Prediction of Transmission Distortion for Wireless Video Communication: Part I: Analysis

Zhifeng Chen and Dapeng Wu
Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611

Abstract

In the wireless video communication system, transmission distortion, caused by packet transmission errors, is a non-linear time-variant function of video frame statistics, channel condition and system parameters. By modeling the transmission distortion process as output of the random video sequence and channel error processes, the system can be modeled as an equivalent non-linear time-variant system. In this paper, for the first time, we identify the governing law that describes how the transmission distortion process evolves over time, and analytically derive it as a closed-form function of frame statistics, channel condition, and system parameters through a divide-and-conquer approach. Besides deriving the transmission distortion formula, this paper also identifies two important properties of transmission distortion for the first time. The first property is that the clipping noise, produced by non-linear clipping, causes decay of propagated error. The second property is that the correlation between motion vector concealment error and propagated error is negative, and has dominant impact on transmission distortion among all other correlations. In this paper, we also identify the relationship between our result and existing models, and specify the conditions, under which those models are accurate.

Index Terms

Wireless video, transmission distortion, clipping noise, slice data partitioning, Unequal Error Protection (UEP), time-varing channel.

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.

I. INTRODUCTION

Both multimedia technology and mobile communications have experienced massive growth and commercial success in recent years. As the two technologies converge, wireless video, such as video phone and mobile TV in 3G/4G systems, is expected to achieve unprecedented growth and worldwide success. However, different from the traditional video coding system, transmitting video over wireless with good quality or low end-to-end distortion is particularly challenging since the received video is subject to not only quantization error but also transmission error. In a wireless video communication system, end-to-end distortion consists of two parts: quantization distortion and transmission distortion. Quantization distortion is caused by quantization errors during the encoding process, and has been extensively studied in rate distortion theory [1], [2]. Transmission distortion is caused by packet errors during the transmission of a video sequence, and it is the major part of the end-to-end distortion in delay-sensitive wireless video communication[1] under high packet error probability (PEP), e.g., in a wireless fading channel.

The capability of predicting transmission distortion at the transmitter can assist in designing video encoding and transmission schemes that achieve maximum video quality under resource constraints. Specifically, transmission distortion prediction can be used in the following three applications in video encoding and transmission: 1) mode selection, which is to find the best intra/inter-prediction mode for encoding a macroblock (MB) with the minimum rate-distortion (R-D) cost given the instantaneous PEP, 2) cross-layer encoding rate control, which is to control the instantaneously encoded bit rate for a real-time encoder to minimize the frame-level end-to-end distortion given the instantaneous PEP, e.g., in video conferencing, 3) packet scheduling, which chooses a subset of packets of the pre-coded video to transmit and intentionally discards the remaining packets to minimize the GOP-level (Group of Picture) end-to-end distortion given the average PEP and average burst length, e.g., in streaming pre-coded video over networks. All the three applications require a formula for predicting how transmission distortion is affected by their respective control policy, in order to choose the optimal mode or encoding rate or transmission schedule.

However, predicting transmission distortion poses a great challenge due to the spatio-temporal correlation inside the input video sequence, the nonlinearity of both the encoder and the decoder, and varying PEP in time-varying channels. In a typical video codec, the temporal correlation among consecutive frames and the spatial correlation among the adjacent pixels of one frame are exploited to improve the

[1]Delay-sensitive wireless video communication usually does not allow retransmission to correct packet errors since retransmission may cause long delay.

coding efficiency. Nevertheless, such a coding scheme brings much difficulty in predicting transmission distortion because a packet error will degrade not only the video quality of the current frame but also the following frames due to error propagation. In addition, as we will see in Section IV, the nonlinearity of both the encoder and the decoder makes the instantaneous transmission distortion not equal to the sum of distortions caused by individual error events. Furthermore, in a wireless fading channel, the PEP is time-varying, which makes the error process a non-stationary random process and hence, as a function of the error process, the distortion process is also a non-stationary random process.

In this paper, we derived the transmission distortion formulae for wireless video communication systems. With consideration of spatio-temporal correlation, nonlinear codec and time-varying channel, our distortion prediction formulae provide, for the first time, the following capabilities: 1) support of prediction at different levels (e.g., pixel/frame/GOP level), 2) support of prediction for multi-reference motion compensation, 3) support of prediction under slice data partitioning, 4) support of prediction under arbitrary slice-level packetization with Flexible Macroblock Ordering (FMO) mechanism, 5) being applicable to time-varying channels, 6) one unified formula for both I-MB and P-MB, and 7) support of prediction for both low motion and high motion video sequences. Besides deriving the transmission distortion formulae, this paper also identified two important properties of transmission distortion for the first time: 1) clipping noise, produced by non-linear clipping, causes decay of propagated error; 2) the correlation between motion vector concealment error and propagated error is negative, and has dominant impact on transmission distortion, among all the correlations between any two of the four components in transmission error. We also discussed the relationship between our formula and existing models.

The rest of the paper is organized as follows. Section II reviews existing works for estimating distortion caused by packet transmission error. Section III presents the preliminaries of our system model under study to facilitate the derivations in the later sections, and illustrates the limitations of existing transmission distortion models. In Section IV, we derive the transmission distortion formula as a function of frame statistics, channel condition, and system parameters. Section V discusses the relationship between our formula and the existing models. In Section VI, we extend formulae for PTD and FTD from single-reference to multi-reference. Section VII concludes the paper.

II. Related Works

According to the aforementioned three applications, the existing algorithms for estimating transmission distortion can be categorized into the following three classes: 1) pixel-level or block-level algorithms (applied to mode selection), e.g., Recursive Optimal Per-pixel Estimate (ROPE) algorithm [3] and Law

of Large Number (LLN) algorithm [4], [5]; 2) frame-level or packet-level or slice-level algorithms (applied to cross-layer encoding rate control) [6], [7], [8], [9]; 3) GOP-level or sequence-level algorithms (applied to packet scheduling) [10], [11], [12], [13], [14]. Although the existing distortion estimation algorithms 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. In this paper, we use the divide-and-conquer approach to decompose complicated transmission distortion into four components, and analyze their effects on transmission distortion individually. This divide-and-conquer approach enables us to identify the governing law that describes how the transmission distortion process evolves over time.

Stuhlmuller et al. [6] observed that the distortion caused by the propagated error decays over time due to spatial filtering and intra coding of MBs, and analytically derived a formula for estimating transmission distortion under spatial filtering and intra coding. The effect of spatial filtering is analyzed under the implicit assumption that MVs are always correctly received at the receiver, while the effect of intra coding is modeled as a linear decay under another implicit assumption that the I-MBs are also always correctly received at the receiver. However, these two assumptions are usually not valid in realistic delay-sensitive wireless video communication. To address this, this paper derives the transmission distortion formula under the condition that both I-MBs and MVs may be erroneous at the receiver. In addition, we observe an interesting phenomenon that even without using the spatial filtering and intra coding, the distortion caused by the propagated error still decays! We identify, for the first time, that this decay is caused by non-linear clipping, which is used to clip those out-of-range[2] reconstructed pixel after motion compensation; this is the first of the two properties identified in this paper. While such out-of-range values produced by the inverse transform of quantized transform coefficients is negligible at the encoder, its counterpart produced by transmission error at the decoder has significant impact on transmission distortion.

Some existing works [6], [7] estimate transmission distortion based on a linear time-invariant (LTI) system model, which regards packet error as input and transmission distortion as output. The LTI model simplifies the analysis of transmission distortion. However, it sacrifices accuracy in distortion estimation since it neglects the effect of correlation between newly induced error and propagated error. Liang et al. [14] studied the effect of correlation and observed that the LTI models [6], [7] underestimate transmission distortion due to the positive correlation between two adjacent erroneous frames; however,

[2]A reconstructed pixel value may be out of the range of the original pixel value, e.g., [0, 255].

they did not consider the effect of motion vector (MV) error on transmission distortion and their algorithm was not tested with high motion videos. To address these issues and find the cause of the under-estimation, this paper classifies the transmission reconstructed error into three individual random errors, namely, Residual Concealment Error (RCE), MV Concealment Error (MVCE), and propagated error; the first two types of error are called *newly induced error*. We identify, for the first time, that MVCE is negatively correlated with propagated error and this correlation has dominant impact on transmission distortion, among all the correlations between any two of the three error types, for high motion videos; this is the second of the two properties identified in this paper. For this reason, as long as MV transmission errors exist in high motion videos, the LTI model over-estimates transmission distortion. We also quantifies the effect of individual error types and their correlations on transmission distortion in this paper. Thanks to the analysis of correlation effect, our distortion formula is accurate for both low motion video and high motion video as verified by experimental results. Another merit of considering the effect of MV error on transmission distortion is the applicability of our results to video communication with slice data partitioning, where the residual and MV could be transmitted under Unequal Error Protection (UEP).

Refs. [3], [4], [8], [9] proposed some models to estimate transmission distortion under the consideration that both MV and I-MB may experience transmission errors. However, the parameters in the linear models [8], [9] can only be acquired by experimentally curve-fitting over multiple frames, which forbids the models from estimating instantaneous distortion. In addition, the linear models [8], [9] still assume there is no correlation between the newly induced error and propagated error. In Ref. [3], the ROPE algorithm considers the correlation between MV concealment error and propagated error by recursively calculating the second moment of the reconstructed pixel value. However, ROPE neglects the non-linear clipping function and therefore over-estimates the distortion. In addition, the extension of ROPE algorithm [15] to support the averaging operations, such as interpolation and deblocking filtering in H.264, requires intensive computation of correlation coefficients due to the high correlation between reconstructed values of adjacent pixels, and thereby prohibiting it from applying to H.264. In H.264 reference code JM14.0 [16], the LLN algorithm [4] is adopted since it is capable of supporting both clipping and averaging operations. However, in order to predict transmission distortion, all possible error events for each pixel in all frames should be simulated at the encoder, which significantly increases the complexity of the encoder. Different from Refs. [3], [4], the divide-and-conquer approach in this paper enables our formula to provide not only more accurate prediction but also lower complexity and higher degree of extensibility. The multiple reference picture motion compensated prediction extended from the single reference is analyzed in Section VI, and, for the first time, the effect of multiple references on

transmission distortion is quantified. In addition, the transmission distortion formula derived in this paper is unified for both I-MBs and P-MBs, in contrast to two different formulae in Refs. [3], [8], [9].

Different from wired channels, wireless channels suffer from multipath fading, which can be regarded as multiplicative random noise. Fading leads to time-varying PEP and burst errors in wireless video communication. Ref. [6] uses a two-state stationary Markov chain to model burst errors. However, even if the channel gain is stationary, packet error process is a non-stationary random process. Specifically, since PEP is a function of the channel gain [17], which is not constant in a wireless fading channel, instantaneous PEP is also not constant. This means the probability distribution of packet error state is time-varying in wireless fading channels, that is, the packet error process is a non-stationary random process. Hence the Markov chain in Ref. [6] is neither stationary, nor ergodic for wireless fading channel. As a result, averaging the burst length and PEP as in Ref. [6] cannot accurately predict instantaneous distortion. To address this, this paper derives the formula for Pixel-level Transmission Distortion (PTD) by considering non-stationarity over time. Regarding the Frame-level Transmission Distortion (FTD), since two adjacent MBs may be assigned to two different packets, under the slice-level packetization and FMO mechanism in H.264 [18], [19], their error probability could be different. However, existing frame-level distortion models [6], [7], [8], [9] assume all pixels in the same frame experience the same channel condition. As a result, the applicable scope of those models are limited to video with small resolution. In contrast, this paper derives the formula for FTD by considering non-stationarity over space. Due to consideration of non-stationarity over time and over space, our formula provides an accurate prediction of transmission distortion in a time-varying channel.

III. System Description

A. Structure of a Wireless Video Communication System

Fig. 1 shows the structure of a typical wireless video communication system. It consists of an encoder, two channels and a decoder where residual packets and MV packets are transmitted over their respective channels. If residual packets or MV packets are erroneous, the error concealment module will be activated. In typical video encoders such as H.263/264 and MPEG-2/4 encoders, the functional blocks can be divided into two classes: 1) basic parts, such as predictive coding, transform, quantization, entropy coding, motion compensation, and clipping; and 2) performance-enhanced parts, such as interpolation filtering, deblocking filtering, B-frame, multi-reference prediction, etc. Although the up-to-date video encoder includes more and more performance-enhanced parts, the basic parts do not change. In this paper, we use the structure in Fig. 1 for transmission distortion analysis.

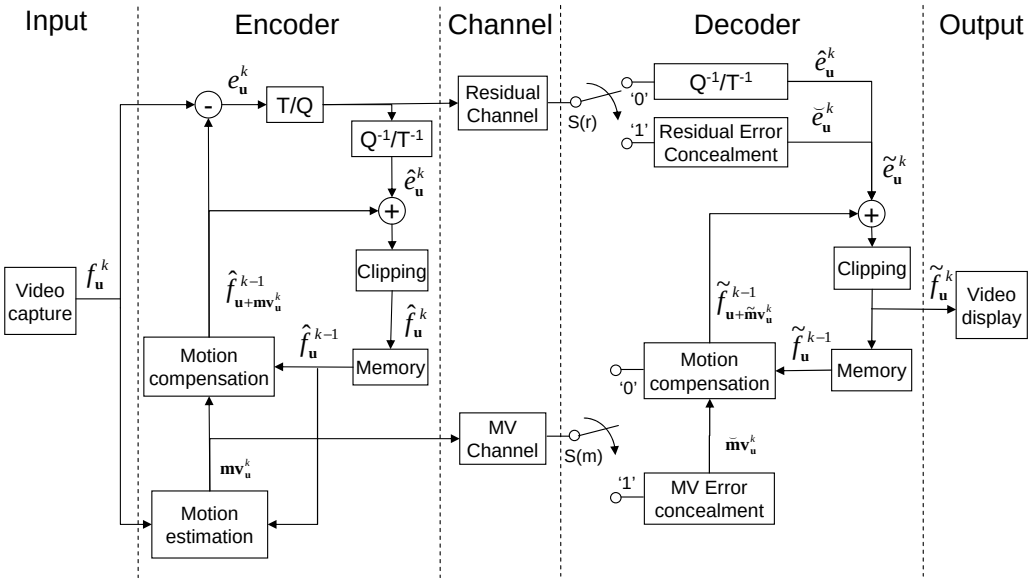

Fig. 1. System structure, where T, Q, Q^{-1}, and T^{-1} denote transform, quantization, inverse quantization, and inverse transform, respectively.

Note that in this system, both residual channel and MV channel are application-layer channels; specifically, both channels consist of entropy coding and entropy decoding, networking layers[3], and physical layer (including channel encoding, modulation, wireless fading channel, demodulation, channel decoding). Although the residual channel and MV channel usually share the same physical-layer channel, the two application-layer channels may have different parameter settings (e.g., different channel code-rate) for the slice data partitioning under UEP. For this reason, our formula obtained from the structure in Fig. 1 can be used to estimate transmission distortion for an encoder with slice data partitioning.

B. Clipping Noise

In this subsection, we examine the effect of clipping noise on the reconstruction pixel value along each pixel trajectory over time (frames). All pixel positions in a video sequence form a three-dimensional spatio-temporal domain, i.e., two dimensions in spatial domain and one dimension in temporal domain. Each pixel can be uniquely represented by $\mathbf{u}^k$ in this three-dimensional time-space, where k means the k-th frame in temporal domain and $\mathbf{u}$ is a two-dimensional vector in spatial domain. The philosophy

[3]Here, networking layers can include any layers other than physical layer.

behind inter-coding of a video sequence is to represent the video sequence by virtual motion of each pixel, i.e., each pixel recursively moves from position $\mathbf{v}^{k-1}$ to position $\mathbf{u}^k$. The difference between these two positions is a two-dimensional vector called MV of pixel $\mathbf{u}^k$, i.e., $\mathbf{mv}_{\mathbf{u}}^k = \mathbf{v}^{k-1} - \mathbf{u}^k$. The difference between the pixel values of these two positions is called residual of pixel $\mathbf{u}^k$, that is, $e_{\mathbf{u}}^k = f_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1}$[4]. Recursively, each pixel in the k-th frame has one and only one reference pixel trajectory backward towards the latest I-frame.

At the encoder, after transform, quantization, inverse quantization, and inverse transform for the residual, the reconstructed pixel value for $\mathbf{u}^k$ may be out-of-range and should be clipped as

$$\hat{f}_{\mathbf{u}}^k = \Gamma(\hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1} + \hat{e}_{\mathbf{u}}^k), \tag{1}$$

where $\Gamma(\cdot)$ function is a clipping function defined by

$$\Gamma(x) = \begin{cases} \gamma_L, & x < \gamma_L \\ x, & \gamma_L \le x \le \gamma_H \\ \gamma_H, & x > \gamma_H, \end{cases} \tag{2}$$

where γ_L and γ_H are user-specified low threshold and high threshold, respectively. Usually, $\gamma_L = 0$ and $\gamma_H = 255$.

