

Rate-Distortion Optimized Cross-layer Rate Control in Wireless Video Communication

Zhifeng Chen and Dapeng Wu

Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611

Abstract—A wireless video communication system can be designed based on the rate-distortion (R-D) criterion, i.e., minimizing the end-to-end distortion (which includes quantization distortion and transmission distortion) subject to the transmission bit-rate constraint. The minimization can be achieved by adjusting the source encoding parameters and channel encoding parameters. This R-D optimization (RDO) is usually done for each video frame individually in a real-time video communication system, e.g. video calls or videoconferencing. To achieve this, an accurate bit-rate model and distortion model for each frame can be used to reduce the RDO complexity. In this paper, we derive a source bit-rate model and quantization distortion model; we also improve the performance bound for channel coding under a convolutional code and a Viterbi decoder, and derive its performance bound under a Rayleigh block fading channel. Given the instantaneous channel condition, e.g., signal-to-noise ratio (SNR) and transmission bit-rate constraint, we design an R-D optimized cross-layer rate control (CLRC) algorithm by jointly choosing quantization step size in source coding and code rate in channel coding. Experimental results show that our proposed R-D models are more accurate than the existing R-D models. Experimental results also showed that the rate control under our models has more stable R-D performance than the existing rate control algorithms; using the channel estimation, CLRC can further achieve remarkable R-D performance gain over that without channel estimation. Another important result is that the subjective quality of our CLRC algorithm is much better than the existing algorithms due to its intelligent reference frame selection.

Index Terms—Cross-layer rate control (CLRC), end-to-end distortion, rate-distortion optimization (RDO), fading channel, wireless video, joint source channel coding (JSCC)

I. INTRODUCTION

Under the prevalence of 3G/4G network and smart phones nowadays, real-time mobile video applications, e.g., video-phone calls, are becoming more and more popular. However, transmitting video over mobile phone with good quality is particularly challenging since the mobile channels are subject to multipath fading, and therefore the channel condition may change from frame to frame. Given the instantaneous channel condition, e.g., SNR and transmission bit-rate, the minimum end-to-end distortion under the transmission bit-rate constraint can be achieved by optimally allocating the transmission bit-rate between source bit-rate and redundant bit-rate incurred by

channel coding. In a practical wireless video communication system, this can be achieved by jointly adjusting the source encoding parameters, e.g. quantization step size, in the video encoder, and channel encoding parameters, e.g., code rate, in the channel encoder. Since both the video statistics and channel condition vary with time, we need to adapt these parameters for each frame in real-time video encoding and packet transmission. Therefore, we need to estimate the bit-rate and distortion for each possible combination of parameters before encoding each frame. As a result, accurate bit-rate model and distortion model will be very helpful to achieve the minimum end-to-end distortion with low complexity.

Many works were intended to address this problem in the past. While most of them derive the end-to-end distortion as functions of bit-rate and packet error rate [1], [2], others use operational rate-distortion (R-D) functions [3]. The analytical models are more desirable since it is very difficult for the video encoder to get all operational functions for different video statistics and channel conditions before actual encoding. However, the existing analytical models are still not accurate enough to accommodate the time-varying channel condition. On the other hand, to obtain tractable formulae in those analytical models [1], [2], the authors all assume that specific block codes, i.e., Reed-Solomon codes, are adopted as the forward error correction (FEC) scheme. Based on that FEC scheme, the distortion is derived as a function of code rate in channel coding and bit error rate. However, this assumption has two limitations: 1) most up-to-date video communication systems use rate compatible punctured convolutional (RCPC) codes or more advanced codes, e.g., turbo codes, for physical layer channel coding due to their flexible choice of code rate without the need of changing the channel encoder structure; 2) in the cross-layer optimization problem, allocating source bit-rate and redundant bit-rate subject to a bit error rate constraint achieves poorer performance than allocating source bit-rate and redundant bit-rate subject to the given instantaneous channel condition, e.g., SNR and transmission bit-rate. In this paper, we aim to solve the cross-layer optimization problem by deriving more accurate bit-rate model and end-to-end distortion model, which consists of two parts, that is, quantization distortion model and transmission distortion model.

Many bit-rate models have been developed in existing literature. Most existing works derive bit-rate as a function of video statistics and quantization step size [4], [5], [6], [7], while others model bit-rate as a function of video statistics and other parameters such as parameter ρ [8]. In Refs. [6], [8],

Please direct all correspondence to Prof. Dapeng Wu, University of Florida, Dept. of Electrical & Computer Engineering, P.O.Box 116130, Gainesville, FL 32611, USA. Tel. (352) 392-4954. Fax (352) 392-0044. Email: wu@ece.ufl.edu. Homepage: <http://www.wu.ece.ufl.edu>. This work was supported in part by the US National Science Foundation under grant ECCS-1002214. Copyright (c) 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

the bit-rate models include some model parameters which can only be estimated from previous frames. However, in theory, the instantaneous frame bit rate should be independent of previous frames given instantaneous residual histogram and quantization step size. In addition, the estimation error of those model parameters may have a significant impact on the model accuracy, which can be observed in the H.264/AVC JM reference software [9] and will be explained in detail in the experimental section of this paper. In Refs. [7], [10], [11], the bit-rate models are functions of only quantization step size and parameters, e.g., variance, of a source distribution. However, they share some limitations: 1) the assumed residual probability distribution, e.g., Laplacian distribution, may deviate significantly from the true histogram; 2) the implicit assumption of all transform coefficients being identically distributed is not valid since different coefficient show different variances as shown in our experiment. In this paper, we improve the bit-rate model by considering the characteristics of different variances of different coefficients. We also compensate the mismatch between the true histogram and the assumed Laplacian distribution in bit-rate models by utilizing the deviation of estimation in previous frames. Experimental results show that our method achieves a more accurate estimate of bit-rate compared to existing models.

Quantization distortion is caused by quantization error under lossy source coding and it has been extensively explored since the seminal work of Shannon's rate distortion theory first proposed in Ref. [12] and later proved in Ref. [13]. The quantization distortion are studied either as a function of bit-rate and the source probability distribution, e.g., the classical R-D function for Gaussian source [14], [15], or as a function of the number of quantization levels and the source probability distribution given a certain quantizer, e.g., uniform scalar quantizer for Gaussian source [16]. In the case of memoryless Gaussian sources, uniform quantizers achieve entropies that exceed the rate-distortion function by approximately 0.25 bits/sample at relatively high rates [17]. In Ref. [18], the performance of optimum quantizers for a wide class of memoryless non-Gaussian sources is investigated, and it is shown that uniform quantizers perform as effectively as optimum quantizers. For this reason, a uniform quantizer is usually adopted in a practical video encoder, e.g., H.264 [19]. For a uniform quantizer, the quantization distortion have been derived as a function of quantization step size and video frame statistics [8], [6], [10], [7]. Although these quantization distortion models have achieved quite accurate result, there is still room for improvement due to the source distribution model inaccuracy. In this paper, we improve the accuracy of estimating quantization distortion by compensating the mismatch between the true histogram and the assumed distribution in quantization distortion models [10], [7]. Experimental results show that our quantization distortion model is more accurate than the existing models.

Transmission distortion is caused by transmission error under error-prone channels. Predicting transmission distortion at the transmitter poses a great challenge due to the spatio-temporal correlation inside the input video sequence, the non-linearity of video codec, and varying packet error probability

(PEP) in time-varying channels. The existing transmission distortion models can be categorized into the following three classes: 1) pixel-level or block-level models (applied to prediction mode selection) [20], [21]; 2) frame-level or packet-level or slice-level models (applied to cross-layer encoding rate control) [1], [2], [22]; 3) GOP-level or sequence-level models (applied to packet scheduling) [23], [24]. Although different transmission distortion models work at different levels, they share some common properties, which come from the inherent characteristics of wireless video communication system, that is, spatio-temporal correlation, nonlinear codec and time-varying channel. However, none of these existing works analyzed the effect of non-linear clipping noise on the transmission distortion, and therefore cannot provide accurate transmission distortion estimation. In Ref. [25], we analytically derive, for the first time, the transmission distortion formula as a closed-form function of packet error probability (PEP), video frame statistics, and system parameters; and then in Ref. [26], we design the RMPC¹ algorithm to predict the transmission distortion with low complexity and high accuracy. In this paper, we will further derive PEP and transmission distortion as functions of SNR, transmission rate, and code rate for cross-layer optimization.

Channel coding can be considered as the embedding of signal constellation points in a higher dimensional signaling space than is needed for communications. By mapping to a higher dimensional space, the distance between points increases, which provides better error correction and detection performance [27]. In general, the performance of soft-decision decoding is about 2-3dB better than hard-decision decoding [27]. Since convolutional decoders have efficient soft-decision decoding algorithms, such as Viterbi algorithm [28], we choose convolutional codes for physical layer channel coding in this paper². In addition, Rate-compatible punctured convolutional (RCPC) codes can adaptively change the code rate without changing the encoder structure, which makes convolutional codes an appropriate method in real-time video communication over wireless fading channels. In this paper we improve the performance bound of convolutional codes by adding a threshold for low SNR case, and extend it to support a more flexible SNR threshold for transmitters with channel estimation. For transmitters without channel estimation, we also derive the expected PEP as a simple function of convolutional encoder structure and channel condition under Rayleigh block fading channel.

Given the bit-rate function, quantization distortion function and transmission distortion function, minimizing end-to-end distortion becomes an optimization problem under the transmission bit-rate constraint. In this paper, we also apply our bit-rate model, quantization distortion model and transmission distortion model to cross-layer rate control with rate-distortion optimization (RDO). Due to the discrete characteristics and the

¹R means residual concealment error; M means motion vector concealment error; P means propagated error with clipping noise; C means correlation among these three types of errors.

²Our algorithm can also be used for other channel codes, e.g. block codes, Turbo codes, and LDPC codes, given their performance for different code rates.

possibility of non-convexity of distortion function [29], the traditional Lagrange multiplier solution for continuous convex function optimization is infeasible in a video communication system. The discrete version of Lagrangian optimization is first introduced in Ref. [30], and then first used in a source coding application in Ref. [29]. Due to its simplicity and effectiveness, this optimization method is de facto adopted by the practical video codec, e.g., H.264 reference code JM [9]. In this paper, we will use the same method to solve our optimization problem.

