}{3} - \log_2(12)\right) - \log_2\left(U_{k_\tau}\right) \\ &+ \frac{\sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} \log_2\left(\alpha\, \sigma_{i_t}^2\, U_{i_t}\right)}{N_{Tot}}\end{aligned} \tag{31}$$

To make delta quantizers appear, we consider the case of a GOP encoded with a unique quantization parameter, named QP. We inject (30) into (27) and simplify $QP_{k_\tau} = QP, \forall k_\tau$ to obtain

$$\frac{2R_{Tot}}{N_{Tot}} = \frac{4-QP}{3} + \log_2(12) + \frac{\sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} \log_2\left(\alpha\, \sigma_{i_t}^2\right)}{N_{Tot}}. \tag{32}$$

Since the AQ is designed to be neutral with regards to the average GOP rate, and assuming residue variances are kept unchanged, we can mix (31) and (32) to exhibit the optimal delta quantizer $dQP_{k_\tau} = (QP_{k_\tau} - QP)$ of the CU k_τ.

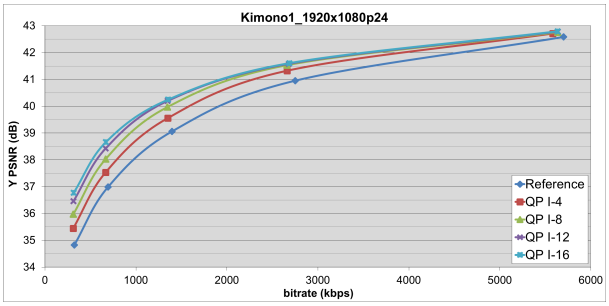
$$dQP_{k_\tau} = -str\left(log_2\left(U_{k_\tau}\right) - \frac{1}{N_{Tot}} \sum_{t=1}^{T}\sum_{i=1}^{N} c_{i_t} log_2\left(U_{i_t}\right)\right) \tag{33}$$

We note that str is called the *strength* and its theoretical optimal value is $str = 3$, coming from the relationship between QP_{k_τ} and Δ_{k_τ}. Increasing or decreasing this value may stretch the quantizers dynamic range and thus modify the R-D efficiency and the TBD. Given the multiple approximations, we have tried several str values and the best results were obtained with $str = 2$. Consequently, we set $str = 2$ for all experiments as a good trade-off between R-D gains and TBD.

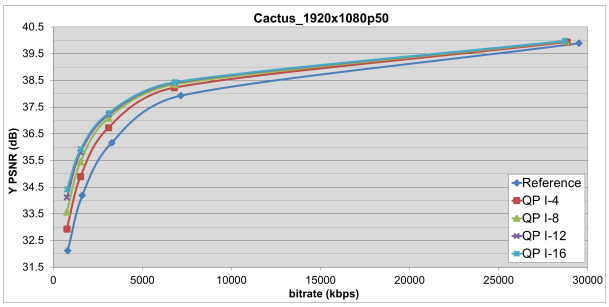
The RDSTQ algorithm is based on the temporal propagation model presented in Section III. In considering such a model into the local quantization problem presented in (11), we are able to efficiently improve the overall R-D coding efficiency. Thanks to the analytical solution, optimal delta quantizers are easily estimated based on the look-ahead and do not require extensive multi-pass analysis. Moreover, as shown in the next section, the range of delta quantizers is bounded and controllable.

V. Model Discussion

This section aims to provide justification and validation for some of the simplifications or assumptions made in the previous section. It is divided into five subsections. First, the independence of rates hypothesis considered during the analytical solution development is validated through experiments. Second, the estimation of inter probability is discussed with



(a)



(b)

Fig. 1. R-D curves of non-Intra frames according to Intra QP offsets with (a) *Kimono* and (b) *Cactus* sequences

the support of ground truth data extracted from off-line encodings. Third, the Skip mode probability is discussed. Then, the look-ahead which provides necessary input parameters for the RDSTQ to compute delta quantizer is discussed, with details of its implementations into x265 and HM. Most notably, we demonstrate that the range of delta quantizers is bounded and can be controlled beforehand.

A. Experimental proof of rates independence

The theoretical explanation is given appendix B. An experiment was conducted in order to evaluate the correctness of the rate independence assumption. To do so, R-D curves of non-Intra frames in a GOP are generated with a fixed QP configuration while different QP offsets are set on the Intra frame, in the set $[0; -4; -8; -12; -16]$. The *Reference* R-D curve corresponds to the 0 offset case. The experiment was conducted into HM encoder with Intra coding disabled in non-Intra frames. The objective of this experiment is to confirm that increasing quality on Intra frame shifts the R-D curves of depending frames towards less distortion without rate deviation.

Fig. 1 shows experimental results for *Kimono* (a) and *Cactus* (b) video sequences in RA configuration with hierarchical 3-B. These curves show that R-D points are aligned along the rate axis whatever the QP offset on the Intra frames. Consequently, temporal dependency between CUs only impacts distortions and not rates. This validates the assumption of rate independence applied in Section IV-B.

B. Inter Probability Estimation

In this section, we present the Inter probability estimators considered in [14] and in [15]. $\omega_{i_t}^{Intra} > 0$ and $\omega_{i_t}^{Inter} > 0$

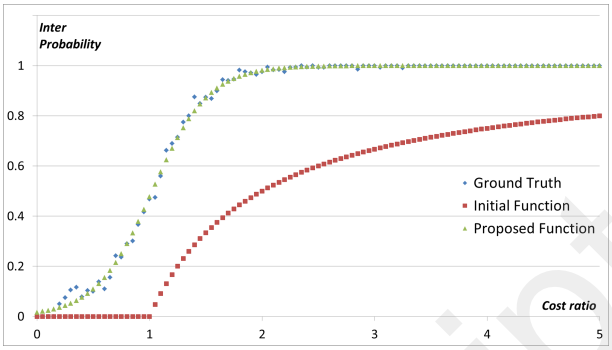


Fig. 2. Inter Probability p_{i_t} according to cost ratio r_{i_t} estimated by (34) for the initial function and (35) for the proposed function.

are defined as the SATD prediction costs of Intra and Inter modes, respectively. The SATD costs are estimated in the look-ahead analysis. Probability of Inter prediction mode is defined as a function of the ratio $r_{i_t} = \omega_{i_t}^{Intra}/\omega_{i_t}^{Inter}$. The Inter probability estimator used in [14] is given by

$$p_{i_t} = 1 - min(1; \frac{1}{r_{i_t}}). \tag{34}$$

This formula implies that if SATD costs are equivalent, i.e. $r_{i_t} = 1$, Inter probability should be null and there is no propagation, i.e. $p_{i_t} = 0$. However, close Intra/Inter prediction costs should intuitively lead to equiprobable Intra and Inter modes. Moreover, neither theoretical nor experimental proof of the correctness of (34) has been given. *Bichon et al.* propose in [15] to improve the Inter probability estimation.

Based on statistical inference, Inter probability p is estimated, from an off-line RDO analysis, as the Likelihood function $L(r|mode) \propto P(mode = Inter|r)$. Fig. 2 compares both functions p_{i_t} and p. r_{i_t} is the prior information known beforehand while the event for a CU i_t to be Inter coded is the evidence. We observe in Fig. 2 that (34) is quite far from the ground truth. Consequently, another function defined by (35) is proposed, which is a sigmoid distribution fitting the ground truth curve. The fitting is done off-line.

$$p_{i_t} = \frac{1}{1 + a\,\mathrm{e}^{-b\,r_{i_t}}} \tag{35}$$

Function (35) is plotted on Fig. 2. The performance of this function is discussed in Section VI. a and b are model parameters set for our experiments to $a = 0.5651$ and $b = 3.6064$.