The residual and MV at the decoder may be different from their counterparts at the encoder because of channel impairments. Denote $\widetilde{\mathbf{mv}}_{\mathbf{u}}^k$ and $\widetilde{e}_{\mathbf{u}}^k$ the MV and residual at the decoder, respectively. Then, the reference pixel position for $\mathbf{u}^k$ at the decoder is $\widetilde{\mathbf{v}}^{k-1} = \mathbf{u}^k + \widetilde{\mathbf{mv}}_{\mathbf{u}}^k$, and the reconstructed pixel value for $\mathbf{u}^k$ at the decoder is

$$\widetilde{f}_{\mathbf{u}}^k = \Gamma(\widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{e}_{\mathbf{u}}^k). \tag{3}$$

In error-free channels, the reconstructed pixel value at the receiver is exactly the same as the reconstructed pixel value at the transmitter, because there is no transmission error and hence no transmission distortion. However, in error-prone channels, we know from (3) that $\widetilde{f}_{\mathbf{u}}^k$ is a function of three factors: the received residual $\widetilde{e}_{\mathbf{u}}^k$, the received MV $\widetilde{\mathbf{mv}}_{\mathbf{u}}^k$, and the propagated error $\widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$. The received residual $\widetilde{e}_{\mathbf{u}}^k$ depends on three factors, namely, 1) the transmitted residual $\hat{e}_{\mathbf{u}}^k$, 2) the residual packet error state, which depends on instantaneous residual channel condition, and 3) the residual error concealment algorithm if the received residual packet is erroneous. Similarly, the received MV $\widetilde{\mathbf{mv}}_{\mathbf{u}}^k$ depends on 1) the transmitted $\mathbf{mv}_{\mathbf{u}}^k$, 2) the MV packet error state, which depends on instantaneous MV channel condition, and 3) the

[4]For simplicity of notation, we move the superscript k of $\mathbf{u}$ to the superscript k of f whenever $\mathbf{u}$ is the subscript of f.

TABLE I

NOTATIONS

$\mathbf{u}^k$	: Three-dimensional vector that denotes a pixel position in an video sequence
$f_{\mathbf{u}}^k$	: Value of the pixel $\mathbf{u}^k$
$e_{\mathbf{u}}^k$	: Residual of the pixel $\mathbf{u}^k$
$\mathbf{mv}_{\mathbf{u}}^k$	: MV of the pixel $\mathbf{u}^k$
$\Delta_{\mathbf{u}}^k$	: Clipping noise of the pixel $\mathbf{u}^k$
$\varepsilon_{\mathbf{u}}^k$	: Residual concealment error of the pixel $\mathbf{u}^k$
$\xi_{\mathbf{u}}^k$	: MV concealment error of the pixel $\mathbf{u}^k$
$\zeta_{\mathbf{u}}^k$	: Transmission reconstructed error of the pixel $\mathbf{u}^k$
$S_{\mathbf{u}}^k$	: Error state of the pixel $\mathbf{u}^k$
$P_{\mathbf{u}}^k$	: Error probability of the pixel $\mathbf{u}^k$
$D_{\mathbf{u}}^k$	: Transmission distortion of the pixel $\mathbf{u}^k$
D^k	: Transmission distortion of the k-th frame
$\mathcal{V}^k$	: Set of all the pixels in the k-th frame
$\lvert\mathcal{V}\rvert$	: Number of elements in set $\mathcal{V}$ (cardinality of $\mathcal{V}$)
α^k	: Propagation factor of the k-th frame
β^k	: Percentage of I-MBs in the k-th frame
λ^k	: Correlation ratio of the k-th frame
$w^k(j)$	: pixel percentage of using frame $k-j$ as reference in the k-th frame

MV error concealment algorithm if the received MV packet is erroneous. The propagated error $\widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$ includes the error propagated from the reference frames, and therefore depends on all samples in the previous frames indexed by $i < k$ and their reception error states as well as concealment algorithms.

The non-linear clipping function within the pixel trajectory makes the distortion estimation more challenging. However, it is interesting to observe that clipping actually reduces transmission distortion. In Section IV, we will quantify the effect of clipping on transmission distortion.

Table I lists notations used in this paper. All vectors are in bold font. Note that the encoder needs to reconstruct the compressed video for predictive coding; hence the encoder and the decoder have a similar structure for pixel value reconstruction. To distinguish the variables in the reconstruction module of the encoder from those in the reconstruction module of the decoder, we add $\hat{}$ onto the variables at the encoder and add $\widetilde{}$ onto the variables at the decoder.

C. Definition of Transmission Distortion

In this subsection, we define PTD and FTD to be derived in Section IV. To calculate FTD, we need some notations from set theory. In a video sequence, all pixel positions in the k-th frame form a two-dimensional vector set $\mathcal{V}^k$, and we denote the number of elements in set $\mathcal{V}^k$ by $|\mathcal{V}^k|$. So, for any pixel at position $\mathbf{u}$ in the k-th frame, i.e., $\mathbf{u} \in \mathcal{V}^k$, its reference pixel position is chosen from set $\mathcal{V}^{k-1}$ for single-reference.

For a transmitter with feedback acknowledgement of whether a packet is correctly received at the receiver (called acknowledgement feedback), $\widetilde{f}_{\mathbf{u}}^k$ at the decoder side can be perfectly reconstructed by the transmitter, as long as the transmitter knows the error concealment algorithm used by the receiver. Then, the transmission distortion for the k-th frame can be calculated by mean squared error (MSE) as

$$MSE^k = \frac{1}{|\mathcal{V}^k|} \cdot \sum_{\mathbf{u} \in \mathcal{V}^k} [(\hat{f}_{\mathbf{u}}^k - \widetilde{f}_{\mathbf{u}}^k)^2]. \tag{4}$$

For the encoder, every pixel intensity $f_{\mathbf{u}}^k$ of the random input video sequence is a random variable. For any encoder with hybrid coding (see Fig. 1), the residual $\hat{e}_{\mathbf{u}}^k$, MV $\mathbf{mv}_{\mathbf{u}}^k$, and reconstructed pixel value $\hat{f}_{\mathbf{u}}^k$ are functions of $f_{\mathbf{u}}^k$; so they are also random variables before motion estimation[5]. Given the Probability Mass Function (PMF) of $\hat{f}_{\mathbf{u}}^k$ and $\widetilde{f}_{\mathbf{u}}^k$, we define the transmission distortion for pixel $\mathbf{u}^k$ or PTD by

$$D_{\mathbf{u}}^k \triangleq E[(\hat{f}_{\mathbf{u}}^k - \widetilde{f}_{\mathbf{u}}^k)^2], \tag{5}$$

and we define the transmission distortion for the k-th frame or FTD by

$$D^k \triangleq E[\frac{1}{|\mathcal{V}^k|} \cdot \sum_{\mathbf{u} \in \mathcal{V}^k} (\hat{f}_{\mathbf{u}}^k - \widetilde{f}_{\mathbf{u}}^k)^2]. \tag{6}$$

It is easy to prove that the relationship between FTD and PTD is characterized by

$$D^k = \frac{1}{|\mathcal{V}^k|} \cdot \sum_{\mathbf{u} \in \mathcal{V}^k} D_{\mathbf{u}}^k. \tag{7}$$

If the number of bits used to compress a frame is too large to be contained in one packet, the bits of the frame are split into multiple packets. In a time-varying channel, different packet of the same frame may experience different packet error probability (PEP). If pixel $\mathbf{u}^k$ and pixel $\mathbf{v}^k$ belong to different packets, the PMF of $\widetilde{f}_{\mathbf{u}}^k$ may be different from the PMF of $\widetilde{f}_{\mathbf{v}}^k$ even if $\hat{f}_{\mathbf{u}}^k$ and $\hat{f}_{\mathbf{v}}^k$ are identically distributed. In other words, $D_{\mathbf{u}}^k$ may be different from $D_{\mathbf{v}}^k$ even if pixel $\mathbf{u}^k$ and pixel $\mathbf{v}^k$ are in the neighboring MBs when

[5]In applications such as cross-layer encoding rate control, distortion estimation for rate-distortion optimized bit allocation is required before motion estimation.

FMO is activated. As a result, FTD D^k in (7) may be different from PTD $D^k_{\mathbf{u}}$ in (5). For this reason, we will derive formulae for both PTD and FTD, respectively. Note that most existing frame-level distortion models [6], [7], [8], [9] assume that all pixels in the same frame experience the same channel condition and simply use (5) for FTD; however this assumption is not valid for high-resolution/high-quality video transmission over a time-varying channel.

In fact, (7) is a general form for distortions of all levels. If $|\mathcal{V}| = 1$, (7) reduces to (5). For slice/packet-level distortion, $\mathcal{V}$ is the set of the pixels contained in a slice/packet. For GOP-level distortion, $\mathcal{V}$ is the set of the pixels contained in a GOP. In this paper, we only show how to derive formulae for PTD and FTD. Our methodology is also applicable to deriving formulae for slice/packet/GOP-level distortion by using appropriate $\mathcal{V}$.

D. Limitations of the Existing Transmission Distortion Models

In this subsection, we show that clipping noise has significant impact on transmission distortion, and neglect of clipping noise in existing models results in inaccurate estimation of transmission distortion. We define the clipping noise for pixel $\mathbf{u}^k$ at the encoder as

$$\hat{\Delta}^k_{\mathbf{u}} \triangleq (\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}}) - \Gamma(\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}}), \tag{8}$$

and the clipping noise for pixel $\mathbf{u}^k$ at the decoder as

$$\widetilde{\Delta}^k_{\mathbf{u}} \triangleq (\widetilde{f}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{e}^k_{\mathbf{u}}) - \Gamma(\widetilde{f}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{e}^k_{\mathbf{u}}). \tag{9}$$

Using (1), Eq. (8) becomes

$$\hat{f}^k_{\mathbf{u}} = \hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}} - \hat{\Delta}^k_{\mathbf{u}}, \tag{10}$$

and using (3), Eq. (9) becomes

$$\widetilde{f}^k_{\mathbf{u}} = \widetilde{f}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{e}^k_{\mathbf{u}} - \widetilde{\Delta}^k_{\mathbf{u}}, \tag{11}$$

where $\hat{\Delta}^k_{\mathbf{u}}$ only depends on the video content and encoder structure, e.g., motion estimation, quantization, mode decision and clipping function; and $\widetilde{\Delta}^k_{\mathbf{u}}$ depends on not only the video content and encoder structure, but also channel conditions and decoder structure, e.g., error concealment and clipping function.

In most existing works, both $\hat{\Delta}^k_{\mathbf{u}}$ and $\widetilde{\Delta}^k_{\mathbf{u}}$ are neglected, i.e., these works assume $\hat{f}^k_{\mathbf{u}} = \hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}}$ and $\widetilde{f}^k_{\mathbf{u}} = \widetilde{f}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{e}^k_{\mathbf{u}}$. However, this assumption is only valid for stored video or error-free communication. For error-prone communication, decoder clipping noise $\widetilde{\Delta}^k_{\mathbf{u}}$ has a significant impact on transmission distortion and hence should not be neglected. To illustrate this, Table II shows an example for the system in Fig. 1, where only the residual packet in the $(k-1)$-th frame is erroneous at the decoder (i.e.,

TABLE II

AN EXAMPLE THAT SHOWS THE EFFECT OF CLIPPING NOISE ON TRANSMISSION DISTORTION.

Encoder	Transmitted	$\hat{f}_{\mathbf{w}}^{k-2}=250$	$\hat{e}_{\mathbf{v}}^{k-1}=-50$ (erroneous)	$\hat{e}_{\mathbf{u}}^{k}=50$
	Reconstructed	$\hat{f}_{\mathbf{w}}^{k-2}=250$	$\hat{f}_{\mathbf{v}}^{k-1}=\Gamma(\hat{f}_{\mathbf{w}}^{k-2}+\hat{e}_{\mathbf{v}}^{k-1})=200$	$\hat{f}_{\mathbf{u}}^{k}=\Gamma(\hat{f}_{\mathbf{v}}^{k-1}+\hat{e}_{\mathbf{u}}^{k})=250$
Decoder	Received	$\widetilde{f}_{\mathbf{w}}^{k-2}=250$	$\widetilde{e}_{\mathbf{v}}^{k-1}=0$ (concealed)	$\widetilde{e}_{\mathbf{u}}^{k}=50$
	Reconstructed	$\widetilde{f}_{\mathbf{w}}^{k-2}=250$	$\widetilde{f}_{\mathbf{v}}^{k-1}=\Gamma(\widetilde{f}_{\mathbf{w}}^{k-2}+\widetilde{e}_{\mathbf{v}}^{k-1})=250$	$\widetilde{f}_{\mathbf{u}}^{k}=\Gamma(\widetilde{f}_{\mathbf{v}}^{k-1}+\widetilde{e}_{\mathbf{u}}^{k})=255$
	Clipping noise	$\widetilde{\Delta}_{\mathbf{w}}^{k-2}=0$	$\widetilde{\Delta}_{\mathbf{v}}^{k-1}=0$	$\widetilde{\Delta}_{\mathbf{u}}^{k}=45$
	Distortion	$D_{\mathbf{w}}^{k-2}=(\hat{f}_{\mathbf{w}}^{k-2}-\widetilde{f}_{\mathbf{w}}^{k-2})^2=0$	$D_{\mathbf{v}}^{k-1}=(\hat{f}_{\mathbf{v}}^{k-1}-\widetilde{f}_{\mathbf{v}}^{k-1})^2=2500$	$D_{\mathbf{u}}^{k}=(\hat{f}_{\mathbf{u}}^{k}-\widetilde{f}_{\mathbf{u}}^{k})^2=25$
Prediction without clipping	Received	$\widetilde{f}_{\mathbf{w}}^{k-2}=250$	$\widetilde{e}_{\mathbf{v}}^{k-1}=0$ (concealed)	$\widetilde{e}_{\mathbf{u}}^{k}=50$
	Reconstructed	$\widetilde{f}_{\mathbf{w}}^{k-2}=250$	$\widetilde{f}_{\mathbf{v}}^{k-1}=\widetilde{f}_{\mathbf{w}}^{k-2}+\widetilde{e}_{\mathbf{v}}^{k-1}=250$	$\widetilde{f}_{\mathbf{u}}^{k}=\widetilde{f}_{\mathbf{v}}^{k-1}+\widetilde{e}_{\mathbf{u}}^{k}=300$
	Distortion	$\hat{D}_{\mathbf{w}}^{k-2}=(\hat{f}_{\mathbf{w}}^{k-2}-\widetilde{f}_{\mathbf{w}}^{k-2})^2=0$	$\hat{D}_{\mathbf{v}}^{k-1}=(\hat{f}_{\mathbf{v}}^{k-1}-\widetilde{f}_{\mathbf{v}}^{k-1})^2=2500$	$\hat{D}_{\mathbf{u}}^{k}=(\hat{f}_{\mathbf{u}}^{k}-\widetilde{f}_{\mathbf{u}}^{k})^2=2500$

$\hat{e}_{\mathbf{v}}^{k-1}$ is erroneous), and all other residual packets and all the MV packets are error-free. Suppose the trajectory of pixel $\mathbf{u}^k$ in the $(k-1)$-th frame and $(k-2)$-th frame is specified by $\mathbf{v}^{k-1}=\mathbf{u}^k+\mathbf{mv}_{\mathbf{u}}^k$ and $\mathbf{w}^{k-2}=\mathbf{v}^{k-1}+\mathbf{mv}_{\mathbf{v}}^{k-1}$. Since $\hat{e}_{\mathbf{v}}^{k-1}$ is erroneous, the decoder needs to conceal the error; a simple concealment scheme is to let $\widetilde{e}_{\mathbf{v}}^{k-1}=0$. From this example, we see that neglect of clipping noise (e.g., $\widetilde{\Delta}_{\mathbf{u}}^{k}=45$) results in highly inaccurate estimate of distortion, e.g., the estimated distortion $\hat{D}_{\mathbf{u}}^{k}=2500$ (without considering clipping) is much larger than the true distortion $D_{\mathbf{u}}^{k}=25$. Note that if an MV is erroneous at the decoder, the pixel trajectory at the decoder will be different from the trajectory at the encoder; then the resulting clipping noise $\widetilde{\Delta}_{\mathbf{u}}^{k}$ may be much larger than 45 as in this example, and hence the distortion estimation of the existing models without considering clipping may be much more inaccurate.

On the other hand, the encoder clipping noise $\hat{\Delta}_{\mathbf{u}}^{k}$ has negligible effect on quantization distortion and transmission distortion. This is due to two reasons: 1) the probability that $\hat{\Delta}_{\mathbf{u}}^{k}=0$, is close to one, since the probability that $\gamma_L\le\hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1}+\hat{e}_{\mathbf{u}}^{k}\le\gamma_H$, is close to one; 2) in case that $\hat{\Delta}_{\mathbf{u}}^{k}\neq 0$, $\hat{\Delta}_{\mathbf{u}}^{k}$ usually takes a value that is much smaller than the residuals. Since $\hat{\Delta}_{\mathbf{u}}^{k}$ is negligible, the clipping function can be removed at the encoder if only quantization distortion needs to be considered, e.g., for stored video or error-free communication. Since $\hat{\Delta}_{\mathbf{u}}^{k}$ is very likely to be a very small value, we would neglect it and assume $\hat{\Delta}_{\mathbf{u}}^{k}=0$ in deriving our formula for transmission distortion.

IV. Transmission Distortion Formulae

In this section, we derive formulae for PTD and FTD. The section is organized as below. Section IV-A presents an overview of our approach to analyzing PTD and FTD. Then we elaborate on the derivation details in Section IV-B through Section IV-E. Specifically, Section IV-B quantifies the effect of RCE on transmission distortion; Section IV-C quantifies the effect of MVCE on transmission distortion; Section IV-D quantifies the effect of propagated error and clipping noise on transmission distortion; Section IV-E quantifies the effect of correlations (between any two of the error sources) on transmission distortion. Finally, Section IV-F summarizes the key results of this paper, i.e., the formulae for PTD and FTD.

