On the Choice of Perception Loss Function for Learned Video Compression

Sadaf Salehkalaibar[*]
ECE Department
University of Toronto
sadafs@ece.utoronto.ca

Buu Phan[*]
ECE Department
University of Toronto
truong.phan@mail.utoronto.ca

Jun Chen
ECE Department
McMaster University
chenjun@mcmaster.ca

Wei Yu
ECE Department
University of Toronto
weiyu@ece.utoronto.ca

Ashish Khisti
ECE Department
University of Toronto
akhisti@ece.utoronto.ca

Abstract

We study causal, low-latency, sequential video compression when the output is subjected to both a mean squared-error (MSE) distortion loss as well as a perception loss to target realism. Motivated by prior approaches, we consider two different perception loss functions (PLFs). The first, PLF-JD, considers the joint distribution (JD) of all the video frames up to the current one, while the second metric, PLF-FMD, considers the framewise marginal distributions (FMD) between the source and reconstruction. Using information theoretic analysis and deep-learning based experiments, we demonstrate that the choice of PLF can have a significant effect on the reconstruction, especially at low-bit rates. In particular, while the reconstruction based on PLF-JD can better preserve the temporal correlation across frames, it also imposes a significant penalty in distortion compared to PLF-FMD and further makes it more difficult to recover from errors made in the earlier output frames. Although the choice of PLF decisively affects reconstruction quality, we also demonstrate that it may not be essential to commit to a particular PLF during encoding and the choice of PLF can be delegated to the decoder. In particular, encoded representations generated by training a system to minimize the MSE (without requiring either PLF) can be *near universal* and can generate close to optimal reconstructions for either choice of PLF at the decoder. We validate our results using (one-shot) information-theoretic analysis, detailed study of the rate-distortion-perception tradeoff of the Gauss-Markov source model as well as deep-learning based experiments on moving MNIST and KTH datasets. Code will be available at https://github.com/truongbuu/URDP_flow.

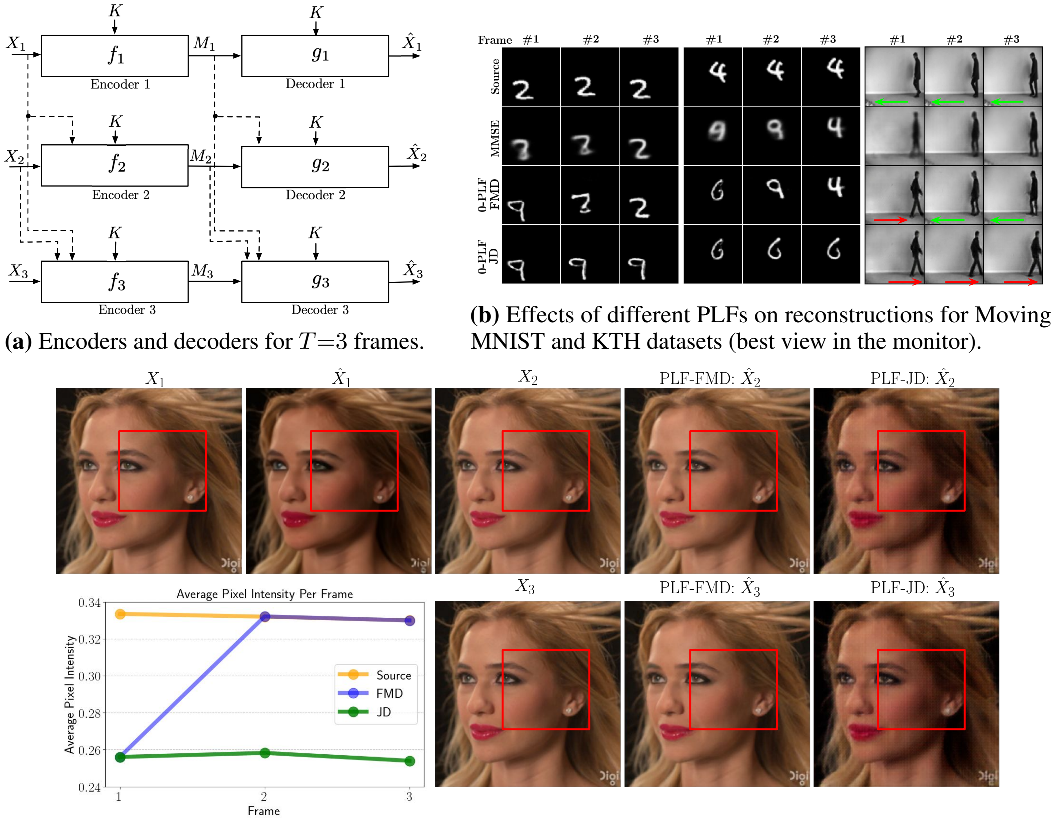
1 Introduction

There is an increasing demand for video compression algorithms that are able to generate visually pleasing videos at low bitrates. Most of the current video codecs use distortion measures such as PSNR [1–4], MSE and MS-SSIM [3–5] to generate reconstructions which tend to be blurry at extremely low bitrates. In recent years, there has been a growing interest (see e.g., [6–10]) in using deep generative models to make the reconstructions look more realistic. Such techniques introduce an additional perception loss function that measures a distance between distributions of the source and reconstruction, with *perfect* perception corresponding requiring that the two distributions be identical.

[*] Equal Contribution

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

https://doi.org/10.52202/075280-2093



(a) Encoders and decoders for $T{=}3$ frames.

(b) Effects of different PLFs on reconstructions for Moving MNIST and KTH datasets (best view in the monitor).

(c) Error permanence on the UVG dataset. The PLF-JD reconstructions propagate the flaws in the color tone from the previous I-frame reconstruction while the decoder based on PLF-FMD is able to fix these flaws.

Figure 1: (a) Proposed System Model (b, c) Error permanence phenomenon under different PLFs. High fidelity but incorrect I-frame reconstruction propagates the error to subsequent P-frames in 0-PLF-JD reconstructions. The MMSE and 0-PLF-FMD reconstructions do not have this problem.