C. Skip probability justification

To achieve an analytical solution, the high rate assumption is used to estimate the quantizer from the distortion model. Despite its mathematical tractability, such assumption is debatable and does not stand for low bitrates. To be more robust to different use cases, we consider the distortion formula of *Xu et al.* proposed in [33] that is given in (36), with $\sigma_{i_t}^2$ being the variance of the input sample.

$$D_{i_t} = \frac{\sigma_{i_t}^2 \Delta_{i_t}^2}{12\sigma_{i_t}^2 + \Delta_{i_t}^2} = \frac{\Delta_{i_t}^2}{12} \times \underbrace{\frac{12\sigma_{i_t}^2}{12\sigma_{i_t}^2 + \Delta_{i_t}^2}}_{c_{i_t}} \tag{36}$$

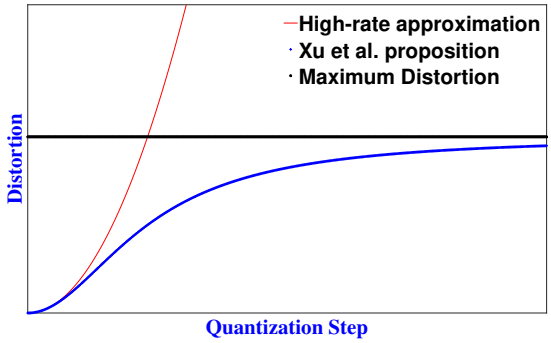


Fig. 3. High-Rate approximation and *Xu et al.* [33] model of the Distortion-Quantization relationship, with distortion expressed in MSE, against the maximum achievable distortion $\sigma_{i_t}^2$.

This proposal is close to the approximated distortion (29) in the high rate case but does not suffer from the same inaccuracy in the low-rate case, as exposed in Fig. 3. Indeed, the distortion is limited by $\sigma_{i_t}^2$ the input variance, but the high rate approximation suggests overcoming this maximum at some point.

This model makes it difficult to extract the delta quantizer based on the distortion. However, the chosen non-Skip probability c_{i_t} scales the distortion, using the high rate assumption, into the desired distortion, as shown by the equality in (36). Using developments described in (12), (13) and (14) allow to keep the analytical solution simple while using a more robust distortion model.

D. Look-Ahead Design

In this section we give more insights about the look-ahead implementation in the x265 and the look-ahead we developed in the HM.

The analytical solution explained above provides the optimal set of local quantizers to the encoder from an R-D standpoint. However, several input parameters, depending on source characteristics, are required prior to compute these quantizers. The look-ahead is a common sub-process designed to estimate such parameters, based on a pre-analysis which mimics the encoder behavior. This operation requires to spare some of the computational resources, but it can be multi-threaded and the obtained data can drive the encoder to faster convergence towards optimal decisions. Due to the algorithm requirements, a look-ahead was used in both x265 and HM implementations.

The x265 encoder already encloses an efficient look-ahead. Videos are first down-sampled in order to divide the height and width of original pictures by 2. Low-resolution frames are partitioned into 8x8 blocks and each block is analyzed in Intra and Inter modes. Intra and Inter modes are compared based on SATD costs. For both Intra mode and Inter motion estimation, fast analysis is used and based on dichotomous approaches.

In the HM encoder, no look-ahead is currently available. Taking advantage of available tools in the HM, we successfully emulated a look-ahead to extract the necessary information. Our look-ahead is configured as follow:

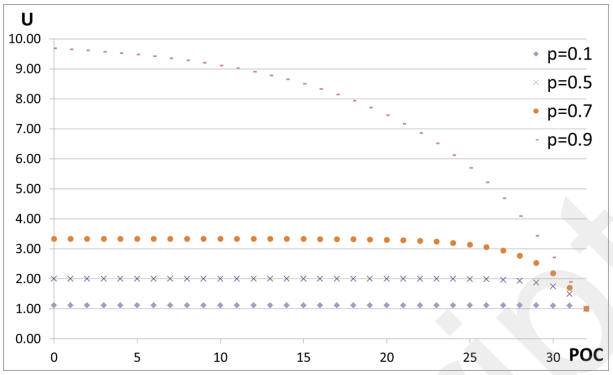


Fig. 4. U value evolution within a GOP of 32 frames for different values of p, with $\Psi_{i_t} = 1\,\forall i, t$.

- No QuadTree: only 16x16 CUs are used
- All modes are analyzed in SATD and use source signal for reference prediction
- No bit stream is actually written since no reconstructed data are required
- All necessary values are stored in a look-ahead file

The HM look-ahead is finally achieved by parsing this look-ahead file. The proposed look-ahead in HM is more accurate, in terms of correlation with the actual encoder decisions, compared to the x265 one. Consequently, better R-D efficiency is observed for the HM, as shown in Section VI-A. The computational overhead of this pre-analysis is around 30% of the HM encoding complexity. However, this complexity increase is usually very manageable for real industrial implementations, first thanks to the efficient use of multi-threading, and second by leveraging on look-ahead informations to speed up the main encode decisions.

E. Quantizer dynamic range

In this section, the dynamic range of delta quantizers is analyzed. We prove that delta quantizers obtained through the model are bounded. The output dynamic range of delta quantizers is predictable before to encode. This property helps to prevent from any conformance issue or boundary defect.

Let assume a sequence as temporally stable, i.e. probability of Inter mode is equal for all CUs with identical spatial positions in different frames. We have seen in Section III that accumulation factor is recursively defined by (10).

For the sake of simplicity, let $\Psi_{i_t} = 1\,\forall i, t$. If we assume all probabilities in the same spatial area to be equal to p, i.e. within the temporally stable part of a picture, then we obtain

$$U_{i_t} = \sum_{k=t}^{T} p^{T-k}. \tag{37}$$

Under the assumption that $p_{i_t} = p$, $\forall i, t$, U_{i_t} is a geometrical series. Knowing that $p \in [0...1]$, we finally obtain

$$U_{max} = \lim_{(t,T)\to(0,\infty)} U_{i_t} = \frac{1}{1-p}. \tag{38}$$

This equation suggests that, by design, U_{i_t} converges at a maximum value U_{max}, that depends on source characteristics,

TABLE I
CODING EFFICIENCY OVER NO LOCAL QUANTIZATION IN *HM*.