A. Overview of the Approach to Analyzing PTD and FTD

To analyze PTD and FTD, we take a divide-and-conquer approach. We first divide transmission reconstructed error into four components: three random errors (RCE, MVCE and propagated error) due to their different physical causes, and clipping noise, which is a non-linear function of these three random errors. This error decomposition allows us to further decompose transmission distortion into four terms, i.e., distortion caused by 1) RCE, 2) MVCE, 3) propagated error plus clipping noise, and 4) correlations between any two of the error sources, respectively. This distortion decomposition facilitates the derivation of a simple and accurate closed-form formula for each of the four distortion terms. Next, we elaborate on error decomposition and distortion decomposition.

Define transmission reconstructed error for pixel $\mathbf{u}^k$ by $\widetilde{\zeta}_{\mathbf{u}}^k \triangleq \hat{f}_{\mathbf{u}}^k - \widetilde{f}_{\mathbf{u}}^k$. From (10) and (11), we obtain

$$\begin{aligned}\widetilde{\zeta}_{\mathbf{u}}^k &= (\hat{e}_{\mathbf{u}}^k + \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1} - \hat{\Delta}_{\mathbf{u}}^k) - (\widetilde{e}_{\mathbf{u}}^k + \widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} - \widetilde{\Delta}_{\mathbf{u}}^k) \\ &= (\hat{e}_{\mathbf{u}}^k - \widetilde{e}_{\mathbf{u}}^k) + (\hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1} - \hat{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}) + (\hat{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} - \widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}) - (\hat{\Delta}_{\mathbf{u}}^k - \widetilde{\Delta}_{\mathbf{u}}^k).\end{aligned} \tag{12}$$

Define RCE $\widetilde{\varepsilon}_{\mathbf{u}}^k$ by $\widetilde{\varepsilon}_{\mathbf{u}}^k \triangleq \hat{e}_{\mathbf{u}}^k - \widetilde{e}_{\mathbf{u}}^k$, and define MVCE $\widetilde{\xi}_{\mathbf{u}}^k$ by $\widetilde{\xi}_{\mathbf{u}}^k \triangleq \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1} - \hat{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$. Note that $\hat{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} - \widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} = \widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$, which is the transmission reconstructed error of the concealed reference pixel in the reference frame; we call $\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$ propagated error. As mentioned in Section III-D, we assume $\hat{\Delta}_{\mathbf{u}}^k = 0$. Therefore, (12) becomes

$$\widetilde{\zeta}_{\mathbf{u}}^k = \widetilde{\varepsilon}_{\mathbf{u}}^k + \widetilde{\xi}_{\mathbf{u}}^k + \widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^k. \tag{13}$$

(13) is our proposed **error decomposition**.

Combining (5) and (13), we have

$$\begin{aligned} D_{\mathbf{u}}^k &= E[(\widetilde{\varepsilon}_{\mathbf{u}}^k + \widetilde{\xi}_{\mathbf{u}}^k + \widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^k)^2] \\ &= E[(\widetilde{\varepsilon}_{\mathbf{u}}^k)^2] + E[(\widetilde{\xi}_{\mathbf{u}}^k)^2] + E[(\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^k)^2] \\ &\quad + 2E[\widetilde{\varepsilon}_{\mathbf{u}}^k \cdot \widetilde{\xi}_{\mathbf{u}}^k] + 2E[\widetilde{\varepsilon}_{\mathbf{u}}^k \cdot (\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^k)] + 2E[\widetilde{\xi}_{\mathbf{u}}^k \cdot (\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^k)]. \end{aligned} \tag{14}$$

Denote $D_{\mathbf{u}}^k(r) \triangleq E[(\widetilde{\varepsilon}_{\mathbf{u}}^k)^2]$, $D_{\mathbf{u}}^k(m) \triangleq E[(\widetilde{\xi}_{\mathbf{u}}^k)^2]$, $D_{\mathbf{u}}^k(P) \triangleq E[(\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^k)^2]$ and $D_{\mathbf{u}}^k(c) \triangleq 2E[\widetilde{\varepsilon}_{\mathbf{u}}^k \cdot \widetilde{\xi}_{\mathbf{u}}^k] + 2E[\widetilde{\varepsilon}_{\mathbf{u}}^k \cdot (\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^k)] + 2E[\widetilde{\xi}_{\mathbf{u}}^k \cdot (\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^k)]$. Then, (14) becomes

$$D_{\mathbf{u}}^k = D_{\mathbf{u}}^k(r) + D_{\mathbf{u}}^k(m) + D_{\mathbf{u}}^k(P) + D_{\mathbf{u}}^k(c). \tag{15}$$

(15) is our proposed **distortion decomposition** for PTD. The reason why we combine propagated error and clipping noise into one term (called clipped propagated error) is because clipping noise is mainly caused by propagated error and such decomposition will simplify the formulae.

There are three major reasons for our decompositions in (13) and (15). First, if we directly substitute the terms in (5) by (10) and (11), it will produce 5 second moments and 10 cross-correlation terms (assuming $\hat{\Delta}_{\mathbf{u}}^k = 0$); since there are 8 possible error events due to three individual random errors, there are a total of $8 \times (5+10) = 120$ terms for PTD, making the analysis highly complicated. In contrast, our decompositions in (13) and (15) significantly simplify the analysis. Second, each term in (13) and (15) has a clear physical meaning, which leads to accurate estimation algorithms with low complexity. Third, such decompositions allow our formulae to be easily extended for supporting advanced video codec with more performance-enhanced parts, e.g., multi-reference prediction and interpolation filtering.

To derive the formula for FTD, from (7) and (15), we obtain

$$D^k = D^k(r) + D^k(m) + D^k(P) + D^k(c), \tag{16}$$

where

$$D^k(r) = \frac{1}{|\mathcal{V}|} \cdot \sum_{\mathbf{u} \in \mathcal{V}} D_{\mathbf{u}}^k(r), \tag{17}$$

$$D^k(m) = \frac{1}{|\mathcal{V}|} \cdot \sum_{\mathbf{u} \in \mathcal{V}} D_{\mathbf{u}}^k(m), \tag{18}$$

$$D^k(P) = \frac{1}{|\mathcal{V}|} \cdot \sum_{\mathbf{u} \in \mathcal{V}} D_{\mathbf{u}}^k(P), \tag{19}$$

$$D^k(c) = \frac{1}{|\mathcal{V}|} \cdot \sum_{\mathbf{u} \in \mathcal{V}} D_{\mathbf{u}}^k(c). \tag{20}$$

(16) is our proposed distortion decomposition for FTD.

Usually, the set $\mathcal{V}^k$ in a video sequence is the same for all frame k, i.e., $\mathcal{V}^1 = \cdots = \mathcal{V}^k$ for all $k > 1$. Hence, we remove the frame index k and denote the set of pixel positions of an arbitrary frame by $\mathcal{V}$. Note that in H.264, a reference pixel may be in a position out of picture boundary; however, the set of reference pixels, which is larger than the input pixel set, is still the same for all frame k.

B. Analysis of Distortion Caused by RCE

In this subsection, we first derive the pixel-level residual caused distortion $D_{\mathbf{u}}^k(r)$. Then we derive the frame-level residual caused distortion $D^k(r)$.

1) Pixel-level Distortion Caused by RCE: We denote $S_{\mathbf{u}}^k$ as the state indicator of whether there is transmission error for pixel $\mathbf{u}^k$ after channel decoding. Note that as mentioned in Section III-A, both the residual channel and the MV channel contain channel decoding; hence in this paper, the transmission error in the residual channel or the MV channel is meant to be the error uncorrectable by the channel decoding. To distinguish the residual error state and the MV error state, here we use $S_{\mathbf{u}}^k(r)$ to denote the residual error state for pixel $\mathbf{u}^k$. That is, $S_{\mathbf{u}}^k(r) = 1$ if $\hat{e}_{\mathbf{u}}^k$ is received with error, and $S_{\mathbf{u}}^k(r) = 0$ if $\hat{e}_{\mathbf{u}}^k$ is received without error. At the receiver, if there is no residual transmission error for pixel $\mathbf{u}$, $\widetilde{e}_{\mathbf{u}}^k$ is equal to $\hat{e}_{\mathbf{u}}^k$. However, if the residual packets are received with error, we need to conceal the residual error at the receiver. Denote $\check{e}_{\mathbf{u}}^k$ the concealed residual when $S_{\mathbf{u}}^k(r) = 1$, and we have,

$$\widetilde{e}_{\mathbf{u}}^k = \begin{cases} \check{e}_{\mathbf{u}}^k, & S_{\mathbf{u}}^k(r) = 1 \\ \hat{e}_{\mathbf{u}}^k, & S_{\mathbf{u}}^k(r) = 0. \end{cases} \tag{21}$$

Note that $\check{e}_{\mathbf{u}}^k$ depends on $\hat{e}_{\mathbf{u}}^k$ and the residual concealment method, but does not depend on the channel condition. From the definition of $\widetilde{\varepsilon}_{\mathbf{u}}^k$ and (21), we have

$$\begin{aligned} \widetilde{\varepsilon}_{\mathbf{u}}^k &= (\hat{e}_{\mathbf{u}}^k - \check{e}_{\mathbf{u}}^k) \cdot S_{\mathbf{u}}^k(r) + (\hat{e}_{\mathbf{u}}^k - \hat{e}_{\mathbf{u}}^k) \cdot (1 - S_{\mathbf{u}}^k(r)) \\ &= (\hat{e}_{\mathbf{u}}^k - \check{e}_{\mathbf{u}}^k) \cdot S_{\mathbf{u}}^k(r). \end{aligned} \tag{22}$$

$\hat{e}_{\mathbf{u}}^k$ depends on the input video sequence and the encoder structure, while $S_{\mathbf{u}}^k(r)$ depends on communication system parameters such as delay bound, channel coding rate, transmission power, channel gain of the wireless channel. Under our framework shown in Fig. 1, the input video sequence and the encoder structure are independent of communication system parameters. Since $\hat{e}_{\mathbf{u}}^k$ and $S_{\mathbf{u}}^k(r)$ are solely caused by independent sources, we assume $\hat{e}_{\mathbf{u}}^k$ and $S_{\mathbf{u}}^k(r)$ are independent. That is, we make the following assumption.

Assumption 1: $S_{\mathbf{u}}^k(r)$ is independent of $\hat{e}_{\mathbf{u}}^k$.

Assumption 1 means that whether $\hat{e}_{\mathbf{u}}^k$ will be correctly received or not, does not depend on the value of $\hat{e}_{\mathbf{u}}^k$. Denote $\varepsilon_{\mathbf{u}}^k \triangleq \hat{e}_{\mathbf{u}}^k - \check{e}_{\mathbf{u}}^k$; we have $\widetilde{\varepsilon}_{\mathbf{u}}^k = \varepsilon_{\mathbf{u}}^k \cdot S_{\mathbf{u}}^k(r)$. Denote $P_{\mathbf{u}}^k(r)$ as the residual pixel error probability (XEP) for pixel $\mathbf{u}^k$, that is, $P_{\mathbf{u}}^k(r) \triangleq P\{S_{\mathbf{u}}^k(r) = 1\}$. Then, from (22) and Assumption 1, we have

$$D_{\mathbf{u}}^k(r) = E[(\widetilde{\varepsilon}_{\mathbf{u}}^k)^2] = E[(\varepsilon_{\mathbf{u}}^k)^2] \cdot E[(S_{\mathbf{u}}^k(r))^2] = E[(\varepsilon_{\mathbf{u}}^k)^2] \cdot (1 \cdot P_{\mathbf{u}}^k(r)) = E[(\varepsilon_{\mathbf{u}}^k)^2] \cdot P_{\mathbf{u}}^k(r). \tag{23}$$

Hence, our formula for the pixel-level residual caused distortion is

$$D_{\mathbf{u}}^k(r) = E[(\varepsilon_{\mathbf{u}}^k)^2] \cdot P_{\mathbf{u}}^k(r). \tag{24}$$

2) Frame-level Distortion Caused by RCE: To derive the frame-level residual caused distortion, the encoder needs to know the second moment of RCE for each pixel in that frame. However, if encoder knows the characteristics of residual process and concealment method, the formulae will be much simplified. One simple concealment method is to let $\check{e}_{\mathbf{u}}^k = 0$ for all erroneous pixels. A more general concealment method is to use the neighboring pixels to conceal an erroneous pixel. So we make the following assumption.

Assumption 2: The residual $\hat{e}_{\mathbf{u}}^k$ is stationary with respect to 2D variable $\mathbf{u}$ in the same frame. In addition, $\check{e}_{\mathbf{u}}^k$ only depends on $\{\hat{e}_{\mathbf{v}}^k : \mathbf{v} \in \mathcal{N}_{\mathbf{u}}\}$ where $\mathcal{N}_{\mathbf{u}}$ is a fixed neighborhood of $\mathbf{u}$.

In other words, Assumption 2 assumes that 1) $\hat{e}_{\mathbf{u}}^k$ is a 2D stationary stochastic process and the distribution of $\hat{e}_{\mathbf{u}}^k$ is the same for all $\mathbf{u} \in V^k$, and 2) $\check{e}_{\mathbf{u}}^k$ is also a 2D stationary stochastic process since it only depends on the neighboring $\hat{e}_{\mathbf{u}}^k$. Hence, $\hat{e}_{\mathbf{u}}^k - \check{e}_{\mathbf{u}}^k$ is also a 2D stationary stochastic process, and its second moment $E[(\hat{e}_{\mathbf{u}}^k - \check{e}_{\mathbf{u}}^k)^2] = E[(\varepsilon_{\mathbf{u}}^k)^2]$ is the same for all $\mathbf{u} \in V^k$. Therefore, we can drop $\mathbf{u}$ from the notation, and let $E[(\varepsilon^k)^2] = E[(\varepsilon_{\mathbf{u}}^k)^2]$ for all $\mathbf{u} \in V^k$.

Denote $N_i^k(r)$ as the number of pixels contained in the i-th residual packet of the k-th frame; denote $P_i^k(r)$ as PEP of the i-th residual packet of the k-th frame; denote $N^k(r)$ as the total number of residual packets of the k-th frame. Since for all pixels in the same packet, the residual XEP is equal to its PEP, from (17) and (24), we have

$$\begin{aligned} D^k(r) &= \frac{1}{|\mathcal{V}|} \sum_{\mathbf{u} \in \mathcal{V}^k} E[(\varepsilon_{\mathbf{u}}^k)^2] \cdot P_{\mathbf{u}}^k(r) && (25) \\ &= \frac{1}{|\mathcal{V}|} \sum_{\mathbf{u} \in \mathcal{V}^k} E[(\varepsilon^k)^2] \cdot P_{\mathbf{u}}^k(r) && (26) \\ &\overset{(a)}{=} \frac{E[(\varepsilon^k)^2]}{|\mathcal{V}|} \sum_{i=1}^{N^k(r)} (P_i^k(r) \cdot N_i^k(r)) && (27) \\ &\overset{(b)}{=} E[(\varepsilon^k)^2] \cdot \bar{P}^k(r). && (28) \end{aligned}$$

where (a) is due to $P^k_{\mathbf{u}}(r) = P^k_i(r)$ for pixel $\mathbf{u}$ in the i-th residual packet; (b) is due to

$$\bar{P}^k(r) \triangleq \frac{1}{|\mathcal{V}|} \sum_{i=1}^{N^k(r)} (P^k_i(r) \cdot N^k_i(r)). \tag{29}$$

$\bar{P}^k(r)$ is a weighted average over PEPs of all residual packets in the k-th frame, in which different packets may contain different numbers of pixels. Hence, our formula for the frame-level residual caused distortion is

$$D^k(r) = E[(\varepsilon^k)^2] \cdot \bar{P}^k(r). \tag{30}$$

C. Analysis of Distortion Caused by MVCE

Similar to the derivations in Section IV-B1, in this subsection, we derive the formula for the pixel-level MV caused distortion $D^k_{\mathbf{u}}(m)$, and the frame-level MV caused distortion $D^k(m)$.

1) Pixel-level Distortion Caused by MVCE: Denote the MV error state for pixel $\mathbf{u}^k$ by $S^k_{\mathbf{u}}(m)$, and denote the concealed MV by $\check{\mathbf{mv}}^k_{\mathbf{u}}$ when $S^k_{\mathbf{u}}(m) = 1$. Therefore, we have

$$\widetilde{\mathbf{mv}}^k_{\mathbf{u}} = \begin{cases} \check{\mathbf{mv}}^k_{\mathbf{u}}, & S^k_{\mathbf{u}}(m) = 1 \\ \mathbf{mv}^k_{\mathbf{u}}, & S^k_{\mathbf{u}}(m) = 0. \end{cases} \tag{31}$$

Here, we use the temporal error concealment [20] to conceal MV errors. Denote $\xi^k_{\mathbf{u}} \triangleq \hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} - \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}}$, where $\xi^k_{\mathbf{u}}$ depends on the accuracy of MV concealment, and the spatial correlation between reference pixel and concealed reference pixel at the encoder. We also make the following assumption.