Our contributions in this paper are: 1) we derive a source bit-rate model and quantization distortion model, which are more accurate than the existing models; 2) we improve the performance bound for channel coding with a RCPC encoder and a Viterbi decoder, and derive its performance under Rayleigh block fading channel; 3) with PEP function, the transmission distortion can be expressed as a function of instantaneous video frame statistics, system parameters, and channel conditions; 4) we design a R-D optimized cross-layer rate control (CLRC) algorithm by jointly choosing quantization step size in source coding and code rate in channel coding based on the given instantaneous channel condition, e.g., SNR and transmission bit-rate.

The rest of this paper is organized as follows. In Section II, we describe our system model and formulate the cross-layer optimization problem. In Section III, we derive our bit-rate model, quantization distortion model, and transmission distortion model. In Section IV, we propose a practical cross-layer rate control algorithm to achieve minimum end-to-end distortion under the given SNR and transmission bit-rate. Section V shows the experimental results, which demonstrates both the higher accuracy of our models and the better performance of our algorithm over existing algorithms. Section VI concludes the paper.

II. SYSTEM DESCRIPTION AND PROBLEM FORMULATION

Fig. 1 shows the structure of a typical wireless video communication system. Note that in this system, the channel is an application-layer channel, which includes entropy coding, channel coding, modulation and a radio channel as shown in Fig. 2, where R_s is the output rate of the entropy encoder, R_c is the code rate of the channel encoder, and R_t is the output rate of the channel encoder.

The general RDO problem in a wireless video communication system can be formulated as

$$\begin{aligned} \min \quad & D_{ETE}^k \\ \text{s.t.} \quad & R_t^k \leq R_{con}^k, \end{aligned} \quad (1)$$

where D_{ETE}^k is the end-to-end distortion of the k -th frame, R_t^k is the bit-rate for transmitting the k -th frame, R_{con}^k is the bit-rate constraint for the k -th frame, which depends on the channel condition.

From the definition, we have

$$D_{ETE}^k \triangleq E\left[\frac{1}{|\mathcal{V}^k|} \sum_{\mathbf{u} \in \mathcal{V}^k} (f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k)^2\right], \quad (2)$$

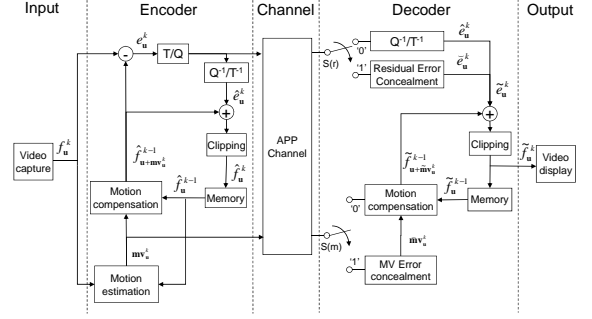


Fig. 1. System structure, where T, Q, Q^{-1} , and T^{-1} denote transform, quantization, inverse quantization, and inverse transform, respectively.

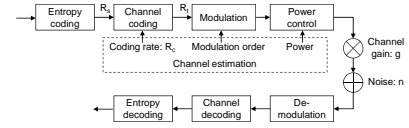


Fig. 2. Components in an application-layer channel.

where \mathcal{V}^k is the set of pixels in the k -th frame; $f_{\mathbf{u}}^k$ is the original pixel value for pixel \mathbf{u} in the k -th frame; $\hat{f}_{\mathbf{u}}^k$ is the reconstructed pixel value for the corresponding pixel at the decoder.

Define quantization error as $f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k$ and transmission error as $\hat{f}_{\mathbf{u}}^k - \tilde{f}_{\mathbf{u}}^k$, where $\hat{f}_{\mathbf{u}}^k$ is the reconstructed pixel value for pixel \mathbf{u} in the k -th frame at the encoder. While $f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k$ depends only on the quantization parameter (QP) ³, $\hat{f}_{\mathbf{u}}^k - \tilde{f}_{\mathbf{u}}^k$ mainly depends on the PEP and the error concealment scheme. In addition, experimental results show that $f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k$ is zero-mean statistically. Therefore, we make the following assumption.

Assumption 1: $f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k$ and $\hat{f}_{\mathbf{u}}^k - \tilde{f}_{\mathbf{u}}^k$ are uncorrelated, and $E[f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k] = 0$.

Under Assumption 1, from (2), we obtain

$$\begin{aligned} D_{ETE}^k &= E\left[\frac{1}{|\mathcal{V}^k|} \sum_{\mathbf{u} \in \mathcal{V}^k} (f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k)^2\right] + E\left[\frac{1}{|\mathcal{V}^k|} \sum_{\mathbf{u} \in \mathcal{V}^k} (\hat{f}_{\mathbf{u}}^k - \tilde{f}_{\mathbf{u}}^k)^2\right] \\ &\quad + \frac{2}{|\mathcal{V}^k|} \sum_{\mathbf{u} \in \mathcal{V}^k} E[(f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k)E[(\hat{f}_{\mathbf{u}}^k - \tilde{f}_{\mathbf{u}}^k)]] \\ &= E\left[\frac{1}{|\mathcal{V}^k|} \sum_{\mathbf{u} \in \mathcal{V}^k} (f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k)^2\right] + E\left[\frac{1}{|\mathcal{V}^k|} \sum_{\mathbf{u} \in \mathcal{V}^k} (\hat{f}_{\mathbf{u}}^k - \tilde{f}_{\mathbf{u}}^k)^2\right] \\ &= D_Q^k + D_T^k, \end{aligned} \quad (3)$$

where D_Q^k is called frame-level quantization distortion (FQD) and defined as $D_Q^k \triangleq E\left[\frac{1}{|\mathcal{V}^k|} \sum_{\mathbf{u} \in \mathcal{V}^k} (f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}}^k)^2\right]$; and D_T^k is called frame-level transmission distortion (FTD) and defined as $D_T^k \triangleq E\left[\frac{1}{|\mathcal{V}^k|} \sum_{\mathbf{u} \in \mathcal{V}^k} (\hat{f}_{\mathbf{u}}^k - \tilde{f}_{\mathbf{u}}^k)^2\right]$.

In a typical video codec, the spatial correlation and temporal correlation are first removed/reduced by intra prediction and inter prediction. Then the residual is transformed and quantized. Given the uniform quantizer, D_Q^k only depends on the quantization step size Q^k and the video frame statistics

³In the rate control algorithm design, quantization offset is often fixed.

ϕ_f^k . Therefore, we can express D_Q^k as a function of Q^k and ϕ_f^k , i.e., $D_Q(Q^k, \phi_f^k)$, where $D_Q(\cdot)$ is independent from the frame index k . In Ref. [25], we have derived D_T^k as a function of PEP, video frame statistics ϕ_f^k and system parameters ϕ_s^k , i.e., $D_T(PEP^k, \phi_f^k, \phi_s^k)$. Since PEP^k depends on SNR $\gamma(t)$, transmission bit-rate R_t^k , and code rate R_c^k , D_T^k also depends on R_c^k . On one hand, the higher code rate R_c^k in channel coding, i.e. the fewer number of redundant bits, the higher PEP^k [28], resulting in a larger D_T^k . On the other hand, under the same constraint on R_t^k , the higher code rate also means the higher source bit-rate, leading to a smaller D_Q^k . In order to design the optimum Q^k and R_c^k to achieve the minimum D_{ETE}^k , we need to have PEP^k as a function of SNR $\gamma(t)$, transmission rate R_t^k , and R_c^k , i.e., $P(\gamma(t), R_t^k, R_c^k)$. Denote ϕ_c^k the channel statistics, i.e. $\phi_c^k = \{\gamma(t), R_t^k\}$, we can express D_T^k as a function of R_c^k , ϕ_c^k , ϕ_f^k , and ϕ_s^k , i.e., $D_T(R_c^k, \phi_c^k, \phi_f^k, \phi_s^k)$. On the other hand, $R_t^k = \frac{R_s^k}{R_c^k}$ where R_s^k denote the source bit-rate and it is a function of the quantization step size Q^k and video frame statistics ϕ_f^k , i.e., $R_s(Q^k, \phi_f^k)$.

Therefore, if we can derive the closed-form functions for $D_Q(Q^k, \phi_f^k)$, $D_T(PEP^k, \phi_f^k, \phi_s^k)$ and $R_s(Q^k, \phi_f^k)$, (1) can be solved by finding the best parameter pair $\{Q^k, R_c^k\}$. In other words, the problem in (1) is equivalent to

$$\begin{aligned} \min \quad & D_Q(Q^k, \phi_f^k) + D_T(R_c^k, \phi_c^k, \phi_f^k, \phi_s^k) \\ \text{s.t.} \quad & \sum_{i=1}^{N^k} \frac{R_s(Q^k, \phi_{f,i}^k)}{R_{c,i}^k} \leq R_{con}^k, \end{aligned} \quad (4)$$

where N^k is the total number of packets that contain the bits of the k -th frame, and i is the packet index. In summary, our problem in (4) is “given the system structure ϕ_c^k , time-varying video frame statistics ϕ_f^k and time-varying channel statistics ϕ_c^k , how to minimize D_{ETE}^k by jointly adjusting the parameter pair $\{Q^k, R_{c,i}^k\}$.”

III. DERIVATION OF BIT RATE FUNCTION, QUANTIZATION DISTORTION FUNCTION AND TRANSMISSION DISTORTION FUNCTION

In this section, we derive the source rate function $R_s(Q^k, \phi_f^k)$, quantization distortion function $D_Q(Q^k, \phi_f^k)$, and transmission distortion function $D_T(PEP^k, \phi_f^k, \phi_s^k)$.