In compression systems, improving realism comes at the price of increasing distortion. The work of Blau and Michaeli [11] establishes the theoretical rate-distortion-perception (RDP) tradeoff which has also been validated in [12–15]. Furthermore *universal* encoded representations were proposed in [16] where the representation is fixed at the encoder and the decoder is adapted to achieve a performance near the optimal RDP tradeoff curve. The extension of these works to video compression involves many challenges. First, the compression system must not only account for spatial redundancy as in image compression, but also exploit the temporal redundancy across video frames, making the system design more complex. Secondly, unlike the case of image compression, there may be no clear choice of the perception loss function (PLF). Indeed, some prior works [7] consider PLF that preserves framewise marginal distribution (PLF-FMD) between the source and reconstruction, while other works consider joint distribution (PLF-JD) across multiple frames [9].

As illustrated in Fig. 1a, we study causal, low-latency, sequential video compression when the output is subjected to both a mean squared-error (MSE) distortion loss and either a PLF-JD or PLF-FMD metric for perception loss. Our contributions are as follows:

- *Differences in reconstruction quality based on the choice of PLF*: We demonstrate that the choice of PLF can decisively affect the reconstruction quality especially in the low bit-rate regime. We approximately characterize the operational RDP region on a per-frame basis for a first-order Markov source model and analyze the special case of Gauss-Markov sources in detail. We show that there is a significant penalty in distortion when using PLF-JD in the low-rate regime. On the experimental side, we demonstrate that while PLF-JD preserves better temporal consistency across video frames, it suffers from the *permanence of error* phenomenon in which the mistakes in reconstructions propogate to future frames. On the other hand, the PLF-FMD metric shows

https://doi.org/10.52202/075280-2093

more capability in correcting mistakes across frames (see Fig. 1b for visualizations involving three-frame videos). On the other hand, if the first frame is transmitted at high bit-rate, we demonstrate that PLF-JD performs better than PLF-FMD.

- *Universality of minimum mean square error (MMSE) reconstructions*: We demonstrate that encoded representations generated from an encoder trained to minimize MSE reconstruction (without considering any PLF) suffice to produce close-to-optimal reconstructions for either choice of PLF. For general sources, we show that when using PLF-FMD, the MMSE reconstruction can be transformed to a reconstruction satisfying perfect perceptual quality by increasing the distortion at most by a *factor of two*. While a similar result does not hold for PLF-JD in general, it is satisfied for a special class of encoders which operate in the low-rate regime. For the Gauss-Markov source model we demonstrate exact universality i.e., encoded representation for the MMSE reconstruction can be adapted to achieve any other reconstruction in the RDP region. We also use deep learning based experiments to provide experimental evidence of these results.

  We note that the above notion of universal encoded representations based on MSE reconstruction can have significant advantages in practice. First, although the reconstructions associated with different choices of PLFs can be visually very different, universal representations delegate the choice of PLF to the decoder rather than requiring the encoder to commit to a specific PLF. Secondly, the universality of MMSE reconstructions is far more significant in the context of learned video compression. Given that perceptual reconstruction frequently generates novel details not present in the source frame, compressing motion flow vectors between the current frame and the prior perceptual reconstruction necessitates a higher bit allocation compared to utilizing the MMSE reconstruction. Hence, a recommended approach is to train end-to-end compression exclusively to minimize MSE and use our proposed scheme to achieve a (near-optimal) tradeoff between the distortion and perception losses as desired by the user.

The study of RDP region for learned video compression is considerably more challenging than the study of RDP function for a single frame in prior works ([11, 16]). This is because of the fact that the RDP region (for first-order Markov sources) involves a tradeoff between the compression rate assigned to each frame imposing a Markov structure on the reconstructions. For Gaussian sources, the proof of optimality of Gaussian reconstructions does not use the closed form of the RDP region due to its complexity. As a result, the study of RDP region for various operating regimes is quite more involved. Furthermore, for the proof of universality, one has to consider the achievability of the entire RDP region as opposed to just points on the boundary of RDP function in [16]. Finally the results on the fixed-encoder setup are more general than prior works ([11, 16]) that required a characterization of the information RDP region, which we do not require.

Related Work

Perceptual Lossy Video Compression. Distribution preserving framework using Generative Adversarial Networks (GAN) has been widely adopted as a surrogate metric for perceptual quality in image [12, 17–19] and, recently, video compression [6–9]. Unlike image compression, where the choice of PLF is straightforward, there is currently no agreed-upon objective for lossy video compression. For instance, DVC-P [6] employs PLF-FMD to improve the visual quality per frame, ignoring the temporal coherence. Similarly, Mentzer et al. [7] utilize the per-frame metric with a conditional GAN model, and found no significant differences when using a GAN objective with multiple frames. Other works target temporal consistency by incorporating multiple frames in their GAN objective. This includes the work by Yang et al. [8], where every two consecutive frames are included, and by Veerabadran et al. [9], where they employ the PLF-JD metric in a non-causal setting. Unlike previous works, we study the impact of two different perception objectives, i.e. PLF-JD and PLF-FMD, on reconstructions in the causal setting, presenting theoretical properties that are verified by deep learning experiments. For the PLF-JD metric, we demonstrate the *error permanence phenomenon*, which, unlike the error propagation issue [18, 20], cannot be resolved by increasing the code rate assigned to the P frames.

RDP Tradeoff and Principle of Universality. Targeting the distribution preserving framework in lossy image compression, several theoretical works have shown the presence of RDP tradeoff [11, 21–23], where perfect perception comes at a cost of increasing distortion by at most a factor of 2. Furthermore, an encoder that generates universal representations exists [16, 24, 25], which enables the decoder to freely choose the level of distortion-perception tradeoff it desires. Our work explores these avenues

https://doi.org/10.52202/075280-2093

in the context of causal video compression. As discussed previously we show that the MMSE representation can be used as a universal representation for both perception metrics which has several advantages in the context of video compression..