		Rate Distortion Temporal Quantization (RDTQ)				RDSTQ			
BD-BR	Probability Model	Initial	Initial + skip	Proposed	Proposed + skip	Initial	Initial + skip	Proposed	Proposed + skip
	Class A (8bits)	-10.35%	-10.16%	-14.15%	-14.02%	-19.26%	-18.76%	-28.02%	-26.45%
	Class B	-7.55%	-7.57%	-12.02%	-12.97%	-13.42%	-13.26%	-22.30%	-23.24%
	Class C	-15.24%	-15.25%	-19.01%	-19.20%	-21.97%	-21.62%	-30.37%	-30.26%
	Class D	-13.95%	-13.48%	-16.52%	-16.08%	-23.65%	-22.67%	-30.93%	-29.28%
	Class E	-14.08%	-13.30%	-22.67%	-21.03%	-14.28%	-13.37%	-23.78%	-21.81%
	Average	**-12.08%**	**-11.83%**	**-16.58%**	**-16.51%**	**-18.38%**	**-17.84%**	**-26.89%**	**-26.26%**
	Best	**-21.04%**	**-21.51%**	**-26.57%**	**-26.68%**	**-30.38%**	**-29.89%**	**-40.90%**	**-39.65%**
	Worst	**-3.36%**	**-7.65%**	**-7.59%**	**-7.86%**	**-10.71%**	**-9.61%**	**-18.65%**	**-17.77%**
		RDTQ				RDSTQ			
TBD	Probability Model	Initial	Initial + skip	Proposed	Proposed + skip	Initial	Initial + skip	Proposed	Proposed + skip
	Class A (8bits)	20.32%	5.86%	43.82%	12.40%	10.28%	4.39%	32.65%	9.57%
	Class B	22.12%	12.92%	64.98%	31.98%	9.89%	5.38%	49.57%	19.43%
	Class C	16.69%	5.36%	41.46%	12.93%	9.52%	3.39%	33.93%	8.26%
	Class D	23.53%	7.65%	50.56%	15.09%	18.96%	4.09%	46.51%	7.28%
	Class E	44.65%	23.84%	113.77%	42.87%	13.11%	4.10%	75.56%	16.54%
	Average	**24.78%**	**11.10%**	**62.33%**	**23.63%**	**12.40%**	**4.33%**	**47.87%**	**12.67%**

under the assumption that T is large enough. We also notice that the lower the p value, the faster U converges. We report in Table II, for a given value of p, the maximum achievable weight U_{max} reached once the number of frames in the GOP equals or exceed N_{conv}. The convergence is assumed with a two decimal places precision.

We can see the consequence of such convergence on Figure 4. If one increases the size of the stack T, as long as the sequence is temporally stable, reference frames ultimately have an equal level of importance within the GOP. We note that such convergence is most likely to occur for small p value, i.e. sequences difficult to predict temporally.

In the special case of $p = 1$, the U_{max} value only depends on the GOP length: $U_{max} = T$. Once U_{max} is estimated, the dynamic range of delta quantizer rng_{dQP} is given by

$$rng_{dQP} = -str\left(log_2\left(U_{max}\right)\right) \tag{39}$$

with str being the strength mentioned in Section IV-B. Based on this formula, one may choose to control the dynamic of the delta quantizers by directly modifying the strength value. In our experiments $str = 2$.

VI. EXPERIMENTS

This section aims to validate the coding efficiency of the proposed solution, assess the TBD reduction and confirm the expected behavior of the model. First, influences of the proposed probability from [15], described in Section V-B, and the Skip probability consideration are evaluated. Second, the rate distribution between frames of the GOP is observed. Third, the positive impact of Ψ function on overspent rate situation is confirmed. Finally, the method is compared to state-of-the-art methods, thanks to the proposed HM implementation.

TABLE II
THEORETICAL NUMBER OF FRAMES N_{conv} REQUIRED FOR U CONVERGENCE BASED ON p VALUES AND RELATED U_{max}

p value	N_{conv}	U_{max}
0.1	3	1.11
0.2	4	1.25
0.3	5	1.43
0.4	8	1.67
0.5	9	2
0.6	13	2.5
0.7	17	3.33
0.8	31	5
0.9	73	10
1.0	NaN	T

The x265 software HEVC encoder [6] is used in the experiments in order to have similar test conditions as [14] and [15]. The HM encoder [5] is also used to confirm results in a different encoder. Common Test Conditions (CTC) defined by the JCT-VC [34] have been followed. Videos are encoded in RA coding configuration, with 3 hierarchical B, for five QP values $\in \{22, 27, 32, 37, 42\}$. The QP value of 42 was added to highlight the Skip mode influence since it is statistically more used at low bitrate.

When no psycho-visual function is considered, i.e. $\Psi_{i_t} = 1, \forall i, t$, the model is simply called RDTQ, since spatial criteria is ignored, and the BD-BR is computed using the PSNR metric. Otherwise, the model is called RDSTQ and the BD-BR is computed using the SSIM metric, that is better correlated with HVS perception of quality. In the case of RDSTQ we set $\Psi_{i_t} = 1/\sigma^2_{i_t}$, with $\sigma^2_{i_t}$ being the local variance of source luminance pixels of the block i_t. *Yeo et al.* [35] proved that weighting a MSE distortion by the inverse of local pixel block variance specifically optimizes the SSIM metric, which explains our choice of Ψ function.

A. Coding Efficiency

Coding performance is measured using the BD-BR metric [36]. A negative BD-BR value reflects the percentage of bitrate savings achieved at equivalent YUV distortion, between the anchor and the proposed solution. The BD-BR results and the corresponding target bitrate deviations, TBD, averaged on the considered QP values are presented in Table I for HM and in Table III for x265.

The anchors are respectively the x265 and HM encoders without AQ algorithm. The two Inter probability models,

TABLE III
CODING EFFICIENCY OVER NO LOCAL QUANTIZATION IN *x265*.

		RDTQ (No psycho-visual function)				RDSTQ			
BD-BR	Probability Model	Initial	Initial + skip	Proposed	Proposed + skip	Initial	Initial + skip	Proposed	Proposed + skip
	Class A (8bits)	-7.96%	-7.90%	-10.13%	-9.92%	-20.00%	-19.34%	-25.55%	-23.87%
	Class B	-6.90%	-6.93%	-7.36%	-8.48%	-16.79%	-16.38%	-20.43%	-20.96%
	Class C	-13.97%	-13.69%	-15.08%	-15.12%	-24.12%	-23.27%	-28.71%	-27.88%
	Class D	-11.42%	-10.91%	-12.24%	-11.77%	-25.49%	-23.82%	-30.39%	-27.57%
	Class E	-11.63%	-10.89%	-15.96%	-14.28%	-12.94%	-11.58%	-19.20%	-16.43%
	Average	**-10.38%**	**-10.08%**	**-11.90%**	**-11.81%**	**-20.07%**	**-19.09%**	**-24.85%**	**-23.53%**
	Best	**-19.00%**	**-19.01%**	**-22.38%**	**-22.08%**	**-30.87%**	**-30.11%**	**-39.01%**	**-36.60%**
	Worst	**+0.24%**	**-1.59%**	**+1.48%**	**-3.23%**	**-10.58%**	**-8.62%**	**-14.18%**	**-12.11%**
		RDTQ (No psycho-visual function)				RDSTQ			
TBD	Probability Model	Initial	Initial + skip	Proposed	Proposed + skip	Initial	Initial + skip	Proposed	Proposed + skip
	Class A (8bits)	28.04%	7.16%	50.96%	11.75%	12.41%	5.13%	33.09%	7.55%
	Class B	37.17%	10.63%	80.63%	22.19%	14.90%	8.13%	53.24%	14.89%
	Class C	23.30%	4.72%	47.71%	8.08%	10.37%	8.87%	34.20%	10.89%
	Class D	31.09%	5.32%	60.65%	9.82%	21.85%	8.57%	51.47%	9.12%
	Class E	71.03%	25.82%	138.29%	35.90%	14.89%	5.51%	69.26%	8.09%
	Average	**37.37%**	**10.28%**	**75.19%**	**17.43%**	**15.16%**	**7.62%**	**49.05%**	**10.77%**

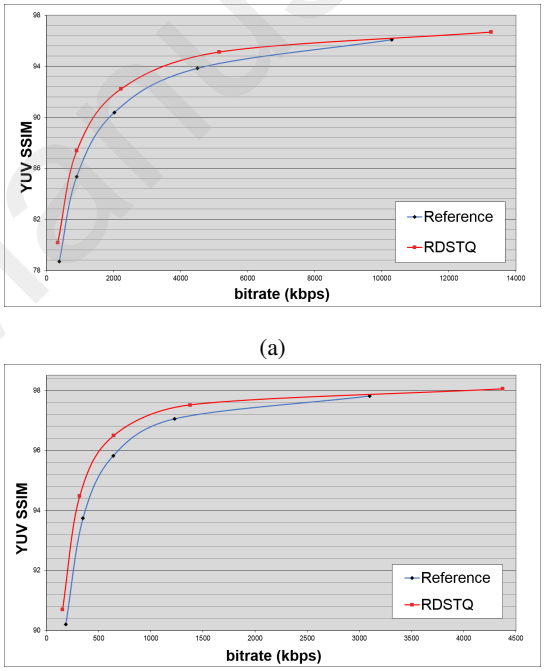
defined in (34) and (35) are compared and respectively named *Initial probability* and *Proposed probability*.