A. Derivation of Source Coding Bit Rate Function

1) *Source coding bit-rate estimation for the H.264 encoder:* For a hybrid video coder with block-based coding scheme, e.g., H.264 encoder, the encoded bit-rate R_s consists of residual bits R_{resi} , motion information bits R_{mv} , prediction mode bits R_{mode} , and syntax bits R_{syntax} . That is,

$$R_s^k = \hat{H}^k \cdot N_{resolution} \cdot N_{fps} + R_{mv}^k + R_{mode}^k + R_{syntax}^k, \quad (5)$$

where $N_{resolution}$ is the normalized video resolution considering color components, and N_{fps} denotes the number of frames per second (fps). Compared to R_{resi}^k , values R_{mv}^k , R_{mode}^k , and R_{syntax}^k are less affected by Q . Therefore, in this paper

we only improve the residual bit rate model from existing works [10], [7], [11] to estimate a more accurate Q for a given bit rate in rate control. In this paper, we follow the same assumption from Ref. [10], [7], [11] that the residual follows a Laplacian distribution. It is noted that some other distributions are proposed for estimating the true residual histogram, e.g., Cauchy distribution [31], [32] and Generalized Gaussian distribution [33], [34]. The authors in Ref. [35] also claimed that the luminance and chrominance sources should be separately modeled. However, in all of them, the transform coefficients are assumed independent and identically distributed (i.i.d.). Based on our analysis of Laplacian distribution, the extension to other distributions is straightforward.

2) *Entropy of quantized transform coefficients for i.i.d. zero-mean Laplacian source under uniform quantizer:* Following the similar deriving process as in Ref. [10], [7], [11], it is easy to prove that for i.i.d. zero-mean Laplacian source under uniform quantizer with quantization step size Q and quantization offset θ_2 , the entropy of quantized transform coefficients is

$$\begin{aligned} H = & -P_0 \cdot \log_2 P_0 + (1 - P_0) \\ & \cdot \left(\frac{\theta_1 \cdot \log_2 e}{1 - e^{-\theta_1}} - \log_2(1 - e^{-\theta_1}) - \theta_1 \cdot \theta_2 \cdot \log_2 e + 1 \right), \end{aligned} \quad (6)$$

where

$$\theta_1 = \frac{\sqrt{2} \cdot Q}{\sigma}; \quad (7)$$

Q is the quantization step size; σ is the standard deviation of the Laplacian distribution; θ_2 is the quantization offset; $P_0 = 1 - e^{-\theta_1 \cdot (1 - \theta_2)}$ is the probability of quantized transform coefficient being zero. (6) is proved in Appendix A.

3) *Improvement by considering different variances of different coefficients:* In a video encoder, the transform coefficients are actually not i.i.d. [36], [37]. Although we may assume the DCT transform or integer transform [19] highly de-correlates the correlation among neighboring pixels, different transform coefficients have very different variances. For example, in a 4x4 integer transform, the 16 coefficients show a decreasing variance in the well-known zigzag scan order as used in H.264. As a result, the coefficients of higher frequency have higher probability of being zeroes after quantization. On the other hand, the coefficients of lower frequency have larger variances after quantization. From information theory, we know the concavity of entropy, i.e., entropy is a concave function of the distribution (Theorem 2.7.3 in Ref. [14]). Therefore, characterizing different coefficients with the same variance as in Refs. [10], [7], [11] will over-estimate the joint entropy of the transform coefficients. Hence, different coefficients should be modeled by random variables of different variances.

To derive the joint entropy for 16 coefficients with different variances, we need to model the variance relationship among those 16 coefficients. Having done extensive experiments, we find an interesting phenomenon of their relationship from samples in one frame or one GOP as shown in Fig. 3. That is, the variance is approximately a function of position in the two-dimensional transform domain. Based on this observation, we use (8) to model the variance relationship among the different

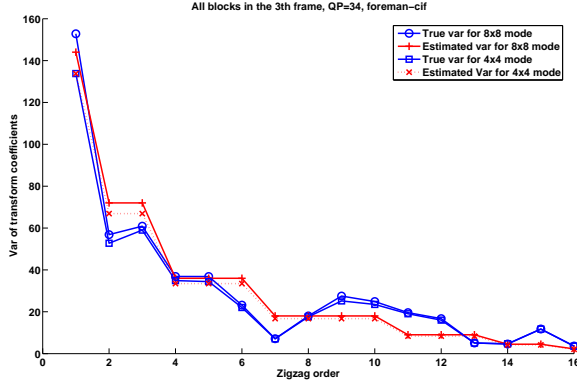


Fig. 3. Variance model.

coefficients in 4x4 integer transform⁴.

$$\sigma_{(x,y)}^2 = 2^{-(x+y)} \cdot \sigma_0^2, \quad (8)$$

where x and y is the position in the two-dimensional transform domain, and σ_0^2 is the variance of the coefficient at position $(0,0)$. With (8), we can derive the variance $\sigma_{(x,y)}^2$ for all positions given the average variance σ^2 as in Appendix B.

Fig. 3 shows the true variances and estimated variances by (28) for all transform coefficients before quantization. The statistics of each inter prediction mode are collected during the mode decision process. For each prediction mode, the mean of variances of all 16 coefficients and the average deviation of iid-coefficient assumption is shown as a reference. From Fig. 3 and the online table⁵, we see this variance model is much more accurate than the i.i.d. assumption. However, we also notice that in intra modes the true variance of DC component is much larger than estimated variance by (28). After careful analysis, we find that this is due to the high correlation among all coefficients in intra prediction modes. The more accurate variance model for intra modes will be investigated in our future work.

With (28), the estimated joint entropy of 16 non-identical transform coefficients by considering the different variances of different coefficients is

$$H_{rlc} = \frac{1}{16} \sum_{x=0}^3 \sum_{y=0}^3 H_{(x,y)}, \quad (9)$$

where $H_{(x,y)}$ is the entropy for coefficient position (x,y) , and can be calculated by (28), (7) and (6) with their own $\sigma_{(x,y)}^2$ and $\theta_{1(x,y)}$.

4) *Practical consideration of Laplacian assumption*: Statistically speaking, (9) is only valid for sufficiently large sample sizes. When there are not enough samples or the sample variance is very small, e.g., $Q > 3\sigma$, the Laplacian assumption for individual coefficients is not accurate. In such cases, we

⁴For 8x8 transform in H.264 Fidelity Range Extensions [38] and larger transform in HEVC [39], different models may be expected.

⁵Due to the page limitation, we list the average deviation of our proposed variance model for different sequences with different QP settings and prediction modes in a table at www.mcn.ece.ufl.edu/public/zhifeng/papers/CLRC/table_variance.pdf.

may use average variance in (6) as the estimate instead of (9). That is,

$$H^k = \begin{cases} \text{estimate by (9) with } \{\sigma_{(x,y)}^2\}, & Q \leq 3\sigma \\ \text{estimate by (6) with average } \sigma^2, & \text{otherwise.} \end{cases} \quad (10)$$

We will see in Section V that (10) provides more accurate estimation of bit rate than using the same variance for all coefficients.

5) *Improvement under the consideration of model inaccuracy*: The assumed residual probability distribution, e.g., Laplacian distribution, may deviate significantly from the true histogram especially when the number of samples are not sufficient. Therefore, we need to compensate the mismatch between the true residual histogram and assumed Laplacian distribution to obtain a better estimate. Denote H_l as the entropy for the case with a Laplacian distribution, H_t as the true rate of residual and $\nu = \frac{H_l}{H_t}$. In a video sequence, the changes of residual statistics and quantization step size between adjacent frames have almost the same effect on H_l and H_t . Therefore, we may use the previous frame statistics to compensate the estimated result from (10). Assume $\nu^k = \nu^{k-1}$. As a result, (10) can be further compensated as

$$\hat{H}^k = \frac{H_t^{k-1} \cdot H^k}{H_l^{k-1}}. \quad (11)$$

Although very simple, (9) and (11) significantly improve the estimation accuracy of residual entropy as shown in Section V-A.

B. Derivation of Quantization Distortion Function

In this subsection, we improve the accuracy of estimating quantization distortion by utilizing the same techniques in Section III-A. In Ref. [10], [7], authors derive the distortion for zero-mean Laplacian residual distribution under uniform quantizer as

$$D_Q = \frac{Q^2 \cdot (\theta_1 \cdot e^{\theta_2 \cdot \theta_1} \cdot (2 + \theta_1 - 2 \cdot \theta_2 \cdot \theta_1) + 2 - 2 \cdot e^{\theta_1})}{\theta_1^2 \cdot (1 - e^{\theta_1})}, \quad (12)$$

Since the coefficients after transform is not identical in distribution, we need to derive the overall quantization distortion function by considering each coefficient individually. Using the variance relationship among coefficients in (8), we have

$$D_{overall} = \frac{1}{16} \sum_{x=0}^3 \sum_{y=0}^3 D_{(x,y)}, \quad (13)$$

where $D_{(x,y)}$ is the distortion for coefficient position (x,y) , and can be calculated by (28), (7) and (12) with their own $\sigma_{(x,y)}^2$ and $\theta_{1(x,y)}$.

When there are not enough samples or the sample variance is very small, e.g., $Q > 3\sigma$, the Laplacian assumption for individual coefficients is not accurate. In such cases, we may use average variance in (12) as the estimate instead of (13). That is,

$$D_Q^k = \begin{cases} \text{estimate by (13) with } \{\sigma_{(x,y)}^2\}, & Q \leq 3\sigma \\ \text{estimate by (12) with average } \sigma^2, & \text{otherwise.} \end{cases} \quad (14)$$

Similarly, we need to compensate the mismatch between the true residual histogram and assumed Laplacian distribution for quantization distortion estimation. Denote $D_{Q,l}$ as quantization distortion for the case with a Laplacian distribution, $D_{Q,t}$ as true quantization distortion of residual and $\mu = \frac{D_{Q,l}}{D_{Q,t}}$. Assume $\mu^k = \mu^{k-1}$; (14) can be compensated as

$$\hat{D}_Q^k = \frac{D_{Q,t}^{k-1} \cdot D_Q^k}{D_{Q,l}^{k-1}}, \quad (15)$$

where D_Q^k is calculated from (14). Using (13) and (15) significantly improve the estimation accuracy of quantization distortion as shown in Section V-A.

C. Derivation of Transmission Distortion Function

In this subsection, we derive the FTD as a function of SNR, transmission rate, and code rate.

1) *Transmission distortion as a function of PEP*: In Ref. [25], we derived the FTD formula under single-reference motion compensation and no slice data partitioning as

$$D_T^k = \bar{P}^k \cdot (E[(\varepsilon^k)^2] + \lambda^k \cdot E[(\xi^k)^2] + D^{k-1}) + (1 - \bar{P}^k) \cdot \alpha^k \cdot D^{k-1} \cdot (1 - \beta^k). \quad (16)$$

\bar{P}^k is the weighted average PEP of all packets in the k -th frame; ε^k is the residual concealment error; ξ^k is the MV concealment error; β^k is the percentage of encoded I-MBs in the k -th frame; both the propagation factor α^k and the correlation ratio λ^k depend on video statistics, channel condition and codec structure, and are therefore called *system parameters*; D^{k-1} is the transmission distortion of the $k-1$ frame, which can be iteratively calculated by (16).