Assumption 3: $S^k_{\mathbf{u}}(m)$ is independent of $\xi^k_{\mathbf{u}}$.

Denote $P^k_{\mathbf{u}}(m)$ as the MV XEP for pixel $\mathbf{u}^k$, that is, $P^k_{\mathbf{u}}(m) \triangleq P\{S^k_{\mathbf{u}}(m) = 1\}$, and following the same derivation process in Section IV-B1, we can obtain

$$D^k_{\mathbf{u}}(m) = E[(\xi^k_{\mathbf{u}})^2] \cdot P^k_{\mathbf{u}}(m). \tag{32}$$

Also note that in H.264 specification, there is no slice data partitioning for an instantaneous decoding refresh (IDR) frame [21]; so $S^k_{\mathbf{u}}(r)$ and $S^k_{\mathbf{u}}(m)$ are fully correlated in an IDR-frame, that is, $S^k_{\mathbf{u}}(r) = S^k_{\mathbf{u}}(m)$, and hence $P^k_{\mathbf{u}}(r) = P^k_{\mathbf{u}}(m)$. This is also true for MB without slice data partitioning. For P-MB with slice data partitioning in H.264, $S^k_{\mathbf{u}}(r)$ and $S^k_{\mathbf{u}}(m)$ are partially correlated. In other words, if the MV packet is lost, the corresponding residual packet cannot be decoded even if it is correctly received, since there is no slice header in the residual packet. Therefore, the residual channel and the MV channel in Fig. 1 are actually correlated if the encoder follows H.264 specification. In this paper, we

study transmission distortion in a more general case where $S_{\mathbf{u}}^k(r)$ and $S_{\mathbf{u}}^k(m)$ can be either independent or correlated.[6]

2) Frame-level Distortion Caused by MVCE: To derive the frame-level MV caused distortion, we also make the following assumption.

Assumption 4: The second moment of $\xi_{\mathbf{u}}^k$ is the same for all $\mathbf{u} \in V^k$.

Under Assumption 4, we can drop $\mathbf{u}$ from the notation, and let $E[(\xi^k)^2] = E[(\xi_{\mathbf{u}}^k)^2]$ for all $\mathbf{u} \in V^k$. Denote $N_i^k(m)$ as the number of pixels contained in the i-th MV packet of the k-th frame; denote $P_i^k(m)$ as PEP of the i-th MV packet of the k-th frame; denote $N^k(m)$ as the total number of MV packets of the k-th frame. Following the same derivation process in Section IV-B2, we obtain the frame-level MV caused distortion for the k-th frame as

$$D^k(m) = E[(\xi^k)^2] \cdot \bar{P}^k(m), \tag{33}$$

where $\bar{P}^k(m) \triangleq \frac{1}{|\mathcal{V}|} \sum_{i=1}^{N_m^k} (P_i^k(m) \cdot N_i^k(m))$, a weighted average over PEPs of all MV packets in the k-th frame, in which different packets may contain different numbers of pixels.

D. Analysis of Distortion Caused by Propagated Error Plus Clipping Noise

In this subsection, we derive the distortion caused by error propagation in a non-linear decoder with clipping. We first derive the pixel-level propagation and clipping caused distortion $D_{\mathbf{u}}^k(P)$. Then we derive the frame-level propagation and clipping caused distortion $D^k(P)$.

1) Pixel-level Distortion Caused by Propagated Error Plus Clipping Noise: First, we analyze the pixel-level propagation and clipping caused distortion $D_{\mathbf{u}}^k(P)$ in P-MBs. From the definition, we know $D_{\mathbf{u}}^k(P)$ depends on propagated error and clipping noise; and clipping noise is a function of RCE, MVCE and propagated error. Hence, $D_{\mathbf{u}}^k(P)$ depends on RCE, MVCE and propagated error. Let r, m, p denote the event of occurrence of RCE, MVCE and propagated error respectively, and let $\bar{r}, \bar{m}, \bar{p}$ denote logical NOT of r, m, p respectively (indicating no error). We use a triplet to denote the joint event of three types of error; e.g., $\{r, m, p\}$ denotes the event that all the three types of errors occur, and $\mathbf{u}^k\{\bar{r}, \bar{m}, \bar{p}\}$ denotes the pixel $\mathbf{u}^k$ experiencing none of the three types of errors.

When we analyze the condition that several error events may occur, the notation could be simplified by the principle of formal logic. For example, $\widetilde{\Delta}_{\mathbf{u}}^k\{\bar{r}, \bar{m}\}$ denotes the clipping noise under the condition that

[6]To achieve this, we change the H.264 reference code JM14.0 by allowing residual packets to be used for decoder without the corresponding MV packets being correctly received, that is, $\hat{e}_{\mathbf{u}}^k$ can be used to reconstruct $\widetilde{f}_{\mathbf{u}}^k$ even if $\mathbf{mv}_{\mathbf{u}}^k$ is not correctly received.

there is neither RCE nor MVCE for pixel $\mathbf{u}^k$, while it is not certain whether the reference pixel has error. Correspondingly, denote $P^k_{\mathbf{u}}\{\bar{r},\bar{m}\}$ as the probability of event $\{\bar{r},\bar{m}\}$, that is, $P^k_{\mathbf{u}}\{\bar{r},\bar{m}\} = P\{S^k_{\mathbf{u}}(r) = 0 \text{ and } S^k_{\mathbf{u}}(m) = 0\}$. From the definition of $P^k_{\mathbf{u}}(r)$, the marginal probability $P^k_{\mathbf{u}}\{r\} = P^k_{\mathbf{u}}(r)$ and the marginal probability $P^k_{\mathbf{u}}\{\bar{r}\} = 1 - P^k_{\mathbf{u}}(r)$. The same, $P^k_{\mathbf{u}}\{m\} = P^k_{\mathbf{u}}(m)$ and $P^k_{\mathbf{u}}\{\bar{m}\} = 1 - P^k_{\mathbf{u}}(m)$.

Define $D^k_{\mathbf{u}}(p) \triangleq E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r},\bar{m}\})^2]$; and define $\alpha^k_{\mathbf{u}} \triangleq \frac{D^k_{\mathbf{u}}(p)}{D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}}$, which is called propagation factor for pixel $\mathbf{u}^k$. The propagation factor $\alpha^k_{\mathbf{u}}$ defined in this paper is different from the propagation factor [9], leakage [6], or attenuation factor [14], which are modeled as the effect of spatial filtering or intra update; our propagation factor $\alpha^k_{\mathbf{u}}$ is also different from the fading factor [7], which is modeled as the effect of using fraction of referenced pixels in the reference frame for motion prediction. Note that $D^k_{\mathbf{u}}(p)$ is only a special case of $D^k_{\mathbf{u}}(P)$ under the error event of $\{\bar{r},\bar{m}\}$ for pixel $\mathbf{u}^k$. However, most existing models inappropriately use their propagation factor, obtained under the error event of $\{\bar{r},\bar{m}\}$, to replace $D^k_{\mathbf{u}}(P)$ of all other error events directly.

To calculate $E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}})^2]$ in (14), we need to analyze $\widetilde{\Delta}^k_{\mathbf{u}}$ in four different error events for pixel $\mathbf{u}^k$: 1) both residual and MV are erroneous, denoted by $\mathbf{u}^k\{r,m\}$; 2) residual is erroneous but MV is correct, denoted by $\mathbf{u}^k\{r,\bar{m}\}$; 3) residual is correct but MV is erroneous, denoted by $\mathbf{u}^k\{\bar{r},m\}$; and 4) both residual and MV are correct, denoted by $\mathbf{u}^k\{\bar{r},\bar{m}\}$. So,

$$\begin{aligned} D^k_{\mathbf{u}}(P) = {} & P^k_{\mathbf{u}}\{r,m\}\cdot E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{r,m\})^2] + P^k_{\mathbf{u}}\{r,\bar{m}\}\cdot E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{r,\bar{m}\})^2] \\ & + P^k_{\mathbf{u}}\{\bar{r},m\}\cdot E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r},m\})^2] + P^k_{\mathbf{u}}\{\bar{r},\bar{m}\}\cdot E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r},\bar{m}\})^2]. \end{aligned} \tag{34}$$

Note that the concealed pixel value should be in the clipping function range, that is, $\Gamma(\widetilde{f}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \check{e}^k_{\mathbf{u}}) = \widetilde{f}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \check{e}^k_{\mathbf{u}}$, so $\widetilde{\Delta}^k_{\mathbf{u}}\{r\} = 0$. Also note that if the MV channel is independent of the residual channel, we have $P^k_{\mathbf{u}}\{r,m\} = P^k_{\mathbf{u}}(r)\cdot P^k_{\mathbf{u}}(m)$. However, as mentioned in Section IV-C1, in H.264 specification, these two channels are correlated. In other words, $P^k_{\mathbf{u}}\{\bar{r},m\} = 0$ and $P^k_{\mathbf{u}}\{\bar{r},\bar{m}\} = P^k_{\mathbf{u}}\{\bar{r}\}$ for P-MBs with slice data partitioning in H.264. In such a case, (34) is simplified to

$$D^k_{\mathbf{u}}(P) = P^k_{\mathbf{u}}\{r,m\}\cdot D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + P^k_{\mathbf{u}}\{r,\bar{m}\}\cdot D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + P^k_{\mathbf{u}}\{\bar{r}\}\cdot D^k_{\mathbf{u}}(p). \tag{35}$$

In a more general case, where $P^k_{\mathbf{u}}\{\bar{r},m\} \neq 0$, Eq. (35) is still valid. This is because $P^k_{\mathbf{u}}\{\bar{r},m\} \neq 0$ only happens under slice data partitioning condition, where $P^k_{\mathbf{u}}\{\bar{r},m\} \ll P^k_{\mathbf{u}}\{\bar{r},\bar{m}\}$ and $E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r},\bar{m}\})^2] \approx E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r},m\})^2]$ under UEP. Therefore, the last two terms in (34) is almost equal to $P^k_{\mathbf{u}}\{\bar{r}\}\cdot D^k_{\mathbf{u}}(p)$.

Note that for P-MB without slice data partitioning, we have $P^k_{\mathbf{u}}\{r,\bar{m}\} = P^k_{\mathbf{u}}\{\bar{r},m\} = 0$, $P^k_{\mathbf{u}}\{r,m\} = P^k_{\mathbf{u}}\{r\} = P^k_{\mathbf{u}}\{m\} = P^k_{\mathbf{u}}$, and $P^k_{\mathbf{u}}\{\bar{r},\bar{m}\} = P^k_{\mathbf{u}}\{\bar{r}\} = P^k_{\mathbf{u}}\{\bar{m}\} = 1 - P^k_{\mathbf{u}}$. Therefore, (35) can be further

simplified to

$$D_{\mathbf{u}}^{k}(P) = P_{\mathbf{u}}^{k} \cdot D_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1} + (1 - P_{\mathbf{u}}^{k}) \cdot D_{\mathbf{u}}^{k}(p). \tag{36}$$

Also note that for I-MB, there will be no transmission distortion if it is correctly received, that is, $D_{\mathbf{u}}^{k}(p) = 0$. So (36) can be further simplified to

$$D_{\mathbf{u}}^{k}(P) = P_{\mathbf{u}}^{k} \cdot D_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}. \tag{37}$$

Comparing (37) with (36), we see that I-MB is a special case of P-MB with $D_{\mathbf{u}}^{k}(p) = 0$, that is, the propagation factor $\alpha_{\mathbf{u}}^{k} = 0$ according to the definition. It is important to note that $D_{\mathbf{u}}^{k}(P) > 0$ for I-MB. In other words, I-MB also contains the distortion caused by propagation error since $P_{\mathbf{u}}^{k} \neq 0$. However, existing LTI models [6], [7] assume that there is no distortion caused by propagation error for I-MB, which under-estimates the transmission distortion.

In the following part of this subsection, we derive the propagation factor $\alpha_{\mathbf{u}}^{k}$ for P-MB and prove some important properties of clipping noise. To derive $\alpha_{\mathbf{u}}^{k}$, we first give Lemma 1 as below.

Lemma 1: Given the PMF of the random variable $\widetilde{\zeta}_{\mathbf{u}+\mathbf{m}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}$ and the value of $\hat{f}_{\mathbf{u}}^{k}$, $D_{\mathbf{u}}^{k}(p)$ can be calculated at the encoder by $D_{\mathbf{u}}^{k}(p) = E[\Phi^2(\widetilde{\zeta}_{\mathbf{u}+\mathbf{m}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}, \hat{f}_{\mathbf{u}}^{k})]$, where $\Phi(x, y)$ is called error reduction function and defined by

$$\Phi(x, y) \triangleq y - \Gamma(y - x) = \begin{cases} y - \gamma_L, & y - x < \gamma_L \\ x, & \gamma_L \leq y - x \leq \gamma_H \\ y - \gamma_H, & y - x > \gamma_H. \end{cases} \tag{38}$$

Lemma 1 is proved in Appendix A. In fact, we have found in our experiments that in any error event, $\widetilde{\zeta}_{\mathbf{u}+\mathbf{m}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}$ approximately follows Laplacian distribution with zero mean. If we assume $\widetilde{\zeta}_{\mathbf{u}+\mathbf{m}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}$ follows Laplacian distribution with zero mean, the calculation for $D_{\mathbf{u}}^{k}(p)$ becomes simpler since the only unknown parameter for PMF of $\widetilde{\zeta}_{\mathbf{u}+\mathbf{m}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}$ is its variance. Under this assumption, we have the following proposition.

Proposition 1: The propagation factor α for propagated error with Laplacian distribution of zero-mean and variance σ^2 is given by

$$\alpha = 1 - \frac{1}{2} e^{-\frac{y-\gamma_L}{b}} \left(\frac{y - \gamma_L}{b} + 1\right) - \frac{1}{2} e^{-\frac{\gamma_H - y}{b}} \left(\frac{\gamma_H - y}{b} + 1\right), \tag{39}$$

where y is the reconstructed pixel value, and $b = \frac{\sqrt{2}}{2}\sigma$.

Proposition 1 is proved in Appendix B. In the zero-mean Laplacian case, $\alpha_{\mathbf{u}}^{k}$ will only be a function of $\hat{f}_{\mathbf{u}}^{k}$ and the variance of $\widetilde{\zeta}_{\mathbf{u}+\mathbf{m}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}$, which is equal to $D_{\mathbf{u}+\mathbf{m}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}$ in this case. Since $D_{\mathbf{u}+\mathbf{m}\mathbf{v}_{\mathbf{u}}^{k}}^{k-1}$ has already

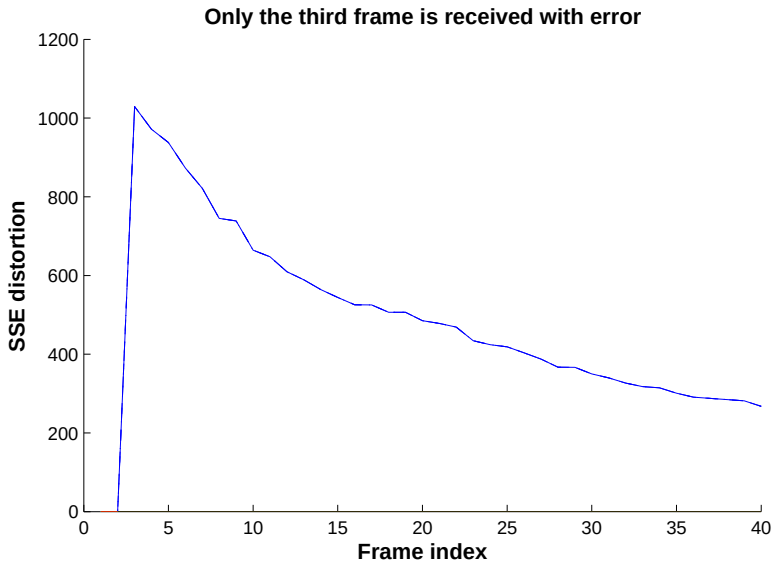

Fig. 2. The effect of clipping noise on distortion propagation.

been calculated during the phase of predicting the $(k-1)$-th frame transmission distortion, $D_{\mathbf{u}}^{k}(p)$ can be calculated by $D_{\mathbf{u}}^{k}(p) = \alpha_{\mathbf{u}}^{k} \cdot D_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^{k}}^{k-1}$ via the definition of $\alpha_{\mathbf{u}}^{k}$. Then we can recursively calculate $D_{\mathbf{u}}^{k}(P)$ in (35) since both $D_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^{k}}^{k-1}$ and $D_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^{k}}^{k-1}$ have been calculated previously for the $(k-1)$-th frame.

Next, we prove an important property of the non-linear clipping function in the following proposition.

Proposition 2: Clipping reduces propagated error, that is, $D_{\mathbf{u}}^{k}(p) \leq D_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^{k}}^{k-1}$, or $\alpha_{\mathbf{u}}^{k} \leq 1$.