2 System Model

Let $(X_1, \ldots, X_T) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_T$ be T frames in a video (with each $\mathcal{X}_i \subseteq \mathbb{R}^d$) distributed according to $P_{X_1 \ldots X_T}$. The frames are available for encoding sequentially; X_1 is available first, then X_2 arrives, followed by X_3 and so on. There is a shared randomness $K \in \mathcal{K}$ which is available at all encoders and decoders. The following (possibly stochastic) mappings define the encoding and decoding functions:

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

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

where $\mathcal{M}_j \in \{0,1\}^\star$ denotes the set of (variable-length) messages assigned by the jth encoder and $\hat{\mathcal{X}}_j \subseteq \mathbb{R}^d$ is the j-th reconstruction alphabet (see Fig. 1a). Let $P_{\hat{X}_1 \ldots \hat{X}_T | X_1 \ldots X_T}$ be the conditional distribution of the reconstructed video given the original video which is basically determined by the mappings $\{f_j\}_{j=1}^T$ and $\{g_j\}_{j=1}^T$. The above setting is a *one-shot* setup as only a single source sample is compressed at a time. For each frame j, a distortion metric is imposed on the output, which we assume throughout is the mean squared-error (MSE) function i.e. $d(x_j, \hat{x}_j) = ||x - \hat{x}_j||^2$, which is commonly used in many applications. From a perceptual point of view, for given probability distributions $P_{X_1 \ldots X_j}$ and $P_{\hat{X}_1 \ldots \hat{X}_j}$ on the original and reconstructed frame j, let $\phi_j(P_{X_1 \ldots X_j}, P_{\hat{X}_1 \ldots \hat{X}_j})$ be the perception function capturing the difference between them. Note that the function ϕ_j is defined based on the joint distribution of all first j frames. We call this metric as *perception loss function based on joint distribution (PLF-JD)*. Note that when $\phi_j(P_{X_1 \ldots X_j}, P_{\hat{X}_1 \ldots \hat{X}_j}) = 0$, we have:

$$P_{X_1 \ldots X_j} = P_{\hat{X}_1 \ldots \hat{X}_j}, \qquad j = 1, \ldots, T. \tag{3}$$

We refer to this case as *zero-perception loss function based on joint distribution (*0*-PLF-JD)*. Alternatively, the *perception loss function based on framewise marginal distribution (PLF-FMD)* is denoted by $\xi_j(P_{X_j}, P_{\hat{X}_j})$ and is based on only the marginal distribution of the j-th frame. In particular, note that 0-PLF-FMD implies that $P_{X_j} = P_{\hat{X}_j}$ for each j. In most of the paper, for simplicity of presentation, we provide some of our results for $T = 3$ frames. In that case, we use the shorthand notation X to denote the tuple (X_1, X_2, X_3), e.g., $\mathsf{M} := (M_1, M_2, M_3)$, $\mathsf{D} := (D_1, D_2, D_3)$, $\mathsf{f} := (f_1, f_2, f_3)$.

3 Distortion Analysis for a Fixed Encoder and Zero-perception Loss

In this section, we assume that the encoding functions f are fixed, but the decoding functions g can be optimized to generate different reconstructions. Equivalently, the distribution $P_{\mathsf{M}|\mathsf{X}K} := \mathbb{1}\{\mathsf{M} = \mathsf{f}(\mathsf{X}, K)\}$ is fixed, while by varying the reconstruction distribution $P_{\hat{\mathsf{X}}|\mathsf{M}K} := \mathbb{1}\{\hat{\mathsf{X}} = \mathsf{g}(\mathsf{M}, K)\}$, one attains different reconstructions $\hat{\mathsf{X}}$, where $\mathbb{1}\{.\}$ denotes the indicator function. Furthermore defining $D_j := \mathbb{E}_P[\|X_j - \hat{X}_j\|^2]$, we denote D as the achievable distortion tuple associated with $P_{\hat{\mathsf{X}}|\mathsf{M}K}$.

One natural choice of reconstructions is the minimum mean squared error (MMSE) reconstruction function. At step j, the reconstruction, which we denote in this case by $\tilde{X}_j$, is obtained by taking the conditional expectation of X_j given all information at the decoder up to time j i.e., $\tilde{X}_j := \mathbb{E}_P[X_j | M_1 \ldots M_j, K]$ for each $j = 1, 2, 3$. It is well known that the MMSE reconstruction functions minimize the reconstruction distortion i.e., if we define the set

$$\Phi_{\mathsf{D}^{\min}}(P_{\mathsf{M}|\mathsf{X}K}) = \{\mathsf{D} : D_j \geq \mathbb{E}_P[\|X_j - \tilde{X}_j\|^2], \qquad j = 1, 2, 3\} \tag{4}$$

then the distortion tuple D associated with any reconstruction $P_{\hat{\mathsf{X}}|\mathsf{M}K}$ satisfies $\mathsf{D} \in \Phi_{\mathsf{D}^{\min}}(P_{\mathsf{M}|\mathsf{X}K})$.

The main result of this section is that assuming fixed encoder, the achievable distortions under 0-PLF-FMD is at most twice of that under the MMSE distortion loss alone. The same conclusion also holds for 0-PLF-JD for a class of encoders operating at low rate. We first consider the case of 0-PLF-FMD.

Definition 1 (0-PLF-FMD Distortion) *For an encoder $P_{\mathsf{M}|\mathsf{X}K}$, the set $\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$ denotes the set of all distortion tuples D for which there exists a reconstruction $P_{\hat{\mathsf{X}}|\mathsf{M}K}$ satisfying $P_{X_j} = P_{\hat{X}_j}$ for each $j \in \{1, 2, 3\}$.*

Theorem 1 *The set $\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$ is characterized as follows:*

$$\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K}) = \{\mathsf{D} : D_j \geq \mathbb{E}_P[\|X_j - \tilde{X}_j\|^2] + W_2^2(P_{\tilde{X}_j}, P_{X_j}),\ j = 1, 2, 3\}, \tag{5}$$

where $W_2^2(P_{X_j}, P_{\hat{X}_j})$ denotes the Wasserstein-2 distance between the two distributions [26]. Furthermore, we also have that:

$$\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K}) \supseteq \{\mathsf{D} : D_j \geq 2\mathbb{E}_P[\|X_j - \tilde{X}_j\|^2], \quad j = 1, 2, 3\}, \tag{6}$$

i.e., minimum achievable distortion with 0-PLF-FMD is at most twice the MMSE distortion.

Proof: See Appendix A. ■

We remark that the proof of Theorem 1, operationally demonstrates that the MMSE reconstruction can be converted to another reconstruction satisfying 0-PLF-FMD with at-most a factor of 2 increase in distortion, generalizing the result in [16] for the single frame scenario (see also [21]).