\bar{P}^k is defined as $\bar{P}^k \triangleq \frac{1}{|Y^k|} \sum_{i=1}^{N^k} (P_i^k \cdot N_i^k)$, where N_i^k is the number of pixels contained in the i -th packet of the k -th frame; P_i^k is PEP of the i -th packet of the k -th frame; N^k is the total number of packets of the k -th frame. The other video frame statistics and system parameters can be easily estimated as described in Ref. [26]. We will describe how to estimate PEP in the following subsections.

2) *PEP as a function of SNR, transmission rate, and code rate in a fading channel*: Below, we analyze the conditional PEP for convolution coding scheme under wireless fading channel, given SNR. Since convolutional codes are linear codes, the probability of error can be obtained by assuming that the all-zero sequence is transmitted, and determining the probability that the decoder decides in favor of a different sequence [27]. The probability of mistaking the true transmitted sequence for a sequence, which is Hamming distance d away, is called pairwise error probability, and denoted as $P_2(d)$. With soft decision, if the coded symbols produced by the convolutional encoder are sent over an AWGN channel using coherent BPSK modulation with SNR γ , then it can be shown that [16]

$$P_2(d) = Q(\sqrt{2\gamma \cdot d}). \quad (17)$$

Before calculating the PEP, we need to analyze the first error probability, which is defined as the probability that another path that merges with the all-zero path at a given node has a

metric that exceeds the metric of the all-zero path for the first time [16]. According to the definition, the first error probability can be approximated by its upper bound, i.e., the probability of mistaking the all-zero path for another path through the trellis, as

$$P_{fe} \leq \sum_{d=d_{free}}^{d_{max}} W_d * P_2(d), \quad (18)$$

where W_d is the number of codewords with weight d ; d_{free} is the free distance of the specific convolutional code; d_{max} is the maximum distance between the transmitted sequence and decoded sequence⁶. As a result, the PEP for a block of L decoded bits and for a given SNR can be upper-bounded as below [40], [41]

$$PEP(\gamma) \leq 1 - (1 - P_{fe}(\gamma))^L \approx L \cdot P_{fe}(\gamma). \quad (19)$$

However, both upper bounds in (18) and (19) are tight only when γ is large. When γ is small such as in a fading channel, the resulting bound may be much larger than 1, i.e., $L \cdot P_{fe}(\gamma) \gg 1$. From our experimental results, we find that $PEP(\gamma)$ follows waterfall shape when γ increases, that is, there exist a threshold γ_{th} such that, when $\gamma > \gamma_{th}$, the bound is quite tight, and when $\gamma < \gamma_{th}$, the bound becomes very loose and exceeds 1 quickly. Therefore, we improve the performance bound by using the following formula.

$$PEP(\gamma) \approx \begin{cases} \frac{R_t \cdot R_c}{N^k \cdot N_{fps}} \cdot \sum_{d=d_{free}}^{d_{max}} W_d * P_2(d, \gamma), & \gamma \geq \gamma_{th} \\ 1, & \gamma < \gamma_{th}, \end{cases} \quad (20)$$

where γ_{th} can be numerically calculated from (21) given the convolutional encoder structure (W_d , d_{free} , and d_{max}), code rate (R_c) and modulation scheme ($P_2(d)$). Note that W_d , d_{free} , and d_{max} in (20) are functions of R_c in RCPC. (20) is quite accurate as shown in Fig. 6, where $PEP_{th} = 1$.

$$\sum_{d=d_{free}}^{d_{max}} \frac{R_t \cdot R_c}{N^k \cdot N_{fps}} \cdot W_d \cdot P_2(d, \gamma_{th}) = PEP_{th}. \quad (21)$$

Note that a change in the modulation and demodulation technique used to transmit the coded information sequence affects only the computation of $P_2(d)$ [16]. Therefore, (20) is general for any modulation scheme.

In a real-time video communication system, if the estimated $PEP(\gamma)$ is larger than a threshold value, i.e. $PEP(\gamma) > PEP_{th}$, transmitter may discard this packet instead of transmitting it.⁷ The benefit of doing this is three-folded: 1) if $PEP(\gamma)$ is large, it is a waste of energy and time to transmit the packet; therefore, using (20) saves transmission energy; 2) in cross-layer rate control, since video encoder has the knowledge of channel condition, video encoder will skip encoding current frame when the channel gain is very low, which saves the encoding energy; 3) if the current frame is

⁶(18) is different from the formula in Ref. [16] since our codeword is truncated to fit the packet length, which improve the upper bound. But the effect on performance is negligible when packet length is large [16].

⁷In non-real-time applications, e.g., streaming video, a buffer is used to hold packets when the channel condition is poor. In such cases, packets will be dropped at the transmitter only when the buffer is full or the delay bound is violated, which will decrease the PEP.

skipped, the video encoder may intelligently choose previous encoded frames, except the current frame, as references for encoding the following frames, which highly reduces the reference error propagation. The details of intelligent reference frame selection will be discussed in Section V-B3.

(20) is derived under the condition that γ is known at the transmitter with channel estimation. In some wireless system, γ is unknown by the transmitter, e.g., without feedback channel from the receiver to the transmitter. In such a case, the expected *PEP*, i.e., $E_\gamma[PEP]$, instead of *PEP* is used for estimating transmission distortion given the probability distribution of channel gain. Proposition 1 gives the formula of expected *PEP* under a Rayleigh block fading channel.

Proposition 1: Under a Rayleigh block fading channel, the expected *PEP* is given by

$$E_\gamma[PEP] = \frac{\gamma_{th}}{\bar{\gamma}} e^{-\frac{\gamma_{th}}{\bar{\gamma}}} \left(1 + \frac{1}{d_{free} \gamma_{th}}\right), \quad (22)$$

where γ_{th} is defined by (21).

Proposition 1 is proved in Appendix C. We see from (30) that if $\gamma_{th} \geq \bar{\gamma}$, $E_\gamma[PEP] \geq 1 - e^{-\frac{\gamma_{th}}{\bar{\gamma}}} \geq 1 - e^{-1} \approx 0.63$. So, to control the *PEP* under a reasonable level, the transmitter should set its transmission power so that $\bar{\gamma} \gg \gamma_{th}$ before transmitting the packet.

3) *Transmission distortion as a function of SNR, transmission rate, and code rate in a fading channel:* In case of adaptive modulation, adaptive transmission power and adaptive bandwidth (subcarrier) allocation, $P_2(d)$ is a function of modulation order M , transmission power P_t and passband bandwidth B . In this paper, we study the case that modulation, power and bandwidth are all given during the cross-layer rate control. Under such conditions, transmission bit-rate R_t is fixed and SNR γ only depends on the channel gain g and noise power spectral density $\frac{N_0}{2}$. For example, with Nyquist pulse-shaping, $R_t = B \cdot \log_2(M)$ and $\gamma = \frac{P_t \cdot g}{N_0 \cdot B}$. As a result, both *PEP* and D_t depend only on the tuning parameter R_c .

IV. RATE-DISTORTION OPTIMIZED CROSS-LAYER RATE CONTROL AND ALGORITHM DESIGN

In this section, we apply our models derived in Section III to cross-layer rate control application. We adopt the discrete version of Lagrange multiplier as used in JM [9] to achieve the R-D optimized parameter pair $\{Q^k, R_{c,i}^k\}$. We also design a practical cross-layer rate control algorithm.

A. Optimization of Cross-layer Rate Control Problem

To solve (4), we may either use the Lagrangian approach or dynamic programming approach [42]. In terms of complexity, the Lagrangian approach is preferable, since it can be run independently in each coding unit, whereas dynamic programming requires a tree to be grown. Note that the complexity of the dynamic programming approach can grow exponentially with the number of coding units considered, while the Lagrangian approach's complexity only grow linearly [42]. By using the theorem in Ref. [30], [29], we may use the Lagrangian

approach for the i -th packet in the k -th frame independently as

$$(Q_i^k, R_{c,i}^k)^* = \arg \min \left\{ D_Q(Q_i^k) + D_T(R_{c,i}^k) + \Lambda \cdot \frac{R_s(Q_i^k)}{R_{c,i}^k} \right\}, \quad (23)$$

where Λ is the preset Lagrange multiplier, which can be determined either by bi-section search [29], [43] or by modeling [44], [9].

For some special cases, e.g., video conference, the frame size is usually small. In such a case, each frame is fully contained in one packet, and therefore, the bit allocation problem can be simplified. To be more specific, since all bits are allocated into one packet, given γ and R_t , every value of R_c have a corresponding value of R_s ; every value of R_s has corresponding values of Q and D_Q (by (5), (11), (6) and (14)). As mentioned in Section III-C3, D_T is also a function of R_c . In other words, the end-to-end distortion D_{ETE}^k only depends on R_c . Therefore, there exists an optimum R_c^k , such that D_{ETE} is minimized. As a result, the Lagrange multiplier can be omitted. That is, the optimum R_c^k can be achieved by comparing D_{ETE} for all possible R_c , and the optimum Q^k can be calculated by the corresponding R_c^k .

B. Algorithm Design

In this subsection, we propose a practical algorithm for cross-layer rate-distortion optimization in wireless video communication.

Algorithm 1: Cross-layer optimized determination of quantization step size Q and code rate R_c for the k -th frame.

1) **Input:** R_t, γ, PEP_{th} .

2) Initialization of Q^k and R_c^k for the first frame, i.e., $k = 1$. If $k > 1$, go to 3).

3a) If $N^k > 1$.