From Table I we can observe higher bitrate savings for the Proposed probability (35) over the Initial probability (34), whether the Skip mode consideration is enabled or not. The Proposed probability (35) saves in average -4.5% PSNR-based BD-BR compared to the Initial probability with RDTQ and -8.61% SSIM-based BD-BR compared to the Initial probability with RDSTQ. When Skip is considered, performances suffer from an average bitrate increase between 0.07% and 0.25% for RDTQ and between 0.54% and 0.63% for RDSTQ.

The TBD shows 2 to 4 times higher deviation when using the Proposed probability compared to Initial probability. Indeed, Proposed probability induces larger propagation and consequently smaller delta quantizers, i.e. more rates, on reference frames. The consideration of Skip probability efficiently reduces the TBD, and then the Proposed probability provides similar TBD as the Initial probability while maintaining BD-BR gains. The average TBD for RDTQ and RDSTQ is equal to 23.63% and 12.67%, respectively.

Despite the significant TBD decrease enabled by using both spatial criteria and Skip probability, average values remain quite high. Class B and Class E sequences suffer a deviation above 15%. R-D curves of two sequences, *ParkScene* and *FourPeople*, are presented in Fig. 5. We observe that TBD is significantly higher at high rates than low rates. The TBD at low rates is well managed by the Skip probability consideration, but the RDSTQ model behavior is not as good for high rate. At high bitrate, two observations can be made:

1) The Skip probability has no more influence (i.e. $c_{i_t} \rightarrow 1$). The delta quantizer dynamic over the GOP (i.e. rate allocation) depends only on source statistics and strength parameter, not on the target average quantizer set for GOP (i.e. target rate). The strength is set only once for the entire test set and range of target QPs.
2) Based on R-D Shannon bound, a small variation of distortion induces a large rate variation. Typically, a small additional improvement in quality requires a large bitrate increase. In our context of global RDO, to slightly improve the quality on reference frames, the necessary rate increase is too large for significantly improving the global coding efficiency of dependent frames.



(a)

(b)

Fig. 5. RDSTQ and no-AQ R-D curves in HM for (a) *ParkScene* and (b) *FourPeople*

Consequently, to improve the model at high bitrate, we should lower the dynamic of delta quantizers (or rate allocation), and adapt the dynamic based on target quantization (or rate). Several options should be investigated, such as adapting the strength parameter based on target QP or use an Inter probability dependent of the target QP.

To estimate quality gains, we also provide the results for HM encoder, with BD-PSNR (without psycho-visual factor) and with BD-SSIM (with the psycho-visual factor) in Table IV.

(a) I-frame t=0 (b) B-frame t=9 (c) B-frame t=18 (d) P-frame t=23

Fig. 6. Comparisons of subjective visual fidelity on *ParkScene* sequence encoded at 500 kbps. Reference is x265 without adaptive quantization on top. Test is x265 with RDSTQ on bottom.

These results are obtained by using both the Proposed Inter probability and the Skip probability. We can observe average BD-PSNR and BD-SSIM improvements of 0.59 and 1.36, respectively. In addition, we proceeded in visual quality comparisons of RDSTQ encoded bit-streams against reference bit-streams without AQ. The encoded bit-streams based on x265 are available at: https://github.com/MaximeBichon/RDSTQ. The proposed RDSTQ method significantly improves the visual quality, as detailed for one example Fig. 6. Fig. 6 compares multiple pictures of the *ParkScene* sequence for RDSTQ against the reference without AQ, both at identical average bitrates. For every frame, the blocking and pattern artifacts are efficiently reduced on trees and less blurring is also observed on the floor and background. Overall, it results in more temporal quality consistency and sharpness, which is consistent with RDSTQ motivation on better considering temporal distortion (quality) propagation.

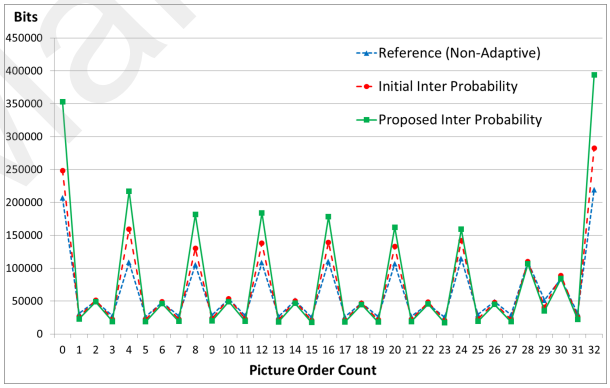
Observations from x265 experiments depicted in Table III leads us to similar conclusions. The average bitrate savings for RDTQ and RDSTQ are respectively -1.43% PSNR-based BD-BR and -3.46% SSIM-based BD-BR. The TBD is however reduced compared to the Initial probability without Skip consideration. The TBD is reduced from 37.37% to 17.43% with RDTQ and from 15.16% to 10.77% with RDSTQ.

As desired, the Proposed probability improves the coding efficiency while the Skip mode consideration efficiently reduces the TBD. We demonstrate in this paper that the model stands whatever the codec implementation or the Ψ scaling factor used.

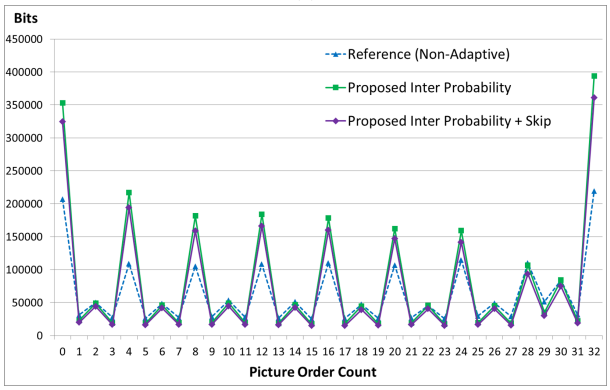
TABLE IV
AQ CODING EFFICIENCY OVER NO AQ IN *HM*.

Sequences	BD-PSNR for RDTQ	BD-SSIM for RDSTQ
Class A (8bits)	0.50	1.17
Class B	0.33	0.86
Class C	0.81	1.98
Class D	0.67	2.12
Class E	0.68	0.47
Average	**0.59**	**1.36**
Best	**1.18**	**2.77**
Worst	**0.18**	**0.35**

B. Frame Rate Distribution



(a)



(b)

Fig. 7. Rate distribution of first GOP frames with sequence *RaceHorses* at $QP = 32$ for (a) RDTQ with Initial Probability and RDTQ with Proposed Probability; for (b) Proposed Probability with and without Skip consideration.