We next consider the case when zero perception loss is satisfied under the PLF-JD metric. Analogous to $\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$ in Definition 1, one can define $\Phi^{\text{joint}}_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$ to be the set of distortions associated with reconstruction functions that satisfy (3). The analysis of $\Phi^{\text{joint}}_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$ is discussed in Appendix B as it is more involved. In general, the *factor of two bound* as in Theorem 1 cannot be realized in this case as demonstrated by a counter-example in Appendix B. Nevertheless, for a special family of encoders we can obtain a counterpart of Theorem 1. In this family of encoders, the source X_j at time j is nearly independent of the encoder outputs up to and including time j, i.e., we can express:

$$P^{\text{noisy}}_{X_j|M_1\ldots M_j K} = (1-\mu)P_{X_j} + \mu Q^{\text{noisy}}_{X_j|M_1\ldots M_j K}, \qquad j = 1, 2, 3. \tag{7}$$

where μ is a sufficiently small constant and the distribution $Q^{\text{noisy}}(\cdot)$ could be arbitrary conditional distribution with same marginal as P_{X_j}. We note that such encoders are studied in a variety of problems in information theory (see e.g., [27]) that correspond to the low rate operating regime. The following result states that the factor-two bound holds approximately for such encoders.

Theorem 2 *For the class of encoders given by* (7), *we have*

$$\Phi^{\textit{joint}}_{\mathsf{D}^0}(P^{\textit{noisy}}_{\mathsf{M}|\mathsf{X}K}) \supseteq \{\mathsf{D} : D_j \geq 2\mathbb{E}_{P^{\textit{noisy}}}[\|X_j - \tilde{X}_j\|^2] + O(\mu), \quad j = 1, 2, 3\}. \tag{8}$$

Proof: See Appendix C. ■

We note that the low-rate operating regime is practically important, as at higher rates MMSE based reconstructions can suffice and the use of PLF metrics may be less relevant.

4 Rate-Distortion-Perception Region

In this section, we assume that both the encoder $P_{\mathsf{M}|\mathsf{X}K}$ as well as the reconstruction $P_{\hat{\mathsf{X}}|\mathsf{M}K}$ can be optimized and study the associated rate-distortion-perception (RDP) tradeoff. We remind the reader that for PLF-JD and PLF-FMD, the PLFs are denoted by $\phi_j(P_{X_1\ldots X_j}, P_{\hat{X}_1\ldots\hat{X}_j})$ and $\xi_j(P_{X_j}, P_{\hat{X}_j})$, respectively. In this case, the operational RDP region in the one-shot setting is defined as follows.

Definition 2 (Operational RDP region) *For a given P_{X}, an RDP tuple $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ is said to be achievable for the one-shot setting if there exist an encoder $P_{\mathsf{M}|\mathsf{X}K}$ and a reconstruction $P_{\hat{\mathsf{X}}|\mathsf{M}K}$ satisfying:*

$$\mathbb{E}[\ell(M_j)] \leq R_j, \quad \mathbb{E}[\|X_j - \hat{X}_j\|^2] \leq D_j, \quad \phi_j(P_{X_1\ldots X_j}, P_{\hat{X}_1\ldots\hat{X}_j}) \leq P_j, \ \ j = 1, 2, 3, \tag{9}$$

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

The region $\mathcal{C}^o_{\mathsf{RDP}}$ cannot be directly computed as it involves all possible one-shot encoders/decoders. But for first-order Markov source, it has a tractable approximation in terms of mutual information.

https://doi.org/10.52202/075280-2093

4.1 RDP Region of First-Order Markov Sources

We first define the first-order Markov sources and then introduce an iRDP region which is computable.

Definition 3 *We call* X *as a first-order Markov source if the Markov chain* $X_1 \to X_2 \to X_3$ *holds.*

Definition 4 (Information RDP region) *For first-order Markov sources, the information RDP (iRDP) region, denoted by* $\mathcal{C}_{\mathsf{RDP}}$*, is the set of all tuples* $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ *which satisfy the following*

$$R_1 \geq I(X_1; X_{r,1}), \qquad R_2 \geq I(X_2; X_{r,2}|X_{r,1}), \qquad R_3 \geq I(X_3; X_{r,3}|X_{r,1}, X_{r,2}) \tag{10}$$

$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j\|^2], \qquad P_j \geq \phi_j(P_{X_1 \ldots X_j}, P_{\hat{X}_1 \ldots \hat{X}_j}), \qquad j = 1, 2, 3, \tag{11}$$

for auxiliary random variables $(X_{r,1}, X_{r,2}, X_{r,3})$ *and* $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ *satisfying the following*

$$\hat{X}_1 = \eta_1(X_{r,1}), \;\; \hat{X}_2 = \eta_2(X_{r,1}, X_{r,2}), \;\; \hat{X}_3 = X_{3,r}, \tag{12}$$

$$X_{r,1} \to X_1 \to (X_2, X_3), \;\; X_{r,2} \to (X_2, X_{r,1}) \to (X_1, X_3), \tag{13}$$

$$X_{r,3} \to (X_3, X_{r,1}, X_{r,2}) \to (X_1, X_2), \tag{14}$$

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

The expression for the iRDP region involves a search over auxiliary random variables X_r and $\hat{\mathsf{X}}$ that satisfy (12)-(14) subject to the constraints in (10)–(11). For first-order Markov sources, the following theorem states that the operational DP rate region can be approximated by the iDP rate region.

Theorem 3 *For first-order Markov sources, a given* (D, P) *and* $\mathsf{R} \in \mathcal{R}(\mathsf{D}, \mathsf{P})$*, we have*

$$\mathsf{R} + \log(\mathsf{R} + 1) + 5 \in \mathcal{R}^o(\mathsf{D}, \mathsf{P}) \subseteq \mathcal{R}(\mathsf{D}, \mathsf{P}). \tag{15}$$

Proof: See Appendix D. ■

From Theorem 3, it follows that $\mathcal{R}(\mathsf{D}, \mathsf{P})$ with overhead $\log(\mathsf{R}+1)+5$ is an inner bound to $\mathcal{R}^o(\mathsf{D}, \mathsf{P})$. On the other hand, $\mathcal{R}(\mathsf{D}, \mathsf{P})$ provides an outer bound to $\mathcal{R}^o(\mathsf{D}, \mathsf{P})$. The two bounds match with each other in high rates. It can be shown that the overhead also vanishes in the large-blocklength setting where multiple symbols are encoded at a time. In the remainder of this paper, we will approximate $\mathcal{R}^o(\mathsf{D}, \mathsf{P})$ with $\mathcal{R}(\mathsf{D}, \mathsf{P})$ and use the latter region for our analysis.