Initialize $\Lambda_j = \Lambda_0$ by the method in Ref. [29]

loop for $\Lambda_j = \Lambda_0, \Lambda_1, \dots, \Lambda^*$,

for packet index i from 1 to $N^k > 1$,

For each packet, loop for all combinations

of $\{Q, R_c\}$ under the given Λ_j

calculate γ_{th} by (21),

estimate P_i^k for all packets by (20),

estimate D_T by (16),

estimate θ_1 by (7),

estimate D_Q by (15),

calculate $D_{ETE}(Q, R_c)$ by (3),

estimate $R_{s,i}^k$ by (6), (9), (11) and (5),

End

obtain the best $\{Q_i^k(\Lambda_j), R_{c,i}^k(\Lambda_j)\}$,

i.e., $\{Q_i^k(\Lambda_j), R_{c,i}^k(\Lambda_j)\}^*$ via (23),

End

estimate R_t^k by $R_{s,i}^k$ and $R_{c,i}^k$,

calculate Λ_{j+1} ,

End

obtain the best $\{Q_i^k, R_{c,i}^k\}$, i.e. $\{Q_i^k(\Lambda^*), R_{c,i}^k(\Lambda^*)\}^*$, for each packet.

3b) If $N^k = 1$.

loop for all code rates

calculate γ_{th} by (21),

estimate PEP for the k -th frame by (20),
 estimate D_T^k by (16),
 estimate \hat{H}^k by (5),
 estimate Q^k by (11), (9), (6) and (7),
 estimate \hat{D}_Q^k by (15),
 calculate $D_{ETE}(R_c)$ by (3),

End

select the best code rate R_c^k and corresponding Q^k
 with minimum end-to-end distortion.

4) **Output:** the best $\{Q_i^k, R_{c,i}^k\}$.

Algorithm 1 is referred to as CLRC. Note that Algorithm 1 considers both multi-packet case and single-packet case of each frame for purpose of general applications of our models. In Section V we only test R-D performance with small resolution sequences, i.e. CIF format, for interactive video communication over fading channels where one frame is in one packet which corresponds to 3b). In the proposed multi-packet case, i.e. 3a), the iterations to acquire the best Lagrange multiplier Λ^* use bi-section search [29], [43], which may cause high complexity. Ref. [44] shows a concrete example about how Lagrangian method can be used to replace bi-section search and derives an accurate function between Lagrange multiplier and Q for error-free channels. To reduce the complexity, we may also use the modeling method [44], [9] instead of the bi-section search to estimate Λ^* . In such a case, R-D optimized $\{Q, R_c\}$ decision is similar to the R-D optimized mode decision in Ref. [9] except three differences: 1) the mode decision in Ref. [9] is replaced by code rate decision given the Lagrange multiplier, 2) the quantization distortion is replaced by the end-to-end distortion, 3) the source coding bit-rate is replaced by the transmission bit-rate. Although simple, an inaccurate function between Lagrange multiplier and Q may cause inaccurate estimation of the best $\{Q_i^k, R_{c,i}^k\}$. Therefore, it is still a research problem to derive a similar function, as in Ref. [44], [9], between Lagrange multiplier and Q for 3a) in Algorithm 1. We will investigate low complexity Lagrangian method for 3a) for real-time encoding applications in our future work.

V. EXPERIMENTAL RESULTS

In Section V-A, we verify the accuracy of our proposed models. Then in Section V-B, we compare the R-D performance of rate control algorithm using our models to those using existing models. To show the performance gain of cross-layer rate control with channel estimation, we also compare our CLRC algorithm to the rate control algorithms without channel estimation.

A. Model Accuracy

In this subsection, we test the accuracy of our bit-rate model in (11), our distortion model in (15), and our PEP formula in (20). The JM16.0 [9] encoder is used to collect the true bit rate, true distortion and required statistics. Since CABAC is more accurate estimation of entropy than CAVLC, we use CABAC to evaluate our model accuracy in Fig. 4. Testing conditions in VCEG-AA10 [45] are followed. Sequences with CIF

resolution are downloaded from <http://trace.eas.asu.edu/yuv/> and sequences with 720p resolution are downloaded from <http://media.xiph.org/video/derf/>.

For bit rate model and distortion model, we compare our proposed models to ‘Rho-domain’ models proposed in Refs. [2] and Laplacian-based models proposed in Refs. [7]. In Refs. [7], authors basically use (6) and (12), with different forms from this paper, to estimate the bit rate and quantization distortion. In order to match the model to the true bit rate encoded by H.264 encoder, authors included some heuristic parameters to improve their bit rate model. In our experiment, we use the same values as suggested in Refs. [7].

1) *Bit rate model:* In order to make different curves distinguishable, Fig. 4 only shows the true residual bit-rate and estimated residual bit-rate for ‘foreman’ and ‘mobile’ for the first 20 frames⁸. In Fig. 4, ‘True bpp’ means the true bit per pixel (bpp) produced by the JM16.0 encoder; ‘without rlc’ means bpp estimated by (6); ‘with rlc’ means bpp estimated by (9); ‘without compensation’ means bpp estimated by (10); ‘with compensation’ means bpp estimated by (10) and (11); ‘Rho-domain’ means bpp estimated by Refs. [2]; ‘Xiang’s model’ means bpp estimated by Refs. [7].

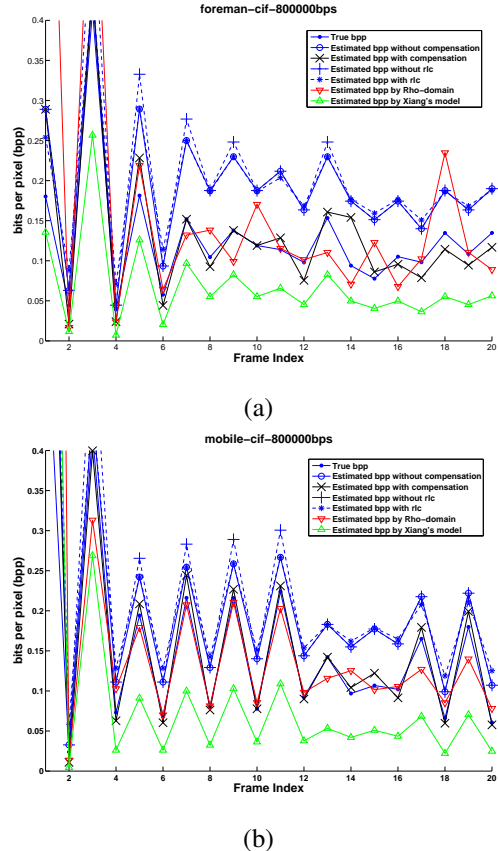


Fig. 4. bpp vs. Frame index: (a) foreman, (b) mobile.

We can see that the estimation accuracy is improved by (9) when true bpp is relatively large. However, when true

⁸Due to the page limitation, the estimation accuracy, measured by average deviation, of our proposed bit rate model comparing to existing models for more video sequences with full length are tabulated in an online table at www.mcn.ece.ufl.edu/public/zhifeng/papers/CLRC/table_bit-rate.pdf.

bpp is small, ‘without rlc’ gives higher estimation accuracy. By utilizing the statistics of the previous frame from (11), the estimation accuracy is further improved. We also find that ‘Rho-domain’ is accurate at low bpp; however, it is not accurate at high bpp. For ‘Xiang’s model’, the estimated bpp is smaller than the true bpp in most cases. Note that we also want to compare the bit-rate model used in JM16.0. However, due to the estimation error of its model parameters, the first few frames may abnormally underestimate the quantization step size Q . Therefore, the rate control algorithm in JM16.0 use three parameters, i.e., RCMInQPPSslice, RCMaxQPPSslice and RCMaxQPChange, to reduce the effect of the estimation error. Their default values are 8, 42, 4, respectively. However, we believe a good rate control algorithm should depend mainly on the model accuracy rather than those manually chosen thresholds. When those manual constraints are removed, i.e. those three parameters being set as 0, 51, 51, the estimated QP could even be 0 in the first few frames. That is, the first few frames consume most of the allocated bits, and there are only few bits available for the remaining frames in JM. Therefore, we do not test its model accuracy in this subsection. Instead, we will plot the R-D performance for it in Section V-B.

2) *Quantization distortion model*: Fig. 5 shows the quantization distortion curve corresponding to each bit-rate curve in Fig. 4⁹. Note that since Refs. [7] directly use (12) to estimate the quantization distortion, ‘without rlc’ means the quantization distortion estimated by both (12) and ‘Xiang’s model’. Similar to Fig. 4, we can see that the estimation accuracy is improved by (13) when θ_1 is small, i.e., when quantization distortion is relatively small. However, when quantization step size is large, (12) is more accurate than (13). Note that, the relativity is for the same video sequence. For different video sequences, since the residual variances are different, in order to achieve the same bit-rate, sequences with larger variance, e.g., ‘mobile’, will use larger quantization step size than sequences with lower variance, e.g., ‘foreman’. Different from the bit-rate model, which depends only on θ_1 , the quantization distortion model in (12) and (13) depends on both Q and θ_1 . Therefore, we cannot use the absolute value of quantization distortion between two sequences for comparing estimation accuracy of (12) and (13). After normalized by the factor Q^2 in (12) and (13), their relative accuracy is valid in most cases. However, in some rare cases, (13) is more accurate than (12) even when $Q > 3\sigma$. This can be observed for frame index from 14 to 17 in foreman sequence. We still need to investigate the reason behind it to further improve our model accuracy. For all cases, the estimation accuracy is improved by utilizing the statistics of the previous frame from (15). Similar to Fig. 4, ‘rho-domain’ is more accurate at large θ_1 , i.e., low bit-rate or relatively large quantization distortion, than at small θ_1 .

3) *PEP formula*: Here we verify the accuracy of the PEP formula in (20). We use the RCPC codes from Table I-VI in Ref. [46]. To be more specific, we choose a typical

⁹The estimation accuracy, measured by average deviation, of our proposed distortion model comparing to existing models for different video sequences with full length are tabulated in an online table at www.mcn.ece.ufl.edu/public/zhifeng/papers/CLRC/table_distortion.pdf.