Proof: First, from Lemma 4, which is presented and proved in Appendix F, we have $\Phi^2(x,\, y) \leq x^2$ for any $\gamma_L \leq y \leq \gamma_H$. In other words, the function $\Phi(x,\, y)$ reduces the energy of propagated error. This is the reason why we call it error reduction function. With Lemma 1, it is straightforward to prove that whatever the PMF of $\widetilde{\zeta}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^{k}}^{k-1}$ is, $E[\Phi^2(\widetilde{\zeta}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^{k}}^{k-1},\, \hat{f}_{\mathbf{u}}^{k})] \leq E[(\widetilde{\zeta}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^{k}}^{k-1})^2]$, that is, $D_{\mathbf{u}}^{k}(p) \leq D_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^{k}}^{k-1}$, which is equivalent to $\alpha_{\mathbf{u}}^{k} \leq 1$. ■

Proposition 2 tells us that if there is no newly induced errors in the k-th frame, transmission distortion decreases from the $(k-1)$-th frame to the k-th frame. Fig. 2 shows the experimental result of transmission distortion propagation for 'bus' sequence in cif format, where transmission errors only occur in the third frame.

In fact, if we consider the more general cases where there may be new error induced in the k-th frame, we can still prove that $E[(\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^{k}}^{k-1} + \widetilde{\Delta}_{\mathbf{u}}^{k})^2] \leq E[(\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^{k}}^{k-1})^2]$ using the proof for the following corollary.

Corollary 1: The correlation coefficient between $\widetilde{\zeta}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^{k}}^{k-1}$ and $\widetilde{\Delta}_{\mathbf{u}}^{k}$ is non-positive. Specifically, they

are negatively correlated under the condition $\{\bar{r}, p\}$, and uncorrelated under other conditions. Corollary 1 is proved in Appendix H. This property is very important for designing a low complexity algorithm to estimate propagation and clipping caused distortion in PTD, which will be presented in the sequel paper [22].

2) Frame-level Distortion Caused by Propagated Error Plus Clipping Noise: In (35), $D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} \neq D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ due to the non-stationarity of the error process over space. However, both the sum of $D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}$ over all pixels in the $(k-1)$-th frame and the sum of $D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ over all pixels in the $(k-1)$-th frame will converge to D^{k-1} due to the randomness of MV. The formula for frame-level propagation and clipping caused distortion is given in Lemma 2.

Lemma 2: The frame-level propagation and clipping caused distortion in the k-th frame is

$$D^k(P) = D^{k-1} \cdot \bar{P}^k(r) + D^k(p) \cdot (1 - \bar{P}^k(r))(1 - \beta^k), \tag{40}$$

where $D^k(p) \triangleq \frac{1}{|\mathcal{V}|} \sum_{\mathbf{u} \in \mathcal{V}^k} D^k_{\mathbf{u}}(p)$ and $\bar{P}^k(r)$ is defined in (29); β^k is the percentage of I-MBs in the k-th frame; D^{k-1} is the transmission distortion in the $(k-1)$-th frame.

Lemma 2 is proved in Appendix C. Define the propagation factor for the k-th frame $\alpha^k \triangleq \frac{D^k(p)}{D^{k-1}}$; then we have $\alpha^k = \frac{\sum_{\mathbf{u} \in \mathcal{V}^k} \alpha^k_{\mathbf{u}} \cdot D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}}{D^{k-1}}$. Note that $D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ may be different for different pixels in the $(k-1)$-th frame due to the non-stationarity of error process over space. However, when the number of pixels in the $(k-1)$-th frame is sufficiently large, the sum of $D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ over all the pixels in the $(k-1)$-th frame will converge to D^{k-1}. Therefore, we have $\alpha^k = \frac{\sum_{\mathbf{u} \in \mathcal{V}^k} \alpha^k_{\mathbf{u}} \cdot D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}}{\sum_{\mathbf{u} \in \mathcal{V}} D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}}$, which is a weighted average of $\alpha^k_{\mathbf{u}}$ with the weight being $D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$. As a result, $D^k(p) \leq D^k(P)$[7]. However, most existing works directly use $D^k(P) = D^k(p)$ in predicting transmission distortion. This is another reason why LTI models [6], [7] under-estimate transmission distortion when there is no MV error. Details will be discussed in Section V-B.

E. Analysis of Correlation Caused Distortion

In this subsection, we first derive the pixel-level correlation caused distortion $D^k_{\mathbf{u}}(c)$. Then we derive the frame-level correlation caused distortion $D^k(c)$.

[7]When the number of pixels in the $(k-1)$-th frame is small, $\sum_{\mathbf{u} \in \mathcal{V}^k} \alpha^k_{\mathbf{u}} \cdot D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ may be larger than D^{k-1} although its probability is small as observed in our experiments.

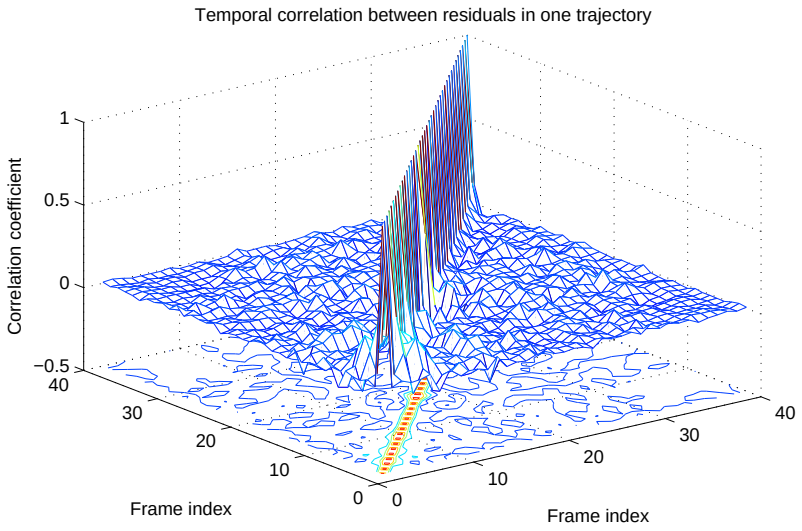

Fig. 3. Temporal correlation between the residuals in one trajectory.

1) Pixel-level Correlation Caused Distortion: We analyze the correlation caused distortion $D^k_{\mathbf{u}}(c)$ at the decoder in four different cases: i) for $\mathbf{u}^k\{\bar{r},\bar{m}\}$, both $\widetilde{\varepsilon}^k_{\mathbf{u}} = 0$ and $\widetilde{\xi}^k_{\mathbf{u}} = 0$, so $D^k_{\mathbf{u}}(c) = 0$; ii) for $\mathbf{u}^k\{r,\bar{m}\}$, $\widetilde{\xi}^k_{\mathbf{u}} = 0$ and $D^k_{\mathbf{u}}(c) = 2E[\varepsilon^k_{\mathbf{u}} \cdot (\widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{m}\mathbf{v}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{r,\bar{m}\})]$; iii) for $\mathbf{u}^k\{\bar{r},m\}$, $\widetilde{\varepsilon}^k_{\mathbf{u}} = 0$ and $D^k_{\mathbf{u}}(c) = 2E[\xi^k_{\mathbf{u}} \cdot (\widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r},m\})]$; iv) for $\mathbf{u}^k\{r,m\}$, $D^k_{\mathbf{u}}(c) = 2E[\varepsilon^k_{\mathbf{u}} \cdot \xi^k_{\mathbf{u}}] + 2E[\varepsilon^k_{\mathbf{u}} \cdot (\widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{r,m\})] + 2E[\xi^k_{\mathbf{u}} \cdot (\widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{r,m\})]$. From Section IV-D1, we know $\widetilde{\Delta}^k_{\mathbf{u}}\{r\} = 0$. So, we obtain

$$\begin{aligned} D^k_{\mathbf{u}}(c) = {} & P^k_{\mathbf{u}}\{r,\bar{m}\} \cdot 2E[\varepsilon^k_{\mathbf{u}} \cdot \widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{m}\mathbf{v}^k_{\mathbf{u}}}] + P^k_{\mathbf{u}}\{\bar{r},m\} \cdot 2E[\xi^k_{\mathbf{u}} \cdot (\widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r},m\})] \\ & + P^k_{\mathbf{u}}\{r,m\} \cdot (2E[\varepsilon^k_{\mathbf{u}} \cdot \xi^k_{\mathbf{u}}] + 2E[\varepsilon^k_{\mathbf{u}} \cdot \widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}] + 2E[\xi^k_{\mathbf{u}} \cdot \widetilde{\zeta}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}]). \end{aligned} \tag{41}$$

In our experiments, we find that in the trajectory of pixel $\mathbf{u}^k$, 1) the residual $\hat{e}^k_{\mathbf{u}}$ is approximately uncorrelated with the residual in all other frames $\hat{e}^i_{\mathbf{v}}$, where $i \neq k$, as shown in Fig. 3 for 'foreman' sequence in cif format[8]; and 2) the residual $\hat{e}^k_{\mathbf{u}}$ is approximately uncorrelated with the MVCE of the corresponding pixel $\xi^k_{\mathbf{u}}$ and the MVCE in all previous frames $\xi^i_{\mathbf{v}}$, where $i < k$, as shown in Fig. 4 for 'foreman' sequence in cif format. Based on the above observations, we further assume that for any $i < k$, $\hat{e}^k_{\mathbf{u}}$ is uncorrelated with $\hat{e}^i_{\mathbf{v}}$ and $\xi^i_{\mathbf{v}}$ if $\mathbf{v}^i$ is not in the trajectory of pixel $\mathbf{u}^k$, and make the following assumption.

Assumption 5: $\hat{e}^k_{\mathbf{u}}$ is uncorrelated with $\xi^k_{\mathbf{u}}$, and is uncorrelated with both $\hat{e}^i_{\mathbf{v}}$ and $\xi^i_{\mathbf{v}}$ for any $i < k$.

[8]All other sequences show the same statistics for Fig. 3, Fig. 4, Fig. 5 and Fig. 6.

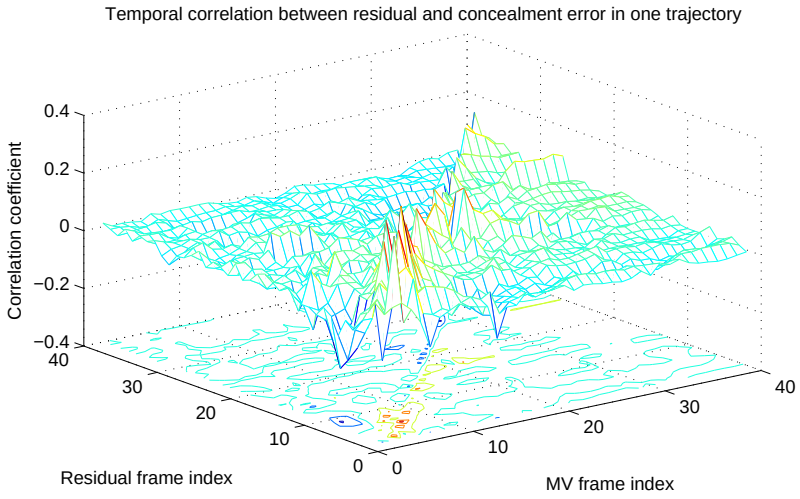

Fig. 4. Temporal correlation matrix between residual and MVCE in one trajectory.

Since $\widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ and $\widetilde{\zeta}^{k-1}_{\mathbf{u}+\breve{\mathbf{m}}\breve{\mathbf{v}}^k_{\mathbf{u}}}$ are the transmission reconstructed errors accumulated from all the frames before the k-th frame, $\varepsilon^k_{\mathbf{u}}$ is uncorrelated with $\widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ and $\widetilde{\zeta}^{k-1}_{\mathbf{u}+\breve{\mathbf{m}}\breve{\mathbf{v}}^k_{\mathbf{u}}}$ due to Assumption 5. Thus, (41) becomes

$$D^k_{\mathbf{u}}(c) = 2P^k_{\mathbf{u}}\{m\} \cdot E[\xi^k_{\mathbf{u}} \cdot \widetilde{\zeta}^{k-1}_{\mathbf{u}+\breve{\mathbf{m}}\breve{\mathbf{v}}^k_{\mathbf{u}}}] + 2P^k_{\mathbf{u}}\{\bar{r}, m\} \cdot E[\xi^k_{\mathbf{u}} \cdot \widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r}, m\}]. \tag{42}$$

However, we observe that in the trajectory of pixel $\mathbf{u}^k$, 1) the residual $\hat{e}^k_{\mathbf{u}}$ is correlated with the MVCE $\xi^i_{\mathbf{v}}$, where $i > k$, as seen in Fig. 4; and 2) the MVCE $\xi^k_{\mathbf{u}}$ is highly correlated with the MVCE $\xi^i_{\mathbf{v}}$ as shown in Fig. 5 for 'foreman' sequence in cif format. This interesting phenomenon could be exploited by an error concealment algorithm and is subject to our future study.

As mentioned in Section IV-D1, for P-MBs with slice data partitioning in H.264, $P^k_{\mathbf{u}}\{\bar{r}, m\} = 0$. So, (42) becomes

$$D^k_{\mathbf{u}}(c) = 2P^k_{\mathbf{u}}\{m\} \cdot E[\xi^k_{\mathbf{u}} \cdot (\hat{f}^{k-1}_{\mathbf{u}+\breve{\mathbf{m}}\breve{\mathbf{v}}^k_{\mathbf{u}}} - \widetilde{f}^{k-1}_{\mathbf{u}+\breve{\mathbf{m}}\breve{\mathbf{v}}^k_{\mathbf{u}}})]. \tag{43}$$

Note that in the more general case that $P^k_{\mathbf{u}}\{\bar{r}, m\} \neq 0$, Eq. (43) is still valid since $\xi^k_{\mathbf{u}}$ is almost uncorrelated with $\widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r}, m\}$ as observed in the experiment.

For MBs without slice data partitioning, since $P^k_{\mathbf{u}}\{r, \bar{m}\} = P^k_{\mathbf{u}}\{\bar{r}, m\} = 0$ and $P^k_{\mathbf{u}}\{r, m\} = P^k_{\mathbf{u}}\{r\} = P^k_{\mathbf{u}}\{m\} = P^k_{\mathbf{u}}$ as mentioned in Section IV-D1, (41) can be simplified to

$$D^k_{\mathbf{u}}(c) = 2P^k_{\mathbf{u}} \cdot (2E[\varepsilon^k_{\mathbf{u}} \cdot \xi^k_{\mathbf{u}}] + 2E[\varepsilon^k_{\mathbf{u}} \cdot \widetilde{\zeta}^{k-1}_{\mathbf{u}+\breve{\mathbf{m}}\breve{\mathbf{v}}^k_{\mathbf{u}}}] + 2E[\xi^k_{\mathbf{u}} \cdot \widetilde{\zeta}^{k-1}_{\mathbf{u}+\breve{\mathbf{m}}\breve{\mathbf{v}}^k_{\mathbf{u}}}]). \tag{44}$$

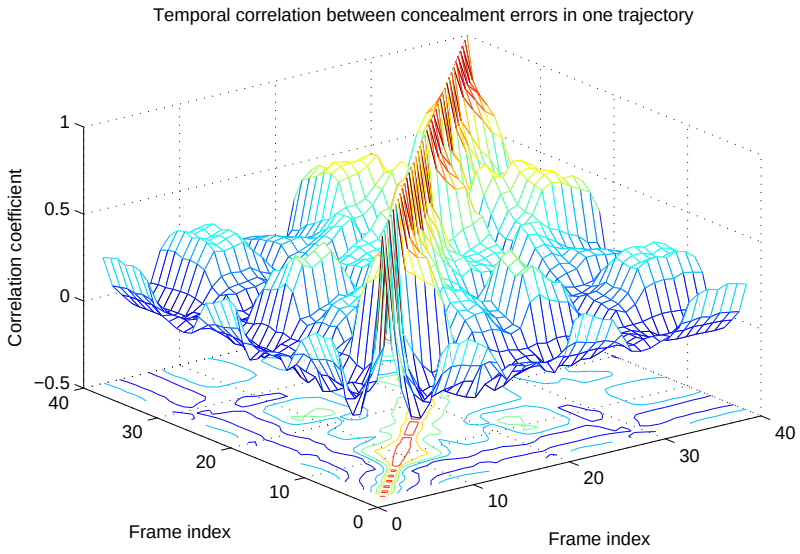

Fig. 5. Temporal correlation matrix between MVCEs in one trajectory.

Under Assumption 5, (44) reduces to (43).

Define $\lambda_{\mathbf{u}}^k \triangleq \frac{E[\xi_{\mathbf{u}}^k \cdot \widetilde{f}_{\mathbf{u}+\breve{\mathbf{m}}\mathbf{v}_{\mathbf{u}}^k}^{k-1}]}{E[\xi_{\mathbf{u}}^k \cdot \hat{f}_{\mathbf{u}+\breve{\mathbf{m}}\mathbf{v}_{\mathbf{u}}^k}^{k-1}]}$; $\lambda_{\mathbf{u}}^k$ is a correlation ratio, that is, the ratio of the correlation between MVCE and concealed reference pixel value *at the decoder*, to the correlation between MVCE and concealed reference pixel value *at the encoder*. $\lambda_{\mathbf{u}}^k$ quantifies the effect of the correlation between the MVCE and propagated error on transmission distortion.