Remark 1 *(Encoded Representations): The proof of the inner bound in Theorem 3 in Appendix D provides an operational interpretation to the auxiliary random variables* $\mathsf{X}_r = (X_{r,1}, X_{r,2}, X_{r,3})$ *defined in iRDP region in Definition 4. In particular,* $X_{r,j}$ *is a lossy version of the source sample* X_j *generated by the encoder in step* j*. It is compressed and transmitted to the decoder at rate* R_j *in* (10)*. We refer to* X_r *as the* encoded representation *of the source* X*. The Markov chains* (13)–(14) *indicate that without loss of optimality, an encoded representation* $X_{r,j}$ *can be computed from the source* X_j *and past reconstructions* $X_{r,1}, \ldots, X_{r,j-1}$ *without using past source samples* $X_1, \ldots, X_{j-1}$*.*

Remark 2 *(Deterministic Reconstructions): Note that the reconstruction functions generating* $\hat{\mathsf{X}}$ *in Definition 4 are deterministic functions of the encoded representations (c.f.* (12)*). In particular, the shared randomness* K *is not required in the reconstruction functions. However, as the proof of the inner bound of Theorem 3 illustrates, the shared randomness is required in the compression and construction of* $X_{r,j}$*. Moreover, by following the arguments in [28], one can set the reconstruction function of the last frame to be identity. Thus, in Definition 4, we have set* $\hat{X}_3 = X_{r,3}$ *in* (12) *where* $T = 3$*. In the sequel, for T frames we will set* $\hat{X}_T = X_{r,T}$*.*

Remark 3 *The result in Theorem 3 also holds for the PLF-FMD. That is, one can replace the PLF in* (9) *and* (11) *by* $\xi_j(P_{X_j}, P_{\hat{X}_j})$ *and get a similar result (see Appendix D for the justification).*

4.2 Gauss-Markov Source Model: RDP Region

In this section, we obtain practical insights through the analysis of the special case of first-order Gauss-Markov sources. We assume that $X_1 \sim \mathcal{N}(0, \sigma_1^2)$,

$$X_2 = \rho_1 \frac{\sigma_2}{\sigma_1} X_1 + N_1, \qquad X_3 = \rho_2 \frac{\sigma_3}{\sigma_2} X_2 + N_2, \tag{16}$$

where N_j is independent of X_j with mean zero and variance $(1-\rho_j^2)\sigma_{j+1}^2$ for $j=1,2$. Note that the model extends naturally to the case of T time-steps. The perception metric is assumed to be the Wasserstein-2 distance, i.e., $\phi_j(P_{X_1\ldots X_j}, P_{\hat{X}_1\ldots\hat{X}_j}) := W_2^2(P_{X_1\ldots X_j}, P_{\hat{X}_1\ldots\hat{X}_j})$. For the PLF-FMD, the perception metric is given by $\xi_j(P_{X_j}, P_{\hat{X}_j}) := W_2^2(P_{X_j}, P_{\hat{X}_j})$.

The following result states that for Gaussian sources, jointly Gaussian reconstructions are optimal. Thus, for a given tuple (D,P), the characterization of $\mathcal{R}(\mathsf{D},\mathsf{P})$ becomes computable.

Theorem 4 *For the Gauss-Markov source model, any tuple* $(\mathsf{R},\mathsf{D},\mathsf{P})\in\mathcal{C}_{\mathsf{RDP}}$ *can be attained by a jointly Gaussian distribution over* X_r *and identity mappings for* $\eta_j(\cdot)$ *in Definition 4.*

Proof: See Appendix E. ■

Generally, the optimized distribution in the above theorem may not admit a simple form. In the special case of $T=2$ frames, the optimal reconstructions are given in Appendix E. To obtain practical insights, we consider various asymptotic operating regimes and provide a detailed analysis in Appendix F for the case of $T=2$ frames and with $\sigma_1^2=\sigma_2^2$. A summary of these results is provided in Table 2 in the same Appendix. We briefly summarize some of these results next.

4.3 Gauss-Markov Source Model: Extremal Rates

One of the key observations of this paper is that the choice of PLF has implication on the rate allocation across different frames. Specifically, first consider the case when both $R_1{=}R_2$ are small i.e., $R_1{=}R_2{=}\epsilon$ (for small enough ϵ). We discuss how each PLF affects the reconstruction in the second step. In the first step, we note that reconstruction in both cases must be identical and of the form $\hat{X}_1^G{=}\sqrt{2\epsilon\ln 2}X_1{+}Z_1$ where $Z_1{\sim}\mathcal{N}(0,(1{-}2\epsilon\ln 2)\sigma^2)$ is independent of X_1; the resulting distortion is given by $D_1{=}2(1{-}\sqrt{2\epsilon\ln 2})\sigma^2$. However, the reconstructions in the second steps will be different for the two measures. For simplicity, we consider the extreme case when $\rho{=}1$ (i.e., when $X_2{=}X_1$). Here, the PLF-JD metric is required to preserve perfect correlation and thus has to set $\hat{X}_2^G{=}\hat{X}_1^G$ and results in $D_2{=}D_1$. In other words, the decoder in the second step is unable to use any information transmitted in the second step as 0-PLF-JD enforces the stringent constraint $\hat{X}_2^G{=}\hat{X}_1^G$. In contrast, for the PLF-FMD metric, it can be shown that the reconstruction in the second step for $\rho{=}1$ reduces to $\hat{X}_2^G{=}\sqrt{2}\sqrt{2\epsilon\ln 2}X_1{+}Z_{\text{FMD}}$ and the associated distortion is given by $D_2{=}2(1{-}\sqrt{4\epsilon\ln 2})\sigma^2$, which is lower than PLF-JD. Extending this example to T steps (with $\rho{=}1$), we note that PLF-JD will always be forced to output $\hat{X}_1$, while the reconstruction using PLF-FMD will successively improve. The following theorem formalizes this observation.