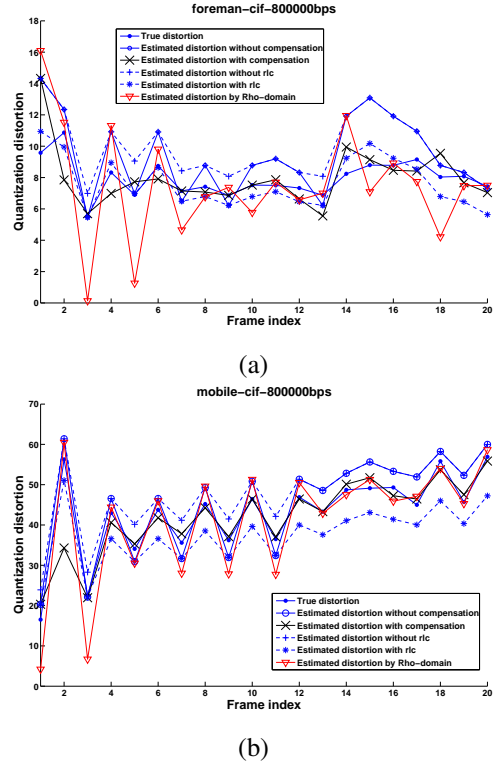


Fig. 5. Quantization vs. Frame index: (a) foreman, (b) mobile.

convolutional encoder structure with constraint length 7, i.e., 6 memory units, $G1 = 133$ and $G2 = 171$. The code rates are $2/3$, $3/4$, $4/5$, $5/6$, $6/7$ and $7/8$. For completeness, we put all encoder parameters in Table I. Viterbi algorithm is used to decode the received bits corrupted by noise. Coherent BPSK modulation is used. Each packet contains 2000 information bits. For each SNR and code rate, there are 1000 packets simulated to collect the true packet error rate (PER).

Fig. 6 shows the true PER and estimated PEP by the upper bound in (20). We can see that the estimated PEP curve is only about 1dB higher than the corresponding true PER curve.¹⁰

B. Performance Comparison

In this subsection, we show both objective performance and subjective performance of different rate control algorithms. We investigate the performance gain achieved by the higher accuracy of our models comparing to the existing models. In order to see the gain achieved by channel estimation, we also compare the performance achieved by (20) and (22). This result may serve as a guideline for system design to balance the performance and cost.

1) *Experiment Setup*: The JM16.0 encoder and decoder [9] are used in the experiments. Testing conditions in VCEG-AA10 [45] are followed. All the tested video sequences are in CIF resolution at 30fps. Each video sequence is encoded

¹⁰From the experimental results, we observe that the estimated PEP curves show a constant offset from the true PER curves given a RCPC encoder structure, and different RCPC encoder structures show different offsets. We may utilize this observation to further improve the PEP formula in our future work.

TABLE I
RCPC ENCODER PARAMETERS

code rate	puncturing matrix	dfree	$\{W_d\}$
2/3	[1 1, 1 0]	6	[1, 16, 48, 158, 642, 2435, 9174, 34701, 131533, 499312]
3/4	[1 1, 1 0, 0 1]	5	[8, 31, 160, 892, 4512, 23297, 120976, 624304, 3229885, 16721329]
4/5	[1 1, 1 0, 1 0, 1 0]	4	[3, 24, 172, 1158, 7408, 48706, 319563, 2094852, 13737566, 90083445]
5/6	[1 1, 1 0, 0 1, 1 0, 0 1]	4	[14, 69, 654, 4996, 39677, 314973, 2503576, 19875546, 157824160, 1253169928]
6/7	[1 1, 1 0, 1 0, 0 1, 1 0, 0 1]	3	[1, 20, 223, 1961, 18084, 168982, 1573256, 14620204, 135966265, 1264590899]
7/8	[1 1, 1 0, 1 0, 1 0, 0 1, 1 0, 0 1]	3	[2, 46, 499, 5291, 56137, 598557, 6371293, 67889502, 723039772, 7701832191]

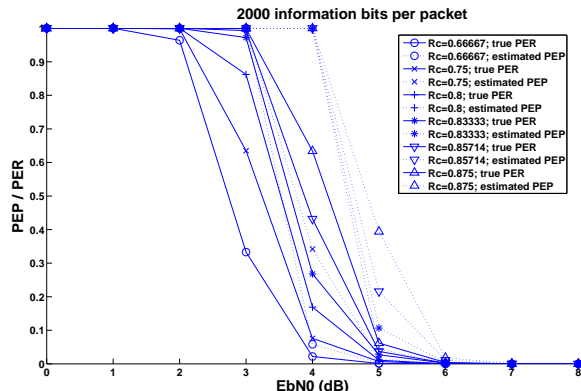


Fig. 6. PEP under different RCPC code rates.

for its first 100 frames, except stefan sequence¹¹, where the first frame is an I-frame and the following frames are P-frames. Each frame is included in one packet. The error concealment method is to copy the pixel value in the same position of the previous frame. The first frame is assumed to be correctly received with enough channel protection or timely acknowledgement feedback. To verify the model accuracy, 4x4 transform is used and CABAC is enabled for entropy coding. For all rate control algorithms, the first frame use a fix QP, i.e., QP=28.

Each encoded video sequence is tested under different Rayleigh fading channels, i.e., different combinations of transmission bit rate R_t from 500kbps to 2Mbps and average SNR from 15dB to 30dB. For each specific Rayleigh fading channel, we simulate 50 SNR samples for each average SNR to reduce the effect of random SNR sample. For each SNR sample, i.e., a specific PEP, there are 30 loops of simulated channel and decoding process to mitigate the effect of random error. RCPC codes and modulation are the same as those in Section V-A3.

2) *PSNR Performance*: Table II shows Y-component PSNR vs. average SNR and transmission bit rate. ‘JM’ represents the performance achieved by the default rate control algorithm in JM16.0 without channel estimation. ‘proposed’ represents the performance achieved by our models without channel estimation, i.e., using (22). We also include the rate control algorithm proposed in Ref. [32], where the bit rate model and distortion model are derived from the assumption of cauchy distribution. For each algorithm, we test several parameter settings of (RCMinQPPSlice, RCMaxQPPSlice, RCMaxQPChange) to verify how stable the resulted performance with those models

are under different manually set thresholds. In Table II, we show the experimental results for two QP-limitation settings, i.e., (8, 42, 4) and (8, 42, 51). The experimental results show that under the same QP-limitation, CLRC achieves remarkable PSNR gain in all sequences over ‘JM’, ‘Ref. [32]’, and ‘proposed’. This is because transmission distortion is the dominant part of the end-to-end distortion in ‘JM’, ‘Ref. [32]’, and ‘proposed’ algorithms. However, in the CLRC algorithm, due to the intelligent reference frame selection, the transmission distortion is negligible and quantization distortion becomes the dominant part of the end-to-end distortion. As a result, we also see that when the transmission bit rate increases the PSNR improves a lot in CLRC but not that much in other algorithms.

We also observe that ‘JM’ and ‘Ref. [32]’ show better results under QP-limitation (8, 42, 4) than less QP-limitation (8, 42, 51). When the QP-limitation is fully removed, i.e. use (0, 51, 51), both ‘JM’ and ‘Ref. [32]’ will consume almost all bits in the first few frames, therefore, the performance is even worse. The reason is because the estimation error of those bit rate models results in large encoded bit rate. In order to compensate the model inaccuracy, (RCMinQPPSlice, RCMaxQPPSlice, RCMaxQPChange) is (8, 42, 4) in default in JM encoder. On the other hand, when the setting of QP-limitation changes in a wide range, our models, especially with CLRC algorithm, show more stable results. Our ‘proposed’ method performs better than ‘JM’ with less strict QP-limitation, especially when the QP-limitation is fully removed, while sometimes worse than ‘JM’ with strict QP-limitation, e.g. the default QP-limitation in JM. The fact that no constraint of strict QP-limitation gives better performance further proves the higher degree of accuracy of our models.

Note that in Ref. [2], authors also propose bit-rate model, quantization distortion model and transmission distortion model for H.263 to solve joint source channel rate control problem. However, in both of their bit-rate model and quantization distortion model, only the model parameter, i.e., ‘rho’, can be estimated from a given bit-rate or quantization distortion. In order to estimated the quantization step size or QP from ‘rho’, those models requires the prior knowledge of residual histogram [47]. This kind of prior knowledge is accessible before determining QP in H.263 since H.263 encoders does not require QP before motion estimation. However, the residual histogram is not available in H.264 encoders before QP is given. In H.264 encoder, R-D cost, i.e. both rate and distortion, instead of only distortion is adopted as the criterion for motion estimation and mode decision. The R-D cost function induces a Lagrange multiplier, which can only be

¹¹For stefan sequence, there is only 90 frames in full.

TABLE II
PSNR VERSUS AVERAGE SNR AND TRANSMISSION BIT RATE

sequence	average SNR (dB)	bit rate (kbps)	JM	JM (QP-limit)	Ref. [32]	Ref. [32] (QP-limit)	proposed	proposed (QP-limit)	CLRC	CLRC (QP-limit)
foreman	20	500	26.32	30.00	19.86	27.21	29.13	26.78	36.97	36.92
		1000	24.54	26.98	15.82	26.01	26.82	27.11	39.54	39.28
		1500	24.63	29.89	24.62	28.94	28.27	27.95	35.51	40.56
	25	2000	26.21	29.81	27.58	27.34	31.04	27.22	41.61	41.36
		500	24.11	27.80	19.40	27.87	27.74	30.53	36.82	36.80
		1000	24.99	31.12	16.64	34.07	31.09	32.01	39.27	39.07
mobile	20	1500	28.22	30.62	18.29	33.28	32.77	33.41	35.51	40.31
		2000	27.66	31.68	31.67	27.84	29.02	30.49	41.27	41.08
		500	18.03	21.14	13.87	22.56	23.23	22.85	28.87	28.70
	25	1000	20.70	22.90	12.27	19.47	22.58	22.48	30.99	31.00
		1500	24.89	24.62	14.21	23.67	23.34	23.13	32.11	31.90
		2000	24.19	24.40	21.47	23.25	25.55	22.92	33.10	32.99
stefan	20	500	17.25	20.52	13.01	23.53	23.28	25.25	28.80	28.70
		1000	19.48	26.02	14.01	23.06	26.16	26.97	31.13	31.03
		1500	28.79	26.47	12.15	28.86	27.88	27.66	32.11	31.95
	25	2000	26.62	28.42	12.87	24.42	24.45	25.85	33.21	33.06
		500	17.91	22.65	14.99	22.01	22.75	21.97	31.00	30.96
		1000	17.44	22.03	14.71	18.52	22.12	22.53	24.87	33.42
stefan	20	1500	22.51	21.98	18.47	23.00	23.06	21.46	34.97	34.62
		2000	21.99	22.74	18.46	23.77	24.28	22.61	35.96	35.72
		500	18.16	23.13	15.15	22.98	23.43	24.46	30.30	30.36
	25	1000	18.05	25.62	14.68	19.72	25.55	26.41	32.53	32.43
		1500	27.68	26.76	14.85	25.81	26.01	25.56	33.49	33.43
		2000	26.32	25.06	26.48	24.21	25.73	25.63	34.25	34.30

determined after QP is known. Therefore, their bit-rate model encounters a chicken-and-egg problem if one tries to apply it for estimating quantization step size in H.264 encoders. Due to this reason, we do not implement those models in Ref. [2] for cross-layer rate control in the H.264 encoder [9]. Note that since the model parameters in Ref. [2] is attainable after real encoding, we still compare their model accuracy in Section V-A. For the accuracy comparison between our transmission distortion model and the transmission distortion model in Ref. [2], please refer to Ref. [26].