Note that although we do not know the exact value of $\lambda_{\mathbf{u}}^k$ at the encoder, its range is

$$\prod_{i=1}^{k-1} P_{\mathbf{T}(i)}^i\{\bar{r}, \bar{m}\} \leq \lambda_{\mathbf{u}}^k \leq 1, \tag{45}$$

where $\mathbf{T}(i)$ is the pixel position of the i-th frame in the trajectory, for example, $\mathbf{T}(k-1) = \mathbf{u}^k + \mathbf{mv}_{\mathbf{u}}^k$ and $\mathbf{T}(k-2) = \mathbf{v}^{k-1} + \mathbf{mv}_{\mathbf{v}}^{k-1}$. The left inequality in (45) holds in the extreme case that any error in the trajectory will cause $\xi_{\mathbf{u}}^k$ and $\widetilde{f}_{\mathbf{u}+\breve{\mathbf{m}}\mathbf{v}_{\mathbf{u}}^k}^{k-1}$ to be uncorrelated, which is usually true for high motion video. The right inequality in (45) holds in another extreme case that all errors in the trajectory do not affect the correlation between $\xi_{\mathbf{u}}^k$ and $\widetilde{f}_{\mathbf{u}+\breve{\mathbf{m}}\mathbf{v}_{\mathbf{u}}^k}^{k-1}$, that is $E[\xi_{\mathbf{u}}^k \cdot \widetilde{f}_{\mathbf{u}+\breve{\mathbf{m}}\mathbf{v}_{\mathbf{u}}^k}^{k-1}] \approx E[\xi_{\mathbf{u}}^k \cdot \hat{f}_{\mathbf{u}+\breve{\mathbf{m}}\mathbf{v}_{\mathbf{u}}^k}^{k-1}]$, which is usually true for low motion video. The details on how to estimate $\lambda_{\mathbf{u}}^k$ will be presented in the sequel paper [22].

Using the definition of $\lambda_{\mathbf{u}}^k$, we have the following proposition.

Proposition 3:

$$D_{\mathbf{u}}^k(c) = (\lambda_{\mathbf{u}}^k - 1) \cdot D_{\mathbf{u}}^k(m). \tag{46}$$

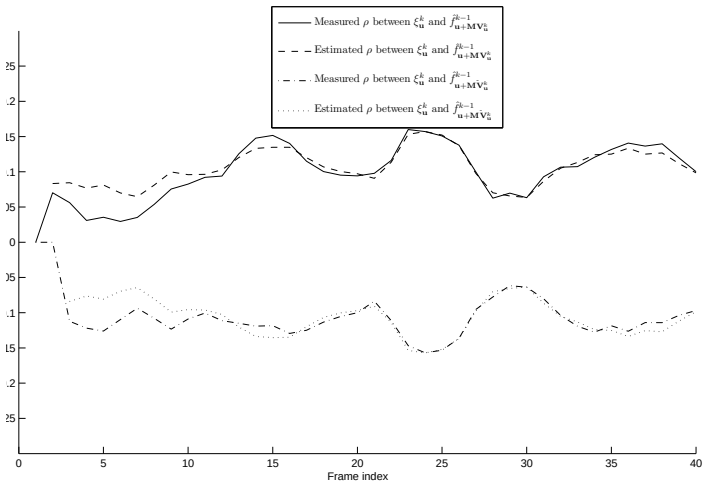

Fig. 6. Comparison between measured and estimated correlation coefficients.

Proposition 3 is proved in Appendix D.

If we assume $E[\xi^k_{\mathbf{u}}] = 0$, we may further derive the correlation coefficient between $\xi^k_{\mathbf{u}}$ and $\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}}$. Denote ρ as their correlation coefficient, we have $\rho = \frac{E[\xi^k_{\mathbf{u}} \cdot \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}}] - E[\xi^k_{\mathbf{u}}] \cdot E[\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}}]}{\sigma_{\xi^k_{\mathbf{u}}} \cdot \sigma_{\hat{f}^k_{\mathbf{u}}}} = -\frac{\sigma_{\xi^k_{\mathbf{u}}}}{2\sigma_{\hat{f}^k_{\mathbf{u}}}}$; similarly, it is easy to prove that the correlation coefficient between $\xi^k_{\mathbf{u}}$ and $\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ is $\frac{\sigma_{\xi^k_{\mathbf{u}}}}{2\sigma_{\hat{f}^k_{\mathbf{u}}}}$. This agrees well with the experimental results shown in Fig. 6. Via the same derivation process, one can obtain the correlation coefficient between $\hat{e}^k_{\mathbf{u}}$ and $\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$, and between $\hat{e}^k_{\mathbf{u}}$ and $\hat{f}^k_{\mathbf{u}}$. One possible application of these correlation properties is error concealment with partial information available.

2) Frame-Level Correlation Caused Distortion: Denote $V^k_i(m)$ the set of pixels in the i-th MV packet of the k-th frame. From (20), (77) and Assumption 4, we obtain

$$\begin{aligned} D^k(c) &= \frac{E[(\xi^k)^2]}{|\mathcal{V}|} \sum_{\mathbf{u}\in\mathcal{V}^k} (\lambda^k_{\mathbf{u}} - 1) \cdot P^k_{\mathbf{u}}(m) \\ &= \frac{E[(\xi^k)^2]}{|\mathcal{V}|} \sum_{i=1}^{N^k(m)} \{P^k_i(m) \sum_{\mathbf{u}\in\mathcal{V}^k_i(m)} (\lambda^k_{\mathbf{u}} - 1)\}. \end{aligned} \tag{47}$$

Define $\lambda^k \triangleq \frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in V^k} \lambda^k_{\mathbf{u}}$; due to the randomness of $\mathbf{mv}^k_{\mathbf{u}}$, $\frac{1}{N^k_i(m)} \sum_{\mathbf{u}\in\mathcal{V}^k_i\{m\}} \lambda^k_{\mathbf{u}}$ will converge to λ^k for any packet that contains a sufficiently large number of pixels. By rearranging (47), we obtain

$$\begin{aligned} D^k(c) &= \frac{E[(\xi^k)^2]}{|\mathcal{V}|} \sum_{i=1}^{N^k(m)} \{P^k_i(m) \cdot N^k_i(m) \cdot (\lambda^k - 1)\} \\ &= (\lambda^k - 1) \cdot E[(\xi^k)^2] \cdot \bar{P}^k(m). \end{aligned} \tag{48}$$

From (33), we know that $E[(\xi^k)^2] \cdot \bar{P}^k(m)$ is exactly equal to $D^k(m)$. Therefore, (48) is further simplified to

$$D^k(c) = (\lambda^k - 1) \cdot D^k(m). \tag{49}$$

F. Summary

In Section IV-A, we decomposed transmission distortion into four terms; we derived a formula for each term in Sections IV-B through IV-E. In this section, we combine the formulae for the four terms into a single formula.

1) Pixel-Level Transmission Distortion:

Theorem 1: Under single-reference prediction, the PTD of pixel $\mathbf{u}^k$ is

$$D^k_{\mathbf{u}} = D^k_{\mathbf{u}}(r) + \lambda^k_{\mathbf{u}} \cdot D^k_{\mathbf{u}}(m) + P^k_{\mathbf{u}}\{r, m\} \cdot D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + P^k_{\mathbf{u}}\{r, \bar{m}\} \cdot D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + P^k_{\mathbf{u}}\{\bar{r}\} \cdot \alpha^k_{\mathbf{u}} \cdot D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}. \tag{50}$$

Proof: (50) can be obtained by plugging (24), (32), (35), and (77) into (15). ■

Corollary 2: Under single-reference prediction and no slice data partitioning, (50) is simplified to

$$D^k_{\mathbf{u}} = P^k_{\mathbf{u}} \cdot (E[(\varepsilon^k_{\mathbf{u}})^2] + \lambda^k_{\mathbf{u}} \cdot E[(\xi^k_{\mathbf{u}})^2] + D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}) + (1 - P^k_{\mathbf{u}}) \cdot \alpha^k_{\mathbf{u}} \cdot D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}. \tag{51}$$

2) Frame-Level Transmission Distortion:

Theorem 2: Under single-reference prediction, the FTD of the k-th frame is

$$D^k = D^k(r) + \lambda^k \cdot D^k(m) + \bar{P}^k(r) \cdot D^{k-1} + (1 - \bar{P}^k(r)) \cdot D^k(p) \cdot (1 - \beta^k). \tag{52}$$

Proof: (52) can be obtained by plugging (30), (33), (40) and (49) into (16). ■

Corollary 3: Under single-reference prediction and no slice data partitioning, the FTD of the k-th frame is simplified to

$$D^k = D^k(r) + \lambda^k \cdot D^k(m) + \bar{P}^k \cdot D^{k-1} + (1 - \bar{P}^k) \cdot D^k(p) \cdot (1 - \beta^k). \tag{53}$$

V. Relationship between Theorem 2 and Existing Transmission Distortion Models

In this section, we will identify the relationship between Theorem 2 and their models, and specify the conditions, under which those models are accurate. Note that in order to demonstrate the effect of non-linear clipping on transmission distortion propagation, we disable intra update, that is, $\beta^k = 0$ for all the following cases.

A. Case 1: Only the $(k-1)$-th Frame Has Error, and the Subsequent Frames are All Correctly Received

In this case, the models proposed in Refs. [6], [9] state that when there is no intra coding and spatial filtering, the propagation distortion will be the same for all the frames after the $(k-1)$-th frame, i.e., $D^n(p) = D^{n-1}$ ($\forall\ n \geq k$). However, this is not true as we proved in Proposition 2. Due to the clipping function, we have $\alpha^n \leq 1$ ($\forall\ n \geq k$), i.e., $D^n \leq D^{n-1}$ ($\forall\ n \geq k$) in case the n-th frame is error-free. Actually, from Appendix F, we know that the equality only holds under a very special case that $\hat{f}^k_{\mathbf{u}} - \gamma_H \leq \widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} \leq \hat{f}^k_{\mathbf{u}} - \gamma_L$ for all pixel $\mathbf{u} \in V^k$.

B. Case 2: Burst Errors in Consecutive Frames

In Ref. [14], authors observe that the transmission distortion caused by accumulated errors from consecutive frames is generally larger than the sum of those distortions caused by individual frame errors. This is also observed in our experiment when there is no MV error. To explain this phenomenon, let us first look at a simple case that residuals in the k-th frame are all erroneous, while the MVs in the k-th frame are all correctly received. In this case, we obtain from (52) that $D^k = D^k(r) + \bar{P}^k(r) \cdot D^{k-1} + (1 - \bar{P}^k(r)) \cdot D^k(p)$, which is larger than the simple sum $D^k(r) + D^k(p)$ as in the LTI model; the under-estimation caused by the LTI model is due to $D^k - (D^k(r) + D^k(p)) = (1 - \alpha^k) \cdot \bar{P}^k(r) \cdot D^{k-1}$.

However, when MV is erroneous, the experimental result is quite different from that claimed in Ref. [14] especially for the high motion video. In other words, the LTI model now causes over-estimation for a burst error channel. In this case, the predicted transmission distortion can be calculated via (52) in Theorem 2 as $D^k_1 = D^k(r) + \lambda^k \cdot D^k(m) + \bar{P}^k(r) \cdot D^{k-1}_1 + (1 - \bar{P}^k(r)) \cdot \alpha^k \cdot D^{k-1}_1$, and by the LTI model as $D^k_2 = D^k(r) + D^k(m) + \alpha^k \cdot D^{k-1}_2$. So, the prediction difference between Theorem 2 and the LTI model is

$$D^k_1 - D^k_2 = (1 - \alpha^k) \cdot \bar{P}^k(r) \cdot D^{k-1}_1 - (1 - \lambda^k) \cdot \bar{P}^k(m) \cdot E[(\xi^k)^2] + \alpha^k \cdot (D^{k-1}_1 - D^{k-1}_2). \quad (54)$$

At the beginning, $D^0_1 = D^0_2 = 0$, and $D^{k-1} << E[(\xi^k)^2]$ when k is small. Therefore, the transmission distortion caused by accumulated errors from consecutive frames will be smaller than the sum of the distortions caused by individual frame errors, that is, $D^k_1 < D^k_2$. We may see from (54) that, due to the propagation of over-estimation $D^{k-1}_1 - D^{k-1}_2$ from the $(k-1)$-th frame to the k-th frame, the accumulated difference between D^k_1 and D^k_2 will become larger and larger as k increases.

C. Case 3: Modeling Transmission Distortion as an Output of an LTI System with PEP as input

In Ref. [7], authors propose an LTI transmission distortion model based on their observations from experiments. This LTI model ignores the effects of correlation between the newly induced error and the propagated error, that is, $\lambda^k = 1$. This is only valid for low motion video. From (52), we obtain

$$D^k = D^k(r) + D^k(m) + (\bar{P}^k(r) + (1 - \bar{P}^k(r)) \cdot \alpha^k) \cdot D^{k-1}. \tag{55}$$

Let $\eta^k = \bar{P}^k(r) + (1 - \bar{P}^k(r)) \cdot \alpha^k$. If 1) there is no slice data partitioning, i.e., $P^k(m) = P^k(r) = P^k$, and 2) $\bar{P}^k(r) = P^k(r)$ (which means one frame is transmitted in one packet, or different packets experience the same channel condition), then (55) becomes $D^k = \{E[(\xi^k)^2] + E[(\varepsilon^k)^2]\} \cdot P^k + \eta^k \cdot D^{k-1}$. Let $E^k \triangleq E[(\xi^k)^2] + E[(\varepsilon^k)^2]$. Then the recursive formula results in

$$D^k = \sum_{l=k-L}^{k} [(\prod_{i=l+1}^{k} \eta^i) \cdot (E^l \cdot P^l)], \tag{56}$$

where L is the time interval between the k-th frame and the latest correctly received frame.

Denote the system by an operator H that maps the error input sequence $\{P^k\}$, as a function of frame index k, to the distortion output sequence $\{D^k\}$. Since generally $D^k(p)$ is a nonlinear function of D^{k-1}, as a ratio of $D^k(p)$ and D^{k-1}, α^k is still a function of D^{k-1}. As a result, η^k is a function of D^{k-1}. That means the operator H is non-linear, i.e., the system is non-linear. In addition, since α^k varies from frame to frame as mentioned in Section IV-D2, the system is time-variant. In summary, H is generally a non-linear time-variant system.

The LTI model assumes that 1) the operator H is linear, that is, $H(a \cdot P_1^k + b \cdot P_2^k) = a \cdot H(P_1^k) + b \cdot H(P_2^k)$, which is valid only when η^k does not depend on D^{k-1}; and 2) the operator H is time-invariant, that is, $D^{k+\delta} = H(P^{k+\delta})$, which is valid only when η^k is constant, i.e., both $P^k(r)$ and α^k are constant. Under these two assumptions, we have $\eta^i = \eta$, and we obtain $\prod_{i=l+1}^{k} \eta^i = (\eta)^{k-l}$. Let $h[k] = (\eta)^k$, where $h[k]$ is the impulse response of the LTI model; then we obtain

$$D^k = \sum_{l=k-L}^{k} [h[k-l] \cdot (E^l \cdot P^l)]. \tag{57}$$

From Proposition 2, it is easy to prove that $0 \leq \eta \leq 1$; so $h[k]$ is a decreasing function of time. We see that (57) is a convolution between the error input sequence and the system impulse response. Actually, if we let $h[k] = e^{-\gamma k}$, where $\gamma = -\log \eta$, it is exactly the formula proposed in Ref. [7]. Note that (57) is a very special case of (52) with the following limitations: 1) the video content has to be of low motion; 2) there is no slice data partitioning or all pixels in the same frame experience the same channel condition; 3) η^k is a constant, that is, both $\bar{P}^k(r)$ and the propagation factor α^k are constant, which requires the

probability distributions of reconstructed pixel values in all frames should be the same. Note that the physical meaning of η^k is not the actual propagation factor, but it is just a notation for simplifying the formula.

VI. PTD and FTD under Multi-Reference Prediction

The PTD and FTD formulae in Section IV are for single-reference prediction. In this section, we extend the formulae to multi-reference prediction.

A. Pixel-level Distortion under Multi-Reference Prediction

If multiple frames are allowed to be the references for motion estimation, the reconstructed pixel value at the encoder in (1) becomes

$$\hat{f}_{\mathbf{u}}^k = \Gamma(\hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j} + \hat{e}_{\mathbf{u}}^k). \tag{58}$$

For the reconstructed pixel value at the decoder in (3), it is a bit different as below.

$$\widetilde{f}_{\mathbf{u}}^k = \Gamma(\widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'} + \widetilde{e}_{\mathbf{u}}^k). \tag{59}$$

If $\mathbf{mv}_{\mathbf{u}}^k$ is correctly received, $\widetilde{\mathbf{mv}}_{\mathbf{u}}^k = \mathbf{mv}_{\mathbf{u}}^k$ and $\widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'} = \widetilde{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j}$. However, if $\mathbf{mv}_{\mathbf{u}}^k$ is received with error, the concealed MV has no difference from the single-reference case, that is, $\widetilde{\mathbf{mv}}_{\mathbf{u}}^k = \check{\mathbf{mv}}_{\mathbf{u}}^k$ and $\widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'} = \widetilde{f}_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$.