In this section, we discuss the distribution of rates for several frames, typically a whole GOP, when some of the models

presented above are enabled. The sequence *RaceHorses* with resolution of 832x480 is used for experiments in this section.

As expected, we observe in Fig. 7 (a) that RDTQ model allocates more rate on the reference frames and lower temporal layer, while it decreases the rate allocated to frames in the highest temporal layers. The Proposed probability, propagating more *weight* on reference frames, tends to stretch even more the bitrate distribution across temporal layers.

When considering the Skip probability, rates are equally decreased for each type of frame as observed in Fig. 7 (b), but it does not alter the delta rates between frames. This behavior is expected since Skip mode consideration aims to limit the overspent rate on the entire GOP.

C. Ψ function and QP spatial distribution

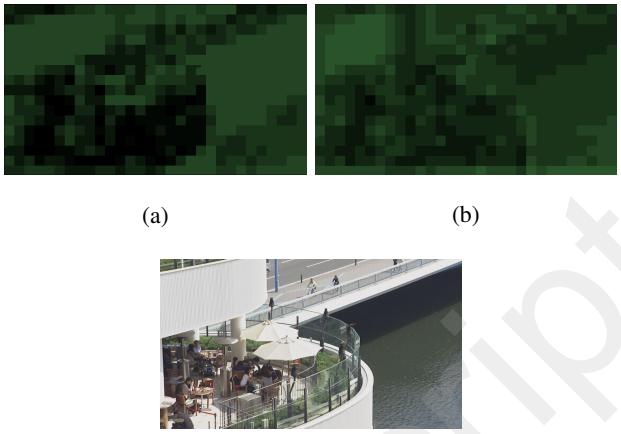
In this section, more insights are given about the impact of the Ψ function on QPs spatial distribution. We observe the distribution of quantizers over an entire frame when the Ψ function is enabled. As early introduced, the psycho-visual factor chosen here is based on local spatial pixel variance, and is dedicated to optimize SSIM score. It has the property to consider spatial masking effect, i.e. the fact that human eyes are less sensible to distortion made on high textured area (high local variance) than on area of low spatial complexity (low local variance). Spatial masking significantly impacts compression artifact perception, as further analyzed by *Rimac-Drlje et al.* [37].

The distribution of quantizers with and without psycho-visual function is shown on Fig. 8 for the frame 128 of *BQTerrace* sequence, with $QP = 22$. The darker blocks have the lowest quantizer (high rate) and the brighter ones have the highest quantizer (low rate). We point out that for this particular sequence encoded at $QP = 22$, almost no block is coded in Skip mode. Hence, we can keep apart the influence of the model Skip estimation in this analysis.

We observe on Fig. 8 (a) that if no psycho-visual function is considered, the terrace is affected with high quality while the water and the roof are quantized more aggressively. The Inter probability is based on the relative difference between Intra mode and Inter mode estimated complexities. The more the Intra complexity is high compared to the Inter one, the more importance is put on reference frames. Given that, *BQTerrace* is highly uniform in terms of temporal complexity but not in terms of spatial complexity, it explains why more quality is affected to spatially complex areas, such as the terrace in this case.

However, the more textured is a block, the less distortions are visible by the human eye. When the psycho-visual function is enabled (Fig. 8 (b)), we observe a more balanced distribution of the quantizers over the frame. Less rate is overspent on the terrace, while the water is subject to a quality improvement, according to the spatial masking effect.

In our experiments we focus on the spatial masking effects based on local pixel variance, that correlates well the SSIM quality metric. However, RDSTQ may be used to optimize any other perceptual criteria based on the selection of a Ψ factor that scale well the MSE. For the interested readers, *Winkler* [38] provides a good overview of possible vision model and perceptual metrics to consider.



(a) (b)

(c)

Fig. 8. QP distribution over the frame 128 of *BQTerrace* sequence (a) without psycho-visual function and (b) with psycho-visual enabled. (c) The source frame

D. Comparison to state of the art

This section compares our method with some state of the art solutions. In order to be comparable to other methods found in the literature, the coding scheme was changed for the 7-B hierarchical and QP values $\in \{22, 27, 32, 37\}$. Other coding parameters remain the same and the reference is the HM encoder without AQ algorithm.

Three methods were chosen for comparison. The first one is proposed by *Gao et al.* [27] and designed for optimizing the PSNR. Two other methods designed for optimizing the SSIM are proposed by *Yeo et al.* [30] and *Xiang et al.* [31]. The proposed solution denoted as *Ours* in the results is the RDSTQ improved by the Proposed Inter and Skip probabilities. Simulation results are presented in Table V, with methods reference numbers, for the HM encoder.

Our proposed solution substantially outperforms the SDTP optimized for RA coding configuration by -11.51% in terms of PSNR BD-BR in average. The main reason is the simplified estimation of the dependencies made in [27] that extrapolates the dependency network instead of building it through a look-ahead as proposed. Consequently, *Gao et al.* solution saves some computational complexity by avoiding the use of a

TABLE V
CODING EFFICIENCY COMPARISON OVER NO AQ IN *HM*.

Classes	BD-BR PSNR		BD-BR SSIM		
	[27]	Ours	[30]	[31]	Ours
Class A	-4.25%	-12.53%	-5.78%	-5.42%	-27.98%
Class B	-4.10%	-9.35%	-4.18%	-3.12%	-20.82%
Class C	-5.60%	-17.56%	-3.90%	-5.13%	-29.78%
Class D	-4.10%	-15.82%	-4.47%	-4.53%	-30.86%
Class E	-8.40%	-25.09%	-2.86%	-0.25%	-27.37%
Average	**-5.18%**	**-15.59%**	**-4.14%**	**-3.69%**	**-26.93%**

TABLE VI
COMPARISON OF ENCODING RUNTIME IN SECONDS WITH NO-AQ AND RDSTQ IN *HM*.

Sequences	No-AQ	RDSTQ		Runtime
	Encoding	Encoding	Lookahead	relative offset
RaceHorses_480	8092.67	7857.36	2081.604	22.81%
	6830.27	6546.99	2081.604	26.33%
	5846.35	5593.69	2081.604	31.28%
	5135.56	4849.16	2081.604	34.96%
	4498.13	4163.41	2081.604	38.84%
BasketballDrill	9725.36	9055.82	3839.134	32.59%
	8431.55	7784.01	3839.134	37.85%
	7418.90	6856.51	3839.134	44.17%
	6635.51	6085.56	3839.134	49.57%
	5973.31	5407.38	3839.134	54.80%
BlowingBubbles	2298.63	2079.30	1200.608	42.69%
	1837.58	1681.99	1200.608	56.87%
	1530.68	1403.72	1200.608	70.14%
	1313.54	1206.92	1200.608	83.29%
	1152.55	1071.54	1200.608	97.14%
BasketballPass	2456.23	2334.28	1396.372	51.89%
	2164.93	2048.26	1396.372	59.11%
	1918.05	1795.33	1396.372	66.40%
	1699.37	1577.95	1396.372	75.03%
	1510.91	1377.15	1396.372	83.57%
KristenAndSara	15247.84	14950.88	7617.788	48.01%
	13175.36	12584.20	7617.788	53.33%
	12380.74	11437.62	7617.788	53.91%
	11934.06	11155.10	7617.788	57.31%
	11605.41	10978.87	7617.788	60.24%
Average	-	-	-	**53.28%**

look-ahead but greatly limits the efficiency of the encoding optimization.