3) *Subjective Performance*: Since PSNR could be less meaningful for error concealment, a much more important performance criterion is the subjective performance, which directly relates to the degree of user's satisfaction. By utilizing the channel information, i.e., SNR and transmission bit-rate, our CLRC algorithm intelligently chooses the reference frames which are transmitted under the best channel conditions and neglects those references frames which experience poor channel conditions. As a result, the well-known error propagation problem is mitigated even during the encoding process.

Fig. 7(a) shows a random channel sample under average SNR=15dB and bit-rate=500kbps; Fig. 7(b) shows Distortion vs. Frame index for foreman_cif under this channel; Fig. 8 shows the corresponding subjective quality of reconstructed frames. 'cauchy-constant-PEP' represents the performance achieved by the rate control algorithm in Ref. [32] without channel estimation, i.e., using (22). We see that due to a low channel SNR during the timeslots of the 4-th frame, the encoder with CLRC skips encoding the 4-th frames to save encoding and transmission energy. Since there are no packets transmitted, the reconstructed picture of the 4-th frames at both the encoder and the decoder is the same as the reconstructed

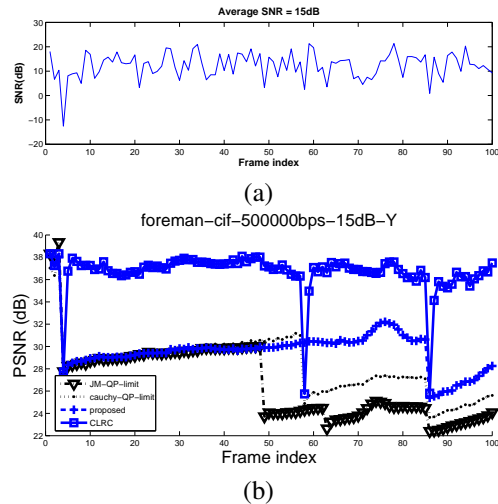


Fig. 7. A random channel sample under average SNR=15dB and bit-rate=500kbps: (a) A random SNR sample, (b) Distortion vs. Frame index for foreman_cif under this channel.

3-th frame. Then, when the channel condition goes well in the 5-th frame, encoder with CLRC uses the 3-th frame as reference to reconstruct the 5-th frame. Since the channel condition is good in the timeslot of the 5-th frame, there are no transmission distortion at the decoder. Therefore, the error propagation is stopped in the following frames.

For the encoder without channel estimation, the 4-th frame is encoded and transmitted. Due to the low channel SNR during the timeslots of the 4-th frame, the packets are received with error at the receiver and therefore, the resulted PSNR is almost the same as that of encoder with CLRC. However,

without channel information, the encoder still use the 4-th frame as one of the references for encoding the 5-th frame. Therefore, although the 5-th frame is correctly received at the receiver due to good channel condition, the reconstructed error in the 4-th frame are propagated into the 5-th frame at the decoder, which causes lower subjective quality and lower PSNR, compared to the encoder with CLRC. In Fig. 8, we also show the results of different algorithms in the 50-th frame to give a visual comparison. The 100-th frame is not shown here since they are highly distorted due to the error propagation in rate control algorithms without channel estimation. In Fig. 8, due to the space limit, we only show the subjective quality for encoder with ‘JM’ and ‘cauchy-constant-PEP’ under default QP limitation range. As we may foresee, the subjective quality for encoder with ‘JM’ and ‘cauchy-constant-PEP’ under maximum QP limitation range are worse.

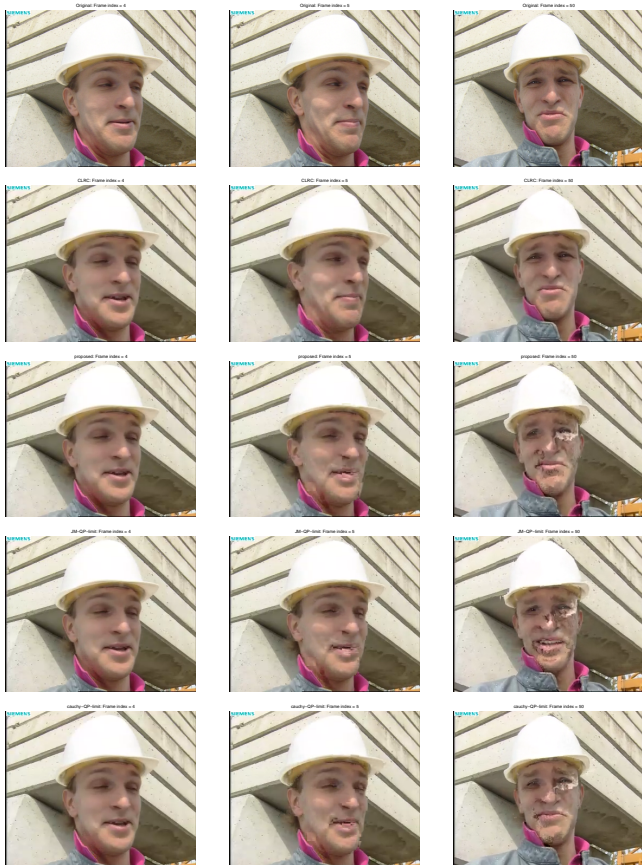


Fig. 8. Subjective result for foreman sequence

VI. CONCLUSION

In this paper, we derived source bit-rate model and quantization distortion model, which are more accurate than the existing models; we also improved the performance bound for channel coding under a RCPC encoder and a Viterbi decoder, and derived its performance bound under a Rayleigh block fading channel. Given the instantaneous channel condition, i.e., SNR and transmission bit-rate, we designed a rate-distortion optimized CLRC algorithm by jointly choosing quantization

step size and code rate. Experimental results showed that our proposed R-D models are much more accurate than the existing R-D models. Experimental results also showed that the rate control under our models has more stable R-D performance than the existing rate control algorithms; using the channel estimation, CLRC can further achieve remarkable R-D performance gain over that without channel estimation. Another important result is that the subjective quality of our CLRC algorithm is much better than the existing algorithms due to its intelligent reference frame selection.

APPENDIX

A. Proof of Equation (6)

Proof: For transform coefficients with i.i.d. zero-mean Laplacian distribution, the probability density function (pdf) is $p(x) = \frac{1}{\sqrt{2} \cdot \sigma} \cdot e^{-\frac{\sqrt{2} \cdot |x|}{\sigma}}$, where σ is the standard deviation. For the uniform quantizer with quantization step size Q and quantization offset θ_2 , the probability of zero after quantization is

$$P_0 = 2 \int_0^{Q \cdot (1 - \theta_2)} p(x) dx = 1 - e^{-\theta_1 \cdot (1 - \theta_2)}, \quad (24)$$

and the probability of level n after quantization is

$$P_n = \int_{Q \cdot (n - \theta_2)}^{Q \cdot (n + 1 - \theta_2)} p(x) dx = \frac{1}{2} (1 - e^{-\theta_1}) \cdot e^{\theta_1 \cdot \theta_2} \cdot e^{-\theta_1 \cdot n}, \quad (25)$$

where $\theta_1 = \frac{\sqrt{2} \cdot Q}{\sigma}$.

As a result,

$$\begin{aligned} H &= -P_0 \cdot \log_2 P_0 - 2 \sum_{n=1}^{\infty} P_n \cdot \log_2 P_n \\ &= -P_0 \cdot \log_2 P_0 + (1 - P_0) \\ &\quad \cdot \left(\frac{\theta_1 \cdot \log_2 e}{1 - e^{-\theta_1}} - \log_2(1 - e^{-\theta_1}) - \theta_1 \cdot \theta_2 \cdot \log_2 e + 1 \right). \end{aligned} \quad (26)$$

■

B. Calculation of Entropy for Different Quantized Transform Coefficients

Proof: For a 4x4 integer transform with average variance σ^2 , the variance for each transform coefficient can be calculate by (8) as

$$\sigma^2 = \frac{1}{16} \sum_{x=0}^4 \sum_{y=0}^4 \sigma_{(x,y)}^2 = \frac{225}{1024} \cdot \sigma_0^2. \quad (27)$$

Therefore, we have

$$\sigma_{(x,y)}^2 = 2^{-(x+y)} \cdot \frac{1024}{225} \cdot \sigma^2. \quad (28)$$

■

C. Proof of Proposition 1

Proof: In a Rayleigh fading channel, the received signal amplitude has the Rayleigh distribution, and received signal power has the exponential distribution. Therefore, SNR in receiver has the exponential distribution [27], that is,

$$P_\gamma(\gamma) = \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}}, \quad (29)$$

where $\bar{\gamma} = \frac{P_t \bar{g}}{N_0 B}$; P_t is transmission power; \bar{g} is the mean of channel gain; $\frac{N_0}{2}$ is noise power spectral density; and B is passband bandwidth.