Theorem 5 *For sufficiently small ϵ, let $R_j=\epsilon$ and suppose that $\rho_j=\rho$ and $\sigma_j=\sigma$, for $j=1,\ldots,T$. The achievable distortions $D_{\mathit{FMD},j}$ (for 0-PLF-FMD) and $D_{\mathit{JD},j}$ (for 0-PLF-JD) are:*

$$D_{\mathit{FMD},j}=2(1-\Delta_{\mathit{FMD},j}\sqrt{2\epsilon\ln 2})\sigma^2,\quad D_{\mathit{JD},j}=2(1-\Delta_{\mathit{JD},j}\sqrt{2\epsilon\ln 2})\sigma^2,\tag{17}$$

where $\Delta_{\mathit{FMD},j}:=\sqrt{1+\rho^2\frac{(2\rho^2)^{j-1}-1}{2\rho^2-1}}$ *and* $\Delta_{\mathit{JD},j}:=\rho^{2(j-1)}+\mathbb{1}\{j\geq 2\}\cdot\sqrt{1-\rho^2}(\sum_{i=0}^{j-2}\rho^{2i})$.

Proof: See Appendix G. ■

In particular, specializing to $\rho=1$, $\Delta_{\text{FMD},j}=2^{\frac{j-1}{2}}$ and $\Delta_{\text{JD},j}=1$. This shows that the decrease in $D_{\text{FMD},j}$ is exponential at each step which implies the ability of decoder based on 0-PLF-FMD in correcting mistakes and not propagating them in future reconstructions. However, as discussed previously the decoder which uses 0-PLF-JD is stuck at $\hat{X}_j=\hat{X}_1$ when $\rho=1$ and results in no subsequent improvement in the distortion. We call this behaviour as *permanence of error*. This phenomenon is magnified in the case when $R_1\to 0$ and $R_2\to\infty$, treated in Table 2 (Appendix F) as the PLF-JD severely constrains the decoder to copy the previous noisy reconstruction while the flexibility provided by PLF-FMD reduces the distortion in the second step.

The case when $R_1\to\infty$ and $R_2=\epsilon$ treated in Table 2 in Appendix F corresponds to the case when X_1 is sent at a sufficiently high rate (as is the case with some I-frames) while X_2 is sent at a low rate. Naturally, we have $\hat{X}_1^G\approx X_1$ for both PLFs. On the other hand, we once again see a qualitatively different behaviour in the reconstruction of X_2. For the case of 0-PLF-FMD, we have $\hat{X}_2^G\approx(1-O(\epsilon))\hat{X}_1^G+O(\epsilon)X_2$, i.e., the decoder essentially copies the previous frame with

https://doi.org/10.52202/075280-2093

little contribution from the second step. In contrast, for the case of 0-PLF-JD, it can be shown that $\hat{X}_2^G \approx \rho_1 \hat{X}_1^G + O(\sqrt{\epsilon})X_2 + Z_{\text{JD}}$, where Z_{JD} is independent Gaussian noise with variance close to $1-\rho^2$. We note that the PLF-JD metric prevents the decoder from simply "copying" the previous frame, but instead forces the decoder to generate a more diverse representation consistent with the joint distribution between the two frames.

4.4 Universal Representations for Gauss-Markov Source Model

In this section, we show that the Gauss-Markov source model admits universal encoded representations. Such representations can be transformed through appropriate reconstruction functions to achieve the entire DP rate region. This is the counterpart of the result for general sources in Theorem 1 where it is shown that the MMSE reconstructions can be transformed to some target reconstructions satisfying the 0-PLF-FMD with at most a factor-2 increase in distortion. In contrast, we demonstrate that the Gauss-Markov model admits *exact universality* i.e., target reconstructions proposed in this section achieve all points in the iDP rate region. Interestingly, the transformation is linear with possibly some additive noise. First, we formalize the notion of universal representations.

Definition 5 (iDP-Tradeoff) *For a given rate tuple* R*, the optimal iDP-tradeoff is the closure of the set of all tuples* (D, P) *such that* $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \mathcal{C}_{\mathsf{RDP}}$ *and is denoted by* $\mathcal{DP}(\mathsf{R})$.

Definition 6 (Universal Representation) *A given encoded representation* X_r *is called universal with respect to rate tuple* R *if it satisfies the rate constraints* (10) *and the Markov chains in* (13)–(14) *and for each* $(\mathsf{D}, \mathsf{P}) \in \mathcal{DP}(\mathsf{R})$*, there exists a reconstruction* $\hat{\mathsf{X}}$ *generated from* $P_{\hat{\mathsf{X}}|\mathsf{X}_r}$ *achieving it.*

For the Gauss-Markov source model, we show that the MMSE reconstruction admits a universal representation. We consider the reconstruction $\hat{\mathsf{X}}_r$ that achieves minimum distortion in the $\mathcal{DP}(\mathsf{R})$ region. This point is explicitly characterized in Appendix H.1. Furthermore, following Theorem 4, since the reconstruction functions $\eta_j(\cdot)$ are identity, the MMSE reconstruction is equivalent to MMSE representation i.e., $\hat{\mathsf{X}}_r = \mathsf{X}_r^{\mathsf{RD}}$.The following theorem establishes that any point in $\mathcal{DP}(\mathsf{R})$ can be achieved from $\hat{\mathsf{X}}_r$.