In terms of SSIM, the proposed solution outperforms the two methods by more than -22% in average. However, an important drawback is that both methods consider rate constraint on a frame basis and not a GOP basis, which forbids bit transfer between frames. Consequently, these AQ methods are more constrained that our proposal, even if *Xiang et al.* [31] implicitly try to consider the temporal dependencies through Inter mode SATD estimation. The large difference in coding efficiency confirms that GOP optimization is much more efficient than frame optimization. Moreover, even if GOP optimization requires a more complex look-ahead than frame optimization, such implementation are very acceptable in industrial applications.

E. Encoding Complexity

We provide a rough comparisons of HM encoding runtime of RDSTQ with Proposed Inter probability and Skip probability for sequences in classes C, D and E in Table VI. Different sequences are tested with $QP \in \{22, 27, 32, 37, 42\}$. We observe that encoding runtime is higher from 53.28% in average. This increase mostly comes from the look-ahead and is more noticeable for low bitrate. We point out that the look-ahead complexity comes from the data writing in a separated file, further read by the HM for RDSTQ. Thus, an embedded look-ahead would not be as complex as the one we proposed. Moreover, in a multi-threaded implementation, such overhead would be neglected.

VII. CONCLUSION

We show through extensive experimentations the benefits of considering both Skip probability and accurate Inter probability estimators for AQ. Our model relies on an analytical solution for delta quantizers, thoroughly demonstrated in this paper. It provides substantial bitrate savings whatever the HEVC encoder implementations. Considering the RDTQ, i.e $\Psi_{i_t} = 1 \, \forall i, t$, we report systematic BD-BR gains of -1.43% and -4.43% PSNR-based in the x265 and the HM encoders, respectively. We obtain these gains against the initial method proposed by *Ropert et al.* [14].

Thanks to the convenient consideration of a psycho-visual factor, the RDSTQ also allows to optimize more perceptually-oriented quality metric, such as the SSIM. When using a psycho-visual factor based on the local pixel variance, that estimates the spatial masking, BD-BR gains based on SSIM are then of -3.46% and -8.61% for the x265 and the HM, respectively. Careful comparison against start-of the art similar approaches is also reported. RDSTQ model outperforms previous techniques with -10.41% PSNR-based and -22.79% SSIM-based average bitrate savings, when reference is without AQ. The main conclusion coming out from these experiments is the higher efficiency of the GOP optimization compared to the frame optimization; GOP optimization being closer to the global optimization bound. Subjective quality assessments demonstrate consistent spatial and temporal quality improvements thanks to the RDSTQ.

We prove that the Skip probability consideration is an efficient way to make the distortion model more robust. It helps to reduce the average TBD of the RDSTQ with Initial probability from 12.4% to 4.33% in the HM and from 15.16% to 7.62% in the x265. Finally, we demonstrate that computed delta quantizers based on the proposed model are bounded. Their output dynamic range is controllable, preventing from any worst case scenario. The TBD can be further reduced, especially at high rate, by using a more complex model taking into account the target quantizer.

However, the proposed model is simplified by assuming all blocks to be of the same size, notably during the look-ahead analysis. In HEVC, the QuadTree partitioning is known to be a key tool in terms of coding efficiency. Obviously, we assume that predicting the QuadTree based non-uniform partitioning into the look-ahead for refining the distortion propagation model and quantizer computation would bring substantial coding gains. Our future work will address this problematic, while taking care of the computational complexity related to the QuadTree partitioning estimation.

APPENDIX A
MULTIPLICATIVE DEPENDENCY MODEL

We define i the current CU and j its reference CU used for temporal prediction. We define ϵ_i as the residue on i before quantization, ϵ_j as the reconstruction error on j and ϵ_{src_i} as the error obtained by motion compensation between original blocks i and j. By definition, residue energy before quantization σ_i^2 is equal to

$$\sigma_i^2 = \mathbb{E}\left[\epsilon_i^2\right] = \mathbb{E}\left[\left(\epsilon_j + \epsilon_{src_i}\right)^2\right], \tag{40}$$

$$\sigma_i^2 = \underbrace{\mathbb{E}\left[\epsilon_j^2\right]}_{D_j} + \underbrace{\mathbb{E}\left[\epsilon_{src_i}^2\right]}_{\sigma_{src_i}^2} + 2\mathbb{E}\left[\epsilon_j \epsilon_{src_i}\right]. \tag{41}$$

We assume that the motion compensation error $\epsilon_{src_i}^2$ on source blocks is not correlated with the reconstruction error ϵ_j^2 on reference images. Then, $\mathbb{E}\left[\epsilon_j \epsilon_{src_i}\right] \approx 0$. If we introduce the R-D Shannon bound (42), we can write (43).

$$R_i = -\frac{1}{2}\log_2\left(\frac{D_i}{\alpha\,\sigma_i^2}\right), \tag{42}$$

$$\begin{aligned} D_i = \alpha\sigma_i^2 2^{-2R_i} &= \underbrace{\alpha\sigma_{src_i}^2 2^{-2R_i}}_{d_i} + \alpha D_j 2^{-2R_i} \\ &= d_i + \frac{\alpha\sigma_{src_i}^2 2^{-2R_i}}{\sigma_{src_i}^2} D_j \\ &= d_i + \frac{d_i}{\sigma_{src_i}^2} D_j \end{aligned} \tag{43}$$

If we assume no dependencies in intra coding, i.e. $D_i = d_i$, and that the probability of a CU to be coded in Inter is equal to p_i, the distortion can be expressed as

$$D_i = d_i\left(1 + \frac{p_i}{\sigma_{src_i}^2} D_j\right). \tag{44}$$

Appendix B
Discussion of the independence of rates

We consider a CU indexed by i and its reference CU with distortion D_{ref}. According to the Shannon R-D function, the rate R_i of the CU i can be expressed as in (45).

$$R_i = -\frac{1}{2}\log_2\left(\frac{D_i}{\sigma_i^2}\right) \tag{45}$$

We consider equations (41) and (44) to express the rate R_i as (46).

$$\begin{aligned} R_i &= -\frac{1}{2}\log_2\left(\frac{d_i\left(1 + \frac{p_i D_{ref}}{\sigma_{src_i}^2}\right)}{\sigma_{src_i}^2 + p_i D_{ref}}\right) \\ &= -\frac{1}{2}\log_2\left(\frac{\frac{d_i}{\sigma_{src_i}^2}\left(\sigma_{src_i}^2 + p_i D_{ref}\right)}{\sigma_{src_i}^2 + p_i D_{ref}}\right) \\ &= -\frac{1}{2}\log_2\left(\frac{d_i}{\sigma_{src_i}^2}\right) \end{aligned} \tag{46}$$

Consequently, since d_i and $\sigma_{src_i}^2$ does not depend on D_{ref}, we can assume the rate R_i is independent from D_{ref}.

