Perception Loss Function Adaptive to Rate for Learned Video Compression

Anonymous Author(s)
Affiliation
Address
email

Abstract

We consider causal, low-latency, sequential video compression, with mean squared-error (MSE) as the distortion loss, and a perception loss function (PLF) to enhance the realism of outputs. Prior works have employed two PLFs: one based on the joint distribution (JD) of all frames up to the current one, and the other based on frame-wise marginal distribution (FMD). We introduce a new PLF, called *adaptive to rate (AR)*, which preserves the joint distribution of the current frame with all previous reconstructions. Through information-theoretic analysis and deep-learning experiments, we show that PLF-AR can rectify past errors in future reconstructions when the initial frame is compressed at a low bitrate. However, in this bitrate scenario, PLF-JD exhibits the error permanence phenomenon, propagating mistakes in subsequent outputs. When the initial frame is compressed at a high bitrate, PLF-AR maintains temporal correlation among frames, preventing error propagation in future reconstructions—unlike PLF-JD, which remains stuck in past mistakes. Furthermore, PLF-FMD does not preserve temporal correlation as effectively as PLF-AR. These characteristics of PLFs are especially apparent in scenarios with sharp frame movements. In contrast, when frame movements are smoother, the three PLFs display slight variations: PLF-AR and PLF-JD yield more diverse outputs, while PLF-FMD tends to replicate the initial frame in all future reconstructions. We validate our findings through information-theoretic analysis of the rate-distortion-perception tradeoff for the Gauss-Markov source model and deep-learning experiments on moving MNIST and UVG datasets.

1 Introduction

In recent years, the topic of lossy compression for videos has received significant attention, driven by the growing demand for producing visually appealing reconstructions even at lower bitrates. Early versions of compression algorithms relied on distortion measures, e.g., MSE, MS-SSIM [1–3] and PSNR [2–5]. However, these metrics often resulted in outputs that were perceived as blurry and lacking *realism*. Consequently, there have been efforts to incorporate *perception*-based loss functions into compression systems to improve visual quality. These loss functions aim to quantify the divergence between the distributions of the source and the reconstruction, where achieving *perfect* perceptual quality means that the two distributions match with each other. Blau and Michaeli [6] explored the rate-distortion-perception (RDP) tradeoff from a theoretical perspective. Subsequently, Zhang et al. [7] introduced universal representations, wherein the representation remains fixed during encoding, and only the decoder can be adjusted to attain near-optimal performance.

Extending image compression algorithms to handle video poses a challenge as they must maintain temporal correlation across frames, alongside spatial correlation preservation. Moreover, with the multitude of frames in a video, defining a unique perception loss function (PLF) becomes nontrivial.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

(a) (b) (c)

Figure 1: (a) The outputs of different PLFs for the MovingMNIST dataset when the first frame is compressed at a low rate. Both PLF-AR and PLF-FMD recover from previous mistakes while PLF-JD suffers from the error permanence phenomenon. **(b)** The outputs of different PLFs for the UVG dataset when the first frame is compressed at a low rate. Both PLF-AR and PLF-FMD are able to preserve the color tone of the output, while PLF-JD propagates mistakes in the color tone. **(c)** The outputs of different PLFs for the MovingMNIST dataset when the first frame is compressed at a high rate. When there are sharp movements in the trajectory, PLF-AR preserves the temporal correlation across different frames. PLF-JD propagates the mistakes. PLF-FMD produces some errors in the reconstruction and does not successfully maintain the temporal correlation.

Some previous studies have approached PLF by considering frame-wise marginal distributions (FMD) of the source and reconstruction [8], as well as joint distribution (JD) of different frames [9]. A recent study [10] explored the rate-distortion-perception (RDP) tradeoff for sequential video compression theoretically. It highlighted that at low bitrates, the PLF-JD encounters the *error permanence phenomenon*, wherein mistakes propagate across all future reconstructions, leaving distortion unchanged across frames.

In this work, we explore a causal, sequential video compression scenario where the mean squared error (MSE) serves as the distortion measure. Introducing a new perception loss function, we propose a metric that maintains the joint distribution of the current frame alongside all previous reconstructions. We refer to this PLF as *Adaptive to Rate (AR)*, and we will explain the reasoning behind this label in the following discussion. Our contributions are as follows:

- *Error correction when the initial frame is compressed at a low bitrate*: We demonstrate that PLF-AR does not suffer from the error permanence phenomenon in low bitrates. On the theoretical side, we use an approximation for the operational RDP region for a first-order Markov source model and specialize it to Gauss-Markov sources. We show that when the first frame is compressed at a low bitrate, given a medium bitrate to the future frames, PLF-AR is able to recover from the previous mistakes in future reconstructions. On the experimental side (see Fig. 1a and Fig. 1b), at low bitrates, PLF-JD suffers from the error permanence phenomenon where mistakes are propagated in future outputs.

- *Maintenance of temporal correlation when the initial frame is compressed at a high bitrate*: Through both theoretical analysis and experimental findings (see Fig. 1c), we demonstrate that when the second frame is allocated a low bitrate, PLF-AR can rectify errors in subsequent reconstructions. This phenomenon is particularly prominent when video frames exhibit rapid movements, resulting in a low correlation coefficient between them. However, PLF-JD tends to remain stuck on errors from previous reconstructions. Additionally, PLF-FMD fails to preserve temporal correlation as effectively as PLF-AR. In cases where frame movements are smoother (indicating a higher correlation coefficient between frames), the three PLFs exhibit slightly different behaviors: PLF-AR and PLF-JD generate more *diverse* outputs, while PLF-FMD tends to *copy* the first frame, resulting in more *static* reconstructions.

Based on the above discussion, PLF-AR does not suffer from the error permanence phenomenon at low bitrates and maintains temporal correlation among frames, especially when the first frame undergoes high-rate compression. Consequently, it leverages the advantages of both metrics (PLF-FMD or PLF-JD) depending on the operational rate regime. This adaptability to varying rates is the rationale behind naming it PLF adaptive to rate.

2 System Model and Preliminaries

Assume that we have T frames of video denoted by $(X_1, \ldots, X_T) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_T$ (where $\mathcal{X}_i \subseteq \mathbb{R}^d$) distributed according to joint distribution $P_{X_1 \ldots X_T}$. The encoders and decoders have access to a shared common randomness $K \in \mathcal{K}$. The (possibly stochastic) jth encoding function gets the sources $(X_1, \ldots, X_j)$ and the key K and outputs a variable length message $M_j \in \mathcal{M}_j (= \{0,1\}^\star)$, i.e.,

$$f_j \colon \mathcal{X}_1 \times \ldots \times \mathcal{X}_j \times \mathcal{K} \to \mathcal{M}_j, \qquad\qquad j = 1, \ldots, T. \tag{1}$$

The jth decoding function receives the messages $(M_1, \ldots, M_j)$ and using the key K, it outputs a reconstruction $\hat{X}_j \in \hat{\mathcal{X}}_j (\subseteq \mathbb{R}^d)$, i.e.,

$$g_j \colon \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_j \times \mathcal{K} \to \hat{\mathcal{X}}_j. \tag{2}$$

The mappings $\{f_j\}_{j=1}^T$ and $\{g_j\}_{j=1}^T$ induce the conditional distribution $P_{\hat{X}_1 \ldots \hat{X}_T | X_1 \ldots X_T}$ for the reconstructed video given the original video. The proposed framework is a *one-shot* setting where a single sample of the source is compressed at a time.

The reconstruction of each frame j should satisfy a certain distortion from the source where the metric is assumed to be the mean squared-error (MSE) function i.e. $d(x_j, \hat{x}_j) = ||x_j - \hat{x}_j||^2$, which is widely used in many applications. From the perceptual perspective, for given probability distributions $P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}$ and $P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j}$, let $\phi_j(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j})$ denote the perception metric capturing the divergence between them. We call this metric as *perception loss function adaptive to rate (PLF-AR)*. If $\phi_j(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j}) = 0$, we get

$$P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j} = P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j}, \qquad j = 1, \ldots, T, \tag{3}$$

which is called as *zero-perception loss function adaptive to rate (0-PLF-AR)*. In the following, we define two other perception metrics which are extensively used in many works. For given probability distributions $P_{X_1 \ldots X_j}$ and $P_{\hat{X}_1 \ldots \hat{X}_j}$, let $\xi_j(P_{X_1 \ldots X_j}, P_{\hat{X}_1 \ldots \hat{X}_j})$ be called as *perception loss function based on joint distribution (PLF-JD)*. Alternatively, the *perception loss function based on framewise marginal distribution (PLF-FMD)* is shown by $\psi_j(P_{X_j}, P_{\hat{X}_j})$. Notice that 0-PLF-JD and 0-PLF-FMD imply that $P_{X_1 \ldots X_j} = P_{\hat{X}_1 \ldots \hat{X}_j}$ and $P_{X_j} = P_{\hat{X}_j}$ for $j = 1, \ldots, T$, respectively.

Definition 1 (Operational RDP region) *An RDP tuple* $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ *is said to be achievable for the one-shot setting if there exist encoders and decoders such that:*

$$\mathbb{E}[\ell(M_j)] \leq R_j, \tag{4}$$

$$\mathbb{E}[\|X_j - \hat{X}_j\|^2] \leq D_j, \tag{5}$$

$$\phi_j(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j}) \leq P_j, \quad j = 1, 2, 3, \tag{6}$$

where $\ell(M_j)$ denotes the length of the message M_j. The operational RDP region, denoted by $\mathcal{RDP}^o$, is the closure of the set of all achievable tuples. Moreover, for a given (D, P), the operational rate region, denoted by $\mathcal{R}^o(\mathsf{D}, \mathsf{P})$, is the closure of the set of all tuples R such that $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \mathcal{RDP}^o$.

Furthermore, we consider Gauss-Markov sources as follows. We assume that $X_1 \sim \mathcal{N}(0, \sigma^2)$ for some $\sigma^2 > 0$,

$$X_2 = \rho X_1 + N_1, \qquad X_3 = \rho X_2 + N_2, \tag{7}$$

for some $0 \leq \rho \leq 1$, where N_j is independent of X_j with mean zero and variance $(1-\rho^2)\sigma^2$ for $j = 1, 2$. The model extends naturally to the case of T time-steps. We assume that the perception metric is Wasserstein-2 distance, i.e.,

$$\phi_j(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j}) := W_2^2(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j}). \tag{8}$$

Using Strong Functional Representation Lemma (SFRL) [11], we find an alternative characterization for the operational RDP region which is more tractable and then investigate it for Gauss-Markov sources (see Appendices A and B for details).

3 Distortion Analysis for Gauss-Markov Sources and Zero-Perception Loss

In this section, we present practical insights from analyzing the Gauss-Markov source model. We consider two extreme compression rates for the first frame: a low rate, denoted as $R_1 = \epsilon$ for very small $\epsilon > 0$, and a high rate where $R_1 \to \infty$.

3.1 Compressing the First Frame at a Low Rate ($R_1 = \epsilon$ for sufficiently small $\epsilon > 0$)

One of the key observations in this section is that how the performances of PLFs vary based on the operating rate regime. In the following result, we assume that the rate of the second step, R_2, can take on any nonnegative value, and we then investigate how each PLF affects the reconstruction in this step. The achievabale distortions for the second frame, $D^0_{2,\text{AR}}$ (for 0-PLF-AR), $D^0_{2,\text{FMD}}$ (for 0-PLF-FMD) and $D^0_{2,\text{JD}}$ (for 0-PLF-JD) are given by (see Appendix C for the proof)

$$D^0_{2,\text{AR}} = 2\sigma^2(1 - \sqrt{1 - 2^{-2R_2}}), \qquad D^0_{2,\text{FMD}} = 2\sigma^2(1 - \sqrt{1 - 2^{-2R_2} + \rho^2 2\epsilon \ln 2}),$$
$$D^0_{2,\text{JD}} = 2\sigma^2(1 - \sqrt{1-\rho^2}\sqrt{1 - 2^{-2R_2}} - \rho^2\sqrt{2\epsilon \ln 2}). \tag{9}$$

We specialize to $\rho = 1$ (see Fig. 2). 0-PLF-JD results in the same distortion across different frames meaning that mistakes in reconstructions are propagated in future frames. This behavior is called *error permanence phenomenon* as introduced in [10]. Both 0-PLF-AR and 0-PLF-FMD do not suffer from the error permanence phenomenon as observed in Fig. 2. For 0-PLF-AR, the reconstructions of two frames are *decoupled*, minimizing the potential for error propagation. For 0-PLF-FMD, the reconstruction of each frame relies on both preceding and current frames, with each frame's portion optimized to minimize distortion. As it can be observed from Fig. 2, for $R_1 = 0.1$ and $R_2 \approx 0.05$, the distortion of the second frame for 0-PLF-AR outperforms that of 0-PLF-JD. For $R_2 \approx R_1$ (i.e. for a large range of R_2), the performance of 0-PLF-AR is close to that of 0-PLF-FMD. This observation implies that 0-PLF-AR is able to adapt its performance to the operating rate regime. Some further results on the third frame are detailed in Appendix C.

In the next section, we will discuss that 0-PLF-AR is able to preserve the temporal correlation of different frames when the first frame is compressed at a high rate.

3.2 Compressing the First Frame at a High Rate ($R_1 \to \infty$)

In this section, we discuss that the choice of PLF significantly affects the temporal correlation across different frames. Specifically, we consider the case where $R_2 = R_3 = \epsilon$ for sufficiently small $\epsilon > 0$. In the first step, the high rate assumption implies that $\hat{X}_1 = X_1$. The achievable reconstructions of all 0-PLFs for the second and third steps are shown in Table 1. As it can be observed from the first row of Table 1, for a sufficiently large correlation coefficient (i.e., $\sqrt{\epsilon} \ll \rho < 1$), the reconstruction based on 0-PLF-FMD for the second frame is given by $\hat{X}_2 \approx (1 - O(\epsilon))\hat{X}_1 + O(\epsilon)X_2$, meaning that the first frame is *copied* in the future reconstruction. A similar argument applies to the reconstruction of the third frame. So, the outputs of 0-PLF-FMD are expected to look more *static* comparing to the other PLFs. This type of static reconstruction mostly happens when the correlation coefficient ρ is large enough, making the movement between frames smooth. However, for a small correlation coefficient (i.e., $0 < \rho \ll \sqrt{\epsilon}$) corresponding to the second row of Table 1, we have $\hat{X}_2 \approx Z'_{2,\text{FMD}}$ where $Z'_{2,\text{FMD}} \sim \mathcal{N}(0, (1 - O(\epsilon))\sigma^2)$ is independent of X_2. Therefore, the decoder based on 0-PLF-FMD

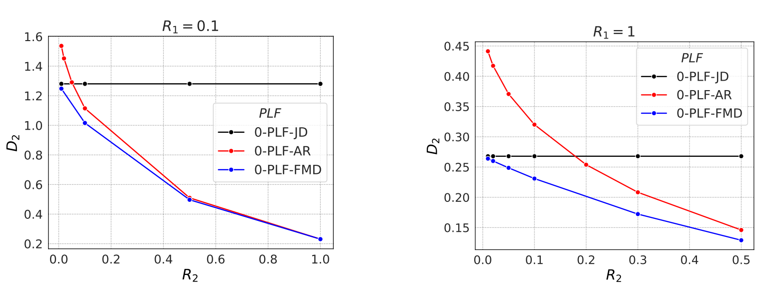


Figure 2: Distortion of the second frame versus its rate for the low-rate regime and $\rho = 1$.

primarily reconstructs the second frame by introducing artificial noise, $Z'_{2,\text{FMD}}$, which could lead to errors in the output. When there is a small correlation coefficient between frames, it means the video will have abrupt movements, and using 0-PLF-FMD might result in random errors in the reconstruction.

The 0-PLF-AR condition in the second frame is expressed as $P_{\hat{X}_1 X_2} = P_{\hat{X}_1 \hat{X}_2}$. When combined with the high compression rate for the initial frame (i.e., $R_1 \to \infty$), it reduces to $P_{X_1 X_2} = P_{\hat{X}_1 \hat{X}_2}$, which is equivalent to the constraint in the 0-PLF-JD framework. According to the third and fourth rows of Table 1, both 0-PLF-AR and 0-PLF-JD are able to get the informative portion of the first frame (i.e., $\approx \rho$) in the second reconstruction. So, both PLFs preserve the temporal correlation between different reconstructions and generate more *diverse* outputs in the sense that they both do not simply copy the first frame in the future reconstruction. Comparing the third and fourth rows of Table 1 for the third step, the decoder based on 0-PLF-JD gets a constant factor ρ of the reconstruction noise in the previous step, i.e., $\rho Z_{2,\text{JD}}$, hence it is more susceptible to propagate false information in future frames. However, 0-PLF-AR experiences a significantly reduced factor of reconstruction noise from the preceding step, approximately $O(\sqrt{\epsilon})Z_{2,\text{AR}}$. This allows for more flexibility in correcting errors by introducing artificial noise, labeled as $Z_{3,\text{AR}}$. In our experiments, we also permit a sufficiently high compression rate for the third frame, making the artificial noise $Z_{3,\text{AR}}$ a reliable approximation of the original frame. Considering the discussion in both this section and the previous one, 0-PLF-AR manages to leverage the benefits of each metric (either 0-PLF-FMD or 0-PLF-JD) based on the operating rate regime. The way 0-PLF-AR behaves inspired us to name it a PLF adaptive to rate.

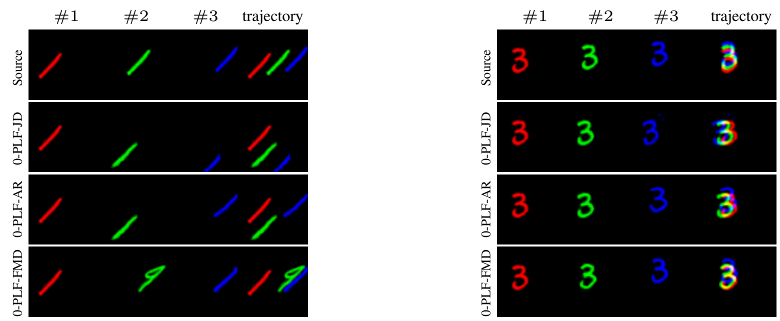
4 Experimental Results

Our theoretical results for PLF-AR show that PLF-AR is a new perceptual metric that inherits advantages in both PLF-JD and PLF-FMD. When the first frame is lossily compressed at a low rate, it does not suffer from the error permanence as in PLF-JD. When the first frame is perfectly transmitted, on the other hand, its reconstruction does not suffer from content modification, which is the phenomenon that happens within PLF-FMD in this rate-regime. In this section, we provide experimental results to validate our proposed theory for learning-based perceptual video compression. Expanding upon the experimental framework established in [10], we merge the scale-space-flow neural video coding architecture introduced by [12] with Wasserstein GANs for perceptual quality enhancement, as proposed in [13]. We employ two datasets: the 1-digit MovingMNIST dataset [14] and UVG dataset [15], offering varying levels of video resolution and scene complexity. The MovingMNIST dataset consists of low-complexity synthetic sequences with dimensions of 64×64, while the UVG dataset comprises high-definition real-life video patches sized at 256×256. The preference for certain deep learning structures and datasets aims at confirming the suggested theory rather than developing the most advanced neural network architectures. We start our experiments by generating the RDP tradeoffs for PLF-AR, PLF-JD, and PLF-FMD. Following that, we validate the low-rate regime presented in Section 3.1. Finally, the complementary high-rate regime described in Section 3.2 is implemented. More details on experiments are provided in Appendix E.

Fig. 1a shows samples of 3-frame MovingMNIST sequences when the first frame is encoded with a low rate. This aligns with the discussion in Section 3.1. For each sample, the first frame reconstruction $\hat{X}_1$ is wrongly decoded. As expected, the 0-PLF-JD reconstructions for the second and third frames (i.e., $\hat{X}_2$ and $\hat{X}_3$) suffer from the error permanence phenomenon. On the other hand, the decoder

Table 1: Achievable reconstructions and distortions for $R_1 \to \infty$ and $R_2 = R_3 = \epsilon$.