Theorem 6 *For the Gauss-Markov source model and a given rate tuple* R *with strictly positive components, let the MMSE representation be denoted as* $\mathsf{X}_r^{RD} = (X_{r,1}^{RD}, X_{r,2}^{RD}, X_{r,3}^{RD})$*. Let* $(\mathsf{D}, \mathsf{P}) \in \mathcal{DP}(\mathsf{R})$ *and let* $\hat{\mathsf{X}} = (\hat{X}_1, \hat{X}_2, \hat{X}_3)$ *be the corresponding reconstruction achieving it. Then there exist* κ_1*,* θ_1*,* θ_2*,* ψ_1*,* ψ_2 *and* ψ_3 *and noise variables* (Z_1, Z_2, Z_3) *independent of* $(X_{r,1}^{RD}, X_{r,2}^{RD}, X_{r,3}^{RD})$*, which satisfy the following*

$$\hat{X}_1 = \kappa_1 X_{r,1}^{RD} + Z_1, \quad \hat{X}_2 = \theta_1 X_{r,1}^{RD} + \theta_2 X_{r,2}^{RD} + Z_2, \quad \hat{X}_3 = \psi_1 X_{r,1}^{RD} + \psi_2 X_{r,2}^{RD} + \psi_3 \hat{X}_{r,3}^{RD} + Z_3.$$

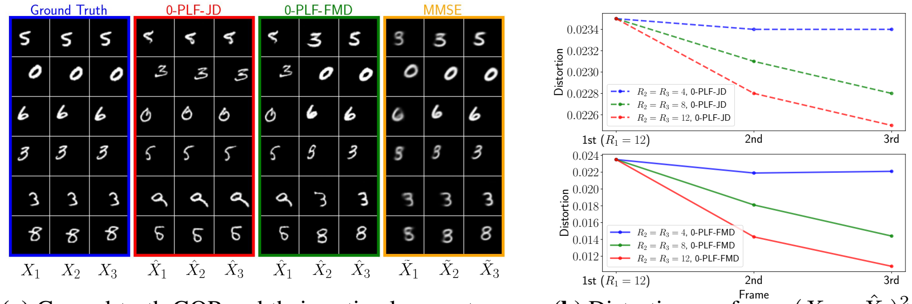
Proof: See Appendix H.2. ■

The above theorem indicates that the MMSE representation can be linearly transformed to achieve any point in $\mathcal{DP}(\mathsf{R})$. In general the MMSE representation may have to be degraded through additional noise terms. In the proof of Theorem 6 we identify conditions when such degradation is not needed.

As discussed in Appendix H.2, Theorem 6 holds for both PLFs. This suggests the idea that one can train an encoder to get MMSE representations which are oblivious to the choice of PLF. Then, the decoder can generate a reconstruction which satisfies either of PLFs by simply applying a linear transformation to the MMSE representation. Thus, the task of choosing the right PLF can be assigned to the decoder based on distortion and perception requirements. We conclude by noting that Appendix H.3 provides an example where the coefficients in Theorem 6 can be computed explicitly.

5 Experimental Results

We conduct experiments on the MovingMNIST dataset [29] (with 1 digit) using Wasserstein GAN [30], to verify the implications of our theoretical claims to perceptual video compression. Additional results on the KTH dataset [31] are available in Appendix J.3. Our compression network is built on the scale-space flow model [32] and conditional module [33]. For a given rate and PLF, we obtain different distortion-perception tradeoff points by optimizing the weighted sum between distortion



(a) Ground-truth GOP and their optimal reconstructions with different PLFs for R_1=R_2=R_3=12 bits.

(b) Distortion per frame $(X_i - \hat{X}_i)^2$ for different rates with $i = 1, 2, 3$.

Figure 2: Permanence of Error Phenomenon. In (a), we visually compare the reconstructions. Note that $\hat{X}_1$ is the same for both 0-PLF-JD and 0-PLF-PMD. In (b), we show the framewise distortion for different (R_2, R_3).

and perception losses. Details about the architecture and training procedure are available in the Appendix J.1. The experimental setup is focused on validating our theory, rather than proposing state-of-the-art neural network architectures. Accordingly, we begin by (1) validating Theorems 1 and 2, which characterize the factor-of-two bounds on the distortion of 0-PLF reconstructions (2) empirically demonstrating the *error permanence* phenomenon of the PLF-JD in Section 4.3 and (3) computing the DP tradeoff function experimentally as well as confirming that the MMSE reconstruction provide near universal representations, as motivated by the results in Section 4.4.

As our first experimental result in Table 1, we validate the *factor of two bounds* in Theorems 1 and 2. We consider the compression of two frames X_1 and X_2 at rates R_1 and R_2 respectively. The compression of X_1 is performed without any prior reference and corresponds to the compression of the "I-frame", while the compression of X_2 corresponds to the "P-frame", using X_1 as the reference. We consider the cases when either R_1=∞ or R_1=12 bits, where the former corresponds to lossless compression of X_1 and the latter corresponds to the low rate regime (see Appendix I for a justification). The average distortion for the first frame when $R_1 = 12$ is 0.0124 for the MMSE reconstruction and 0.0235 for the 0-PLF reconstruction, thus satisfying the factor of two bound. In compression of X_2, we systematically vary the value of the rate $R_2$$\in$$\{4, 8, 12, \infty\}$. Following Table 1b, for 0-PLF-JD reconstruction, only R_2=4 bits (low rate) satisfies the factor of two bounds as expected. Intuitively, even as more bits are acquired, the 0-PLF-JD criteria actively restricts improving the reconstructions, resulting in persistently higher distortion. Even when $R_2 = \infty$, the distortion remains non-zero as the decoder is forced to maintain temporal consistency with $\hat{X}_1$. In contrast, for FMD, the factor of 2 bound holds at all rates, consistent with Theorem 1.

Table 1: Distortions of optimal reconstructions at different regime (✓means factor of 2 holds and ✗means otherwise). Distortion is scaled by 10^{-2}.

R_2	MMSE	0-PLF-FMD	0-PLF-JD
1	1.08 ± 0.01	1.74 ± 0.02 ✓	2.05 ± 0.03 ✓
2	0.88 ± 0.01	1.39 ± 0.03✓	1.46 ± 0.02 ✓
3.17	0.53 ± 0.01	0.76 ± 0.01 ✓	0.79 ± 0.01 ✓

(a) Case 1: R_1=∞ bits

R_2	MMSE	0-PLF-FMD	0-PLF-JD
4	1.23 ± 0.01	2.21 ± 0.04✓	2.36 ± 0.04 ✓
8	1.04 ± 0.01	1.78 ± 0.03 ✓	2.28 ± 0.03 ✗
12	0.89 ± 0.02	1.43 ± 0.02 ✓	2.26 ± 0.03 ✗
∞	0.0	0.0 ✓	2.18 ± 0.02 ✗

(b) Case 2: R_1=12 bits(ϵ).