As a result, (12) becomes

$$\begin{aligned}\widetilde{\zeta}_{\mathbf{u}}^k &= (\hat{e}_{\mathbf{u}}^k + \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j} - \hat{\Delta}_{\mathbf{u}}^k) - (\widetilde{e}_{\mathbf{u}}^k + \widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'} - \widetilde{\Delta}_{\mathbf{u}}^k) \\ &= (\hat{e}_{\mathbf{u}}^k - \widetilde{e}_{\mathbf{u}}^k) + (\hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j} - \hat{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'}) + (\hat{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'} - \widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'}) - (\hat{\Delta}_{\mathbf{u}}^k - \widetilde{\Delta}_{\mathbf{u}}^k).\end{aligned} \tag{60}$$

Following the same derivation process from Section IV-A to Section IV-E, the formulae for PTD under multi-reference prediction are the same as those under single-reference prediction except the following changes: 1) MVCE $\widetilde{\xi}_{\mathbf{u}}^k \triangleq \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j} - \hat{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'}$ and clipping noise $\widetilde{\Delta}_{\mathbf{u}}^k \triangleq (\widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'} + \widetilde{e}_{\mathbf{u}}^k) - \Gamma(\widetilde{f}_{\mathbf{u}+\widetilde{\mathbf{mv}}_{\mathbf{u}}^k}^{k-j'} + \widetilde{e}_{\mathbf{u}}^k)$; 2) $D_{\mathbf{u}}^k(m)$ and $D_{\mathbf{u}}^k(c)$ are given by (32) and (46), respectively, with a new definition of $\xi_{\mathbf{u}}^k \triangleq \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j} - \hat{f}_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$; 3) $D_{\mathbf{u}}^k(p) \triangleq E[(\widetilde{\zeta}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j} + \widetilde{\Delta}_{\mathbf{u}}^k\{\bar{r}, \bar{m}\})^2]$, $\alpha_{\mathbf{u}}^k \triangleq \frac{D_{\mathbf{u}}^k(p)}{D_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j}}$ and

$$D_{\mathbf{u}}^k(P) = P_{\mathbf{u}}^k\{r, m\} \cdot D_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} + P_{\mathbf{u}}^k\{r, \bar{m}\} \cdot D_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-j} + P_{\mathbf{u}}^k\{\bar{r}\} \cdot D_{\mathbf{u}}^k(p), \tag{61}$$

compared to (35). The generalization of PTD formulae to multi-reference prediction is straightforward since the multi-reference prediction case just has a larger set of reference pixels than the single-reference case. Following the same derivation process, we have the following general theorem for PTD.

Theorem 3: Under multi-reference prediction, the PTD of pixel $\mathbf{u}^k$ is

$$D^k_{\mathbf{u}} = D^k_{\mathbf{u}}(r) + \lambda^k_{\mathbf{u}} \cdot D^k_{\mathbf{u}}(m) + P^k_{\mathbf{u}}\{r,m\} \cdot D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}} + P^k_{\mathbf{u}}\{r,\bar{m}\} \cdot D^{k-j}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} + P^k_{\mathbf{u}}\{\bar{r}\} \cdot \alpha^k_{\mathbf{u}} \cdot D^{k-j}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}. \tag{62}$$

Corollary 4: Under multi-reference prediction and no slice data partitioning, (62) is simplified to

$$D^k_{\mathbf{u}} = P^k_{\mathbf{u}} \cdot (E[(\varepsilon^k_{\mathbf{u}})^2] + \lambda^k_{\mathbf{u}} \cdot E[(\xi^k_{\mathbf{u}})^2] + D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}) + (1 - P^k_{\mathbf{u}}) \cdot \alpha^k_{\mathbf{u}} \cdot D^{k-j}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}. \tag{63}$$

B. Frame-level Distortion under Multi-Reference Prediction

Under multi-reference prediction, each block typically is allowed to choose its reference block independently; hence, different pixels in the same frame may have different reference frames. Define $V^k(j) \triangleq \{\mathbf{u}^k : \mathbf{u}^k = \mathbf{v}^{k-j} - \mathbf{mv}^k_{\mathbf{u}}\}$, where $j \in \{1,2,...,J\}$ and J is the number of reference frames; i.e., $V^k(j)$ is the set of the pixels in the k-th frame, whose reference pixels are in the $(k-j)$-th frame. Obviously, $\bigcup_{j=1}^{J} V^k(j) = V^k$ and $\bigcap_{j=1}^{J} V^k(j) = \varnothing$. Define $w^k(j) \triangleq \frac{|\mathcal{V}^k(j)|}{|\mathcal{V}^k|}$. Note that V^k and $V^k(j)$ have the similar physical meanings but only the different cardinalities.

$D^k(m)$ and $D^k(c)$ are given by (33) and (49), respectively, with a new definition of $\xi^k(j) \triangleq \{\xi^k_{\mathbf{u}} : \xi^k_{\mathbf{u}} = \hat{f}^{k-j}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}} - \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}\}$. and $\xi^k = \sum_{j=1}^{J} w^k(j) \cdot \xi^k(j)$. Define the propagation factor of $V^k(j)$ by $\alpha^k(j) \triangleq \frac{\sum_{\mathbf{u}\in V^k(j)} \alpha^k_{\mathbf{u}} \cdot D^{k-j}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}}{\sum_{\mathbf{u}\in V^k(j)} D^{k-j}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}}$. The following lemma gives the formula for $D^k(P)$.

Lemma 3: The frame-level propagation and clipping caused distortion in the k-th frame for the multi-reference case is

$$\begin{aligned} D^k(P) = D^{k-1} \cdot \bar{P}^k\{r,m\} + \sum_{j=1}^{J} (\bar{P}^k(j)\{r,\bar{m}\} \cdot w^k(j) \cdot D^{k-j}) \\ + (1-\beta^k) \cdot \sum_{j=1}^{J} (\bar{P}^k(j)\{\bar{r}\} \cdot w^k(j) \cdot \alpha^k(j) \cdot D^{k-j}), \end{aligned} \tag{64}$$

where β^k is the percentage of I-MBs in the k-th frame; $\bar{P}^k(j)\{r,\bar{m}\}$ is the weighted average of joint PEPs of event $\{r,\bar{m}\}$ for the j-th sub-frame in the k-th frame. $\bar{P}^k(j)\{\bar{r}\}$ is the weighted average of PEP of event $\{\bar{r}\}$ for the j-th sub-frame in the k-th frame.

Lemma 3 is proved in Appendix E. With Lemma 3, we have the following general theorem for FTD.

Theorem 4: Under multi-reference prediction, the FTD of the k-th frame is

$$\begin{aligned} D^k = D^k(r) + \lambda^k \cdot D^k(m) + D^{k-1} \cdot \bar{P}^k\{r,m\} + \sum_{j=1}^{J} (\bar{P}^k(j)\{r,\bar{m}\} \cdot w^k(j) \cdot D^{k-j}) \\ + (1-\beta^k) \cdot \sum_{j=1}^{J} (\bar{P}^k(j)\{\bar{r}\} \cdot w^k(j) \cdot \alpha^k(j) \cdot D^{k-j}). \end{aligned} \tag{65}$$

Proof: (65) can be obtained by plugging (30), (33), (64) and (49) into (16). ■

It is easy to prove that (52) in Theorem 2 is a special case of (65) with $J = 1$ and $w^k(j) = 1$. It is also easy to prove that (62) in Theorem 3 is a special case of (65) with $|\mathcal{V}| = 1$.

Corollary 5: Under multi-reference prediction and no slice data partitioning, (65) is simplified to

$$D^k = D^k(r) + \lambda^k \cdot D^k(m) + D^{k-1} \cdot \bar{P}^k\{r\} + (1-\beta^k) \cdot \sum_{j=1}^{J} (\bar{P}^k(j)\{\bar{r}\} \cdot w^k(j) \cdot \alpha^k(j) \cdot D^{k-j}). \quad (66)$$

VII. Conclusion

In this paper, we derived the transmission distortion formulae for wireless video communication systems. With consideration of spatio-temporal correlation, nonlinear codec and time-varying channel, our formulae provide, for the first time, the following capabilities: 1) support of distortion prediction at different levels (e.g., pixel/frame/GOP level), 2) support of accurate multi-reference prediction, 3) support of slice data partitioning, 4) support of arbitrary slice-level packetization with FMO mechanism, 5) being applicable to time-varying channels, 6) one unified formula for both I-MB and P-MB, and 7) support of both low motion and high motion video sequences. Besides deriving the transmission distortion formulae, this paper also identified two important properties of transmission distortion for the first time: 1) clipping noise, produced by non-linear clipping, causes decay of propagated error; 2) the correlation between motion vector concealment error and propagated error is negative, and has dominant impact on transmission distortion, among all the correlations between any two of the four components in transmission error. We also discussed the relationship between our formula and existing models. In the sequel paper [22], we use the formulae derived in this paper to design algorithms for estimating pixel-level and frame-level transmission distortion and apply the algorithms to video codec design; we also verify the accuracy of the formulae derived in this paper through experiments; the application of these formulae shows superior performance over existing models.

Acknowledgments

This work was supported in part by an Intel gift, the US National Science Foundation under grant DBI-0529012 and CNS-0643731. The authors would like to thank Jun Xu and Qian Chen for many fruitful discussions related to this work and suggestions that helped to improve the presentation of this paper.

APPENDIX

A. Proof of Lemma 1

Proof: From (11) and (13), we obtain $\widetilde{f}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{e}^k_{\mathbf{u}} = \hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}} - \widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}}$,

Together with (9), we obtain

$$\widetilde{\Delta}^k_{\mathbf{u}} = (\hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}} - \widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}}) - \Gamma(\hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}} - \widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}}). \tag{67}$$

So, $\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}} = (\hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}}) - \Gamma(\hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}} - \widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}})$, and

$$D^k_{\mathbf{u}}(P) = E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}})^2] = E[\Phi^2(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}},\, \hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}})]. \tag{68}$$

We know from the definition that $D^k_{\mathbf{u}}(p)$ is a special case of $D^k_{\mathbf{u}}(P)$ under the condition $\{\bar{r}, \bar{m}\}$, which means $\widetilde{e}^k_{\mathbf{u}} = \hat{e}^k_{\mathbf{u}}$, i.e. $\widetilde{\varepsilon}^k_{\mathbf{u}} = 0$, and $\widetilde{\mathbf{mv}}^k_{\mathbf{u}} = \mathbf{mv}^k_{\mathbf{u}}$, i.e. $\widetilde{\xi}^k_{\mathbf{u}} = 0$. Therefore, we obtain

$$D^k_{\mathbf{u}}(p) = E[\Phi^2(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}},\, \hat{f}^k_{\mathbf{u}})]. \tag{69}$$

■

B. Proof of Proposition 1

Proof: The probability density function of the random variable having a Laplacian distribution is $f(x|\mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$. Since $\mu = 0$, we have $E[x^2] = 2b^2$, and from (38), we obtain

$$\begin{aligned} E[x^2] - E[\Phi^2(x,\, y)] &= \int_{y-\gamma_L}^{+\infty} (x^2 - (y-\gamma_L)^2)\frac{1}{2b}e^{-\frac{x}{b}}dx + \int_{-\infty}^{y-\gamma_H} [x^2 - (y-\gamma_H)^2]\frac{1}{2b}e^{\frac{x}{b}}dx \\ &= e^{-\frac{y-\gamma_L}{b}}((y-\gamma_L)\cdot b + b^2) + e^{-\frac{\gamma_H - y}{b}}((\gamma_H - y)\cdot b + b^2). \end{aligned} \tag{70}$$

From the definition of propagation factor, we obtain $\alpha = \frac{E[\Phi^2(x,y)]}{E[x^2]} = 1 - \frac{1}{2}e^{-\frac{y-\gamma_L}{b}}(\frac{y-\gamma_L}{b} + 1) - \frac{1}{2}e^{-\frac{\gamma_H - y}{b}}(\frac{\gamma_H - y}{b} + 1)$. ■

C. Proof of Lemma 2

Proof: For P-MBs with slice data partioning, from (19) and (35) we obtain

$$D^k(P) = \frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{r,m\}\cdot D^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}}) + \frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{r,\bar{m}\}\cdot D^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}) + \frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{\bar{r}\}\cdot D^k_{\mathbf{u}}(p)). \tag{71}$$

Denote $V^k_i\{r,m\}$ the set of pixels in the k-th frame with the same XEP $P^k_i\{r,m\}$; denote $N^k_i\{r,m\}$ the number of pixels in $V^k_i\{r,m\}$; denote $N^k\{r,m\}$ the number of sets with different XEP $P^k_i\{r,m\}$ in the k-th frame.

We have

$$\frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{r,m\}\cdot D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}})=\frac{1}{|\mathcal{V}|}\sum_{i=1}^{N^k\{r,m\}}(P^k_i\{r,m\}\sum_{\mathbf{u}\in\mathcal{V}^k_i\{r,m\}}D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}). \tag{72}$$

For large $N^k_i\{r,m\}$, we have $\frac{1}{N^k_i\{r,m\}}\sum_{\mathbf{u}\in\mathcal{V}^k_i\{r,m\}}D^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}$ converges to D^{k-1}, so the first term in the right-hand side in (71) is $D^{k-1}\cdot\bar{P}^k\{r,m\}$, where $\bar{P}^k\{r,m\}=\frac{1}{|\mathcal{V}|}\sum_{i=1}^{N^k\{r,m\}}(P^k_i\{r,m\}\cdot N^k_i\{r,m\})$.

Following the same process, we obtain the second term in the right-hand side in (71) as $D^{k-1}\cdot\bar{P}^k\{r,\bar{m}\}$, where $\bar{P}^k\{r,\bar{m}\}=\frac{1}{|\mathcal{V}|}\sum_{i=1}^{N^k\{r,\bar{m}\}}(P^k_i\{r,\bar{m}\}\cdot N^k_i\{r,\bar{m}\})$; and

$$\frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{\bar{r}\}\cdot D^k_{\mathbf{u}}(p))=\frac{1}{|\mathcal{V}|}\sum_{i=1}^{N^k\{\bar{r}\}}(P^k_i\{\bar{r}\}\sum_{\mathbf{u}\in\mathcal{V}^k_i\{\bar{r}\}}D^k_{\mathbf{u}}(p)). \tag{73}$$

For large $N^k_i\{\bar{r}\}$, we have $\frac{1}{N^k_i\{\bar{r}\}}\sum_{\mathbf{u}\in\mathcal{V}^k_i\{\bar{r}\}}D^k_{\mathbf{u}}(p)$ converges to $D^k(p)$, so the third term in the right-hand side in (71) is $D^k(p)\cdot(1-\bar{P}^k(r))$.

Note that $P^k_i\{r,m\}+P^k_i\{r,\bar{m}\}=P^k_i\{r\}$ and $N^k_i\{r,m\}=N^k_i\{r,\bar{m}\}$. So, we obtain

$$D^k(P)=D^{k-1}\cdot\bar{P}^k(r)+D^k(p)\cdot(1-\bar{P}^k(r)). \tag{74}$$

For P-MBs without slice data partitioning, it is straightforward to acquire (74) from (36). For I-MBs, from (37), it is also easy to obtain $D^k(P)=D^{k-1}\cdot\bar{P}^k(r)$. So, together with (74), we obtain (40). ■

D. Proof of Proposition 3

Proof: Using the definition of $\lambda^k_{\mathbf{u}}$, (43) becomes

$$D^k_{\mathbf{u}}(c)=2P^k_{\mathbf{u}}\{m\}\cdot(1-\lambda^k_{\mathbf{u}})\cdot E[\xi^k_{\mathbf{u}}\cdot\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}]. \tag{75}$$

Under the condition that the distance between $\mathbf{mv}^k_{\mathbf{u}}$ and $\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}$ is small, for example, inside the same MB, the statistics of $\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}$ and $\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}$ are almost the same. Therefore, we may assume $E[(\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}})^2]=E[(\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}})^2]$.

Since $\xi^k_{\mathbf{u}}=\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}-\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}$, we have

$$\begin{aligned}E[(\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}})^2]&=E[(\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}})^2]\\&=E[(\xi^k_{\mathbf{u}}+\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}})^2],\end{aligned} \tag{76}$$

and therefore $E[\xi^k_{\mathbf{u}}\cdot\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}]=-\frac{E[(\xi^k_{\mathbf{u}})^2]}{2}$[9].

[9]Note that following the same derivation process, we can prove $E[\xi^k_{\mathbf{u}}\cdot\hat{f}^{k-1}_{\mathbf{u}+\mathbf{mv}^k_{\mathbf{u}}}]=\frac{E[(\xi^k_{\mathbf{u}})^2]}{2}$.

Therefore, (75) can be simplify as

$$D^k_{\mathbf{u}}(c) = (\lambda^k_{\mathbf{u}} - 1) \cdot E[(\xi^k_{\mathbf{u}})^2] \cdot P^k_{\mathbf{u}}(m). \tag{77}$$

From (32), we know that $E[(\xi^k_{\mathbf{u}})^2] \cdot P^k_{\mathbf{u}}(m)$ is exactly equal to $D^k_{\mathbf{u}}(m)$. Therefore, (77) is further simplified to (46).

■

E. Proof of Lemma 3

Proof: For P-MBs with slice data partioning, from (19) and (61) we obtain

$$D^k(P) = \frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{r,m\} \cdot D^{k-1}_{\mathbf{u}+\tilde{\mathbf{m}}\mathbf{v}^k_{\mathbf{u}}}) + \frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{r,\bar{m}\} \cdot D^{k-j}_{\mathbf{u}+\mathbf{m}\mathbf{v}^k_{\mathbf{u}}}) + \frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{\bar{r}\} \cdot D^k_{\mathbf{u}}(p)). \tag{78}$$

The first term in the right-hand side in (78) is exactly the same as the first term in the right-hand side in (71), that is, it equal to $D^{k-1} \cdot \bar{P}^k\{r,m\}$, where $\bar{P}^k\{r,m\} = \frac{1}{|\mathcal{V}|}\sum_{i=1}^{N^k\{r,m\}}(P^k_i\{r,m\} \cdot N^k_i\{r,m\})$.