Appendix C
Computing D_{Tot}

Notation are simplified by defining $s_{i_t} = c_{i_t} d_{i_t} + (1 - c_{i_t})\,\sigma_{src_{i_t}}^2$. We start from the distortion defines with the temporal distortion propagation model as below,

$$D_{i_t} = s_{i_t} + p_{i_t} \sum_{j_{t_{ref}} \in Ref(i_t)} r_{j_{t_{ref}}, i_t} D_{j_{t_{ref}}}. \tag{47}$$

We refine the overlapping ratio $r_{j_{t_{ref}}, i_t}$ as follow:

$$r_{j_{t_{ref}}, i_t} = \begin{cases} 0 & \text{if} \quad j_{t_{ref}} \notin Ref(i_t) \\ r_{j_{t_{ref}}, i_t} & \text{if} \quad j_{t_{ref}} \in Ref(i_t) \end{cases} \tag{48}$$

For sake of simplification, the belonging to the reference image is removed from equations to lighten the notation. At the same time, as now the sum occurs on the entire image, given the fact that the referencing of contributing CU is carried by the ratio $r_{j_{t_{ref}}, i_t}$, Then j_{t-1} is a silent index, and can be replaced by i_{t-1}. The adopted notation becomes:

$$D_{i_t} = s_{i_t} + p_{i_t} \sum_{i_{t-1}} r_{i_{t-1}, i_t} D_{i_{t-1}} \tag{49}$$

Then we can write the following:

$$D_{i_1} = s_{i_1} \tag{50}$$

$$D_{i_2} = s_{i_2} + p_{i_2} \sum_{i_1} r_{i_1, i_2} D_{i_1} \tag{51}$$

$$D_{i_2} = s_{i_2} + p_{i_2} \sum_{i_1} r_{i_1, i_2} s_{i_1} \tag{52}$$

$$D_{i_3} = s_{i_3} + p_{i_3} \sum_{i_2} r_{i_2, i_3} D_{i_2} \tag{53}$$

$$D_{i_3} = s_{i_3} + p_{i_3} \sum_{i_2} r_{i_2, i_3} \left(s_{i_2} + p_{i_2} \sum_{i_1} r_{i_1, i_2} s_{i_1}\right) \tag{54}$$

The distortion on the CU i_τ with $\tau > 1$ is expressed as

$$\begin{aligned} D_{i_\tau} = p_{i_\tau} \sum_{i_{\tau-1}} r_{i_{\tau-1}, i_\tau} \Bigg(& p_{i_{\tau-1}} \sum_{i_{\tau-2}} r_{i_{\tau-2}, i_{\tau-1}} \\ & \left(\ldots p_{i_2} \sum_{i_1} r_{i_1, i_2} s_{i_1} + s_{i_2}\right) + \ldots \Bigg) + s_{i_{\tau-1}} + s_{i_\tau}, \end{aligned} \tag{55}$$

and the total distortion D_{Tot} is expressed as:

$$\begin{aligned} D_{Tot} &= \sum_{t=1}^{T} \sum_{i=1}^{N} D_{i_t} \Psi_{i_t} \\ &= \sum_{t=1}^{T} \Bigg(\sum_{i=1}^{N} \Psi_{i_t} \Bigg(p_{i_t} \sum_{i_{t-1}} r_{i_{t-1}, i_t} \Bigg(p_{i_{t-1}} \sum_{i_{t-2}} r_{i_{t-2}, i_{t-1}} \\ & \left(\ldots p_{i_2} \sum_{i_1} r_{i_1, i_2} s_{i_1} + s_{i_2}\right) + \ldots \Bigg) + s_{i_{t-1}} + s_{i_t} \Bigg) \Bigg) \end{aligned} \tag{56}$$

From (56), D_{Tot} can be written as a linear combination of U_{i_t} and s_{i_t}, then

$$D_{Tot} = \sum_{t=1}^{T} \sum_{i=1}^{N} s_{i_t} U_{i_t}. \tag{57}$$

U_{i_t} is the s_{i_t} contribution to D_{Tot} and can be computed as the partial derivative of D_{Tot} with respect to s_{i_t}. After calculation and rearranging we obtain:

$$\begin{aligned} U_{n_\tau} &= \frac{\partial D_{Tot}}{\partial s_{n_\tau}} \\ &= \Psi_{n_\tau} + \sum_{t=\tau+1}^{T} \left(\sum_{i_t} \sum_{i_{t-1}} \dots \sum_{i_{\tau+1}} \Psi_{i_t} p_{i_t} r_{i_{t-1},i_t} \right. \\ &\quad \left. \Psi_{i_{t-1}} p_{i_{t-1}} r_{i_{t-2},i_{t-1}} \dots p_{i_{\tau+1}} r_{n_\tau,i_{\tau+1}} \right) \end{aligned} \tag{58}$$

Appendix D
Accumulation factor in recursive form

From (58) written at the rank $\tau-1$, after some manipulations we obtain the expression in (59).

$$\begin{aligned} U_{n_{\tau-1}} = \Psi_{n_{\tau-1}} + \sum_{i_\tau} p_{i_\tau} r_{n_{\tau-1},i_\tau} \left(\Psi_{i_\tau} \right. \\ + \sum_{t=\tau+1}^{T} \left(\sum_{i_t} \sum_{i_{t-1}} \dots \sum_{i_{\tau+1}} \Psi_{i_t} p_{i_t} r_{i_{t-1},i_t} \right. \\ \left. \left. \Psi_{i_{t-1}} p_{i_{t-1}} r_{i_{t-2},i_{t-1}} \dots p_{i_{\tau+1}} r_{i_\tau,i_{\tau+1}} \right) \right) \end{aligned} \tag{59}$$

It can be expressed as the recursive function:

$$U_{n_{\tau-1}} = \Psi_{n_{\tau-1}} + \sum_{i_\tau} \left(p_{i_\tau} r_{n_{\tau-1},i_\tau} U_{i_\tau} \right). \tag{60}$$

Trivially, when $\tau = T$, we obtain:

$$U_{n_T} = \Psi_{n_T} \tag{61}$$

It demonstrates the recursive form of the accumulation factor U as summarized in (62)

$$\begin{cases} U_{j_T} &= \Psi_{j_T} \\ U_{j_{t-1}} &= \sum_{i_t} p_{i_t} r_{j_{t-1},i_t} U_{i_t} + \Psi_{j_{t-1}}. \end{cases} \tag{62}$$

Appendix E
Computing the Lagrangian multiplier

$$\frac{\partial J_{Tot}}{\partial \Delta_{i_t}} = \frac{\partial d_{i_t}}{\partial \Delta_{i_t}} c_{i_t} U_{i_t} + \lambda \frac{\partial R_{i_t}}{\partial \Delta_{i_t}} c_{i_t} = 0 \tag{63}$$

The minimization of J_{Tot} is independent of c_{i_t}, according to (14). c_{i_t} is removed from equations. Then we obtain the λ as

$$\lambda \frac{\partial R_{i_t}}{\partial \Delta_{i_t}} = -U_{i_t} \frac{\partial d_{i_t}}{\partial \Delta_{i_t}}, \tag{64}$$

$$\lambda = -U_{i_t} \frac{\frac{\partial d_{i_t}}{\partial \Delta_{i_t}}}{\frac{\partial R_{i_t}}{\partial \Delta_{i_t}}} = -U_{i_t} \frac{\partial d_{i_t}}{\partial R_{i_t}} = -U_{i_t} \frac{\partial D_{i_t}}{\partial R_{i_t}}. \tag{65}$$

By using the R-D Shannon bound $R_{i_t} = -\frac{1}{2} log_2 \left(\frac{D_{i_t}}{\alpha \sigma_{i_t}^2} \right)$, we obtain

$$\frac{\partial R_{i_t}}{\partial D_{i_t}} = \frac{1}{2ln(2) D_{i_t}}. \tag{66}$$

Finally, the optimal λ is defined by

$$\lambda = 2ln(2) U_{i_t} D_{i_t}. \tag{67}$$

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 (TCSVT)*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] 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, Dec. 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6317156

[3] T. K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J.-R. Ohm, and G. J. Sullivan, "Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 76–90, Jan. 2016. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7254155

[4] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[5] K. McCann, C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G. Sullivan, "JCTVC-R1002: High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description," Jul. 2014.

[6] x265, "[Online]. Available: https://bitbucket.org/multicoreware/x265."