By using the well-known upper bound as approximation for Q function, i.e., $Q(x) \approx \frac{1}{2} \cdot e^{-\frac{x^2}{2}}$ [48], from (20) we have

$$\begin{aligned} E_\gamma[PEP] &= \int_0^\infty PEP(\gamma) P_\gamma(\gamma) d\gamma \\ &\approx \int_0^{\gamma_{th}} \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}} d\gamma + \int_{\gamma_{th}}^\infty \left(\sum_{d=d_{free}}^{d_{max}} \frac{1}{2} LW_d e^{-\gamma d} \right) \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}} d\gamma \\ &= 1 - e^{-\frac{\gamma_{th}}{\bar{\gamma}}} + \sum_{d=d_{free}}^{d_{max}} \frac{1}{2} LW_d \int_{\gamma_{th}}^\infty \frac{1}{\bar{\gamma}} e^{-\gamma d - \frac{\gamma}{\bar{\gamma}}} d\gamma \\ &= 1 - e^{-\frac{\gamma_{th}}{\bar{\gamma}}} + \sum_{d=d_{free}}^{d_{max}} \frac{1}{2} LW_d e^{-d\gamma_{th}} \frac{1}{1 + d\bar{\gamma}} e^{-\frac{\gamma_{th}}{\bar{\gamma}}}, \end{aligned} \quad (30)$$

where γ_{th} is defined by (21).

let $f(d) = \frac{1}{2} LW_d e^{-d\gamma_{th}}$ and $PEP_{th} = 1$, from (21) we have $\sum_{d=d_{free}}^{d_{max}} f(d) = 1$. If we regard $f(d)$ as a pmf for d and further let $g(d) = \frac{1}{1+d\bar{\gamma}} e^{-\frac{\gamma_{th}}{\bar{\gamma}}}$, the third term in (30) can be regarded as a expected value of $g(d)$ with pmf $f(d)$. Since $f(d)$ decays exponentially with the increase of d , $g(d)$ can be approximated by an upper bound $\frac{1}{1+d_{free}\bar{\gamma}} e^{-\frac{\gamma_{th}}{\bar{\gamma}}}$. Therefore, (30) becomes

$$\begin{aligned} E_\gamma[PEP] &\approx 1 - e^{-\frac{\gamma_{th}}{\bar{\gamma}}} + \frac{e^{-\frac{\gamma_{th}}{\bar{\gamma}}}}{1 + d_{free}\bar{\gamma}} \sum_{d=d_{free}}^{d_{max}} \frac{1}{2} LW_d \cdot e^{-d\gamma_{th}} \\ &= 1 - e^{-\frac{\gamma_{th}}{\bar{\gamma}}} + \frac{e^{-\frac{\gamma_{th}}{\bar{\gamma}}}}{1 + d_{free}\bar{\gamma}}, \end{aligned} \quad (31)$$

In a practical communication system, $d_{free}\bar{\gamma} \gg 1$. On the other hand, since $\bar{\gamma} \gg \gamma_{th}$ as mentioned in Section III-C2 and $e^{-x} \approx 1 - xe^{-x}$ at small x , we may approximate $1 - e^{-\frac{\gamma_{th}}{\bar{\gamma}}}$ by $\frac{\gamma_{th}}{\bar{\gamma}} e^{-\frac{\gamma_{th}}{\bar{\gamma}}}$. Therefore, we have

$$\begin{aligned} E_\gamma[PEP] &\approx \frac{\gamma_{th}}{\bar{\gamma}} e^{-\frac{\gamma_{th}}{\bar{\gamma}}} + \frac{e^{-\frac{\gamma_{th}}{\bar{\gamma}}}}{d_{free}\bar{\gamma}} \\ &= \frac{\gamma_{th}}{\bar{\gamma}} e^{-\frac{\gamma_{th}}{\bar{\gamma}}} \left(1 + \frac{1}{d_{free}\bar{\gamma}} \right) \end{aligned} \quad (32)$$

Note that xe^{-x} increases as x increases in the interval $0 < x < 1$. Therefore, $E_\gamma[PEP]$ decreases as $\bar{\gamma}$ increases. ■

REFERENCES

- [1] K. Stuhlmüller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1012–1032, Jun. 2000.
- [2] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Transactions on Circuits and Systems for Video Technology, special issue on wireless video*, vol. 12, pp. 511–523, Jun. 2002.
- [3] L. Kondi, F. Ishtiaq, and A. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 1043–1052, 2002.
- [4] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172–185, 1999.
- [5] T. Chiang and Y. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–250, 1997.
- [6] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H. 264/AVC video coding and its application to rate control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, p. 1533, 2005.
- [7] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 2, pp. 193–205, 2009.
- [8] Z. He and S. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ -domain source modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 840–849, 2002.
- [9] "H.264/AVC reference software JM16.0," Jul. 2009. [Online]. Available: <http://iphome.hhi.de/suehring/tml/download>
- [10] H. Hang and J. Chen, "Source model for transform video coder and its application. I. Fundamental theory," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 287–298, 1997.
- [11] F. Moscheni, F. Dufaux, and H. Nicolas, "Entropy criterion for optimal bit allocation between motion and prediction error information," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 2094, 1993, pp. 235–242.
- [12] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
- [13] —, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec. Part*, vol. 4, pp. 142–163, 1959.
- [14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.
- [15] T. Berger, *Rate distortion theory: A mathematical basis for data compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
- [16] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2001.
- [17] I. QUANTIZERS and R. ROBERTS, "Analog Source Digitization: A Comparison of Theory and Practice," *IEEE Transactions on Information Theory*, 1967.
- [18] N. Farvardin and J. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless sources," *IEEE Transactions on Information Theory*, vol. 30, no. 3, p. 485, 1984.
- [19] *ITU-T Series H: Audiovisual and Multimedia Systems, Advanced video coding for generic audiovisual services*, Nov. 2007.
- [20] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, Jun. 2000.
- [21] T. Stockhammer, M. Hannuksela, and T. Wiegand, "H. 264/AVC in wireless environments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 657–673, 2003.
- [22] Y. Wang, Z. Wu, and J. M. Boyce, "Modeling of transmission-loss-induced distortion in decoded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 6, pp. 716–732, Jun. 2006.
- [23] C. Zhang, H. Yang, S. Yu, and X. Yang, "GOP-level transmission distortion modeling for mobile streaming video," *Signal Processing: Image Communication*, 2007.
- [24] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 861–874, Jul. 2008.
- [25] Z. Chen and D. Wu, "Prediction of Transmission Distortion for Wireless Video Communication: Part I: Analysis," 2010, <http://www.wu.ece.ufl.edu/mypapers/journal-1.pdf>.

- [26] —, “Prediction of Transmission Distortion for Wireless Video Communication: Algorithm and Application,” *Journal of Visual Communication and Image Representation*, vol. 21, no. 8, pp. 948–964, 2010.
- [27] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [28] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, 1983.
- [29] Y. Shoham and A. Gersho, “Efficient bit allocation for an arbitrary set of quantizers,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 9, pp. 1445–1453, 1988.
- [30] H. Everett III, “Generalized Lagrange multiplier method for solving problems of optimum allocation of resources,” *Operations Research*, vol. 11, no. 3, pp. 399–417, 1963.
- [31] Y. Altunbasak and N. Kamaci, “An analysis of the DCT coefficient distribution with the H. 264 video coder,” in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP’04). IEEE International Conference on*, vol. 3. IEEE, 2004, pp. iii–177.
- [32] N. Kamaci, Y. Altunbasak, and R. Mersereau, “Frame bit allocation for the H. 264/AVC video coder via Cauchy-density-based rate and distortion models,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 8, pp. 994–1006, 2005.
- [33] F. Muller, “Distribution shape of two-dimensional DCT coefficients of natural images,” *Electronics Letters*, vol. 29, no. 22, pp. 1935–1936, 1993.
- [34] G. Yovanof and S. Liu, “Statistical analysis of the DCT coefficients and their quantization error,” in *Signals, Systems and Computers, 1996. 1996 Conference Record of the Thirtieth Asilomar Conference on*, vol. 1. IEEE, 1996, pp. 601–605.
- [35] Z. Chen and K. Ngan, “Towards rate-distortion tradeoff in real-time color video coding,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 2, pp. 158–167, 2007.
- [36] E. Lam and J. Goodman, “A mathematical analysis of the dct coefficient distributions for images,” *Image Processing, IEEE Transactions on*, vol. 9, no. 10, pp. 1661–1666, 2000.
- [37] G. Sullivan and S. Sun, “On dead-zone plus uniform threshold scalar quantization,” in *Proc. SPIE*, vol. 5960, 2005, pp. 1041–1052.
- [38] G. Sullivan, P. Topiwala, and A. Luthra, “The H. 264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions,” in *SPIE Annual Conf. Apps. of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard H.264/AVC*, Aug. 2004, pp. 454–74.
- [39] T. Wiegand, W.-J. Han, B. Bross, J.-R. Ohm, and G. J. Sullivan, *WD1: Working Draft 1 of High-Efficiency Video Coding*, Guangzhou, Oct. 2010, JCTVC-C403, 3rd JCT-VC Meeting.
- [40] E. Malkamaki and H. Leib, “Evaluating the performance of convolutional codes over block fading channels,” *IEEE Transactions on Information Theory*, vol. 45, no. 5, p. 1643, 1999.
- [41] S. Kallel and C. Leung, “Efficient ARQ schemes with multiple copy decoding,” *IEEE Transactions on Communications*, vol. 40, no. 3, pp. 642–650, 1992.
- [42] 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.
- [43] K. Ramchandran and M. Vetterli, “Best wavelet packet bases in a rate-distortion sense,” *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 160–175, 1993.
- [44] G. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74–90, 1998.
- [45] T. Tan, G. Sullivan, and T. Wedi, “Recommended simulation common conditions for coding efficiency experiments,” in *ITU - Telecommunications Standardization Sector STUDY GROUP 16 Question 6 Video Coding Experts Group (VCEG), Doc. VCEG-AA10, 27th Meeting: Nice, France*, 2005.
- [46] G. Begin, D. Haccoun, and C. Paquin, “Further results on high-rate punctured convolutional codes for Viterbi and sequential decoding,” *IEEE Transactions on Communications*, vol. 38, no. 11, pp. 1922–1928, 1990.
- [47] Z. He and S. Mitra, “A linear source model and a unified rate control algorithm for DCT video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 11, pp. 970–982, 2002.
- [48] M. Simon and D. Divsalar, “Some new twists to problems involving the Gaussian probability integral,” *IEEE Transactions on Communications*, vol. 46, no. 2, 1998.

PLACE
PHOTO
HERE

Zhifeng Chen received Ph.D. degree in Electrical and Computer Engineering from the University of Florida, Gainesville, Florida, in 2010. He joined Interdigital Inc. in 2010, where he is currently a staff engineer working on video coding research.

PLACE
PHOTO
HERE

Dapeng Wu (S’98–M’04–SM’6) received Ph.D. in Electrical and Computer Engineering from Carnegie Mellon University, Pittsburgh, PA, in 2003. Currently, he is a professor of Electrical and Computer Engineering Department at University of Florida, Gainesville, FL.