	Second Step	Third Step
0-PLF-FMD ($\sqrt{\epsilon} \ll \rho < 1$)	$\hat{X}_2 = (1 - O(\epsilon))\hat{X}_1 + O(\epsilon)X_2 + Z_{2,\text{FMD}}$ $Z_{2,\text{FMD}} \sim \mathcal{N}(0, O(\epsilon)\sigma^2)$ $D^{\infty}_{2,\text{FMD}} = 2(1 - \rho - O(\epsilon))\sigma^2$ [10, Table 2]	$\hat{X}_3 = (1 - O(\epsilon))\hat{X}_1 + O(\epsilon)X_2 + O(\epsilon)X_3 + Z_{3,\text{FMD}}$ $Z_{3,\text{FMD}} \sim \mathcal{N}(0, O(\epsilon)\sigma^2)$ $D^{\infty}_{3,\text{FMD}} = 2(1 - \rho^2 - O(\epsilon))\sigma^2$ (Appendix D.3)
0-PLF-FMD ($0 < \rho \ll \sqrt{\epsilon}$)	$\hat{X}_2 = O(\sqrt{\epsilon})X_2 + Z'_{2,\text{FMD}}$ $Z'_{2,\text{FMD}} \sim \mathcal{N}(0, (1 - O(\epsilon))\sigma^2)$ $D^{\infty}_{2,\text{FMD}} = 2\sigma^2(1 - O(\sqrt{\epsilon}))$ (Appendix D.3)	$\hat{X}_3 = O(\sqrt{\epsilon})X_3 + Z'_{3,\text{FMD}}$ $Z'_{3,\text{FMD}} \sim \mathcal{N}(0, (1 - O(\epsilon))\sigma^2)$ $D^{\infty}_{3,\text{FMD}} = 2\sigma^2(1 - O(\sqrt{\epsilon}))$ (Appendix D.3)
0-PLF-JD	$\hat{X}_2 = (\rho - O(\sqrt{\epsilon}))\hat{X}_1 + O(\sqrt{\epsilon})X_2 + Z_{2,\text{JD}}$ $Z_{2,\text{JD}} \sim \mathcal{N}(0, (1 - \rho^2 + O(\epsilon))\sigma^2)$ $D^{\infty}_{2,\text{JD}} = 2\sigma^2(1 - \rho^2 - O(\sqrt{\epsilon}))$ [10, Table 2]	$\hat{X}_3 = \rho^2\hat{X}_1 + O(\sqrt{\epsilon})N_1 + O(\sqrt{\epsilon})N_2 + \rho Z_{2,\text{JD}} + Z_{3,\text{JD}}$ $Z_{3,\text{JD}} \sim \mathcal{N}(0, (1 - \rho^2 + O(\epsilon))\sigma^2)$ $D^{\infty}_{3,\text{JD}} = 2\sigma^2(1 - \rho^4 - O(\sqrt{\epsilon}))$ (Appendix D.2)
0-PLF-AR	$\hat{X}_2 = (\rho - O(\sqrt{\epsilon}))\hat{X}_1 + O(\sqrt{\epsilon})X_2 + Z_{2,\text{AR}}$ $Z_{2,\text{AR}} = Z_{2,\text{JD}}$ $D^{\infty}_{2,\text{AR}} = D^{\infty}_{2,\text{JD}}$ (Appendix D.1)	$\hat{X}_3 = \rho^2\hat{X}_1 + O(\sqrt{\epsilon})N_1 + O(\sqrt{\epsilon})N_2 + O(\sqrt{\epsilon})Z_{2,\text{AR}} + Z_{3,\text{AR}}$ $Z_{3,\text{AR}} \sim \mathcal{N}(0, (1 - \rho^4 + O(\epsilon))\sigma^2)$ $D^{\infty}_{3,\text{AR}} = 2\sigma^2(1 - \rho^4 - O(\sqrt{\epsilon}))$ (Appendix D.1)



(a) Sharp movement scenario. **(b)** Slow movement scenario.

Figure 3: MovingMNIST reconstructions for ∞-R_2-R_3 with $R_2 = 2$ bits and $R_3 = 16$ bits. Digits are coloured for easily visualizing the trajectory across frames.

based on 0-PLF-FMD is able to recover from the wrong prediction in $\hat{X}_1$. The proposed 0-PLF-AR also successfully corrects the wrongly predicted $\hat{X}_1$ frame, demonstrating its capability to rectify previous mistakes. Analogous results for the UVG dataset are shown in Fig. 1b. When encoded with a low rate, the first frame reconstruction $\hat{X}_1$ presents an altered overall color tone. This error is propagated to the subsequent frames $\hat{X}_2$, and $\hat{X}_3$ by 0-PLF-JD. However, 0-PLF-FMD, 0-PLF-AR can correct the color tone of the subsequent frame reconstructions $\hat{X}_2$, $\hat{X}_3$.

In Fig. 3, we present experimental results to validate the discussion in Section 3.2, and Table 1. Here, we encode the first frame using a high rate $R_1 = \infty$, while setting $R_2 = 2$ bits and $R_3 = 16$ bits to represent low and medium rates. This configuration ensures $\hat{X}_1 = X_1$ across all PLFs. We evaluate each PLF's performance with reconstructed frames, denoted as $\hat{X}_2$ and $\hat{X}_3$. We analyze the digit trajectory across the three frames, considering both scenarios of sharp movement (see Fig. 3a corresponding to a large correlation coefficient ρ) and slow movement (see Fig. 3b corresponding to a small correlation coefficient ρ). In practice, the sharp and slow movements correspond to the scenarios where the video sampling rate is small and high respectively. It is important to note that, for this rate regime, 0-PLF-JD and 0-PLF-AR produce identical reconstructions in the second frame (refer to the third and fourth rows of Table 1).

For the sharp-movement scenario in Fig. 3a, the digit maintains its motion direction across all three frames. In the second frame, 0-PLF-AR fails to identify the direction correctly, but still provides the perceptual quality as it preserves the content consistency in the first frame. We note that 0-PLF-AR still manages to rectify this error by the third frame, particularly when given a medium compression rate. In contrast, 0-PLF-JD does not benefit from higher rates at $\hat{X}_3$, propagating the error. When comparing 0-PLF-AR and 0-PLF-FMD, the latter metric cannot preserve the temporal correlation as effective as the former one. This is because 0-PLF-FMD tries to reconstruct the low-rate frame $\hat{X}_2$ mainly by introducing some artificial noise. While the limited rate of the second frame lets 0-PLF-FMD decode the position of the digit, but it struggles to correctly identify which digit it is. Intuitively, this is because the 0-PLF-FMD model does not have the incentive to preserve the content in the sequence. As such, when dealing with sharp movement at low bit rates, it will compensate for this high uncertainty by adjusting the content to minimize the distortion. On the other hand, unlike the 0-PLF-FMD, 0-PLF-AR can simultaneously maintain the content and correct errors. The behavior of PLFs slightly differs in the scenario of slow movement depicted in Fig. 3b. 0-PLF-AR can still recover from the wrong prediction in $\hat{X}_2$ by reconstructing $\hat{X}_3$ at a high rate, while 0-PLF-JD propagates errors. 0-PLF-FMD copies the digit's position in the reconstruction of the second frame $\hat{X}_2$. As observed in the previously discussed rate regime, PLF-AR continues to leverage the advantageous features of both PLF-JD and PLF-FMD, adapting effectively to the given rate.

5 Conclusions

In this work, we proposed a new PLF for video compression setting. In low bitrates, this PLF does not suffer from the error permanence phenomenon. When the first frame is compressed at a high bitrate, it is able to preserve the temporal correlation across different frames. So, it adapts itself to the operating rate regime. This behavior motivated us to call this PLF adaptive to rate.

References

[1] A. Golinski, R. Pourreza, Y. Yang, G. Sautiere, and T. S. Cohen, "Feedback recurrent autoencoder for video compression," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[2] O. Rippel, A. G. Anderson, K. Tatwawadi, S. Nair, C. Lytle, and L. Bourdev, "Elf-vc: Efficient learned flexible-rate video coding," 2021. [Online]. Available: https://arxiv.org/abs/2104.14335

[3] J. Li, B. Li, and Y. Lu, "Deep contextual video compression," in *Advances in Neural Information Processing Systems*, 2021, pp. 18 114–18 125.

[4] E. Agustsson, D. Minnen, N. Johnston, J. Ballé, S. J. Hwang, and G. Toderici, "Scale-space flow for end-to-end optimized video compression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8503–8512.

[5] R. Yang, Y. Yang, J. Marino, and S. Mandt, "Hierarchical autoregressive modeling for neural video compression," 2020. [Online]. Available: https://arxiv.org/pdf/2010.10258.pdf

[6] Y. Blau and T. Michaeli, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *International Conference on Machine Learning*. PMLR, 2019, pp. 675–685.

[7] G. Zhang, J. Qian, J. Chen, and A. Khisti, "Universal rate-distortion-perception representations for lossy compression," in *Advances in Neural Information Processing Systems*, 2021, pp. 11 517–11 529.

[8] F. Mentzer, E. Agustsson, J. Ballé, D. Minnen, N. Johnston, and G. Toderici, "Neural video compression using gans for detail synthesis and propagation," in *European Conference on Computer Vision*, 2022.

[9] V. Veerabadran, R. Pourreza, A. Habibian, and T. Cohen, "Adversarial distortion for learned video compression," 2021. [Online]. Available: https://arxiv.org/pdf/2004.09508.pdf

[10] S. Salehkalaibar, B. Phan, J. Chen, W. Yu, and A. Khisti, "On the choice of perception loss function for learned video compression," in *Advances in Neural Information Processing Systems*, 2023.

[11] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Trans. on Info. Theory*, vol. 64, no. 11, pp. 6967–6978, 2018.

[12] E. Agustsson, D. Minnen, N. Johnston, J. Balle, S. J. Hwang, and G. Toderici, "Scale-space flow for end-to-end optimized video compression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8503–8512.

[13] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[14] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using lstms," in *International conference on machine learning*. PMLR, 2015, pp. 843–852.

[15] A. Mercat, M. Viitanen, and J. Vanne, "Uvg dataset: 50/120fps 4k sequences for video codec analysis and development," in *Proceedings of the 11th ACM Multimedia Systems Conference*, ser. MMSys '20. Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3339825.3394937

[16] P. Stavrou, M. Skoglund, and T. Tanaka, "Sequential source coding for stochastic systems subject to finite rate constraints," *IEEE Trans. on Automatic Control*, vol. 67, no. 8, pp. 3822–3835, 2022.

[17] A. Khina, V. Kostina, A. Khisti, and B. Hassibi, "Tracking and control of gauss-markov processes over packet-drop channels with acknowledgments," *IEEE Trans. on Cont. of Net. Systems*, vol. 6, no. 2, pp. 549–560, 2019.

[18] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[19] T. Xue, B. Chen, J. Wu, D. Wei, and W. T. Freeman, "Video enhancement with task-oriented flow," *International Journal of Computer Vision (IJCV)*, vol. 127, no. 8, pp. 1106–1125, 2019.

[20] A. Graves, "Generating sequences with recurrent neural networks," 2014.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

A Operational RDP Region

It is not feasible to compute the region $\mathcal{RDP}^o$ directly since it involves searching over all possible encoding-decoding functions. But, for first-order Markov sources where the Markov chain $X_1 \to X_2 \to X_3$ holds, the following region can be used as an approximation. So, with this motivation, we introduce the information RDP region as follows.

Definition 1 (Information RDP Region) *For first-order Markov sources, let the information RDP region, denoted by $\mathcal{RDP}$, be the set of all tuples $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ which satisfy the following $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ satisfying*

$$
\begin{aligned}
&R_1 \geq I(X_1; X_{r,1}), && (10)\\
&R_2 \geq I(X_2; X_{r,2}|X_{r,1}), && (11)\\
&R_3 \geq I(X_3; X_{r,3}|X_{r,1}, X_{r,2}), && (12)\\
&D_j \geq \mathbb{E}[\|X_j - \hat{X}_j\|^2], \qquad j = 1, 2, 3, && (13)\\
&P_j \geq \phi_j(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j}), \qquad j = 1, 2, 3, && (14)
\end{aligned}
$$

for auxiliary random variables $(X_{r,1}, X_{r,2}, X_{r,3})$ and $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ such that

$$
\begin{aligned}
&\hat{X}_1 = \eta_1(X_{r,1}), \quad \hat{X}_2 = \eta_2(X_{r,1}, X_{r,2}), \quad \hat{X}_3 = X_{r,3}, && (15)\\
&X_{r,1} \to X_1 \to (X_2, X_3), && (16)\\
&X_{r,2} \to (X_2, X_{r,1}) \to (X_1, X_3), && (17)\\
&X_{r,3} \to (X_3, X_{r,1}, X_{r,2}) \to (X_1, X_2), && (18)
\end{aligned}
$$

for some deterministic functions $\eta_1(.)$ and $\eta_2(.,.)$. Moreover, for a given (D, P), the information rate region, denoted by $\mathcal{R}(\mathsf{D}, \mathsf{P})$, is the closure of the set of all tuples R that $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \mathcal{RDP}$.

Proposition 1 *For first-order Markov sources, a given (D, P) and $\mathsf{R} \in \mathcal{R}(\mathsf{D}, \mathsf{P})$, we have*

$$\mathsf{R} + \log(\mathsf{R} + 1) + 5 \in \mathcal{R}^o(\mathsf{D}, \mathsf{P}). \tag{19}$$

Moreover, the following holds:

$$\mathcal{R}^o(\mathsf{D}, \mathsf{P}) \subseteq \mathcal{R}(\mathsf{D}, \mathsf{P}). \tag{20}$$

To prove the above statement, we first discuss the achievable scheme that results in (19). Then, we will provide the proof of outer bound in (20).

Before stating the achievable scheme, we remind the strong functional representation lemma [11]. It states that for jointly distributed random variables X and Y, there exists a random variable U independent of X, and function ϕ such that $Y = \phi(X, U)$. Here, U is not necessarily unique. The strong functional representation lemma states further that there exists a U which has information of Y in the sense that

$$H(Y|U) \leq I(X; Y) + \log(I(X; Y) + 1) + 4. \tag{21}$$

Notice that the strong functional representation lemma can be applied conditionally. Given $P_{XY|W}$, we can represent Y as a function of (X, W, U) such that U is independent of (X, W) and

$$H(Y|W, U) \leq I(X; Y|W) + \log(I(X; Y|W) + 1) + 4. \tag{22}$$

Proof of (19) *(Inner bound)*:

For a given (D, P) and $\mathsf{R} \in \mathcal{R}(\mathsf{D}, \mathsf{P})$, let $\mathsf{X}_r = (X_{r,1}, X_{r,2}, X_{r,3})$ be jointly distributed with $\mathsf{X} = (X_1, X_2, X_3)$ where the Markov chains (16)–(18) hold and the rate constraints in (10)–(12) are satisfied such that there exist $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ for which distortion-perception constraints (13)–(14) hold. Denote the joint distribution of $(\mathsf{X}, \mathsf{X}_r, \hat{\mathsf{X}})$ by $P_{\mathsf{X}\mathsf{X}_r\hat{\mathsf{X}}}$ and notice that according to the Markov chains in (16)–(18), it factorizes as the following

$$
\begin{aligned}
P_{\mathsf{X}\mathsf{X}_r\hat{\mathsf{X}}} = {} & P_{X_1 X_2 X_3} \cdot P_{X_{r,1}|X_1} \cdot P_{X_{r,2}|X_{r,1} X_2} \cdot P_{X_{r,3}|X_{r,2} X_{r,1} X_3} \\
& \cdot \mathbb{1}\{\hat{X}_1 = g_1(X_{r,1})\} \cdot \mathbb{1}\{\hat{X}_2 = g_2(X_{r,1}, X_{r,3})\} \cdot \mathbb{1}\{\hat{X}_3 = X_{r,3}\}.
\end{aligned} \tag{23}
$$

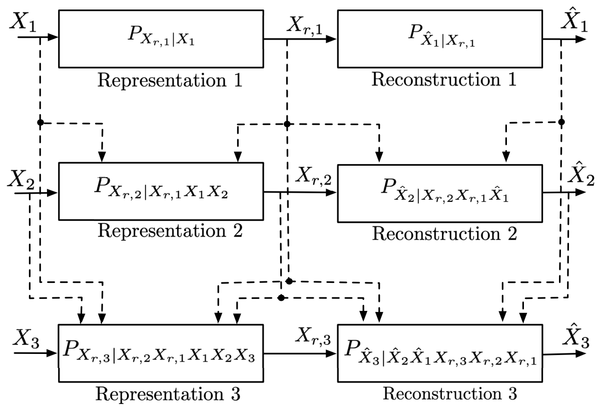


Figure 4: Encoded representations and reconstructions of the iRDP region $\mathcal{RDP}$.

For an illustration of encoded representations X_r and reconstructions $\hat{\mathsf{X}}$ in $\mathcal{R}(\mathsf{D},\mathsf{P})$ which are induced by distribution $P_{\mathsf{X}\mathsf{X}_r\hat{\mathsf{X}}}$, see Fig. 4.

Now, we show that $\mathsf{R} + \log(\mathsf{R}+1) + 5 \in \mathcal{R}(\mathsf{D},\mathsf{P})$. The achievable scheme is as follows. Fix the joint distribution P_{X_r} according to (23) which constructs the codebook, given by

$$P_{\mathsf{X}_r} = P_{X_{r,1}} P_{X_{r,2}|X_{r,1}} P_{X_{r,3}|X_{r,2}X_{r,1}}. \tag{24}$$

From the strong functional representation lemma [11], we know that

- there exist a random variable V_1 independent of X_1 and a deterministic function q_1 such that $X_{r,1} = q_1(X_1, V_1)$ and

$$H(X_{r,1}|V_1) \leq I(X_1; X_{r,1}) + \log(I(X_1; X_{r,1}) + 1) + 4, \tag{25}$$

  which means that the first encoder observes the source X_1 and applies the function q_1 to get $X_{r,1}$ whose distribution needs to be preserved according to (24) (see Fig. 5);

- according to the conditional strong functional representation lemma, there exist a random variable V_2 independent of $(X_2, X_{r,1})$ and a deterministic function q_2 such that $X_{r,2} = q_2(X_{r,1}, X_2, V_2)$ and

$$H(X_{r,2}|X_{r,1}, V_2) \leq I(X_2; X_{r,2}|X_{r,1}) + \log(I(X_2; X_{r,2}|X_{r,1}) + 1) + 4. \tag{26}$$

  At the second step, the representation $X_{r,1}$ is available at the second encoder. So, upon observing the source X_2, it applies the function q_2 to get $X_{r,2}$ whose conditional distribution given $X_{r,1}$ needs to be preserved according to (24) (see Fig. 5);

- according to the conditional strong functional representation lemma, there exist a random variable V_3 independent of $(X_3, X_{r,1}, X_{r,2})$ and a deterministic function q_3 such that $X_{r,3} = q_3(X_{r,1}, X_{r,2}, X_3, V_3)$ and

$$H(X_{r,3}|X_{r,1}, X_{r,2}, V_3) \leq I(X_3; X_{r,3}|X_{r,1}, X_{r,2}) + \log(I(X_3; X_{r,3}|X_{r,1}, X_{r,2}) + 1) + 4. \tag{27}$$

Now, the encoding and decoding are as follows

- With V_1 available at all encoders and decoders, we can have a class of prefix-free binary codes indexed by V_1 with the expected codeword length not larger than $I(X_1; X_{r,1}) + \log(I(X_1; X_{r,1}) + 1) + 5$ to represent $X_{r,1}$, losslessly (see Fig. 5).
- With V_2 available at the encoders and decoders, we can design a set of prefix-free binary codes indexed by $(V_2, X_{r,1})$ with expected codeword length not larger than $I(X_2; X_{r,2}|X_{r,1}) + \log(I(X_2; X_{r,2}|X_{r,1}) + 1) + 5$ to represent $X_{r,2}$, losslessly(see Fig. 5).
- Similarly, one can represent $X_{r,3}$ losslessly with V_3 available at the third encoder and decoder.

- The decoders can use functions $\hat{X}_1 = \eta_1(X_{r,1})$, $\hat{X}_2 = \eta_2(X_{r,1}, X_{r,2})$ and $\hat{X}_3 = X_{r,3}$ to get the reconstruction $\hat{\mathsf{X}}$.

This shows that $\mathsf{R} + \log(\mathsf{R}+1) + 5 \in \mathcal{R}^o(\mathsf{D}, \mathsf{P})$.

Proof of (20) *(Outer Bound)*:

For any (D, P), $\mathsf{R} \in \mathcal{R}^o(\mathsf{D}, \mathsf{P})$, shared randomness K, encoding functions $f_j : \mathcal{X}_1 \times \ldots \times \mathcal{X}_j \times \mathcal{K} \to \mathcal{M}_j$ and decoding functions $g_j : \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_j \times \mathcal{K} \to \hat{\mathcal{X}}_j$ such that

$$R_j \geq \mathbb{E}[\ell(M_j)], \qquad j = 1, 2, 3, \tag{28}$$

and

$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j\|^2], \qquad j = 1, 2, 3, \tag{29}$$
$$P_j \geq \phi_j(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_j}), \qquad j = 1, 2, 3, \tag{30}$$

we lower bound the expected length of the messages. Define

$$X_{r,1} := (M_1, K), \tag{31}$$
$$X_{r,2} := (M_1, M_2, K), \tag{32}$$

and recall that according to the decoding functions, we have

$$\hat{X}_j = g_j(M_1, \ldots, M_j, K), \qquad j = 1, 2, 3. \tag{33}$$

We can write

$$R_1 \geq \mathbb{E}[\ell(M_1)] \geq H(M_1|K) \tag{34}$$
$$= I(X_1; M_1|K) \tag{35}$$
$$= I(X_1; M_1, K) \tag{36}$$
$$= I(X_1; X_{r,1}). \tag{37}$$

Now, consider the following set of inequalities

$$R_2 \geq \mathbb{E}[\ell(M_2)] \geq H(M_2|M_1, K) \tag{38}$$
$$= I(X_1, X_2; M_2|M_1, K) \tag{39}$$
$$= I(X_1, X_2; X_{2,r}|X_{r,1}). \tag{40}$$

Similarly, we have

$$R_3 \geq \mathbb{E}[\ell(M_3)] \geq H(M_3|M_1, M_2, K) \tag{41}$$
$$= I(X_1, X_2, X_3; M_3|M_1, M_2, K) \tag{42}$$
$$\geq I(X_1, X_2, X_3; \hat{X}_3|X_{r,1}, X_{r,2}). \tag{43}$$

Notice that the definitions in (31)–(32) imply the following Markov chains

$$X_{r,1} \to X_1 \to (X_2, X_3), \tag{44}$$
$$X_{r,2} \to (X_1, X_2, X_{r,1}) \to X_3. \tag{45}$$

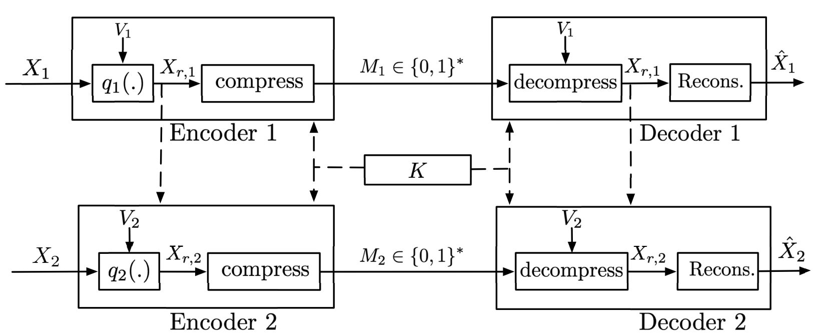


Figure 5: Strong functional representation lemma for $T = 2$ frames.