[7] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, 1998.

[8] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *IEEE Int. Conference on Image Processing (ICIP)*, Thessaloniki, October 2001, pp. 542–545.

[9] J. Chen, E. Alshina, G. J. Sullivan, J. R. Ohm, and J. Boyce, "Algorithm description of joint exploration model 7," in *JVET-G1001*, Jul. 2017.

[10] N. Sidaty, W. Hamidouche, O. Deforges, and P. Philippe, "Compression Efficiency of the Emerging Video Coding Tools," in *IEEE Conference on Image Processing (ICIP)*, September 2017.

[11] H. Schwarz, C. Rudat, M. Siekmann, B. Bross, D. Marpe, and T. Wiegand, "Coding Efficiency / Complexity Analysis of JEM 1.0 coding tools for the Random Access Configuration," in *Document JVET-B0044 3rd 2nd JVET Meeting: San Diego, CA, USA*, February 2016.

[12] E. Alshina, A. Alshin, K. Choi, and M. Park, "Performance of JEM 1 tools analysis," in *Document JVET-B0044 3rd 2nd JVET Meeting: San Diego, CA, USA*, February 2016.

[13] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, 1998.

[14] M. Ropert, J. Le Tanou, M. Bichon, and M. Blestel, "R-D spatio-temporal adaptive quantization based on temporal distortion back-propagation in HEVC," in *IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, 2017.

[15] M. Bichon, J. Le Tanou, M. Ropert, W. Hamidouche, L. Morin, and L. Zhang, "Temporal adaptive quantization using accurate estimations of inter and skip probabilities," in *Picture Coding Symposium*, 2018.

[16] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and mpeg video coders," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 553–545, 1994.

[17] J. Wen, M. Luttrell, and J. Villasenor, "Trellis-based R-D optimal quantization in h.263+," *IEEE Transactions on Image Processing*, vol. 9, no. 8, pp. 1431–1434, 2000.

[18] 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, 2017.

[19] M. Winken, A. Roth, H. Schwarz, and T. Wiegand, "Multi-frame optimized quantization for high efficiency video coding," in *Picture Coding Symposium (PCS)*, 2015.

[20] M. Bichon, J. Le Tanou, M. Ropert, W. Hamidouche, L. Morin, and L. Zhang, "Inter-block dependencies consideration for intra coding in H.264/AVC and HEVC standards," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[21] ——, "Low complexity joint RDO of prediction units couples for HEVC intra coding," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[22] G. Valenzise and A. Ortega, "Improved video coding efficiency exploiting tree-based pixelwise coding dependencies," in *Visual Information Processing and Communication*, 2010.

[23] Y. Li, H. Jia, X. Xie, and T. Huang, "Rate control for consistent video quality with inter-dependent distortion model for HEVC," in *Visual Communications and Image Processing (VCIP)*, 2016.

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

[25] J. Xie, L. Song, R. Xie, Z. Luo, and X. Wang, "Temporal dependent bit allocation scheme for rate control in HEVC," in *IEEE Workshop on Signal Processing Systems (SiPS)*, 2015.

[26] Y. Gao, C. Zhu, and S. Li, "Hierarchical temporal dependent rate-distortion optimization for low-delay coding," in *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, May 2016.

[27] Y. Gao, C. Zhu, S. Li, and T. Yang, "Source distortion temporal propagation analysis for random-access hierarchical video coding optimization," *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[28] J. Garrett-Glaser, "A novel macroblock-tree algorithm for high performance optimization of dependent video coding in H.264/AVC," 2011.

[29] Z. Wang, A.-C. Bonvik, H.-R. Sheikh, and E.-P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[30] C. Yeo, H. L. Tan, and Y. H. Tan, "SSIM-based adaptive quantization in HEVC," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013.

[31] G. Xiang, H. Jia, M. Yang, Y. Li, and X. Xie, "A novel adaptive quantization method for video coding," *Multimedia Tools and Applications*, vol. 77, no. 12, pp. 14 817–14 840, 2017.

[32] T. Wiegand and H. Schwarz, *Source Coding: Part I of Fundamentals of Source and Video Coding*. Foundations and Trends in Signal Processing, 2011.

[33] L. Xu, X. Ji, W. Gao, and D. Zhao, "Laplacian Distortion Model (LDM) for Rate Control in Video Coding," in *Advances in Multimedia Information Processing Pacific Rim Conference on Multimedia (PCM)*, Hong Kong, China, December 2007, pp. 638–646.

[34] F. Bossen, "Common test conditions and software reference configurations," *Tech. Rep. JCTVC-L1100*, Jan. 2013.

[35] C. Yeo, H. L. Tan, and Y. H. Tan, "On rate distortion optimization using ssim," *Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 23, no. 7, pp. 1170–1181, 2013.

[36] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *ITU-T VCEG-M33, Texas, USA*, Austin, Texas, Apr. 2001.

[37] S. Rimac-Drlje, D. agar, and G. Martinovi, "Spatial masking and perceived video quality in multimedia applications," in *Int. Conference on Systems, Signals and Image Processing (IWSSIP)*, Chalkida, Greece, Jun. 2009.

[38] S. Winkler, *Digital Video Quality: Vision Models and Metrics*. John Wiley and Sons, Ltd, 2005.

Maxime Bichon received the degree in engineering from Ecole Suprieure dIngnieurs de Rennes, France, in 2015, and the Ph.D. degree in signal and image processing from the INSA Rennes, France, in 2019. He is currently pursuing the Ph.D. degree with MediaKind, Saint-Jacques-de-la-Lande, France. His current research interests include the rate-distortion optimization, adaptive quantization and dependencies consideration for image and video compression.

Julien Le Tanou is a Senior Engineer at MediaKind. In this role, he is responsible for conducting research into video processing, compression and associated technologies. Since 2012, he has been with Envivio France in charge of research and algorithms design for Envivio's SW video encoding solutions. Prior to Envivio, he has been with Orange Labs in France and Dolby Laboratories in USA, where he conducted research and development for next generation of video coding standards. He received the MS degree in Signal and Image Processing from Telecom ParisTech, France in 2010, and the MSc degree in Computer and Electrical Engineering from Institut Superieure dElectronique de Paris, France in 2008.

Michael Ropert received the Ph.D. degree from Rennes I University in 1995. From 1991 to 1992, he was a Teacher in applied mathematics. He has been with France Tlcom in the areas of image and video compression for multimedia and digital television from 1996 to 1999. He has been with the initial step of Envivio and leading the video team for several years from 2000 to 2015. He is currently a Lead Technology Scientist with MediaKind, France.

Wassim Hamidouche received the Ph. D. Degree in Signal and Image Processing from the University of Poitiers, France in 2010. From 2011 to 2012 he has been a Research Engineer with Canon Research Centre, Rennes, France. He is Associate Professor at INSA Rennes since 2015 and member of the the Institute of Electronics and Telecommunications of Rennes (IETR), UMR CNRS 6164. His research interests focus on video coding, efficient real time and parallel architectures for the new generation video coding standards, multimedia transmission over heterogeneous networks, and multimedia content security.

Luce Morin is currently a Full-Professor with the Electrical and Computer Engineering Department, National Institute of Applied Sciences and a Researcher with the Institute of Electronics and Telecommunications of Rennes (IETR). She leads the VAADER Team in the IETR Laboratory. She has authored or co-authored over 70 scientific papers in international journals and conferences. Her research activities deal with computer vision, 3-D reconstruction, image and video compression, and representations for 3-D videos and multiview videos.