Denote $V^k_i(j)\{r,\bar{m}\}$ the set of pixels using the same reference frame $k-j$ in the k-th frame with the same XEP $P^k_i(j)\{r,\bar{m}\}$; denote $N^k_i(j)\{r,\bar{m}\}$ the number of pixels in $V^k_i(j)\{r,\bar{m}\}$; denote $N^k(j)\{r,\bar{m}\}$ the number of sets with different XEP $P^k_i(j)\{r,\bar{m}\}$ but the same reference frame $k-j$ in the k-th frame.

We have

$$\frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{r,\bar{m}\} \cdot D^{k-j}_{\mathbf{u}+\mathbf{m}\mathbf{v}^k_{\mathbf{u}}}) = \frac{1}{|\mathcal{V}|}\sum_{i=1}^{N^k(j)\{r,\bar{m}\}}(P^k_i(j)\{r,\bar{m}\}\sum_{j=1}^{J}\sum_{\mathbf{u}\in\mathcal{V}^k_i(j)\{r,\bar{m}\}} D^{k-j}_{\mathbf{u}+\mathbf{m}\mathbf{v}^k_{\mathbf{u}}}). \tag{79}$$

For large $N^k_i(j)\{r,\bar{m}\}$, we have $\frac{1}{N^k_i(j)\{r,\bar{m}\}}\sum_{\mathbf{u}\in\mathcal{V}^k_i(j)\{r,\bar{m}\}} D^{k-j}_{\mathbf{u}+\mathbf{m}\mathbf{v}^k_{\mathbf{u}}}$ converges to D^{k-j}, so (79) becomes

$$\frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{r,\bar{m}\} \cdot D^{k-j}_{\mathbf{u}+\mathbf{m}\mathbf{v}^k_{\mathbf{u}}}) = \frac{1}{|\mathcal{V}|}\sum_{i=1}^{N^k(j)\{r,\bar{m}\}}(P^k_i(j)\{r,\bar{m}\}\sum_{j=1}^{J} N^k_i(j)\{r,\bar{m}\} \cdot D^{k-j}). \tag{80}$$

Similar to the definition in (29), we define the weighted average over joint PEPs, of event that residual is received with error and MV is received without error, for the set of pixels using the same reference frame $k-j$ in the k-th frame as

$$\bar{P}^k(j)\{r,\bar{m}\} \triangleq \frac{1}{|\mathcal{V}^k(j)|}\sum_{i=1}^{N^k(j)\{r,\bar{m}\}}(P^k_i(j)\{r,\bar{m}\} \cdot N^k_i(j)\{r,\bar{m}\}). \tag{81}$$

We have

$$\begin{aligned}\frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P^k_{\mathbf{u}}\{r,\bar{m}\} \cdot D^{k-j}_{\mathbf{u}+\mathbf{m}\mathbf{v}^k_{\mathbf{u}}}) &= \frac{1}{|\mathcal{V}|}\sum_{j=1}^{J}(\bar{P}^k(j)\{r,\bar{m}\} \cdot |\mathcal{V}^k(j)| \cdot D^{k-j}) \\ &= \sum_{j=1}^{J}(\bar{P}^k(j)\{r,\bar{m}\} \cdot w^k(j) \cdot D^{k-j}).\end{aligned} \tag{82}$$

Following the same derivation process, we obtain

$$\frac{1}{|\mathcal{V}|}\sum_{\mathbf{u}\in\mathcal{V}^k}(P_{\mathbf{u}}^k\{\bar{r}\}\cdot D_{\mathbf{u}}^k(p)) = \sum_{j=1}^{J}(\bar{P}^k(j)\{\bar{r}\}\cdot w^k(j)\cdot\alpha^k(j)\cdot D^{k-j}), \tag{83}$$

where

$$\bar{P}^k(j)\{\bar{r}\} \triangleq \frac{1}{|\mathcal{V}^k(j)|}\sum_{i=1}^{N^k(j)\{\bar{r}\}}(P_i^k(j)\{\bar{r}\}\cdot N_i^k(j)\{\bar{r}\}) \tag{84}$$

is the weighted average over joint PEPs, of event that residual is received without error, for the set of pixels using the same reference frame $k-j$ in the k-th frame.

Therefore, we obtain

$$D^k(P) = D^{k-1}\cdot\bar{P}^k\{r,m\} + \sum_{j=1}^{J}(\bar{P}^k(j)\{r,\bar{m}\}\cdot w^k(j)\cdot D^{k-j}) + \sum_{j=1}^{J}(\bar{P}^k(j)\{\bar{r}\}\cdot w^k(j)\cdot\alpha^k(j)\cdot D^{k-j}). \tag{85}$$

For P-MBs without slice data partitioning, $\bar{P}^k\{r,m\} = \bar{P}^k\{r\}$ and $\bar{P}^k(j)\{r,\bar{m}\} = 0$, therefore we have

$$D^k(P) = D^{k-1}\cdot\bar{P}^k\{r\} + \sum_{j=1}^{J}(\bar{P}^k(j)\{\bar{r}\}\cdot w^k(j)\cdot\alpha^k(j)\cdot D^{k-j}). \tag{86}$$

For I-MBs, from (37), it is also easy to obtain $D^k(P) = D^{k-1}\cdot\bar{P}^k(r)$. So, together with (85), we obtain (64). ■

F. Lemma 4 and Its Proof

To prove Proposition 2, we need to use the following lemma.

Lemma 4: The error reduction function $\Phi(x,\, y)$ satisfies $\Phi^2(x,\, y) \le x^2$ for any $\gamma_L \le y \le \gamma_H$.

Proof: From the definition in (38), we obtain

$$\Phi^2(x,\, y) - x^2 = \begin{cases} (y-\gamma_L)^2 - x^2, & x > y-\gamma_L \\ 0, & y-\gamma_H \le x \le y-\gamma_L \\ (y-\gamma_H)^2 - x^2, & x < y-\gamma_H. \end{cases} \tag{87}$$

Since $y \ge \gamma_L$, we obtain $(y-\gamma_L)^2 < x^2$ when $x > y-\gamma_L$. Similarly, since $y \le \gamma_H$, we obtain $(y-\gamma_H)^2 < x^2$ when $x < y-\gamma_H$. Therefore $\Phi^2(x,\, y) - x^2 \le 0$ for $\gamma_L \le y \le \gamma_H$. Fig. 7 shows a pictorial example of the case that $\gamma_H = 255$, $\gamma_L = 0$ and $y = 100$. ■

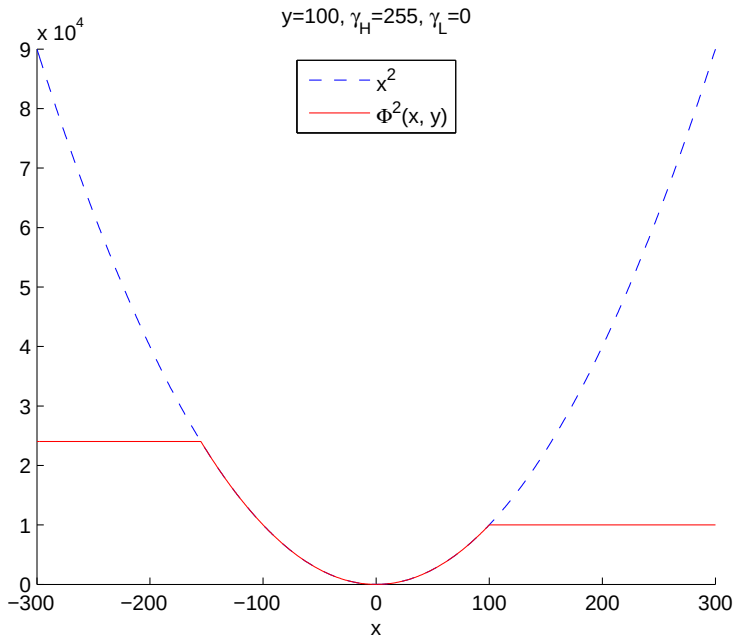

Fig. 7. Comparison of $\Phi^2(x,y)$ and x^2.

G. Lemma 5 and Its Proof

Before presenting the proof, we first give the definition of Ideal Codec.

Definition 1: Ideal Codec: both the true MV and concealed MV are within the search range, and the position pointed by the true MV, i.e., $\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k$, is the best reference pixel, under the MMSE criteria, for $\hat{f}_{\mathbf{u}}^k$ within the whole search range $\mathcal{V}_{SR}^{k-1}$, that is, $\mathbf{v} = \underset{\mathbf{v}\in\mathbf{V}_{SR}^{k-1}}{\arg\min} \{(\hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{v}}^{k-1})^2\}$.

To prove Corollary 1, we need to use the following lemma.

Lemma 5: In an ideal codec, $\widetilde{\Delta}_{\mathbf{u}}^k\{\bar{p}\} = 0$, In other words, if there is no propagated error, the clipping noise for the pixel $\mathbf{u}^k$ at the decoder is always zero no matter what kind of error event occurs in the k-th frame.

Proof: In an ideal codec, we have $(\hat{e}_{\mathbf{u}}^k)^2 = (\hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1})^2 \le (\hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1})^2$. Due to the spatial and temporal continuity of the natural video, we can prove by contradiction that in an ideal codec $\hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1}$ and $\hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$ have the same sign, that is either

$$\hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} \ge \hat{e}_{\mathbf{u}}^k \ge 0, \quad \text{or} \quad \hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1} \le \hat{e}_{\mathbf{u}}^k \le 0. \tag{88}$$

If the sign of $\hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\mathbf{mv}_{\mathbf{u}}^k}^{k-1}$ and $\hat{f}_{\mathbf{u}}^k - \hat{f}_{\mathbf{u}+\check{\mathbf{mv}}_{\mathbf{u}}^k}^{k-1}$ is not the same, then due to the spatial and temporal continuity of the input video, there exists a better position $\mathbf{v} \in \mathcal{V}^{k-1}$ between $\mathbf{mv}_{\mathbf{u}}^k$ and $\check{\mathbf{mv}}_{\mathbf{u}}^k$, and

therefore within the search range, so that $(\hat{e}^k_{\mathbf{u}})^2 \geq (\hat{f}^k_{\mathbf{u}} - \hat{f}^{k-1}_{\mathbf{v}})^2$. In this case, encoder will choose $\mathbf{v}$ as the best reference pixel within the search range. This contradicts the assumption that the best reference pixel is $\mathbf{u} + \mathbf{mv}^k_{\mathbf{u}}$ within the search range.

Therefore, from (88), we obtain

$$\hat{f}^k_{\mathbf{u}} \geq \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}} \geq \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}}, \quad \text{or} \quad \hat{f}^k_{\mathbf{u}} \leq \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}} \leq \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}}. \tag{89}$$

Since both $\hat{f}^k_{\mathbf{u}}$ and $\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}}$ are reconstructed pixel value, they are within the range $\gamma_H \geq \hat{f}^k_{\mathbf{u}}, \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}} \geq \gamma_L$. From (89), we have $\gamma_H \geq \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}} \geq \gamma_L$, and thus $\Gamma(\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}}) = \hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}}$. As a result, we obtain $\widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r}, m, \bar{p}\} = (\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}}) - \Gamma(\hat{f}^{k-1}_{\mathbf{u}+\check{\mathbf{mv}}^k_{\mathbf{u}}} + \hat{e}^k_{\mathbf{u}}) = 0$.

Since $\widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r}, \bar{m}, \bar{p}\} = \hat{\Delta}^k_{\mathbf{u}} = 0$, and from Section IV-D1, we know that $\widetilde{\Delta}^k_{\mathbf{u}}\{r, \bar{p}\} = 0$, hence we obtain $\widetilde{\Delta}^k_{\mathbf{u}}\{\bar{p}\} = 0$. ■

Remark 1: Note that Lemma 5 is proved under the assumption of pixel-level motion estimation. In a practical encoder, block-level motion estimation is adopted with the criterion of minimizing the MSE of the whole block, e.g., in H.263, or minimizing the cost of residual bits and MV bits, e.g., in H.264. Therefore, some reference pixels in the block may not be the best reference pixel within the search range. On the other hand, Rate Distortion Optimization (RDO) as used in H.264 may also cause some reference pixels not to be the best reference pixels. However, the experiment results for all the test video sequences show that the probability of $\widetilde{\Delta}^k_{\mathbf{u}}\{\bar{r}, m, \bar{p}\} \neq 0$ is negligible.

H. Proof of Corollary 1

Proof: From (67), we obtain $\widetilde{\Delta}^k_{\mathbf{u}}\{\bar{p}\} = (\hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}}) - \Gamma(\hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}})$. Together with Lemma 5, which is presented and proved in Appendix G, we have $\gamma_L \leq \hat{f}^k_{\mathbf{u}} - \widetilde{\xi}^k_{\mathbf{u}} - \widetilde{\varepsilon}^k_{\mathbf{u}} \leq \gamma_H$. From Lemma 4 in Appendix F, we have $\Phi^2(x, y) \leq x^2$ for any $\gamma_L \leq y \leq \gamma_H$; together with (68), it is straightforward to prove that $E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}})^2] \leq E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}})^2]$. By expanding $E[(\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} + \widetilde{\Delta}^k_{\mathbf{u}})^2]$, we obtain

$$E[\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}} \cdot \widetilde{\Delta}^k_{\mathbf{u}}] \leq -\frac{1}{2} E[(\widetilde{\Delta}^k_{\mathbf{u}})^2] \leq 0. \tag{90}$$

The physical meaning of (90) is that $\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}}$ and $\widetilde{\Delta}^k_{\mathbf{u}}$ are negatively correlated if $\widetilde{\Delta}^k_{\mathbf{u}} \neq 0$. Since $\widetilde{\Delta}^k_{\mathbf{u}}\{r\} = 0$ as noted in Section IV-D1 and $\widetilde{\Delta}^k_{\mathbf{u}}\{\bar{p}\} = 0$ as proved in Lemma 5, we know that $\widetilde{\Delta}^k_{\mathbf{u}} \neq 0$ is valid only for the error events $\{\bar{r}, m, p\}$ and $\{\bar{r}, \bar{m}, p\}$, and $\widetilde{\Delta}^k_{\mathbf{u}} = 0$ for any other error event. In other words, $\widetilde{\zeta}^{k-1}_{\mathbf{u}+\widetilde{\mathbf{mv}}^k_{\mathbf{u}}}$ and $\widetilde{\Delta}^k_{\mathbf{u}}$ are negatively correlated under the condition $\{\bar{r}, p\}$, and they are uncorrelated under other conditions. ■

REFERENCES

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec. Part*, vol. 4, pp. 142–163, 1959.

[2] T. Berger and J. Gibson, "Lossy source coding," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2693–2723, 1998.

[3] 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.

[4] 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.

[5] T. Stockhammer, T. Wiegand, and S. Wenger, "Optimized transmission of h.26l/jvt coded video over packet-lossy networks," in *IEEE ICIP*, 2002.

[6] K. Stuhlmuller, 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.

[7] J. U. Dani, Z. He, and H. Xiong, "Transmission distortion modeling for wireless video communication," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'05)*, 2005.

[8] 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 System for Video Technology, special issue on wireless video*, vol. 12, pp. 511–523, Jun. 2002.

[9] 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.

[10] J. Chakareski, J. Apostolopoulos, W.-T. Tan, S. Wee, and B. Girod, "Distortion chains for predicting the video distortion for general packet loss patterns," in *Proc. ICASSP*, 2004.

[11] J. Chakareski, J. Apostolopoulos, S. Wee, W.-T. Tan, and B. Girod, "Rate-distortion hint tracks for adaptive video streaming," *IEEE transactions on circuits and systems for video technology*, vol. 15, no. 10, pp. 1257–1269, 2005.

[12] C. Zhang, H. Yang, S. Yu, and X. Yang, "GOP-level transmission distortion modeling for mobile streaming video," *Signal Processing: Image Communication*, 2007.

[13] M. T. Ivrlač, L. U. Choi, E. Steinbach, and J. A. Nossek, "Models and analysis of streaming video transmission over wireless fading channels," *Signal Processing: Image Communication*, vol. 24, no. 8, pp. 651–665, Sep. 2009.

[14] 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.

[15] H. Yang and K. Rose, "Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H. 264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 7, p. 845, 2007.

[16] "H.264/AVC reference software JM14.0," May. 2008. [Online]. Available: http://iphome.hhi.de/suehring/tml/download

[17] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[18] S. Wenger, "H.264/AVC over IP," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 645– 656, Jul. 2003.

[19] 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*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[20] D. Agrafiotis, D. R. Bull, and C. N. Canagarajah, "Enhanced error concealment with mode selection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 8, pp. 960–973, Aug. 2006.

[21] *ITU-T Series H: Audiovidual and Multimedia Systems, Advanced video coding for generic audiovisual services*, Nov. 2007.

[22] Z. Chen and D. Wu, "Prediction of Transmission Distortion for Wireless Video Communication: Part II: Algorithm and Application," 2010, http://www.wu.ece.ufl.edu/mypapers/journal-2.pdf.