On the other hand, the decoding functions of the first and second steps imply that

$$\hat{X}_1 = g_1(M_1, K), \tag{46}$$

$$\hat{X}_2 = g_2(M_1, M_2, K), \tag{47}$$

where together with definitions in (31) and (32), we can write

$$\hat{X}_1 = g_1(M_1, K) := \eta_1(X_{r,1}), \tag{48}$$

$$\hat{X}_2 = g_2(M_1, M_2, K) := \eta_2(X_{r,1}, X_{r,2}), \tag{49}$$

such that $\eta_1(.)$ and $\eta_2(.,.)$ are deterministic functions.

Now, consider the fact that the set of constraints in (29)–(30), (37), (40), (43) with Markov chains in (44)–(45) and deterministic functions in (48)–(49) constitute an iRDP region, denoted by $\overline{\mathcal{RDP}}$, which is the set of all tuples $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ such that

$$R_1 \geq I(X_1; X_{r,1}), \tag{50}$$

$$R_2 \geq I(X_1, X_2; X_{r,2}|X_{r,1}), \tag{51}$$

$$R_3 \geq I(X_1, X_2, X_3; \hat{X}_3|X_{r,1}, X_{r,2}), \tag{52}$$

$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j\|^2], \qquad j = 1, 2, 3, \tag{53}$$

$$P_j \geq \phi_j(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_j}), \qquad j = 1, 2, 3, \tag{54}$$

for auxiliary random variables $(X_{r,1}, X_{r,2})$ and $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ satisfying the following

$$\hat{X}_1 = \eta_1(X_{r,1}), \quad \hat{X}_2 = \eta_2(X_{r,1}, X_{r,2}) \tag{55}$$

$$X_{r,1} \to X_1 \to (X_2, X_3), \tag{56}$$

$$X_{r,2} \to (X_1, X_2, X_{r,1}) \to X_3. \tag{57}$$

for some deterministic functions $\eta_1(.)$ and $\eta_2(.,.)$.

Comparing the two regions $\overline{\mathcal{RDP}}$ and $\mathcal{RDP}$, we identify the following differences. The Markov chain in (16) is more restricted comparing to (57). Moreover, the Markov chain (17) does not exist in $\overline{\mathcal{RDP}}$. The following lemma states that $\overline{\mathcal{RDP}} = \mathcal{RDP}$. Now, for a given (D, P), let $\overline{\mathcal{R}}(\mathsf{D}, \mathsf{P})$ denote the set of rate tuples R such $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \overline{\mathcal{RDP}}$, then this lemma implies that $\overline{\mathcal{R}}(\mathsf{D}, \mathsf{P}) = \mathcal{R}(\mathsf{D}, \mathsf{P})$ which completes the proof of the outer bound.

We conclude this section by the following lemma.

Lemma 1 *For first-order Markov sources, we have*

$$\mathcal{RDP} = \overline{\mathcal{RDP}}. \tag{58}$$

Proof: This result for the scenario without perception constraint has been similarly observed in [16, Eq. (12)]. The proof in this section is provided for completeness.

First, notice that the set of Markov chains in (16)–(18) is more restricted than the ones in (56)–(57), hence $\mathcal{RDP} \subseteq \overline{\mathcal{RDP}}$. Now, it remains to prove that $\overline{\mathcal{RDP}} \subseteq \mathcal{RDP}$. Consider the following facts

1. The distortion constraints in (53) depend only on the joint distribution of $(X_j, \hat{X}_j)$, and thus on the joint distribution of $(X_j, X_{r,1}, \ldots, X_{r,j})$. So, imposing the Markov chain $X_{r,2} \to (X_2, X_{r,1}) \to X_1$ does not affect the expected distortion $\mathbb{E}[\|X_2 - \hat{X}_2\|^2]$ since it does not depend on the joint distribution of X_1 with $(X_{r,1}, X_{r,2}, X_2)$. A similar argument holds for other frames;
2. The perception constraints in (54) depend on the joint distributions $P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}$ and $P_{\hat{X}_1, \ldots, \hat{X}_j}$ (hence on $P_{X_{r,1} \ldots X_{r,j}}$). Thus, imposing $X_{r,2} \to (X_2, X_{r,1}) \to X_1$ does not affect $\phi_2(P_{\hat{X}_1 X_2}, P_{\hat{X}_1 \hat{X}_2})$ since it does not depend on the joint distribution of X_1 with $(X_{r,1}, X_{r,2}, X_2)$. A similar argument holds for other frames;
3. Moreover, the rate constraints in (51) and (52) would be further lower bounded by

$$R_2 \geq I(X_1, X_2; X_{r,2}|X_{r,1}) \geq I(X_2; X_{r,2}|X_{r,1}), \tag{59}$$

$$R_3 \geq I(X_1, X_2, X_3; \hat{X}_3|X_{r,1}, X_{r,2}) \geq I(X_3; \hat{X}_3|X_{r,1}, X_{r,2}). \tag{60}$$

Thus, the set of rate constraints is optimized by the set of Markov chains (16)–(18).

4. The mutual information terms $I(X_1; X_{r,1})$, $I(X_2; X_{r,2}|X_{r,1})$ and $I(X_3; \hat{X}_3|X_{r,1}, X_{r,2})$ depend on distributions $P_{X_1 X_{r,1}}$, $P_{X_{r,1} X_{r,2} X_2}$ and $P_{X_3 \hat{X}_3 X_{r,1} X_{r,2}}$, respectively. So, these distributions should be preserved by the set of Markov chains. The first two distributions are preserved by the choice of (15)–(16). Now, since we have first-order Markov sources, preserving the joint distributions of $P_{X_{r,1} X_1}$ and $P_{X_{r,1} X_{r,2} X_2}$ is sufficient to preserve the distribution $P_{X_{r,1} X_{r,2} X_3}$. So, preserving the joint distribution of $P_{\hat{X}_3 X_{r,1} X_{r,2}}$ is sufficient to keep $I(X_3; \hat{X}_3|X_{r,1}, X_{r,2})$ unchanged.

Considering the above four facts, without loss of optimality, one can impose the following Markov chains

$$X_{r,1} \to X_1 \to (X_2, X_3), \tag{61}$$
$$X_{r,2} \to (X_2, X_{r,1}) \to (X_1, X_3), \tag{62}$$
$$\hat{X}_3 \to (X_3, X_{r,1}, X_{r,2}) \to (X_1, X_2). \tag{63}$$

This concludes the proof of the lemma.

■

B Gauss-Markov Source Model

In this section, we prove that for Gaussian sources, jointly Gaussian reconstructions are optimal.

Proposition 2 *For the Gauss-Markov source model, any tuple* $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \mathcal{RDP}$ *can be attained by a jointly Gaussian distribution over* $(X_{r,1}, X_{r,2}, X_{r,3})$ *and identity mappings for* $\eta_j(\cdot)$ *in Definition 1.*

First, notice that a proof for the setting without perception constraint is provided in [17]. The following proof is different from [17] in some steps and also involves the perception constraint.

For a given tuple $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \mathcal{RDP}$, let $X^*_{r,1}$, $X^*_{r,2}$, $\hat{X}^*_1 = \eta_1(X^*_{r,1})$, $\hat{X}^*_2 = \eta_2(X^*_{r,1}, X^*_{r,2})$ and $\hat{X}^*_3$ be random variables satisfying (15)–(17). Let $P_{\hat{X}^G_1|X_1}$, $P_{\hat{X}^G_2|\hat{X}^G_1 X_2}$ and $P_{\hat{X}^G_3|\hat{X}^G_1 \hat{X}^G_2 X_3}$ be jointly Gaussian distributions such that the following conditions are satisfied.

$$\text{cov}(\hat{X}^G_1, X_1) = \text{cov}(\hat{X}^*_1, X_1), \tag{64}$$
$$\text{cov}(\hat{X}^G_1, \hat{X}^G_2, X_2) = \text{cov}(\hat{X}^*_1, \hat{X}^*_2, X_2), \tag{65}$$
$$\text{cov}(\hat{X}^G_1, \hat{X}^G_2, \hat{X}^G_3, X_3) = \text{cov}(\hat{X}^*_1, \hat{X}^*_2, \hat{X}^*_3, X_3), \tag{66}$$

In general, the Gaussian random variables which satisfy the constraints in (64)–(66) can be written in the following format

$$X_1 = \nu \hat{X}^G_1 + Z_1, \tag{67}$$
$$\hat{X}^G_2 = \omega_1 \hat{X}^G_1 + \omega_2 X_2 + Z_2, \tag{68}$$
$$\hat{X}^G_3 = \tau_1 \hat{X}^G_1 + \tau_2 \hat{X}^G_2 + \tau_3 X_3 + Z_3, \tag{69}$$

for some real $\nu, \omega_1, \omega_2, \tau_1, \tau_2, \tau_3$ where $\hat{X}^G_1 \sim \mathcal{N}(0, \sigma^2_{\hat{X}^G_1})$, $\hat{X}^G_2 \sim \mathcal{N}(0, \sigma^2_{\hat{X}^G_2})$, Z_1, Z_2 and Z_3 are Gaussian random variables with zero mean and variances $\alpha_1^2, \alpha_2^2, \alpha_3^2$, independent of $\hat{X}^G_1$, $(\hat{X}^G_1, X_2)$ and $(\hat{X}^G_1, \hat{X}^G_2, X_3)$, respectively.

We explicitly derive the coefficients ν, ω_1, ω_2, τ_1, τ_2 and τ_3 in the following. Multiplying both sides of (67) by $\hat{X}^G_1$ and taking an expectation, we get

$$\mathbb{E}[X_1 \hat{X}^G_1] = \nu \sigma^2_{\hat{X}^G_1}. \tag{70}$$

According to (64), the above equation can be written as follows

$$\mathbb{E}[X_1 \hat{X}^*_1] = \nu \mathbb{E}[\hat{X}^{*2}_1]. \tag{71}$$

Multiplying both sides of (68) by the vector $[\hat{X}^G_1 \;\; X_2]$ and taking an expectation, we have

$$[\mathbb{E}[\hat{X}^G_1 \hat{X}^G_2] \;\; \mathbb{E}[X_2 \hat{X}^G_2]] = [\omega_1 \;\; \omega_2] \begin{pmatrix} \sigma^2_{\hat{X}^G_1} & \mathbb{E}[X_2 \hat{X}^G_1] \\ \mathbb{E}[X_2 \hat{X}^G_1] & \sigma^2_2 \end{pmatrix} \tag{72}$$

Considering the fact that $\mathbb{E}[X_2\hat{X}_1^G] = \rho_1\mathbb{E}[X_1\hat{X}_1^G]$ and according to (65), the above equation can be written as follows

$$[\mathbb{E}[\hat{X}_1^*\hat{X}_2^*] \;\; \mathbb{E}[X_2\hat{X}_2^*]] = [\omega_1 \;\; \omega_2] \begin{pmatrix} \mathbb{E}[\hat{X}_1^{*2}] & \rho_1\mathbb{E}[X_1\hat{X}_1^*] \\ \rho_1\mathbb{E}[X_1\hat{X}_1^*] & \sigma_2^2 \end{pmatrix}. \tag{73}$$

Similarly, multiplying both sides of (69) by the vector $[\hat{X}_1^G \;\; \hat{X}_2^G \;\; X_3]$, taking an expectation and considering (66), we get

$$[\mathbb{E}[\hat{X}_1^*\hat{X}_3^*] \;\; \mathbb{E}[\hat{X}_2^*\hat{X}_3^*] \;\; \mathbb{E}[X_3\hat{X}_3^*]] = [\tau_1 \;\; \tau_2 \;\; \tau_3] \begin{pmatrix} \mathbb{E}[\hat{X}_1^{*2}] & \mathbb{E}[\hat{X}_1^*\hat{X}_2^*] & \rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] \\ \mathbb{E}[\hat{X}_1^*\hat{X}_2^*] & \mathbb{E}[\hat{X}_2^{*2}] & \rho_2\mathbb{E}[X_2\hat{X}_2^*] \\ \rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] & \rho_2\mathbb{E}[X_2\hat{X}_2^*] & \mathbb{E}[\hat{X}_3^{*2}] \end{pmatrix}. \tag{74}$$

Solving equations (71), (73) and (74), we get

$$\sigma^2_{\hat{X}_1^G} = \mathbb{E}[\hat{X}_1^{*2}], \tag{75}$$

$$\nu = \frac{\mathbb{E}[X_1\hat{X}_1^*]}{\mathbb{E}[\hat{X}_1^{*2}]}, \tag{76}$$

$$\alpha_1^2 = \sigma_1^2 - \frac{\mathbb{E}[X_1\hat{X}_1^*]}{\mathbb{E}[\hat{X}_1^{*2}]}, \tag{77}$$

$$\omega_1 = \frac{\nu\rho_1\mathbb{E}[\hat{X}_1^*\hat{X}_2^*] - \mathbb{E}[X_2\hat{X}_2^*]}{\nu^2\rho_1^2\sigma^2_{\hat{X}_1^G} - \sigma_2^2}, \tag{78}$$

$$\omega_2 = \frac{\nu\rho_1\sigma^2_{\hat{X}_1^G}\mathbb{E}[X_2\hat{X}_2^*] - \sigma_2^2\mathbb{E}[\hat{X}_1^*\hat{X}_2^*]}{\nu^2\rho_1^2\sigma^4_{\hat{X}_1^G} - \sigma_2^2\sigma^2_{\hat{X}_1^G}}, \tag{79}$$

$$\alpha_2^2 = \mathbb{E}[\hat{X}_2^{*2}] - \alpha_2^2\sigma^2_{\hat{X}_1^G} - \omega_2^2\sigma_2^2 - 2\omega_1\omega_2\rho_1\nu\sigma^2_{\hat{X}_1^G}. \tag{80}$$

For the third step, the coefficients and noise variance of (69) are given as follows

$$\begin{aligned} &[\tau_1 \;\; \tau_2 \;\; \tau_3] \\ &\quad = [\mathbb{E}[\hat{X}_1^*\hat{X}_3^*] \;\; \mathbb{E}[\hat{X}_2^*\hat{X}_3^*] \;\; \mathbb{E}[X_3\hat{X}_3^*]] \begin{pmatrix} \mathbb{E}[\hat{X}_1^{*2}] & \mathbb{E}[\hat{X}_1^*\hat{X}_2^*] & \rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] \\ \mathbb{E}[\hat{X}_1^*\hat{X}_2^*] & \mathbb{E}[\hat{X}_2^{*2}] & \rho_2\mathbb{E}[X_2\hat{X}_2^*] \\ \rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] & \rho_2\mathbb{E}[X_2\hat{X}_2^*] & \mathbb{E}[\hat{X}_3^{*2}] \end{pmatrix}^{-1}, \end{aligned} \tag{81}$$

$$\begin{aligned} \alpha_3^2 = \mathbb{E}[\hat{X}_3^{*2}] - \tau_1^2\mathbb{E}[\hat{X}_1^{*2}] - \tau_2^2\mathbb{E}[\hat{X}_2^{*2}] - \tau_3^2\mathbb{E}[X_3^2] \\ -2\tau_1\tau_2\mathbb{E}[\hat{X}_1^*\hat{X}_2^*] - 2\tau_1\tau_3\rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] - 2\tau_2\tau_3\rho_2\mathbb{E}[X_2\hat{X}_2^*], \end{aligned} \tag{82}$$

where $(.)^{-1}$ denotes the inverse of a matrix.

Now, we look at the rate constraints.

Rate Constraints:

Consider the rate constraint of the first step as follows

$$\begin{aligned} R_1 &\geq I(X_1; X^*_{r,1}) && (83) \\ &= H(X_1) - H(X_1|X^*_{r,1}) && (84) \\ &\geq H(X_1) - H(X_1|\hat{X}_1^*) && (85) \\ &= H(X_1) - H(X_1 - \mathbb{E}[X_1|\hat{X}_1^*]|\hat{X}_1^*) && (86) \\ &\geq H(X_1) - H(X_1 - \mathbb{E}[X_1|\hat{X}_1^*]) && (87) \\ &\geq H(X_1) - H(X_1 - \mathbb{E}[X_1|\hat{X}_1^G]) && (88) \\ &= H(X_1) - H(X_1 - \mathbb{E}[X_1|\hat{X}_1^G]|\hat{X}_1^G) && (89) \\ &= I(X_1; \hat{X}_1^G) && (90) \end{aligned}$$

where

- (85) follows because $\hat{X}_1^*$ is a function of $X_{r,1}^*$;
- (88) follows because for a given covariance matrix in (64), the Gaussian distribution maximizes the differential entropy;
- (89) follows because the MMSE is uncorrelated from the data and since the random variables are Gaussian, the MMSE would be independent of the data.

Next, consider the rate constraint of the second step as the following

$$
\begin{aligned}
R_2 &\geq I(X_2; X_{r,2}^* | X_{r,1}^*) && (91)\\
&= H(X_2|X_{r,1}^*) - H(X_2|X_{r,1}^*, X_{r,2}^*) && (92)\\
&\geq H(X_2|X_{r,1}^*) - H(X_2|\hat{X}_1^*, \hat{X}_2^*) && (93)\\
&\geq H(X_2|X_{r,1}^*) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (94)\\
&= H(\rho_1 X_1 + N_1|X_{r,1}^*) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (95)\\
&\geq \frac{1}{2}\log\left(\rho_1^2 2^{2H(X_1|X_{r,1}^*)} + 2^{2H(N_1)}\right) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (96)\\
&\geq \frac{1}{2}\log\left(\rho_1^2 2^{-2R_1} 2^{2H(X_1)} + 2^{2H(N_1)}\right) - H(X_2|\hat{X}_1^G, \hat{X}_2^G), && (97)
\end{aligned}
$$

where

- (93) follows because $\hat{X}_1^*$ and $\hat{X}_2^*$ are deterministic functions of $X_{r,1}^*$ and $(X_{r,1}^*, X_{r,2}^*)$, respectively;
- (94) follows because for a given covariance matrix in (65), the Gaussian distribution maximizes the differential entropy;
- (96) follows from entropy power inequality (EPI) [18, pp. 22];
- (97) follows from (84).

Similarly, consider the rate constraint of the third frame as the following,

$$
\begin{aligned}
R_3 &\geq I(X_3; \hat{X}_3^* | X_{r,1}^*, X_{r,2}^*) && (98)\\
&= H(X_3|X_{r,1}^*, X_{r,2}^*) - H(X_3|X_{r,1}^*, X_{r,2}^*, \hat{X}_3^*) && (99)\\
&\geq H(X_3|X_{r,1}^*, X_{r,2}^*) - H(X_3|\hat{X}_1^*, \hat{X}_2^*, \hat{X}_3^*) && (100)\\
&\geq H(X_3|X_{r,1}^*, X_{r,2}^*) - H(X_3|\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) && (101)\\
&= H(\rho_2 X_2 + N_2|X_{r,1}^*, X_{r,2}^*) - H(X_3|\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) && (102)\\
&\geq \frac{1}{2}\log\left(\rho_2^2 2^{2H(X_2|X_{r,1}^*, X_{r,2}^*)} + 2^{2H(N_2)}\right) - H(X_3|\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) && (103)\\
&\geq \frac{1}{2}\log\left(\rho_2^2 2^{-2R_2} 2^{2H(X_2|X_{r,1}^*)} + 2^{2H(N_2)}\right) - H(X_3|\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) && (104)\\
&\geq \frac{1}{2}\log\left(\rho_1^2\rho_2^2 2^{-2R_1-2R_2} 2^{2H(X_1)} + \rho_2^2 2^{-2R_2} 2^{2H(N_1)} + 2^{2H(N_2)}\right) - H(X_3|\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) && (105)
\end{aligned}
$$

Next, we look at the distortion constraint.

<u>*Distortion Constraint*</u>: The choices in (64)–(66) imply that

$$
D_j \geq \mathbb{E}[\|X_j - \hat{X}_j^*\|^2] = \mathbb{E}[\|X_j - \hat{X}_j^G\|^2], \qquad j = 1, 2, 3. \tag{106}
$$

Finally, we look at the perception constraint

<u>*Perception Constraint*</u>:

Define the following distribution

$$
P_{U^*V^*} := \arg \inf_{\substack{\tilde{P}_{UV}: \\ \tilde{P}_U = P_{X_1} \\ \tilde{P}_V = P_{\hat{X}_1^*}}} \mathbb{E}_{\tilde{P}}[\|U - V\|^2]. \tag{107}
$$

Now, define $P_{U^G V^G}$ to be a Gaussian joint distribution with the following covariance matrix

$$\text{cov}(U^G, V^G) = \text{cov}(U^*, V^*). \tag{108}$$

Then, we have the following set of inequalities:

$$\begin{aligned}
P_1 \geq W_2^2(P_{X_1}, P_{\hat{X}_1^*}) &= \inf_{\substack{\tilde{P}_{UV}:\\ \tilde{P}_U = P_{X_1}\\ \tilde{P}_V = P_{\hat{X}_1^*}}} \mathbb{E}_{\tilde{P}}[\|U - V\|^2] && (109)\\
&= \mathbb{E}[\|U^* - V^*\|^2] && (110)\\
&= \mathbb{E}[\|U^G - V^G\|^2] && (111)\\
&\geq W_2^2(P_{U^G}, P_{V^G}) && (112)\\
&= \inf_{\substack{\hat{P}_{UV}:\\ \hat{P}_U = P_{U^G}\\ \hat{P}_V = P_{V^G}}} \mathbb{E}_{\hat{P}}[\|U - V\|^2] && (113)\\
&= \inf_{\substack{\hat{P}_{UV}:\\ \hat{P}_U = P_{X_1}\\ \hat{P}_V = P_{\hat{X}_1^G}}} \mathbb{E}_{\hat{P}}[\|U - V\|^2] && (114)\\
&= W_2^2(P_{X_1}, P_{\hat{X}_1^G}), && (115)
\end{aligned}$$

where

- (110) follows from the definition in (107);
- (111) follows from (108) which implies that (U^*, V^*) and (U^G, V^G) have the same second-order statistics;
- (114) follows because $P_{V^G} = P_{\hat{X}_1^G}$ which is justified in the following. First, notice that both P_{V^G} and $P_{\hat{X}_1^G}$ are Gaussian distributions. Denote the variance of V^G by $\sigma_{V^G}^2$ and recall that the variance of $\hat{X}_1^G$ is denoted by $\sigma_{\hat{X}_1^G}^2$. According to (108), $\sigma_{V^G}^2$ is equal to the variance of V^*. Also, from (107), we know that $P_{V^*} = P_{\hat{X}_1^*}$, hence the variances of V^* and $\hat{X}_1^*$ are the same. On the other side, according to (64), we know that the variance of $\hat{X}_1^*$ is equal to $\sigma_{\hat{X}_1^G}^2$. Thus, we conclude that $\sigma_{\hat{X}_1^G}^2 = \sigma_{V^G}^2$, which yields $P_{V^G} = P_{\hat{X}_1^G}$. A similar argument shows that $P_{U^G} = P_{X_1}$.

A similar argument holds for the perception constraint of the second and third steps for both PLFs.

Thus, we have proved the set of Gaussian auxiliary random variables $(\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G)$ given in (67)–(69) where the coefficients are chosen according to distortion-perception constraints, provides an outer bound to $\mathcal{RDP}$ which is the set of all tuples $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ such that

$$R_1 \geq I(X_1; \hat{X}_1^G), \tag{116}$$

$$R_2 \geq \frac{1}{2} \log \left(\rho_1^2 2^{-2R_1} 2^{2H(X_1)} + 2^{2H(N_1)} \right) - H(X_2 | \hat{X}_1^G, \hat{X}_2^G), \tag{117}$$

$$R_3 \geq \frac{1}{2} \log \left(\rho_1^2 \rho_2^2 2^{-2R_1 - 2R_2} 2^{2H(X_1)} + \rho_2^2 2^{-2R_2} 2^{2H(N_1)} + 2^{2H(N_2)} \right) - H(X_3 | \hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G), \tag{118}$$

$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j^G\|^2], \qquad j = 1, 2, 3 \tag{119}$$

$$P_j \geq W_2^2(P_{X_1 \ldots X_j}, P_{\hat{X}_1^G \ldots \hat{X}_j^G}). \tag{120}$$

Now, we need to show that the above RDP region is also an inner bound to $\mathcal{RDP}$. This is simply verified by the following choice. In iRDP region of (10)–(18), choose the following:

$$X_{r,j} = \hat{X}_j = \hat{X}_j^G, \qquad j = 1, 2, 3, \tag{121}$$

where $(\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G)$ satisfy (67)–(69) with coefficients chosen according to distortion-perception constraints. The lower bounds on distortion and perception constraints in (119) and (120) are immediately achieved by this choice. Now, we will look at the rate constraints. The achievable rate constraint of the first step can be written as follows

$$R_1 \geq I(X_1; \hat{X}_1^G), \tag{122}$$

which immediately coincides with (116). The achievable rate of the second step can be written as follows

$$\begin{aligned}
R_2 &\geq I(X_2; \hat{X}_2^G | \hat{X}_1^G) && (123)\\
&= H(X_2|\hat{X}_1^G) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (124)\\
&= H(\rho_1 X_1 + N_1 | \hat{X}_1^G) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (125)\\
&= \frac{1}{2}\log(\rho_1^2 2^{2H(X_1|\hat{X}_1^G)} + 2^{2H(N_1)}) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (126)\\
&\geq \frac{1}{2}\log\left(\rho_1^2 2^{-2R_1} 2^{2H(X_1)} + 2^{2H(N_1)}\right) - H(X_2|\hat{X}_1^G, \hat{X}_2^G), && (127)
\end{aligned}$$

where

- (126) follows because EPI holds with "equality" for jointly Gaussian distributions [18, pp. 22];
- (127) follows from (117).

Thus, the bound in (127) coincides with (97). A similar argument holds for the achievable rate of the third frame.

Notice that the above proof (both converse and achievability) can be extended to T frames using the sequential analysis that was presented. Thus, without loss of optimality, one can restrict to the jointly Gaussian distributions and identity functions $\eta_1(.)$ and $\eta_2(.,.)$ in iRDP region $\mathcal{RDP}$.

C Low-rate Regime for the First Frame

In this section, we prove the following theorem when the first frame is compressed at a low rate. The rate of the second frame is an arbitrary nonnegative value.

Theorem 1 *Let $R_1 = \epsilon$ for a sufficiently small $\epsilon > 0$ and R_2 be an arbitrary nonnegative rate. The achievabale distortions for the second frame, $D^0_{2,AR}$ (for 0-PLF-AR), $D^0_{2,FMD}$ (for 0-PLF-FMD) and $D^0_{2,JD}$ (for 0-PLF-JD) are given by*

$$D^0_{2,AR} = 2\sigma^2(1 - \sqrt{1 - 2^{-2R_2}}), \qquad D^0_{2,FMD} = 2\sigma^2(1 - \sqrt{1 - 2^{-2R_2} + \rho^2 2\epsilon \ln 2}),$$
$$D^0_{2,JD} = 2\sigma^2(1 - \sqrt{1-\rho^2}\sqrt{1 - 2^{-2R_2}} - \rho^2\sqrt{2\epsilon \ln 2}). \tag{128}$$

To prove the above theorem, we first remind the RDP region of the Gauss-Markov source model. Then, we will look at each PLF separately; 0-PLF-AR, 0-PLF-FMD, and 0-PLF-JD. For each of these PLFs, we discuss the second step and provide the analysis of the third step for completeness.

Recall the RDP region of the Gauss-Markov model which is the set of all tuples $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ such that

$$R_1 \geq I(X_1; \hat{X}_1), \;\; R_2 \geq I(X_2; \hat{X}_2 | \hat{X}_1), \;\; R_3 \geq I(X_3; \hat{X}_3 | \hat{X}_1, \hat{X}_2) \tag{129a}$$
$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j\|^2], \quad P_j \geq \phi_j(P_{\hat{X}_1 \ldots \hat{X}_{j-1} X_j}, P_{\hat{X}_1 \ldots \hat{X}_{j-1} \hat{X}_j}), \qquad j = 1,2,3, \tag{129b}$$

for some auxiliary random variables $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ which satisfy the following Markov chains

$$\hat{X}_1 \to X_1 \to (X_2, X_3), \; \hat{X}_2 \to (X_2, \hat{X}_1) \to (X_1, X_3), \; \hat{X}_3 \to (X_3, \hat{X}_1, \hat{X}_2) \to (X_1, X_2). \tag{130}$$

For the Gauss-Markov source model, the reconstructions that satisfy the Markov chains in (130) can be generally written as follows

$$\hat{X}_1 = \nu X_1 + Z_1, \tag{131}$$
$$\hat{X}_2 = \omega_1 \hat{X}_1 + \omega_2 X_2 + Z_2, \tag{132}$$
$$\hat{X}_3 = \tau_1 \hat{X}_1 + \tau_2 \hat{X}_2 + \tau_3 X_3 + Z_3, \tag{133}$$

where $\hat{X}_j \sim \mathcal{N}(0,\hat{\sigma}_j^2)$ for $j = 1,2$, Z_1, Z_2 and Z_3 are independent of X_1, $(\hat{X}_1, X_2)$ and $(\hat{X}_1,\hat{X}_2,X_3)$, respectively.

According to (129), the optimization program of the first step is as follows

$$\begin{aligned}\min_{P_{\hat{X}_1|X_1}} \quad & \mathbb{E}[\|X_1-\hat{X}_1\|^2]\\ \text{s.t.} \quad & I(X_1;\hat{X}_1)\leq R_1,\\ & \phi_1(P_{X_1},P_{\hat{X}_1})\leq P_1.\end{aligned} \tag{134}$$

Using the choice in (131), the optimization program of the first step for $P_1=0$ simplifies as follows

$$\min_{\nu} \; 2\sigma^2(1-\nu), \tag{135a}$$

$$\text{s.t.} \qquad \nu^2 \leq (1-2^{-2R_1}), \tag{135b}$$

When $R_1=\epsilon$ for a sufficiently small $\epsilon>0$, the solution of the above program is as follows

$$D_1^0 = 2\sigma^2(1-\sqrt{2\epsilon\ln 2})+O(\epsilon), \tag{136}$$

where the optimal choice of ν is given by

$$\nu=\sqrt{1-2^{-2R_1}}=\sqrt{2\epsilon\ln 2}+O(\epsilon). \tag{137}$$

Next, consider the optimization programs for different steps and PLFs as follows.

C.1 0-PLF-AR

In this section, we provide the optimization programs for different steps of 0-PLF-AR. For the second step, we are able to provide an approximate solution for the low compression rate, i.e., $R_1=\epsilon$. For the third step, we plot the tradeoff in Fig. 6.

Second Step:

The optimization program of the second step is given as follows.

Proposition 3 *The optimization program of* 0*-PLF-AR for the second step can be written as*

$$\min_{\omega_1,\omega_2} \quad 2\sigma^2-2\omega_1\rho\nu\sigma^2-2\omega_2\sigma^2, \tag{138a}$$

$$\textit{s.t.} \qquad \omega_2^2(1-\rho^2\nu^2 2^{-2R_2}) \leq (1-\omega_1^2-2\omega_1\omega_2\rho\nu)(1-2^{-2R_2}), \tag{138b}$$

$$\omega_1+\nu\omega_2\rho=\rho\nu, \tag{138c}$$

$$\nu=\sqrt{1-2^{-2R_1}}. \tag{138d}$$

Proof: According to (129), the optimization problem of the second step is as follows,

$$\begin{aligned}\min_{P_{\hat{X}_2|X_2\hat{X}_1}} \quad & \mathbb{E}[\|X_2-\hat{X}_2\|^2]\\ \text{s.t.} \quad & I(X_2;\hat{X}_2|\hat{X}_1)\leq R_2,\\ & P_{\hat{X}_1X_2}=P_{\hat{X}_1\hat{X}_2}.\end{aligned} \tag{139}$$

We proceed with simplifying the rate constraint as follows,

$$\begin{aligned}
R_2 &\geq I(X_2;\hat{X}_2|\hat{X}_1) && (140)\\
&= h(\hat{X}_2|\hat{X}_1)-h(Z_2) && (141)\\
&= h(\omega_2X_2+Z_2|\hat{X}_1)-h(Z_2) && (142)\\
&= \frac{1}{2}\log 2^{-2h(Z_2)}\left(\omega_2^2 2^{2h(X_2|\hat{X}_1)}+2^{2h(Z_2)}\right) && (143)\\
&= \frac{1}{2}\log 2^{-2h(Z_2)}\left(\omega_2^2 2^{2h(\rho X_1+N_1|\hat{X}_1)}+2^{2h(Z_2)}\right) && (144)\\
&= \frac{1}{2}\log 2^{-2h(Z_2)}\left(\omega_2^2(\rho^2 2^{2h(X_1|\hat{X}_1)}+2^{2h(N_1)})+2^{2h(Z_2)}\right) && (145)\\
&= \frac{1}{2}\log 2^{-2h(Z_2)}\left(\omega_2^2(\rho^2 2^{2h(X_1|\hat{X}_1)}+(1-\rho^2)\sigma^2)+2^{2h(Z_2)}\right) && (146)\\
&\geq \frac{1}{2}\log 2^{-2h(Z_2)}\left(\omega_2^2(\rho^2\sigma^2 2^{-2R_1}+(1-\rho^2)\sigma^2)+2^{2h(Z_2)}\right), && (147)
\end{aligned}$$

where

- (141) and (142) follow from (132);
- (143) and (145) follow because Entropy Power Inequality (EPI) [18, pp. 22] holds with equality for Gaussian sources;
- (144) follows from (7) where $X_2 = \rho X_1 + N_1$;
- (147) follows from the rate constraint of the first step, i.e., $R_1 \geq I(X_1; \hat{X}_1)$.

Inequality (147) can be further simplified as follows,

$$(\omega_2^2(\rho^2\sigma^2 2^{-2R_1} + (1-\rho^2)\sigma^2))2^{-2R_2} \geq (1-2^{-2R_2})2^{2h(Z_2)} \tag{148}$$

$$= (1-2^{-2R_2}) \cdot (1-\omega_1^2-\omega_2^2-2\omega_1\omega_2\nu\rho)\sigma^2. \tag{149}$$

Considering that $\nu = \sqrt{1-2^{-2R_1}}$ and re-arranging the terms in the above inequality, we get to constraint in (138b).

The objective function in (138a) can be obtained as follows,

$$\mathbb{E}[\|X_2 - \hat{X}_2\|^2] = 2\sigma^2 - 2\mathbb{E}[X_2\hat{X}_2] \tag{150}$$

$$= 2\sigma^2 - 2(\rho\nu\omega_1 + \omega_2)\sigma^2, \tag{151}$$

where the last equality follows from (131) and (132).

The derivation of the constraint in (138c) is as follows. We multiply both sides of (131) and (132) by X_2 and $\hat{X}_1$, respectively, and take an expectation from both sides. Thus, we have

$$\mathbb{E}[X_2\hat{X}_1] = \nu\mathbb{E}[X_1X_2] = \nu\rho\sigma^2, \tag{152}$$

$$\mathbb{E}[\hat{X}_1\hat{X}_2] = \omega_1\sigma^2 + \omega_2\mathbb{E}[X_2\hat{X}_1]. \tag{153}$$

Notice that the perception constraint $P_{X_2\hat{X}_1} = P_{\hat{X}_2\hat{X}_1}$ implies that $\mathbb{E}[\hat{X}_1\hat{X}_2] = \mathbb{E}[X_2\hat{X}_1]$ which together with (152) and (153) yields the constraint in (138c). ■

Now, we provide an approximate solution for the optimization program when the first frame is compressed at a low rate, i.e., $R_1 = \epsilon$ where ϵ is sufficiently small. In this case, we have

$$1 - 2^{-2R_1} = 2\epsilon\ln 2 + O(\epsilon^2), \tag{154}$$

$$\nu = \sqrt{2\epsilon\ln 2} + O(\epsilon), \tag{155}$$

so the optimization program of the second step in (138) simplifies as follows

$$\min_{\omega_1,\omega_2} \quad 2\sigma^2 - 2\omega_1\rho\sigma^2\sqrt{2\epsilon\ln 2 + O(\epsilon^2)} - 2\omega_2\sigma^2, \tag{156a}$$

$$\text{s.t.} \quad \omega_2^2(1-\rho^2 2^{-2R_2}(2\epsilon\ln 2 + O(\epsilon^2))) \leq (1-\omega_1^2 - 2\omega_1\omega_2\rho(2\epsilon\ln 2 + O(\epsilon^2)))(1-2^{-2R_2}), \tag{156b}$$

$$\omega_1 + \nu\omega_2\rho = \rho\nu. \tag{156c}$$

Notice that (156c) and (155) imply that $\omega_1 = \Theta(\sqrt{\epsilon})$ which together with (156b) yields the following

$$\omega_2 \leq \sqrt{1-2^{-2R_2}} + O(\sqrt{\epsilon}). \tag{157}$$

On the other side, plugging (156c) into (156a), the program in (156) is upper bounded by the following

$$\min_{\omega_2} \quad 2\sigma^2 - 2\omega_2\sigma^2 + O(\sqrt{\epsilon}) \tag{158}$$

$$\text{s.t.} \quad \omega_2 \leq \sqrt{1-2^{-2R_2}} + O(\sqrt{\epsilon}). \tag{159}$$

The solution of the above program is given by

$$\omega_2 = \sqrt{1-2^{-2R_2}} + O(\sqrt{\epsilon}). \tag{160}$$

Plugging the above into (156c), we get

$$\omega_1 = \rho\sqrt{2\epsilon\ln 2}(1-\sqrt{1-2^{-2R_2}}) + O(\epsilon). \tag{161}$$

Thus, we have

$$\hat{X}_2 = \rho\sqrt{2\epsilon\ln 2}(1-\sqrt{1-2^{-2R_2}})\hat{X}_1 + \sqrt{1-2^{-2R_2}}X_2 + Z_2, \tag{162}$$

where $Z_2 \sim \mathcal{N}(0, (2^{-2R_2} - \rho^2(1-\sqrt{1-2^{-2R_2}})^2(2\epsilon \ln 2))\sigma^2)$ and the solution of optimization program is as follows

$$D^0_{2,\mathrm{AR}} := 2\sigma^2(1-\sqrt{1-2^{-2R_2}}) + O(\sqrt{\epsilon}). \tag{163}$$

Third Step:

For the third step, we have the following optimization program.

Proposition 4 *The optimization program of* 0*-PLF-AR for the third step can be written as follows*

$$\min_{\tau_1,\tau_2,\tau_3} 2\sigma^2 - 2\tau_3\sigma^2 - 2\tau_2\omega_2\rho\sigma^2 - 2\tau_2\omega_1\nu\rho^2\sigma^2 - 2\tau_1\nu\rho^2\sigma^2 \tag{164a}$$

$$\text{s.t.} : \tag{164b}$$

$$\tau_3^2\sigma^2(1-2^{-2R_3}(\rho^4 2^{-2R_1-2R_2} + \rho^2(1-\rho^2)2^{-2R_2} - \rho^2)) \leq$$
$$(1-2^{-2R_3})(1-\tau_1^2-\tau_2^2-2\tau_1\tau_2\omega_1\nu - 2\tau_1\tau_2\omega_2\nu\rho - 2\tau_2\tau_3\omega_1\nu\rho^2 - 2\tau_2\tau_3\omega_2\rho - 2\tau_1\tau_3\nu\rho^2)\sigma^2, \tag{164c}$$

$$\rho^2\nu = \tau_1 + \tau_2\rho\nu + \tau_3\rho^2\nu, \tag{164d}$$

$$\omega_1\rho^2\nu + \rho\omega_2 = \tau_1\rho\nu + \tau_2 + \tau_3(\omega_1\rho^2\nu + \rho\omega_2), \tag{164e}$$

$$\nu = \sqrt{1-2^{-2R_1}}. \tag{164f}$$

Proof: According to (129), the optimization program of the third step is given as follows

$$\begin{aligned}
&\min_{P_{\hat{X}_3|X_3\hat{X}_1\hat{X}_2}} \mathbb{E}[\|X_3-\hat{X}_3\|^2] \\
&\text{s.t.} \quad I(X_3;\hat{X}_3|\hat{X}_1,\hat{X}_2) \leq R_3, \\
&\qquad P_{\hat{X}_1\hat{X}_2X_3} = P_{\hat{X}_1\hat{X}_2\hat{X}_3}.
\end{aligned} \tag{165}$$

Using the above program, we first derive the rate expression in (164c). Consider the following set of inequalities

$$\begin{aligned}
R_3 &\geq I(X_3;\hat{X}_3|\hat{X}_1,\hat{X}_2) && (166)\\
&= h(\hat{X}_3|\hat{X}_1,\hat{X}_2) - h(Z_3) && (167)\\
&= h(\tau_3X_3 + Z_3|\hat{X}_1,\hat{X}_2) - h(Z_3) && (168)\\
&= \frac{1}{2}\log 2^{-2h(Z_3)}\left(\tau_3^2 2^{2h(X_3|\hat{X}_1,\hat{X}_2)} + 2^{2h(Z_3)}\right) && (169)\\
&= \frac{1}{2}\log 2^{-2h(Z_3)}\left(\tau_3^2 2^{2h(\rho X_2+N_2|\hat{X}_1,\hat{X}_2)} + 2^{2h(Z_3)}\right) && (170)\\
&= \frac{1}{2}\log 2^{-2h(Z_3)}\left(\tau_3^2(\rho^2 2^{2h(X_2|\hat{X}_1,\hat{X}_2)} + 2^{2h(N_2)}) + 2^{2h(Z_3)}\right) && (171)\\
&= \frac{1}{2}\log 2^{-2h(Z_3)}\left(\tau_3^2(\rho^2 2^{2h(X_2|\hat{X}_1,\hat{X}_2)} + (1-\rho^2)\sigma^2) + 2^{2h(Z_3)}\right) && (172)\\
&\geq \frac{1}{2}\log 2^{-2h(Z_3)}\left(\tau_3^2(\rho^2 2^{2h(X_2|\hat{X}_1)}2^{-2R_2} + (1-\rho^2)\sigma^2) + 2^{2h(Z_3)}\right) && (173)\\
&= \frac{1}{2}\log 2^{-2h(Z_3)}\left(\tau_3^2(\rho^2 2^{2h(\rho X_1+N_1|\hat{X}_1)}2^{-2R_2} + (1-\rho^2)\sigma^2) + 2^{2h(Z_3)}\right) && (174)\\
&= \frac{1}{2}\log 2^{-2h(Z_3)}\left(\tau_3^2(\rho^4 2^{-2R_2}2^{2h(X_1|\hat{X}_1)} + \rho^2(1-\rho^2)2^{-2R_2}\sigma^2 + (1-\rho^2)\sigma^2) + 2^{2h(Z_3)}\right) && (175)\\
&\geq \frac{1}{2}\log 2^{-2h(Z_3)}\left(\tau_3^2(\rho^4\sigma^2 2^{-2R_1-2R_2} + \rho^2(1-\rho^2)2^{-2R_2}\sigma^2 + (1-\rho^2)\sigma^2) + 2^{2h(Z_3)}\right), && (176)
\end{aligned}$$

where

- (170) follows from (7) where $X_3 = \rho X_2 + N_2$;

- (171) and (175) follow from Entropy Power Inequality (EPI) [18, pp. 22] which holds which equality for Gaussian sources;
- (173) follows from the rate constraint $I(X_2;\hat{X}_2|\hat{X}_1) \leq R_2$ which yields $h(X_2|\hat{X}_2,\hat{X}_1) \geq h(X_2|\hat{X}_1) - R_2$;
- (176) follows from the rate constraint $I(X_1;\hat{X}_1) \leq R_1$ which yields $h(X_1) \geq h(X_1|\hat{X}_1) - R_1$.

Thus, re-arranging the terms in (176), we have

$$\begin{aligned}
&(\tau_3^2(\rho^2(1-\rho^2)\sigma^2 2^{-2R_2} + (1-\rho^2)\sigma^2))2^{-2R_3} \\
&\geq (1-2^{-2R_3})2^{2h(Z_3)} \qquad (177)\\
&= (1-2^{-2R_3})\cdot \\
&\qquad (1-\tau_1^2-\tau_2^2-\tau_3^2-2\tau_1\tau_2\omega_1-2\tau_1\tau_2\omega_2\rho-2\tau_2\tau_3\omega_1\rho^2-2\tau_2\tau_3\omega_2\rho-2\tau_1\tau_3\rho^2)\sigma^2. \qquad (178)
\end{aligned}$$

The above constraint can be simplified as follows

$$\begin{aligned}
&\tau_3^2\sigma^2(1-\rho^2 2^{-2R_3} + \rho^2(1-\rho^2)2^{-2R_2}2^{-2R_3}) \\
&\geq (1-2^{-2R_3})(1-\tau_1^2-\tau_2^2-2\tau_1\tau_2\omega_1-2\tau_1\tau_2\omega_2\rho-2\tau_2\tau_3\omega_1\rho^2-2\tau_2\tau_3\omega_2\rho-2\tau_1\tau_3\rho^2)\sigma^2, \qquad (179)
\end{aligned}$$

which is the rate expression in (164c).

The derivation of the perception constraint in (164d) is given in the following.

$$\begin{aligned}
\rho^2\nu\sigma^2 &= \mathbb{E}[X_3\hat{X}_1] && (180)\\
&= \mathbb{E}[\hat{X}_3\hat{X}_1] && (181)\\
&= \tau_1\sigma^2 + \tau_2\mathbb{E}[\hat{X}_2\hat{X}_1] + \tau_3\mathbb{E}[X_3\hat{X}_1] && (182)\\
&= \tau_1\sigma^2 + \tau_2\mathbb{E}[X_2\hat{X}_1] + \tau_3\rho^2\mathbb{E}[X_1\hat{X}_1] && (183)\\
&= \tau_1\sigma^2 + \tau_2\rho\mathbb{E}[X_1\hat{X}_1] + \tau_3\rho^2\mathbb{E}[X_1\hat{X}_1] && (184)\\
&= \tau_1\sigma^2 + \tau_2\rho\nu\sigma^2 + \tau_3\rho^2\nu\sigma^2, && (185)
\end{aligned}$$

where

- (181) follows from 0-PLF-AR condition, i.e., $P_{\hat{X}_3\hat{X}_2\hat{X}_1} = P_{X_3\hat{X}_2\hat{X}_1}$ which implies that $\mathbb{E}[X_3\hat{X}_1] = \mathbb{E}[\hat{X}_3\hat{X}_1]$ for the Gauss-Markov source model;
- (182) follows from (133) where we multiply both sides with $\hat{X}_1$ and take an expectation over the distribution;
- (183) follows from the 0-PLF-AR condition which implies that $\mathbb{E}[\hat{X}_2\hat{X}_1] = \mathbb{E}[X_2\hat{X}_1]$ and also from (7), we have $X_3 = \rho^2X_1 + \rho N_1 + N_2$ where (N_1, N_2) are independent of $\hat{X}_1$;
- (184) follows from (7) where $X_2 = \rho X_1 + N_1$ and N_1 is independent of $\hat{X}_1$.

Similarly, for derivation of (164e), we have

$$\begin{aligned}
\omega_1\rho^2\nu\sigma^2 + \rho\omega_2\sigma^2 &= \mathbb{E}[\hat{X}_2X_3] && (186)\\
&= \mathbb{E}[\hat{X}_2\hat{X}_3] && (187)\\
&= \tau_1\mathbb{E}[\hat{X}_1\hat{X}_2] + \tau_2\sigma^2 + \tau_3\mathbb{E}[X_3\hat{X}_2] && (188)\\
&= \tau_1\mathbb{E}[\hat{X}_1X_2] + \tau_2\sigma^2 + \tau_3\mathbb{E}[X_3\hat{X}_2] && (189)\\
&= \tau_1\rho\nu\sigma^2 + \tau_2\sigma^2 + \tau_3(\omega_1\rho^2\nu\sigma^2 + \rho\omega_2\sigma^2). && (190)
\end{aligned}$$

The distortion term in (164a) can be derived as follows

$$\begin{aligned}
\mathbb{E}[\|X_3-\hat{X}_3\|^2] &= \mathbb{E}[X_3^2] + \mathbb{E}[\hat{X}_3^2] - 2\mathbb{E}[X_3\hat{X}_3] && (191)\\
&= 2\sigma^2 - 2\mathbb{E}[X_3\hat{X}_3] && (192)\\
&= 2\sigma^2 - 2(\tau_1\mathbb{E}[\hat{X}_1X_3] + \tau_2\mathbb{E}[\hat{X}_2X_3] + \tau_3\sigma^2) && (193)\\
&= 2\sigma^2 - 2(\tau_1\rho^2\mathbb{E}[\hat{X}_1X_1] + \tau_2\rho\mathbb{E}[\hat{X}_2X_2] + \tau_3\sigma^2) && (194)\\
&= 2\sigma^2 - 2(\tau_1\rho^2\nu\sigma^2 + \tau_2\rho(\rho\nu\omega_1 + \omega_2)\sigma^2 + \tau_3\sigma^2), && (195)
\end{aligned}$$

where

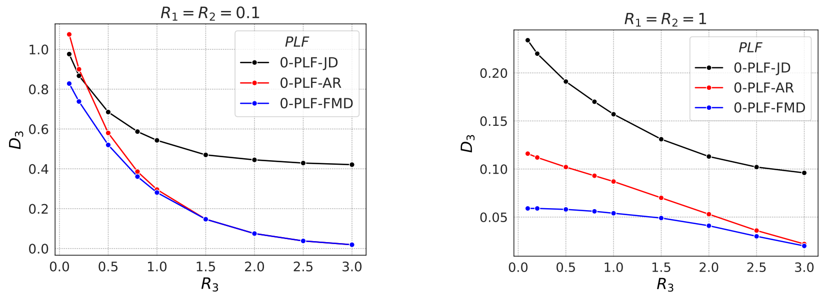


Figure 6: Distortion of the third frame versus its rate for the low-rate regime and $\rho = 1$.

- (192) follows because 0-PLF-AR condition implies that $P_{X_3} = P_{\hat{X}_3}$;
- (193) follows from (133) where $X_3 = \tau_1 \hat{X}_1 + \tau_2 \hat{X}_2 + \tau_3 X_3 + Z_3$;
- (194) follows from (7);
- (195) follows from (131) and (132).

This concludes the proof. ■

The solution of the optimization program in Proposition 4 is plotted in Fig. 6 for some values of the parameters.

C.2 0-PLF-FMD

In this section, we propose the optimization program of 0-PLF-FMD for the second and third steps. We analytically solve the optimization problem of the second step and provide some numerical evaluations for the program of the third step.

Second Step:

The optimization program of the second step is similar to that of Proposition 4 but with a difference that the condition (138c) which preserves the joint distribution of $(\hat{X}_1, \hat{X}_2)$ is not needed for 0-PLF-FMD where only marginal distributions are fixed. We also use the following approximation for the rate of the first frame

$$1 - 2^{-2R_1} = 2\epsilon \ln 2 + O(\epsilon^2). \tag{196}$$

Thus, the optimization problem of the second step for 0-PLF-FMD is as follows

$$\min_{\omega_1, \omega_2} \; 2\sigma^2 - 2\omega_1 \rho \sigma^2 \sqrt{2\epsilon \ln 2 + O(\epsilon^2)} - 2\omega_2 \sigma^2, \tag{197a}$$

$$\text{s.t.} \quad \omega_2^2 (1 - \rho^2 2^{-2R_2} (2\epsilon \ln 2 + O(\epsilon^2))) \leq (1 - \omega_1^2 - 2\omega_1 \omega_2 \rho (2\epsilon \ln 2 + O(\epsilon^2)))(1 - 2^{-2R_2}). \tag{197b}$$

Now, we proceed with solving the above optimization program analytically. Ignoring the small terms of (197b), this condition reduces to the following

$$\omega_2^2 \leq (1 - \omega_1^2)(1 - 2^{-2R_2}). \tag{198}$$

Thus, the optimization program of (197) with considering the dominant terms reduces to the following

$$\min_{\omega_1, \omega_2} \; 2\sigma^2 - 2\omega_1 \rho \sigma^2 \sqrt{2\epsilon \ln 2} - 2\omega_2 \sigma^2, \tag{199a}$$

$$\text{s.t.} \quad \omega_2^2 \leq (1 - \omega_1^2)(1 - 2^{-2R_2}). \tag{199b}$$

The above program is convex and the solution is obtained on the boundary, i.e.,

$$\omega_2^2 = (1 - \omega_1^2)(1 - 2^{-2R_2}). \tag{200}$$

Plugging the above into (197a), we get

$$\min_{\omega_1} 2\sigma^2 (1 - \rho \omega_1 \sqrt{2\epsilon \ln 2} - \sqrt{1 - \omega_1^2} \sqrt{1 - 2^{-2R_2}}) \tag{201}$$

Taking the derivative of the above expression with respect to ω_1, we have

$$\frac{\omega_1}{\sqrt{1-\omega_1^2}}\sqrt{1-2^{-2R_2}} = \rho\sqrt{2\epsilon\ln 2}, \tag{202}$$

which yields

$$\omega_1 = \frac{\rho\sqrt{2\epsilon\ln 2}}{\sqrt{1-2^{-2R_2}+\rho^2 2\epsilon\ln 2}}, \tag{203}$$

and

$$\omega_2 = \frac{1-2^{-2R_2}}{\sqrt{1-2^{-2R_2}+\rho^2(2\epsilon\ln 2)}}. \tag{204}$$

Thus, we get

$$\hat{X}_2 = \frac{\rho\sqrt{2\epsilon\ln 2}}{\sqrt{1-2^{-2R_2}+\rho^2 2\epsilon\ln 2}}\hat{X}_1 + \frac{1-2^{-2R_2}}{\sqrt{1-2^{-2R_2}+\rho^2(2\epsilon\ln 2)}}X_2 + Z_2, \tag{205}$$

where $Z_2 \sim \mathcal{N}(0,(1-\omega_1^2-\omega_2^2-2\rho\nu\omega_1\omega_2)\sigma^2)$ is a Gaussian random variable independent of $(\hat{X}_1, X_2)$, and the optimal distortion is given by

$$D^0_{2,\text{FMD}} := 2\sigma^2(1-\sqrt{1-2^{-2R_2}+\rho^2 2\epsilon\ln 2}) + O(\epsilon). \tag{206}$$

Third Step:

The optimization program of the third step for 0-PLF-FMD is similar to that of (164) with a difference that the conditions (164d) and (164e) that preserve the joint distributions of $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ are not needed since for 0-PLF-FMD, only the marginal distributions are fixed. Thus, we have the following optimization program for the third step

$$\min_{\tau_1,\tau_2,\tau_3} 2\sigma^2 - 2\tau_3\sigma^2 - 2\tau_2\omega_2\rho\sigma^2 - 2\tau_2\omega_1\nu\rho^2\sigma^2 - 2\tau_1\nu\rho^2\sigma^2 \tag{207a}$$

$$\text{s.t.}: \quad \tau_3^2\sigma^2(1-2^{-2R_3}(\rho^4 2^{-2R_1-2R_2}+\rho^2(1-\rho^2)2^{-2R_2}-\rho^2)) \leq$$
$$(1-2^{-2R_3})(1-\tau_1^2-\tau_2^2-2\tau_1\tau_2\omega_1\nu-2\tau_1\tau_2\omega_2\nu\rho-2\tau_2\tau_3\omega_1\nu\rho^2-2\tau_2\tau_3\omega_2\rho-2\tau_1\tau_3\nu\rho^2)\sigma^2. \tag{207b}$$

The solution of the above optimization program is plotted for some values of parameters in Fig. 6.

C.3 0-PLF-JD

In this section, we propose the optimization programs of 0-PLF-JD for the second and third steps. We analytically solve the optimization problem of the second frame and provide some numerical evaluations for the third step.

Second Step:

The optimization program of the second step is similar to that of Proposition 3 with a difference that the condition in (138c) is replaced by the corresponding condition of 0-PLF-JD which is $P_{X_1X_2} = P_{\hat{X}_1\hat{X}_2}$. This constraint implies that $\mathbb{E}[X_1X_2] = \mathbb{E}[\hat{X}_1\hat{X}_2]$ which together with (131) and (132) yields

$$\omega_1 + \nu\omega_2\rho = \rho. \tag{208}$$

Thus, the optimization problem of the second step for 0-PLF-JD when $R_1 = \epsilon$ is as follows

$$\min_{\omega_1,\omega_2} \ 2\sigma^2 - 2\omega_1\rho\sigma^2\sqrt{2\epsilon\ln 2 + O(\epsilon^2)} - 2\omega_2\sigma^2, \tag{209a}$$

$$\text{s.t.} \quad \omega_2^2(1-\rho^2 2^{-2R_2}(2\epsilon\ln 2+O(\epsilon^2))) \leq (1-\omega_1^2-2\omega_1\omega_2\rho\sqrt{2\epsilon\ln 2+O(\epsilon^2)})(1-2^{-2R_2}), \tag{209b}$$

$$\omega_1 + \nu\omega_2\rho = \rho. \tag{209c}$$

The constraint (209c) implies that

$$\omega_1 = \rho - \rho\omega_2\sqrt{2\epsilon\ln 2} + O(\epsilon). \tag{210}$$

Plugging the above into (209a) and (209b), we get

$$\min_{\omega_2} \quad 2\sigma^2(1-\rho^2\sqrt{2\epsilon\ln 2}-\omega_2)+O(\epsilon) \tag{211a}$$

$$\text{s.t. : } \omega_2 \le \sqrt{1-\rho^2}\sqrt{1-2^{-2R_2}}+O(\sqrt{\epsilon}). \tag{211b}$$

The solution of the above program is given by

$$\omega_2 = \sqrt{1-\rho^2}\sqrt{1-2^{-2R_2}}+O(\sqrt{\epsilon}). \tag{212}$$

Thus, we have

$$\hat{X}_2 = (\rho-\rho\omega_2\sqrt{2\epsilon\ln 2})\hat{X}_1+\sqrt{1-\rho^2}\sqrt{1-2^{-2R_2}}X_2+Z_2, \tag{213}$$

where $Z_2 \sim \mathcal{N}(0,((1-\rho^2)2^{-2R_2}-\rho^2\sqrt{1-\rho^2}\sqrt{1-2^{-2R_2}}\sqrt{2\epsilon\ln 2})\sigma^2)$ is a Gaussian random variable independent of $(\hat{X}_1, X_2)$ and the optimal distortion is given by

$$D^0_{2,\text{JD}} := 2\sigma^2(1-\sqrt{1-\rho^2}\sqrt{1-2^{-2R_2}}-\rho^2\sqrt{2\epsilon\ln 2})+O(\epsilon). \tag{214}$$

Third Step:

The optimization program of the third step for 0-PLF-JD is similar to (164) but with a difference that the conditions in (164d) and (164e) are replaced by the corresponding conditions of 0-PLF-JD which is $P_{X_1X_2X_3} = P_{\hat{X}_1\hat{X}_2\hat{X}_3}$. This constraint implies that

$$\mathbb{E}[X_1X_3] = \mathbb{E}[\hat{X}_1\hat{X}_3], \tag{215}$$

$$\mathbb{E}[X_2X_3] = \mathbb{E}[\hat{X}_2\hat{X}_3]. \tag{216}$$

Considering (131)–(133) together with the above conditions, we get

$$\rho^2 = \tau_1+\tau_2\rho+\tau_3\rho^2\nu, \tag{217}$$

$$\rho = \tau_1\rho+\tau_2+\tau_3(\omega_1\rho^2\nu+\rho\omega_2). \tag{218}$$

Thus, we have the following optimization program for the third step

$$\min_{\tau_1,\tau_2,\tau_3} 2\sigma^2-2\tau_3\sigma^2-2\tau_2\omega_2\rho\sigma^2-2\tau_2\omega_1\nu\rho^2\sigma^2-2\tau_1\nu\rho^2\sigma^2 \tag{219a}$$

$$\begin{aligned}\text{s.t. : } \quad & \tau_3^2\sigma^2(1-2^{-2R_3}(\rho^4 2^{-2R_1-2R_2}+\rho^2(1-\rho^2)2^{-2R_2}-\rho^2)) \le \\ & (1-2^{-2R_3})(1-\tau_1^2-\tau_2^2-2\tau_1\tau_2\omega_1\nu-2\tau_1\tau_2\omega_2\nu\rho-2\tau_2\tau_3\omega_1\nu\rho^2-2\tau_2\tau_3\omega_2\rho-2\tau_1\tau_3\nu\rho^2)\sigma^2,\end{aligned} \tag{219b}$$

$$\rho^2 = \tau_1+\tau_2\rho+\tau_3\rho^2\nu, \tag{219c}$$

$$\rho = \tau_1\rho+\tau_2+\tau_3(\omega_1\rho^2\nu+\rho\omega_2). \tag{219d}$$

The solution of the above program is plotted in Fig. 6 for some values of parameters. For the case $R_1 = R_2 = 0.1$ (low compression rates) and a large range of rates R_3, the performances of 0-PLF-AR and 0-PLF-FMD are almost the same. For $R_1 = R_2 = 1$ (low compression rates), the distortion of 0-PLF-AR is significantly smaller than that of 0-PLF-JD for all values of R_3, and for a large enough R_3, it performs similar to 0-PLF-FMD.

D High-Rate Regime for the First Frame

In this section, we first prove the following theorem where the first frame is compressed at a high rate, i.e., $R_1 \to \infty$. The rates of all subsequent frames are assumed to be small, i.e., $R_j = \epsilon$ for sufficiently small $\epsilon > 0$ and $j \in \{2,\dots,T\}$. Then, we provide proofs for the achievable reconstructions of 0-PLF-FMD as outlined in Table 1.

Theorem 2 *Let $R_1 \to \infty$ and $R_j = \epsilon$ for sufficiently small $\epsilon > 0$ and $j \in \{2,\dots,T\}$. An achievable reconstruction of 0-PLF-AR in jth frame ($j \in \{1,\dots,T\}$) is given by*

$$\hat{X}_j = \rho^{j-1}\hat{X}_1+\sum_{i=1}^{j-1}O(\sqrt{\epsilon})N_i+\sum_{i=2}^{j-2}O(\sqrt{\epsilon})Z_{i,AR}+O(\sqrt{\epsilon})Z_{j-1,AR}+Z_{j,AR}, \tag{220}$$

where $Z_{j,AR}$ is a Gaussian random noise independent of $(\{N_i\}_{i=1}^{j-1}, \{Z_{i,AR}\}_{i=2}^{j-1})$, with mean zero and variance $(1-\rho^{2(j-1)}+O(\epsilon))\sigma^2$, and the distortion is as follows

$$D_{j,AR}^{\infty} = 2(1-\rho^{2(j-1)} - O(\sqrt{\epsilon}))\sigma^2 + O(\epsilon), \tag{221}$$

and an achievable reconstruction of 0*-PLF-JD in* j*th frame is given by*

$$\hat{X}_j = \rho^{j-1}\hat{X}_1 + \sum_{i=1}^{j-1} O(\sqrt{\epsilon})N_i + \sum_{i=2}^{j-2} O(\sqrt{\epsilon})Z_{i,JD} + \rho Z_{j-1,JD} + Z_{j,JD}, \tag{222}$$

where $Z_{j,JD}$ is a Gaussian random noise independent of $(\{N_i\}_{i=1}^{j-1}, \{Z_{i,JD}\}_{i=2}^{j-1})$ with mean zero and variance given in Section D.2, and the distortion is as follows

$$D_{j,JD}^{\infty} = 2\left(1-\rho^{2(j-1)} - O(\sqrt{\epsilon})\right)\sigma^2 + O(\epsilon). \tag{223}$$

To prove the above theorem, we consider each PLF separately. We provide the analysis for the second, third and fourth frames. We then use an induction to derive the achievable reconstruction for jth frame. Notice that the solutions for the second and third frames are also presented in Table 1.

D.1 0-PLF-AR

In this section, we introduce the optimization programs of the second, third and fourth steps for 0-PLF-AR and provide the solutions for them. The results are further extended to T frames. Similar to (132)–(133), we write the achievable reconstructions of the second and third steps as follows

$$\hat{X}_2 = \omega_1\hat{X}_1 + \omega_2 X_2 + Z_{2,\mathrm{AR}}, \tag{224}$$
$$\hat{X}_3 = \tau_1\hat{X}_1 + \tau_2\hat{X}_2 + \tau_3 X_3 + Z_{3,\mathrm{AR}}, \tag{225}$$

where $Z_{2,\mathrm{AR}}$ and $Z_{3,\mathrm{AR}}$ are Gaussian random variables independent of $(\hat{X}_1, X_2)$ and $(\hat{X}_1, \hat{X}_2, X_3)$, respectively.

<u>*Second Step:*</u>

The optimization program of the second step for 0-PLF-AR is similar to that of Proposition 3 but with a difference that $\nu = 1$ since we have a high compression rate for the first frame. Thus, the optimization program of the second step is as follows

$$\min_{\omega_1,\omega_2} \; 2\sigma^2 - 2\omega_1\rho\sigma^2 - 2\omega_2\sigma^2, \tag{226a}$$
$$\text{s.t.} \quad \omega_2^2(1-\rho^2 2^{-2R_2}) \leq (1-\omega_1^2 - 2\omega_1\omega_2\rho)(1-2^{-2R_2}), \tag{226b}$$
$$\omega_1 + \omega_2\rho = \rho. \tag{226c}$$

For the second frame, the achievable reconstruction is given as follows (see [10, Table 2])

$$\hat{X}_2 = (\rho - \rho\sqrt{2\epsilon\ln 2})\hat{X}_1 + \sqrt{2\epsilon\ln 2}X_2 + Z_{2,\mathrm{AR}}, \tag{227}$$

where $Z_{2,\mathrm{AR}} \sim \mathcal{N}(0, (1-\rho^2+O(\epsilon))\sigma^2)$ is independent of $(\hat{X}_1, X_2)$ and $\hat{X}_1 = X_1$ and the distortion is given as follows

$$D_{2,\mathrm{AR}}^{\infty} := 2(1-\rho^2 - (1-\rho^2)\sqrt{2\epsilon\ln 2})\sigma^2. \tag{228}$$

<u>*Third Step:*</u>

The optimization program of the third step is similar to that of Proposition 4 but when $\nu = 1$. Thus, we have the following program

$$\min_{\tau_1,\tau_2,\tau_3} 2\sigma^2 - 2\tau_3\sigma^2 - 2\tau_2\omega_2\rho\sigma^2 - 2\tau_2\omega_1\rho^2\sigma^2 - 2\tau_1\rho^2\sigma^2 \tag{229a}$$

$$\begin{aligned}\text{s.t.}: \; & \tau_3^2(1-2^{-2R_3}(\rho^4 2^{-2R_1-2R_2} + \rho^2(1-\rho^2)2^{-2R_2} - \rho^2)) \leq \\ & (1-2^{-2R_3})(1-\tau_1^2-\tau_2^2-2\tau_1\tau_2\omega_1 - 2\tau_1\tau_2\omega_2\rho - 2\tau_2\tau_3\omega_1\rho^2 - 2\tau_2\tau_3\omega_2\rho - 2\tau_1\tau_3\rho^2),\end{aligned} \tag{229b}$$

$$\rho^2 = \tau_1 + \tau_2\rho + \tau_3\rho^2, \tag{229c}$$
$$\omega_1\rho^2 + \rho\omega_2 = \tau_1\rho + \tau_2 + \tau_3(\omega_1\rho^2 + \rho\omega_2). \tag{229d}$$

For the specific case of $R_2 = R_3 = \epsilon$, we will simplify the program (229) and derive the solution. We consider the following approximation

$$1 - 2^{-2R_j} = 2\epsilon \ln 2 + O(\epsilon^2), \qquad j \in \{2, 3\}. \tag{230}$$

Considering the dominant terms of (229b), this constraint can be written as follows

$$(1-\rho^4)\tau_3^2 \leq (1 - \tau_1^2 - \tau_2^2)(2\epsilon \ln 2). \tag{231}$$

So, the optimization program in (229) simplifies as follows

$$\min_{\tau_1,\tau_2,\tau_3} 2\sigma^2 - 2\tau_3\sigma^2 - 2\tau_2\omega_2\rho\sigma^2 - 2\tau_2\omega_1\rho^2\sigma^2 - 2\tau_1\rho^2\sigma^2 \tag{232a}$$

$$\text{s.t.}: \quad (1-\rho^4)\tau_3^2 \leq (1 - \tau_1^2 - \tau_2^2)(2\epsilon \ln 2), \tag{232b}$$

$$\rho^2 = \tau_1 + \tau_2\rho + \tau_3\rho^2, \tag{232c}$$

$$\omega_1\rho^2 + \rho\omega_2 = \tau_1\rho + \tau_2 + \tau_3(\omega_1\rho^2 + \rho\omega_2). \tag{232d}$$

We write τ_1, τ_2 and τ_3 as $\tau_1 = K_1 + \delta_1\sqrt{2\epsilon \ln 2}$, $\tau_2 = K_2 + \delta_2\sqrt{2\epsilon \ln 2}$ and $\tau_3 = \delta_3\sqrt{2\epsilon \ln 2}$, and plug them into (229c)–(229d) to get the following equations

$$\rho^2 = K_1 + \rho K_2, \tag{233a}$$

$$\rho^3 = K_1\rho + K_2, \tag{233b}$$

$$0 = \delta_1 + \rho\delta_2 + \rho^2\delta_3, \tag{233c}$$

$$-\rho^3 + \rho = \rho\delta_1 + \delta_2 + \rho^3\delta_3. \tag{233d}$$

Notice that (233a)–(233b) yields $K_2 = 0$ and $K_1 = \rho^2$. The constant terms of τ_1 and τ_2 which are K_1 and K_2, contribute to the dominant terms of (231). Plugging the values of K_1 and K_2 into (231), we have the following inequality

$$\delta_3 \leq 1. \tag{234}$$

So, considering the dominant terms, the optimization program in (229) is upper bounded by the following

$$\min_{\delta_1,\delta_2,\delta_3} 2\sigma^2(1 - \rho^4 - (\rho^2\delta_1 + \rho^3\delta_2 + \delta_3)\sqrt{2\epsilon \ln 2}) \tag{235a}$$

$$\text{s.t.}: \quad \delta_3 \leq 1, \tag{235b}$$

$$\delta_1 + \rho\delta_2 + \rho^2\delta_3 = 0, \tag{235c}$$

$$\rho\delta_1 + \delta_2 + \rho^3\delta_3 = -\rho^3 + \rho. \tag{235d}$$

The above optimization program is convex, so the solution is obtained at the boundary of the feasible region where we get

$$\delta_1 = -2\rho^2, \tag{236}$$

$$\delta_2 = \rho, \tag{237}$$

$$\delta_3 = 1. \tag{238}$$

Thus, we get the following achievable reconstruction

$$\hat{X}_3 = (\rho^2 - 2\rho^2\sqrt{2\epsilon \ln 2})\hat{X}_1 + \rho\sqrt{2\epsilon \ln 2}\hat{X}_2 + \sqrt{2\epsilon \ln 2}X_3 + Z_{3,\text{AR}}, \tag{239}$$

where $Z_{3,\text{AR}} \sim \mathcal{N}(0, (1 - \rho^4 + O(\epsilon))\sigma^2)$ and the distortion is given by

$$D_{3,\text{AR}}^{\infty} := 2(1 - \rho^4 - (1-\rho^4)\sqrt{2\epsilon \ln 2})\sigma^2. \tag{240}$$

Plugging (227) into (239) yields the following

$$\hat{X}_3 = (\rho^2 - \rho^2\sqrt{2\epsilon \ln 2})\hat{X}_1 + \sqrt{2\epsilon \ln 2}X_3 + \rho\sqrt{2\epsilon \ln 2}Z_{2,\text{AR}} + Z_{3,\text{AR}}, \tag{241}$$

Using (7), the expression in (241) can be written as the following

$$\hat{X}_3 = \rho^2\hat{X}_1 + \rho\sqrt{2\epsilon \ln 2}N_1 + \sqrt{2\epsilon \ln 2}N_2 + \rho\sqrt{2\epsilon \ln 2}Z_{2,\text{AR}} + Z_{3,\text{AR}}. \tag{242}$$

Fourth Step: We derive the optimization program of the fourth frame and solve it. For the fourth frame, we write the achievable reconstruction as follows

$$\hat{X}_4 = \lambda_1\hat{X}_1 + \lambda_2\hat{X}_2 + \lambda_3\hat{X}_3 + \lambda_4X_4 + Z_{4,\text{AR}}, \tag{243}$$

where $Z_{4,\text{AR}}$ is a Gaussian random variable independent of $(\hat{X}_1, \hat{X}_2, \hat{X}_3, X_4)$ with mean zero and its variance will be determined later.

Proposition 5 *The optimization program of the fourth step for* 0*-PLF-AR when the first frame has a high compression rate, is given as follows*

$$\min_{\lambda_1,\lambda_2,\lambda_3,\lambda_4} 2\sigma^2-2\lambda_4\sigma^2-2\lambda_3\rho\tau_3\sigma^2-2\lambda_3\rho^2\tau_2\omega_2\sigma^2-2\lambda_3\rho^3\tau_2\omega_1\sigma^2-2\lambda_3\rho^3\tau_1\sigma^2$$
$$-2\lambda_2\rho^3\omega_1\sigma^2-2\lambda_2\rho^2\omega_2\sigma^2-2\lambda_1\rho^3\sigma^2 \tag{244a}$$

$$\textit{s.t.}: 2^{-2R_4}(\lambda_4^2\rho^6 2^{-2R_3-2R_2-2R_1}\sigma^2+\lambda_4^2\rho^4 2^{-2R_3-2R_2}(1-\rho^2)\sigma^2+\lambda_4^2\rho^2 2^{-2R_3}(1-\rho^2)\sigma^2$$
$$+\lambda_4^2(1-\rho^2)\sigma^2)\leq 2^{2h(Z_{4,AR})}(1-2^{-2R_4}), \tag{244b}$$

$$\rho^3=\lambda_1+\rho\lambda_2+\rho^2\lambda_3+\rho^3\lambda_4, \tag{244c}$$

$$\rho^2(\rho\omega_1+\omega_2)=\rho\lambda_1+\lambda_2+\rho(\rho\omega_1+\omega_2)\lambda_3+\rho^2(\rho\omega_1+\omega_2)\lambda_4, \tag{244d}$$

$$\rho(\rho^2\tau_1+\rho(\rho\omega_1+\omega_2)\tau_2+\tau_3)=$$
$$\rho^2\lambda_1+\rho(\rho\omega_1+\omega_2)\lambda_2+\lambda_3+\rho(\rho^2\tau_1+\rho(\rho\omega_1+\omega_2)\tau_2+\tau_3)\lambda_4. \tag{244e}$$

Proof: An extension of (129) to the fourth step yields the following optimization program

$$\begin{aligned}\min_{P_{\hat{X}_4|X_4\hat{X}_1\hat{X}_2\hat{X}_3}} \quad & \mathbb{E}[\|X_4-\hat{X}_4\|^2]\\ \text{s.t.} \quad & I(X_4;\hat{X}_4|\hat{X}_1,\hat{X}_2,\hat{X}_3)\leq R_4,\\ & P_{\hat{X}_1\hat{X}_2\hat{X}_3X_4}=P_{\hat{X}_1\hat{X}_2\hat{X}_3\hat{X}_4}.\end{aligned} \tag{245}$$

The perception constraints in (244c)–(244e) are derived based on 0-PLF-AR condition which is $P_{\hat{X}_4\hat{X}_3\hat{X}_2\hat{X}_1}=P_{X_4\hat{X}_3\hat{X}_2\hat{X}_1}$. This implies that $\mathbb{E}[\hat{X}_4\hat{X}_1]=\mathbb{E}[X_4\hat{X}_1]$, $\mathbb{E}[\hat{X}_4\hat{X}_2]=\mathbb{E}[X_4\hat{X}_2]$ and $\mathbb{E}[\hat{X}_4\hat{X}_3]=\mathbb{E}[X_4\hat{X}_3]$. These constraints combined with (131)–(133), (243) yield (244c)–(244e). For the rate constraint, consider the following set of inequalities

$$I(X_4;\hat{X}_4|\hat{X}_1,\hat{X}_2,\hat{X}_3) \tag{246}$$
$$=h(\hat{X}_4|\hat{X}_1,\hat{X}_2,\hat{X}_3)-h(Z_{4,\text{AR}}) \tag{247}$$
$$=h(\lambda_4X_4+Z_{4,\text{AR}}|\hat{X}_1,\hat{X}_2,\hat{X}_3)-h(Z_{4,\text{AR}}) \tag{248}$$
$$=\frac{1}{2}\log 2^{-2h(Z_{4,\text{AR}})}\left(\lambda_4^2 2^{2h(X_4|\hat{X}_1,\hat{X}_2,\hat{X}_3)}+2^{2h(Z_{4,\text{AR}})}\right) \tag{249}$$
$$=\frac{1}{2}\log 2^{-2h(Z_{4,\text{AR}})}\left(\lambda_4^2\rho^2 2^{2h(X_3|\hat{X}_1,\hat{X}_2,\hat{X}_3)}+\lambda_4^2 2^{2h(N_3)}+2^{2h(Z_{4,\text{AR}})}\right) \tag{250}$$
$$\geq\frac{1}{2}\log 2^{-2h(Z_{4,\text{AR}})}\left(\lambda_4^2\rho^2 2^{-2R_3}2^{2h(X_3|\hat{X}_1,\hat{X}_2)}+\lambda_4^2 2^{2h(N_3)}+2^{2h(Z_{4,\text{AR}})}\right) \tag{251}$$
$$=\frac{1}{2}\log 2^{-2h(Z_{4,\text{AR}})}\left(\lambda_4^2\rho^4 2^{-2R_3}2^{2h(X_2|\hat{X}_1,\hat{X}_2)}+\lambda_4^2\rho^2 2^{-2R_3}2^{2h(N_2)}+\lambda_4^2 2^{2h(N_3)}+2^{2h(Z_{4,\text{AR}})}\right) \tag{252}$$
$$\geq\frac{1}{2}\log 2^{-2h(Z_{4,\text{AR}})}\left(\lambda_4^2\rho^4 2^{-2R_3-2R_2}2^{2h(X_2|\hat{X}_1)}+\lambda_4^2\rho^2 2^{-2R_3}2^{2h(N_2)}+\lambda_4^2 2^{2h(N_3)}+2^{2h(Z_{4,\text{AR}})}\right) \tag{253}$$
$$=\frac{1}{2}\log 2^{-2h(Z_{4,\text{AR}})}\Big(\lambda_4^2\rho^6 2^{-2R_3-2R_2}2^{2h(X_1|\hat{X}_1)}+\lambda_4^2\rho^4 2^{-2R_3-2R_2}2^{2h(N_1)}+\lambda_4^2\rho^2 2^{-2R_3}2^{2h(N_2)}$$
$$+\lambda_4^2 2^{2h(N_3)}+2^{2h(Z_{4,\text{AR}})}\Big) \tag{254}$$
$$\geq\frac{1}{2}\log 2^{-2h(Z_{4,\text{AR}})}\Big(\lambda_4^2\rho^6 2^{-2R_3-2R_2-2R_1}\sigma^2+\lambda_4^2\rho^4 2^{-2R_3-2R_2}2^{2h(N_1)}+\lambda_4^2\rho^2 2^{-2R_3}2^{2h(N_2)}$$
$$+\lambda_4^2 2^{2h(N_3)}+2^{2h(Z_{4,\text{AR}})}\Big), \tag{255}$$

where

- (249) follows from EPI [18, pp. 22] which holds with equality for Gaussian sources;
- (251), (253) and (255) follows from the rate constraints $R_3\geq I(X_3;\hat{X}_3|\hat{X}_1,\hat{X}_2)$, $R_2\geq I(X_2;\hat{X}_2|\hat{X}_1)$ and $R_1\geq I(X_1;\hat{X}_1)$, respectively;
- (252) and (254) follow from (7) where $X_3=\rho X_2+N_2$ and $X_2=\rho X_1+N_1$, respectively, and the fact that EPI holds with equality for Gaussian sources.

Re-arranging the terms in (255), we get to the constraint in (257b). The objective function in (244a) is obtained by the expansion of $\mathbb{E}[\|X_4 - \hat{X}_4\|^2]$ using (224), (225) and (243). ■

Now, we provide the solution of the optimization program in (244) when $R_2 = R_3 = R_4 = \epsilon$ for sufficiently small $\epsilon > 0$. Using the following approximation

$$1 - 2^{-2R_j} = 2\epsilon \ln 2 + O(\epsilon^2), \tag{256}$$

and considering the dominant terms of (257b), the solution of the optimization program is upper bounded by

$$\begin{aligned}\min_{\lambda_1,\lambda_2,\lambda_3,\lambda_4} \; & 2\sigma^2 - 2\lambda_4\sigma^2 - 2\lambda_3\rho\tau_3\sigma^2 - 2\lambda_3\rho^2\tau_2\omega_2\sigma^2 - 2\lambda_3\rho^3\tau_2\omega_1\sigma^2 - 2\lambda_3\rho^3\tau_1\sigma^2 \\ & -2\lambda_2\rho^3\omega_1\sigma^2 - 2\lambda_2\rho^2\omega_2\sigma^2 - 2\lambda_1\rho^3\sigma^2 \end{aligned} \tag{257a}$$

$$\text{s.t.} : \lambda_4^2(1-\rho^6) \leq (1 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2)(2\epsilon \ln 2), \tag{257b}$$

$$\rho^3 = \lambda_1 + \rho\lambda_2 + \rho^2\lambda_3 + \rho^3\lambda_4, \tag{257c}$$

$$\rho^2(\rho\omega_1 + \omega_2) = \rho\lambda_1 + \lambda_2 + \rho(\rho\omega_1+\omega_2)\lambda_3 + \rho^2(\rho\omega_1+\omega_2)\lambda_4, \tag{257d}$$

$$\begin{aligned}&\rho(\rho^2\tau_1 + \rho(\rho\omega_1+\omega_2)\tau_2 + \tau_3) = \\ &\qquad \rho^2\lambda_1 + \rho(\rho\omega_1+\omega_2)\lambda_2 + \lambda_3 + \rho(\rho^2\tau_1 + \rho(\rho\omega_1+\omega_2)\tau_2+\tau_3)\lambda_4. \end{aligned} \tag{257e}$$

We proceed with solving the above program. We write $\lambda_j = K_j + \delta_j\sqrt{2\epsilon \ln 2}$ for $j \in \{1,2,3\}$ and $\lambda_4 = \delta_4\sqrt{2\epsilon\ln 2}$ and plug them into (257c)–(257e) to get the following

$$\rho^3 = K_1 + \rho K_2 + \rho^2 K_3, \tag{258}$$

$$\rho^4 = \rho K_1 + K_2 + \rho^3 K_3, \tag{259}$$

$$\rho^5 = \rho^2 K_1 + \rho^3 K_2 + K_3. \tag{260}$$

Solving the above equations, we get $K_1 = \rho^3$, $K_2 = K_3 = 0$. Notice that the constant factors of $\{\lambda_j\}_{j=1}^3$ (i.e., $\{K_j\}_{j=1}^3$) contribute to the dominant terms of (257b) which simplifies to the following

$$\delta_4 \leq 1. \tag{261}$$

So, the optimization problem in (257) with dominant terms simplifies to the following

$$\min_{\delta_1,\delta_2,\delta_3,\delta_4} 2(1 - \rho^6 - (\delta_4 + \rho^5\delta_3 + \rho^4\delta_2 + \rho^3\delta_1)\sqrt{2\epsilon\ln 2})\sigma^2 \tag{262a}$$

$$\text{s.t.} : \delta_4 \leq 1, \tag{262b}$$

$$0 = \delta_1 + \rho\delta_2 + \rho^2\delta_3 + \rho^3\delta_4, \tag{262c}$$

$$\rho^2(1-\rho^2) = \rho\delta_1 + \delta_2 + \rho^3\delta_3 + \rho^4\delta_4, \tag{262d}$$

$$\rho(1-\rho^4) = \rho^2\delta_1 + \rho^3\delta_2 + \delta_3 + \rho^5\delta_4. \tag{262e}$$

Solving the above optimization problem, we get

$$\delta_2 = \rho^2, \qquad \delta_3 = \rho, \qquad \delta_1 = -3\rho^3, \qquad \delta_4 = 1. \tag{263}$$

In summary, we get the following reconstruction

$$\hat{X}_4 = (\rho^3 - 3\rho^3\sqrt{2\epsilon\ln 2})\hat{X}_1 + \rho^2\sqrt{2\epsilon\ln 2}\hat{X}_2 + \rho\sqrt{2\epsilon\ln 2}\hat{X}_3 + \sqrt{2\epsilon\ln 2}X_4 + Z_{4,\text{AR}}. \tag{264}$$

Plugging (227) and (239) into the above expression, we get

$$\hat{X}_4 = \rho^3\hat{X}_1 + \rho^2\sqrt{2\epsilon\ln 2}N_1 + \rho\sqrt{2\epsilon\ln 2}N_2 + N_3 + \rho^2\sqrt{2\epsilon\ln 2}Z_{2,\text{AR}} + \rho\sqrt{2\epsilon\ln 2}Z_{3,\text{AR}} + Z_{4,\text{AR}}, \tag{265}$$

where $Z_{4,\text{AR}}$ has variance $(1-\rho^6 + O(\epsilon))\sigma^2$ and the distortion is given by

$$D_{4,\text{AR}}^{\infty} = 2(1 - \rho^6 - \sqrt{2\epsilon\ln 2}(1-\rho^6))\sigma^2 + O(\epsilon). \tag{266}$$

Now, we use an induction to derive the achievable reconstruction of jth frame.

jth Step:

Using induction and extension of the above analysis to j frames, we get the following achievable reconstruction for jth frame

$$\hat{X}_j = \rho^{j-1}\hat{X}_1 + \sqrt{2\epsilon\ln 2}\sum_{i=1}^{j-1}\rho^{j-1-i}N_i + \sqrt{2\epsilon\ln 2}\sum_{i=2}^{j-1}\rho^{j-i}Z_{i,\text{AR}} + Z_{j,\text{AR}}, \tag{267}$$

where $Z_{j,\text{AR}} \sim \mathcal{N}(0,(1-\rho^{2(j-1)}+O(\epsilon))\sigma^2)$ is a Gaussian random variable independent of $(\hat{X}_1,\{N_i\}_{i=1}^{j-1},\{Z_{i,\text{AR}}\}_{i=2}^{j-1})$ and the distortion is given by

$$D_{j,\text{AR}}^{\infty} = 2(1-\rho^{2(j-1)} - \sqrt{2\epsilon\ln 2}(1-\rho^2)\sum_{i=1}^{j-1}\rho^{2(j-1-i)})\sigma^2 + O(\epsilon). \tag{268}$$

D.2 0-PLF-JD

Second Step: When the first frame is compressed at a high rate, the optimization program of the second step for 0-PLF-JD is similar to that in (226) for 0-PLF-AR and the solution is given in (227).

Third Step:

The optimization program of the third step for 0-PLF-JD is similar to (229) but when the perception constraints in (229c)–(229d) are replaced by

$$\rho^2 = \tau_1 + \tau_2\rho + \tau_3\rho^2, \tag{269}$$

$$\rho = \tau_1\rho + \tau_2 + \tau_3(\omega_1\rho^2 + \rho\omega_2). \tag{270}$$

The above equations come from the fact that $P_{X_1X_2X_3} = P_{\hat{X}_1\hat{X}_2\hat{X}_3}$ which implies that $\mathbb{E}[\hat{X}_1\hat{X}_3] = \mathbb{E}[X_1X_3] = \rho^2\sigma^2$ and $\mathbb{E}[\hat{X}_2\hat{X}_3] = \mathbb{E}[X_2X_3] = \rho\sigma^2$. Thus, we have the following optimization program for the third step of 0-PLF-JD when the first frame is compressed at a high rate,

$$\min_{\tau_1,\tau_2,\tau_3} 2\sigma^2 - 2\tau_3\sigma^2 - 2\tau_2\omega_2\rho\sigma^2 - 2\tau_2\omega_1\rho^2\sigma^2 - 2\tau_1\rho^2\sigma^2 \tag{271a}$$

$$\begin{aligned}\text{s.t.}:\quad & \tau_3^2(1-2^{-2R_3}(\rho^4 2^{-2R_1-2R_2} + \rho^2(1-\rho^2)2^{-2R_2} - \rho^2)) \leq \\ & (1-2^{-2R_3})(1-\tau_1^2-\tau_2^2-2\tau_1\tau_2\omega_1-2\tau_1\tau_2\omega_2\rho-2\tau_2\tau_3\omega_1\rho^2-2\tau_2\tau_3\omega_2\rho-2\tau_1\tau_3\rho^2),\end{aligned} \tag{271b}$$

$$\rho^2 = \tau_1 + \tau_2\rho + \tau_3\rho^2, \tag{271c}$$

$$\rho = \tau_1\rho + \tau_2 + \tau_3(\omega_1\rho^2 + \rho\omega_2). \tag{271d}$$

We solve the optimization program when $R_2 = R_3 = \epsilon$. Similar to (257), we consider the dominant terms of the constraint in (271b) and get the following upper bound on the above optimization problem,

$$\min_{\tau_1,\tau_2,\tau_3} 2\sigma^2 - 2\tau_3\sigma^2 - 2\tau_2\omega_2\rho\sigma^2 - 2\tau_2\omega_1\rho^2\sigma^2 - 2\tau_1\rho^2\sigma^2 \tag{272a}$$

$$\text{s.t.}:\quad (1-\rho^4)\tau_3^2 \leq (1-\tau_1^2-\tau_2^2)(2\epsilon\ln 2), \tag{272b}$$

$$\rho^2 = \tau_1 + \tau_2\rho + \tau_3\rho^2, \tag{272c}$$

$$\rho = \tau_1\rho + \tau_2 + \tau_3(\omega_1\rho^2 + \rho\omega_2). \tag{272d}$$

We write τ_1, τ_2 and τ_3 as $\tau_1 = K_1 + \delta_1\sqrt{2\epsilon\ln 2}$, $\tau_2 = K_2 + \delta_2\sqrt{2\epsilon\ln 2}$ and $\tau_3 = \delta_3\sqrt{2\epsilon\ln 2}$, and plug them into (272c)–(272d) to get the following equations

$$\rho^2 = K_1 + \rho K_2, \tag{273a}$$

$$\rho = K_1\rho + K_2, \tag{273b}$$

$$0 = \delta_1 + \rho\delta_2 + \rho^2\delta_3, \tag{273c}$$

$$0 = \rho\delta_1 + \delta_2 + \rho^3\delta_3. \tag{273d}$$

Equations (273a) and (273b) yield $K_1 = 0$ and $K_2 = \rho$. Notice that the constant terms of $\{\tau_j\}_{j=1}^2$ (i.e., $\{K_j\}_{j=1}^2$) contribute to the dominant terms of the inequality (272b). Thus, we have the following condition

$$\delta_3 \leq \frac{1}{\sqrt{1+\rho^2}}. \tag{274}$$

The optimization program in (272) further simplifies as follows

$$\min_{\delta_1,\delta_2,\delta_3} 2(1-\rho^4-(\delta_3+\delta_1\rho^2+\delta_2\rho^3+\rho^2-\rho^4)\sqrt{2\epsilon\ln 2})\sigma^2, \tag{275a}$$

$$\text{s.t.}: \quad \delta_3 \le \frac{1}{\sqrt{1+\rho^2}}, \tag{275b}$$

$$0=\delta_1+\rho\delta_2+\rho^2\delta_3, \tag{275c}$$

$$0=\rho\delta_1+\delta_2+\rho^3\delta_3. \tag{275d}$$

Solving the above optimization program, we get

$$\delta_2=0, \qquad \delta_1=-\frac{\rho^2}{\sqrt{1+\rho^2}}, \qquad \delta_3=\frac{1}{\sqrt{1+\rho^2}}. \tag{276}$$

Thus, we have

$$\hat{X}_3=\rho\hat{X}_2-\frac{\rho^2}{\sqrt{1+\rho^2}}\sqrt{2\epsilon\ln 2}\hat{X}_1+\frac{1}{\sqrt{1+\rho^2}}\sqrt{2\epsilon\ln 2}X_3+Z_{3,\text{JD}}, \tag{277}$$

where $Z_{3,\text{JD}} \sim \mathcal{N}(0,(1-\rho^2+O(\epsilon))\sigma^2)$ is independent of $(\hat{X}_1,\hat{X}_2,X_3)$. Plugging (227) into the above expression yields the following

$$\hat{X}_3=\left(\rho^2-(\rho^2+\frac{\rho^2}{\sqrt{1+\rho^2}})\sqrt{2\epsilon\ln 2}\right)\hat{X}_1+\rho\sqrt{2\epsilon\ln 2}X_2+\frac{\sqrt{2\epsilon\ln 2}}{\sqrt{1+\rho^2}}X_3+\rho Z_{2,\text{JD}}+Z_{3,\text{JD}}, \tag{278}$$

where the distortion is given as follows

$$D^{\infty}_{3,\text{JD}}:=2(1-\rho^4-(1-\rho^2)(\rho^2+\sqrt{1+\rho^2})\sqrt{2\epsilon\ln 2})\sigma^2+O(\epsilon). \tag{279}$$

Using (7), (278) can be further simplified as follows

$$\hat{X}_3=\rho^2\hat{X}_1+\left(\rho+\frac{\rho}{\sqrt{1+\rho^2}}\right)\sqrt{2\epsilon\ln 2}N_1+\frac{1}{\sqrt{1+\rho^2}}\sqrt{2\epsilon\ln 2}N_2+\rho Z_{2,\text{JD}}+Z_{3,\text{JD}}. \tag{280}$$

469 *Fourth Step:*

The optimization program of the fourth step for 0-PLF-JD is similar to that in Proposition 5 but when conditions (257c)–(257e) are replaced by the corresponding conditions of 0-PLF-JD which are

$$\mathbb{E}[\hat{X}_4\hat{X}_3]=\mathbb{E}[X_4\hat{X}_3], \qquad \mathbb{E}[\hat{X}_4\hat{X}_2]=\mathbb{E}[X_4\hat{X}_2], \qquad \mathbb{E}[\hat{X}_4\hat{X}_1]=\mathbb{E}[X_4\hat{X}_1]. \tag{281}$$

The above conditions are further simplified as follows

$$\rho^3=\lambda_1+\rho\lambda_2+\rho^2\lambda_3+\rho^3\lambda_4, \tag{282}$$

$$\rho^2=\rho\lambda_1+\lambda_2+\rho\lambda_3+\rho^2(\rho\omega_1+\omega_2)\lambda_4, \tag{283}$$

$$\rho=\rho^2\lambda_1+\rho\lambda_2+\lambda_3+\rho(\rho^2\tau_1+\rho(\rho\omega_1+\omega_2)\tau_2+\tau_3)\lambda_4. \tag{284}$$

We study the case of $R_2=R_3=R_4=\epsilon$ for a sufficiently small $\epsilon>0$. Thus, considering the dominant terms, we have the following optimization problem for the fourth step of 0-PLF-JD when the first frame is compressed at a high rate

$$\begin{aligned}\min_{\lambda_1,\lambda_2,\lambda_3,\lambda_4} \; & 2\sigma^2-2\lambda_4\sigma^2-2\lambda_3\rho\tau_3\sigma^2-2\lambda_3\rho^2\tau_2\omega_2\sigma^2-2\lambda_3\rho^3\tau_2\omega_1\sigma^2-2\lambda_3\rho^3\tau_1\sigma^2 \\ & -2\lambda_2\rho^3\omega_1\sigma^2-2\lambda_2\rho^2\omega_2\sigma^2-2\lambda_1\rho^3\sigma^2\end{aligned} \tag{285a}$$

$$\text{s.t.}: \lambda_4^2(1-\rho^6)\le(1-\lambda_1^2-\lambda_2^2-\lambda_3^2+O(\epsilon))(2\epsilon\ln 2), \tag{285b}$$

$$\rho^3=\lambda_1+\rho\lambda_2+\rho^2\lambda_3+\rho^3\lambda_4, \tag{285c}$$

$$\rho^2=\rho\lambda_1+\lambda_2+\rho\lambda_3+\rho^2(\rho\omega_1+\omega_2)\lambda_4, \tag{285d}$$

$$\rho=\rho^2\lambda_1+\rho\lambda_2+\lambda_3+\rho(\rho^2\tau_1+\rho(\rho\omega_1+\omega_2)\tau_2+\tau_3)\lambda_4. \tag{285e}$$

We proceed with solving the above optimization program. We write $\lambda_j = K_j + \delta_j\sqrt{2\epsilon\ln 2}$ for $j\in\{1,2,3\}$ and $\lambda_4 = \delta_4\sqrt{2\epsilon\ln 2}$ and plug them into (285c)–(285e) to get

$$\begin{aligned} \rho^3 &= K_1 + \rho K_2 + \rho^2 K_3, && (286)\\ \rho^2 &= \rho K_1 + K_2 + \rho K_3, && (287)\\ \rho &= \rho^2 K_1 + \rho K_2 + K_3, && (288)\\ 0 &= \delta_1 + \rho\delta_2 + \rho^2\delta_3 + \rho^3\delta_4, && (289)\\ 0 &= \rho\delta_1 + \delta_2 + \rho\delta_3 + \rho^4\delta_4, && (290)\\ 0 &= \rho^2\delta_1 + \rho\delta_2 + \delta_3 + \rho^5\delta_4. && (291) \end{aligned}$$

Thus, we have $K_1 = K_2 = 0$, $K_3 = \rho$. Considering the fact that the constant terms of $\{\lambda_j\}_{j=1}^3$ (i.e., $\{K_j\}_{j=1}^3$) contribute to the dominant terms of (285b) which simplifies to the following

$$\delta_4 \le \sqrt{\frac{1-\rho^2}{1-\rho^6}}. \tag{292}$$

The optimization program in (285) further reduces to the following

$$\begin{aligned} &\min_{\delta_1,\delta_2,\delta_3,\delta_4} 2(1-\rho^6-(\delta_1\rho^3+\delta_2\rho^4+\delta_3\rho^5+\delta_4+\rho^2-\rho^6)\sqrt{2\epsilon\ln 2})\sigma^2, && (293a)\\ &\text{s.t. : } \delta_4 \le \sqrt{\frac{1-\rho^2}{1-\rho^6}}, && (293b)\\ &\quad 0 = \delta_1 + \rho\delta_2 + \rho^2\delta_3 + \rho^3\delta_4, && (293c)\\ &\quad 0 = \rho\delta_1 + \delta_2 + \rho\delta_3 + \rho^4\delta_4, && (293d)\\ &\quad 0 = \rho^2\delta_1 + \rho\delta_2 + \delta_3 + \rho^5\delta_4. && (293e) \end{aligned}$$

Solving the above optimization program, we get

$$\delta_1 = -\rho^3\sqrt{\frac{1-\rho^2}{1-\rho^6}}, \qquad \delta_2=\delta_3=0, \qquad \delta_4 = \sqrt{\frac{1-\rho^2}{1-\rho^6}}. \tag{294}$$

In summary, we get the following achievable reconstruction

$$\hat{X}_4 = -\rho^3\sqrt{\frac{1-\rho^2}{1-\rho^6}}\sqrt{2\epsilon\ln 2}\hat{X}_1 + \rho\hat{X}_3 + \sqrt{\frac{1-\rho^2}{1-\rho^6}}\sqrt{2\epsilon\ln 2}X_4 + Z_{4,\text{JD}}, \tag{295}$$

where $Z_{4,\text{JD}} \sim \mathcal{N}(0,(1-\rho^2+\rho^4-\rho^6+O(\epsilon))\sigma^2)$ is a Gaussian random variable independent of $(\hat{X}_1,\hat{X}_3,X_4)$. Now, we plug (227) and (239) into the above expression and we get

$$\begin{aligned} \hat{X}_4 = \rho^3\hat{X}_1 &+ \left(\rho^2+\rho^2\sqrt{\frac{1-\rho^2}{1-\rho^6}}\right)\sqrt{2\epsilon\ln 2}N_1 + \left(\rho+\rho\sqrt{\frac{1-\rho^2}{1-\rho^6}}\right)\sqrt{2\epsilon\ln 2}N_2\\ &+\sqrt{\frac{1-\rho^2}{1-\rho^6}}\sqrt{2\epsilon\ln 2}N_3 + \rho^2\sqrt{2\epsilon\ln 2}Z_{2,\text{JD}} + \rho Z_{3,\text{JD}} + Z_{4,\text{JD}}, \end{aligned} \tag{296}$$

where the distortion is given by

$$D_{4,\text{JD}}^{\infty} := 2\left(1-\rho^6-\sqrt{2\epsilon\ln 2}(1-\rho^2)\left(\sqrt{\frac{1-\rho^6}{1-\rho^2}}+\rho^2-\rho^6\right)\right)\sigma^2 + O(\epsilon). \tag{297}$$

jth Step:

Using induction and extension of the above analysis for the j-th frame yields the following achievable reconstruction

$$\begin{aligned} \hat{X}_j = \rho^{j-1}\hat{X}_1 &+ \sqrt{2\epsilon\ln 2}\left(1+\sqrt{\frac{1-\rho^2}{1-\rho^{2(j-1)}}}\right)\sum_{i=1}^{j-2}\rho^{j-1-i}N_i\\ &+\sqrt{\frac{1-\rho^2}{1-\rho^{2(j-1)}}}\sqrt{2\epsilon\ln 2}N_{j-1} + \sqrt{2\epsilon\ln 2}\sum_{i=2}^{j-2}\rho^i Z_{j-i,\text{JD}} + \rho Z_{j-1,\text{JD}} + Z_{j,\text{JD}}, \end{aligned} \tag{298}$$

where $Z_{j,\text{JD}}$ is a Gaussian random variable independent of $(\{N_i\}_{i=1}^{j-1}, \{Z_{i,\text{JD}}\}_{i=2}^{j-1})$ with mean zero and the following variance

$$\mathbb{E}[Z_{j,\text{JD}}^2] = \begin{cases} ((1-\rho^2)\sum_{i=0}^{\frac{j}{2}-1}\rho^{4i} + O(\epsilon))\sigma^2 & \text{if } j \text{ is even}, \\ ((1-\rho^2)\sum_{i=0}^{\frac{j-1}{2}-1}\rho^{4i} + O(\epsilon))\sigma^2 & \text{if } j \text{ is odd}, \end{cases} \tag{299}$$

and the distortion is given by

$$D_{j,\text{JD}}^{\infty} := 2\left(1 - \rho^{2(j-1)} - \sqrt{2\epsilon\ln 2}(1-\rho^2)\left(\sqrt{\frac{1-\rho^{2(j-1)}}{1-\rho^2}} + \sum_{i=1}^{j-2}\rho^{2(j-1-i)}\right)\right)\sigma^2 + O(\epsilon). \tag{300}$$

D.3 0-PLF-FMD

In this section, we provide the optimization programs for the second and third steps of 0-PLF-FMD and solve them. These results were presented in the first and second rows of Table 1. Recall that for the Gauss-Markov source model, the reconstructions exploit the structure in (131)–(133).

Second Step:

For the second step, similar to (132), we write the achievable reconstruction as

$$\hat{X}_2 = \omega_1\hat{X}_1 + \omega_2 X_2 + Z_{2,\text{FMD}}, \tag{301}$$

where $Z_{2,\text{FMD}}$ is independent of $(\hat{X}_1, X_2)$ and notice that $\hat{X}_1 = X_1$ since we have high compression rate for the first frame. The optimization program of the second step is similar to that of Proposition 3, but with $\nu = 1$ and when the perception constraint in (138c) (which preserves the joint distribution of $(\hat{X}_1, \hat{X}_2)$) is removed and only the marginal distribution is fixed. Thus, we have the following optimization program for the second step of 0-PLF-FMD

$$\min_{\omega_1,\omega_2} \quad 2\sigma^2 - 2\omega_1\rho\sigma^2 - 2\omega_2\sigma^2, \tag{302a}$$

$$\text{s.t.} \quad \omega_2^2(1-\rho^2 2^{-2R_2}) \le (1-\omega_1^2 - 2\omega_1\omega_2\rho)(1-2^{-2R_2}). \tag{302b}$$

The solution of the above program when $R_2 = \epsilon$ (for a sufficiently small ϵ) is given by (see [10, Table 2])

$$\hat{X}_2 = (1 - \frac{(1+\rho^2)2\epsilon\ln 2}{2\rho^2})\hat{X}_1 + \frac{2\epsilon\ln 2}{\rho}X_2 + Z_{2,\text{FMD}}, \tag{303}$$

where $Z_{2,\text{FMD}} \sim \mathcal{N}(0, (\frac{1-\rho^2}{\rho^2})2\sigma^2\epsilon\ln 2)$ is independent of $(\hat{X}_1, X_2)$.

Notice that when $\rho = \Theta(\sqrt{\epsilon})$, the term $\frac{(1+\rho^2)2\epsilon\ln 2}{2\rho^2}$ becomes a constant. In this case, the approximation in (303) is not valid anymore. This case should be handled separately as follows.

Case of $0 < \rho \ll \sqrt{\epsilon}$: In this case, considering the dominant terms of (302), this program reduces to the following

$$\min_{\omega_1,\omega_2} \quad 2\sigma^2 - 2\omega_2\sigma^2, \tag{304a}$$

$$\text{s.t.} \quad \omega_2^2 \le (1-\omega_1^2)(2\epsilon\ln 2). \tag{304b}$$

The solution of the above program is as follows

$$\omega_1 = 0, \tag{305}$$

$$\omega_2 = \sqrt{2\epsilon\ln 2}. \tag{306}$$

Thus, the reconstruction of the second step can be written as follows

$$\hat{X}_2 = \sqrt{2\epsilon\ln 2}X_2 + Z'_{2,\text{FMD}}, \tag{307}$$

where $Z'_{2,\text{FMD}} \sim \mathcal{N}(0, (1-2\epsilon\ln 2)\sigma^2)$ is independent of X_2.

Third Step:

For the third step, similar to (133), we write the achievable reconstruction as

$$\hat{X}_3 = \tau_1 \hat{X}_1 + \tau_2 \hat{X}_2 + \tau_3 X_3 + Z_{3,\text{FMD}}, \tag{308}$$

where $Z_{3,\text{FMD}}$ is a Gaussian random variable independent of $(\hat{X}_1, \hat{X}_2, X_3)$. The optimization program of the third step is similar to that of Proposition 4 but with $\nu = 1$ and when the constraints in (164d) and (164e) which preserve the joint distribution of $P_{\hat{X}_1\hat{X}_2\hat{X}_3}$ are removed and only the marginal distributions are fixed. Thus, we get the following optimization program

$$\min_{\tau_1,\tau_2,\tau_3} 2\sigma^2 - 2\tau_3\sigma^2 - 2\tau_2\omega_2\rho\sigma^2 - 2\tau_2\omega_1\rho^2\sigma^2 - 2\tau_1\rho^2\sigma^2 \tag{309a}$$

$$\begin{aligned}\text{s.t.}: \ & \tau_3^2\sigma^2(1 - 2^{-2R_3}(\rho^4 2^{-2R_1-2R_2} + \rho^2(1-\rho^2)2^{-2R_2} - \rho^2)) \leq \\ & (1-2^{-2R_3})(1 - \tau_1^2 - \tau_2^2 - 2\tau_1\tau_2\omega_1 - 2\tau_1\tau_2\omega_2\rho - 2\tau_2\tau_3\omega_1\rho^2 - 2\tau_2\tau_3\omega_2\rho - 2\tau_1\tau_3\rho^2)\sigma^2.\end{aligned} \tag{309b}$$

We solve the above program when $R_2 = R_3 = \epsilon$ for a sufficiently small $\epsilon > 0$. We use the following approximation

$$1 - 2^{-2R_j} = 2\epsilon \ln 2 + O(\epsilon^2), \qquad j \in \{2, 3\}. \tag{310}$$

Thus, considering the dominant terms of the constraint in (309b), we have

$$(1 - \tau_1^2 - \tau_2^2 - 2\tau_1\tau_2\omega_1 - 2\tau_1\tau_2\omega_2\rho - 2\tau_2\tau_3\omega_1\rho^2 - 2\tau_2\tau_3\omega_2\rho - 2\tau_1\tau_3\rho^2)(2\epsilon\ln 2) \geq (1-\rho^4)\tau_3^2. \tag{311}$$

For the third frame, we have the following optimization program,

$$\min_{\tau_1,\tau_2,\tau_3} \ 2\sigma^2 - 2\tau_3\sigma^2 - 2\tau_2\omega_2\rho\sigma^2 - 2\tau_2\omega_1\rho^2\sigma^2 - 2\tau_1\rho^2\sigma^2, \tag{312a}$$

$$\begin{aligned}\text{s.t.} \ & (1 - \tau_1^2 - \tau_2^2 - 2\tau_1\tau_2\omega_1 - 2\tau_1\tau_2\omega_2\rho - 2\tau_2\tau_3\omega_1\rho^2 - 2\tau_2\tau_3\omega_2\rho - 2\tau_1\tau_3\rho^2)(2\epsilon\ln 2) \\ & \geq (1-\rho^4)\tau_3^2.\end{aligned} \tag{312b}$$

We write τ_1 and τ_2 as follows

$$\tau_1 = \frac{1}{2} - \delta_1(2\epsilon\ln 2), \tag{313}$$

$$\tau_2 = \frac{1}{2} - \delta_2(2\epsilon\ln 2), \tag{314}$$

$$\tau_3 = \delta_3(2\epsilon\ln 2). \tag{315}$$

for some δ_1, δ_2 and δ_3. Plugging the above into (311), we have

$$(3\delta_1 + 3\delta_2 - 2\delta_3\rho^2 - \frac{1}{4} + \frac{1}{4\rho^2}) \geq (1-\rho^4)\delta_3^2. \tag{316}$$

Thus, the optimization program in (312) reduces to the following

$$\min_{\delta_1,\delta_2,\delta_3} \ 2\sigma^2 - 2\rho^2\sigma^2 - (2\delta_3 + 1 - 2(\delta_1+\delta_2)\rho^2 - \frac{1-\rho^2}{2})(2\epsilon\ln 2) \tag{317}$$

$$\text{s.t.} \ \ (3\delta_1 + 3\delta_2 - 2\delta_3\rho^2 - \frac{1}{4} + \frac{1}{4\rho^2}) \geq (1-\rho^4)\delta_3^2. \tag{318}$$

Optimizing over $\delta_1, \delta_2, \delta_3$, we get

$$\delta_3 = \frac{1 - \frac{2}{3}\rho^4}{\frac{2}{3}\rho^2(1-\rho^4)}, \tag{319}$$

and

$$\delta_1 = \delta_2 = \frac{3 - 4\rho^8}{8\rho^4(1-\rho^4)} + \frac{1-\rho^2}{24\rho^2}. \tag{320}$$

Thus, we have

$$\hat{X}_3 = (\frac{1}{2} - \delta_1(2\epsilon\ln 2))\hat{X}_1 + (\frac{1}{2} - \delta_1(2\epsilon\ln 2))\hat{X}_2 + \delta_3(2\epsilon\ln 2)X_3 + Z_{3,\text{FMD}}, \tag{321}$$

where $Z_{3,\text{FMD}} \sim \mathcal{N}(0, O(\epsilon)\sigma^2)$ is independent of $(\hat{X}_1, \hat{X}_2, X_3)$, where the optimal distortion is given by

$$D^{\infty}_{3,\text{FMD}} := 2\left(1-\rho^2-\left(\delta_3+\frac{1-\rho^2}{4}-(\delta_1+\delta_2)\rho^2\right)2\epsilon\ln 2\right)\sigma^2+O(\epsilon^2). \tag{322}$$

Case of $0 < \rho \ll \sqrt{\epsilon}$: In this case, considering the dominant terms of (312), the program reduces to the following:

$$\min_{\tau_1,\tau_2,\tau_3} \quad 2\sigma^2 - 2\tau_3\sigma^2, \tag{323a}$$

$$\text{s.t. } (1-\tau_1^2-\tau_2^2)(2\epsilon\ln 2) \geq \tau_3^2. \tag{323b}$$

The solution of the above program is simply given by

$$\tau_1 = 0, \tag{324}$$

$$\tau_2 = 0, \tag{325}$$

$$\tau_3 = \sqrt{2\epsilon\ln 2}. \tag{326}$$

Thus, the reconstruction is given by

$$\hat{X}_3 = \sqrt{2\epsilon\ln 2}X_3 + Z'_{3,\text{FMD}}, \tag{327}$$

where $Z'_{3,\text{FMD}} \sim \mathcal{N}(0, (1-2\epsilon\ln 2)\sigma^2)$ is independent of X_3.

E Experimental Setup Details

As described in Section 4, our experimental setup is based on the one proposed in [10]. We briefly describe our setup as follows.

Neural Video Compressor. In this work, we use the version of the scale-space flow model [12] presented in [10] to compress each P-frame. This architecture allows us to efficiently learn the statistical characteristics of the source distribution without using any pre-trained module such as an optical flow estimator. To control the bit rate, we adjust the dimension of the latent representation while fixing the quantization interval to 2. We use dithered quantization to simulate the common randomness in our setting [7]. For each frame X_j, we optimize its corresponding encoder-decoder by using the representation from the optimized encoder-decoder pairs of previous frames.

Distortion and Perception Measurement. Our theoretical results require solving a constrained optimization, which is intractable in practice due to the complexity of neural networks. Instead, we optimize the Lagrange approximations:

$$\min \mathbb{E}[\|X_j - \hat{X}_j\|^2] + \lambda\phi_j(P_{\hat{X}_1\ldots\hat{X}_{j-1}X_j}, P_{\hat{X}_1\ldots\hat{X}_{j-1}\hat{X}_j}),$$

where each λ is adjusted to characterize different constraint levels on the perceptuality. Similar to previous works, we use WGAN [13] to approximate this perception function.

Training Details. MovingMNIST models are trained according to the dataset generation algorithms described in Subsection E.1. The neural architectures tested on UVG are trained on 256×256 patches from the Vimeo-90K dataset [19]. For each MNIST encoder-decoder pair, training takes about one day on a single NVIDIA A100 GPU, with Vimeo-90K training procedures taking around two days. For each rate regime, we first pre-train a model to optimize the MMSE loss before fine-tuning the model with the joint distortion-perception loss, which we found to be more stable than training everything end-to-end. We utilize the *rmsprop* optimizer [20] for our MovingMNIST training procedures and the *Adam* optimizer [21] for Vimeo-90K training runs.

E.1 MovingMNIST Digit Trajectory

This subsection describes the algorithms developed to generate digit trajectories for the MovingMNIST experiments. Section 4 addresses the two main rate regimes discussed in our work. First, we describe our *Random Trajectory* algorithm, utilized when the first frame X_1 is encoded with a low rate (Subsection 3.1). Following that, we discuss *Consistent Trajectory* algorithm, applied to experiments where the first frame X_1 is encoded with a high rate (Subsection 3.2).

Algorithm 1 describes how *Random Trajectory* generates a MovingMNIST sequence. The required inputs are the maximum step size S, sequence length N, frame size F, and digit size D. We first sample the initial digit position (x, y) from a uniform distribution $U(0, F-D)$, generating frame X_1 by placing the digit in the sampled initial position (lines $3-4$). For the subsequent frames $X_2, \ldots, X_N$, we check if the moving digit has reached the frame boundaries (lines $7, 10, 14, 17$). We then sample the vertical and horizontal shifts (d_x, d_y) accordingly (lines $6, 9, 12, 15, 19$). The shift is then applied to the current position (x, y), and the frame is generated by placing the digit in the updated position (lines $21-22$). This conditional sampling strategy guarantees that the digit "bounces" in the opposite direction if the margins are reached, keeping the digit always in-frame. In section 4, we utilize $S=5$, $N=3$, $F=64$, and $D=32$ for the regime with a low rate at the first frame (Subsection 3.1).

Algorithm 1 *Random Trajectory* sequence generation.

1: **inputs:** maximum step size S, sequence length N, frame size F, digit size D.
2: $sequence \leftarrow [\,]$
3: $(x, y) \sim U(0, F-D)$
4: $sequence[1] \leftarrow gen\_frame((x, y))$
5: **for** $frame \in \{2, ..., N\}$ **do**
6: $\quad (d_x, d_y) \sim U(-S, S)$
7: $\quad$ **if** $(y < 0)$ **then**
8: $\quad\quad y \leftarrow 0$
9: $\quad\quad d_y \sim U(0, S)$
10: $\quad$ **else if** $(y > F-D)$ **then**
11: $\quad\quad y \leftarrow F-D$
12: $\quad\quad d_y \sim U(-S, 0)$
13: $\quad$ **end if**
14: $\quad$ **if** $(x < 0)$ **then**
15: $\quad\quad x \leftarrow 0$
16: $\quad\quad d_x \sim U(0, S)$
17: $\quad$ **else if** $(x > F-D)$ **then**
18: $\quad\quad x \leftarrow F-D$
19: $\quad\quad d_x \sim U(-S, 0)$
20: $\quad$ **end if**
21: $\quad (x, y) \leftarrow (x, y) + (d_x, d_y)$
22: $\quad sequence[frame] \leftarrow gen\_frame((x, y))$
23: **end for**
24: **return** $sequence$

Algorithm 2 *Consistent Trajectory* sequence generation.

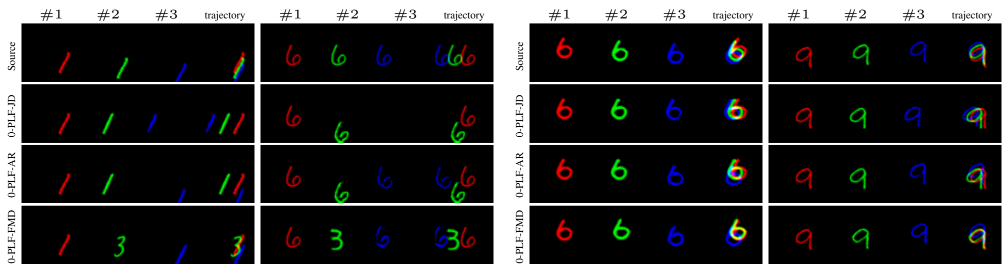
1: **inputs:** maximum step size S, sequence length N, frame size F, digit size D.
2: $sequence \leftarrow [\,]$
3: $(x, y) \sim U(0, F-D)$
4: $(d_x, d_y) \sim U(-S, S)$
5: **for** $frame \in \{1, ..., N\}$ **do**
6: $\quad (x, y) \leftarrow (x, y) + (d_x, d_y)$
7: $\quad sequence[frame] \leftarrow gen\_frame((x, y))$
8: **end for**
9: **return** $sequence$

Algorithm 2 displays the *Constant Trajectory* MovingMNIST sequence generation. Given the same set of inputs as Algorithm 1, we sample a starting position $(x, y) \sim U(0, F-D)$ and a spatial frame-wise shift (d_x, d_y) (lines $3-4$). For every frame, the same pair (d_x, d_y) is applied to the current (x, y) position to generate the next frame (lines $5-7$). The conditional sampling strategy is not utilized, with digits possibly reaching and crossing the frame boundaries. Utilizing the same shift (d_x, d_y) across frames and not applying any direction changes close to the frame edges provide a frame-wise consistent trajectory across the whole sequence. This characteristic enables the trajectory analysis conducted in Section 4 (Figure 3) for the rate regime with X_1 encoded with a high rate

(Subsection 3.2). We utilize sequence length $N = 3$, frame size $F = 64$, and digit size $D = 32$. For sharp movements (Fig. 3a), we have a maximum step size $S = 20$. For slow movements (Fig. 3b), we utilize maximum step size $S = 5$.

E.2 Additional Results

We display additional results to the experimental discussion in (Section 4). Figure 7 contains additional visualizations for the regime where the first frame is encoded with a high rate (Subsection 3.2). The same behavior as the one addressed in Section 4 is observed in both sharp (Figure 7a) and slow (Figure) movement scenarios. In the sharp movement scenario, 0-PLF-JD propagates the wrong trajectory in $\hat{X}_2$ to the following frame $\hat{X}_3$, with 0-PLF-AR being able to recover from the previous mistake. 0-PLF-FMD presents a different behavior for each setup. For sharp movements (low correlation coefficient ρ), $\hat{X}_2$ is reconstructed as the wrong digit. For slow movements (high correlation coefficient ρ), $\hat{X}_2$ is reconstructed as the correct digit, although demonstrating a positional copying behavior.



(a) Sharp movement scenario. **(b)** Slow movement scenario.

Figure 7: MovingMNIST reconstructions for ∞-R_2-R_3 with $R_2 = 2$ bits and $R_3 = 16$ bits. Digits are coloured for easily visualizing the trajectory across frames.

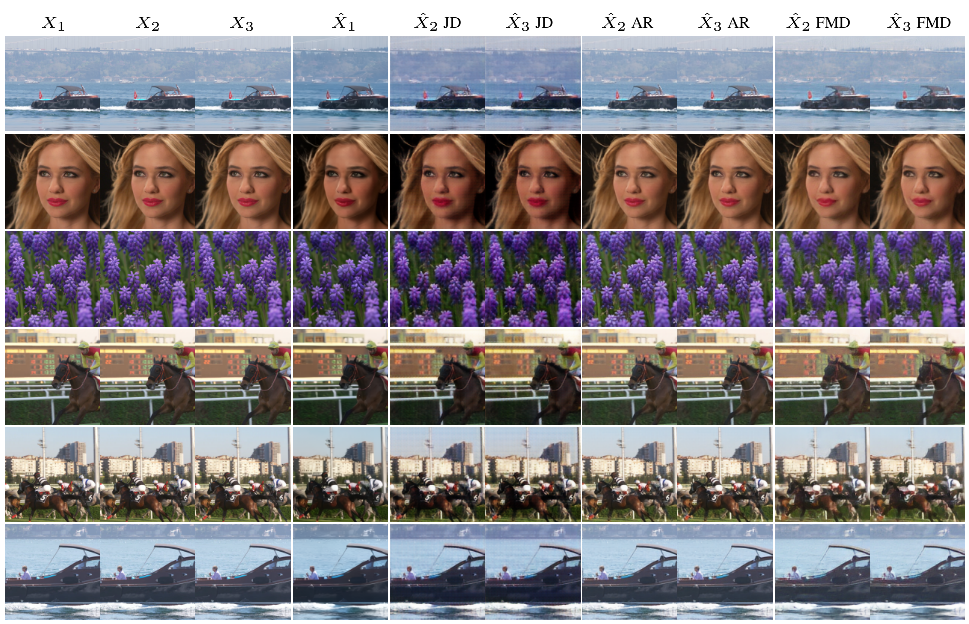


Figure 8: The outputs of different PLFs on the UVG dataset when the first frame is compressed at a low rate. The first reconstructed frame $\hat{X}_1$ is shared across all PLFs, with PLF-JD propagating the distorted color tone.

Figure 8 presents additional results for the regime where the first frame is encoded with a low rate (Subsection 3.1). For each UVG sequence, the first reconstructed frame $\hat{X}_1$ presents a distorted color tone. This compression artifact is propagated by 0-PLF-JD to $\hat{X}_2$ and $\hat{X}_3$ once again. The

0-PLF-AR and 0-PLF-FMD model variants are able to correct the error. Additional MovingMNIST results for the same rate regime sequences are displayed in Figure 9. Here, a wrong digit (i.e., digit class or shape) is reconstructed in $\hat{X}_1$ implied by its low rate. Similarly, 0-PLF-JD propagates the compression artifact to $\hat{X}_2$ and $\hat{X}_3$, with 0-PLF-AR and 0-PLF-FMD correcting the mistake.

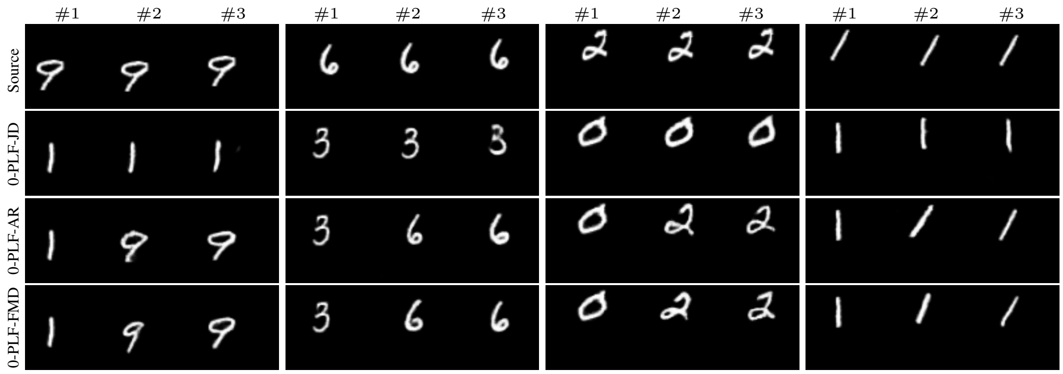


Figure 9: Outputs of different PLFs for the MovingMNIST dataset when the first frame is compressed at a low rate. Both PLF-AR and PLF-FMD recover from previous mistakes while PLF-JD suffers from the error permanence phenomenon.