In Fig. 2, we present our experimental results with a group of pictures (GOP) of size 3 (i.e. one I-frame followed by two P-frames). In Fig. 2a, we visualize sample reconstructions for MSE, 0-PLF-FMD and 0-PLF-JD cases when operating in the low-rate regime with R_j=12 bits for j=$1, 2, 3$. Note that given an incorrect digit reconstruction in $\hat{X}_1$, the decoder with 0-PLF-JD consistently produces incorrect digits (or content) while the 0-PLF-FMD gradually "corrects" it, which confirms the *error permanence phenomenon* discussed in the theoretical analysis in Section 4.3 and Table 2 in Appendix F. We also plot the framewise distortion in Fig. 2b to show the difference in achievable distortion two perception metrics across different values for R_2 and R_3 as a function of the frame index. Consistent with Theorem 5, the achievable distortion decreases much faster for 0-PLF-FMD than 0-PLF-JD for all selection of rates. Finally, we show similar results in Figure 3 for UVG dataset.

In Fig. 4, we plot the tradeoff curves between distortion and perception for the second reconstruction $\hat{X}_2$ for both optimal (end-to-end) and universal representations for two cases: when R_1=∞ and R_1=12 bits and for a range of values for R_2. In general, the curves for both universal and optimal

https://doi.org/10.52202/075280-2093

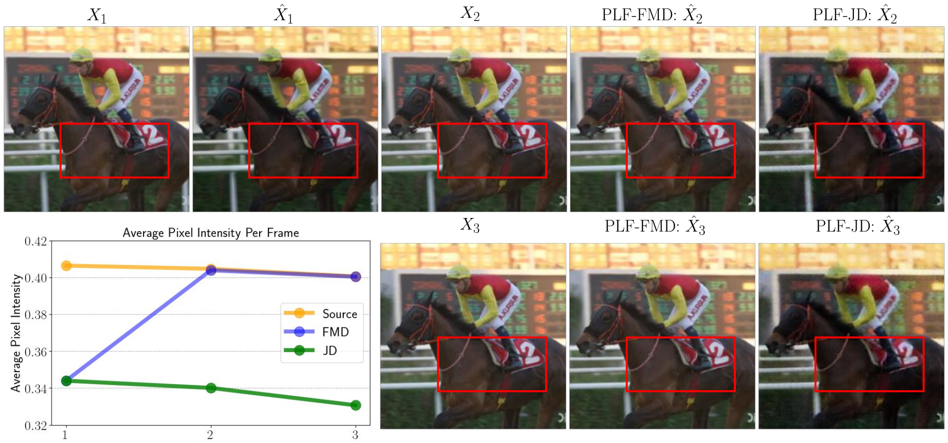
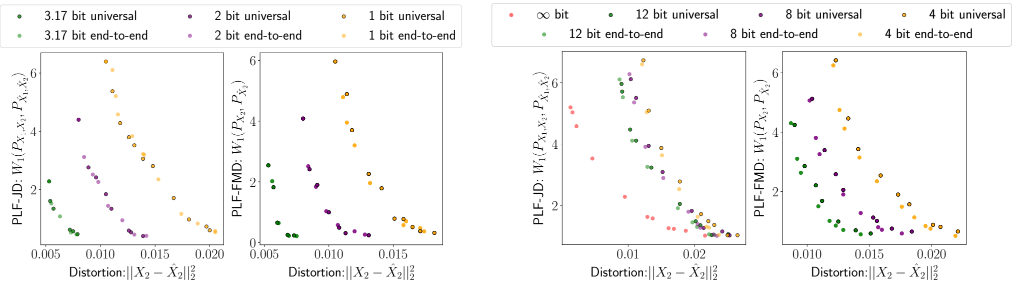


Figure 3: Error permanence phenomenon on the UVG dataset. The PLF-JD reconstructions propagate the flaws in the color tone from the previous I-frame reconstruction while the PLF-FMD is able to fix these flaws. Compression rate for I-frame and P-frame are 0.144bpp (low rate) and 4.632bpp (high rate) respectively.

representations are relatively close to each other at all rate regimes. The general shape of every curve is relatively similar with the exception of the PLF-JD metric in Fig. 4b, where the curves for different rates seemingly converge since increasing the rate does not significantly improve the distortion in this case as noted previously. Finally, as the universal encoders are derived from MMSE solutions, these results imply that one can simply send the MMSE representation to the decoder and the user can flexibly change the DP tradeoff up to their requirements. We further note that even when the end-to-end model targets an operating point different from the MMSE reconstruction, the latter is still required to estimate the motion flow vectors best. The universal representation provides a natural way to reconstruct the MMSE reconstruction from the encoder output. In the plots of Fig. 4, we leverage on established universality results for I-frame compression in prior works [16] to construct the MMSE representation for motion compensation as we have a GOP of size 2.



(a) Case 1: $R_1{=}\infty$, $R_2{=}\{1, 2, 3.17\}$ bits

(b) Case 2: $R_1{=}12$, $R_2{=}\{4, 8, 12, \infty\}$ bits

Figure 4: RDP tradeoff curves for end-to-end and universal models. We plot the tradeoff for the two regimes: $R_1{=}\infty$ and $R_1{=}\epsilon$ in (a) and (b) respectively. The universal and optimal curves are close to each other.

Finally in Appendix J.5, we consider the ability of the decoder to generate diverse reconstructions when operating under either PLF-JD or PLF-FMD. We focus on the case when X_1 is transmitted losslessly and when X_2 is compressed at low rates. Consistent with the theoretical analysis in Section 4.2 and Table 2 in Appendix F, the decoder optimized for PLF-JD is capable of producing diverse reconstructions by mimicking the actual motion between the frames. The PLF-FMD leads to reconstructions that are highly correlated and less desirable.

6 Conclusions

This work examines different perception loss functions for causal video coding, establishing its key theoretical properties such as the operational RDP region and universality principle. Our analysis highlights that while 0-PLF-JD reconstruction preserves temporal correlation, it is susceptible to the error permanence phenomenon. Moreover, our investigation of universality reveals that the encoder can transform the MMSE representation to other points on the DP tradeoffs, irrespective of the PLF. We suggest future research directions such as exploring region-based perceptual metrics [34], incorporating image-aware bits allocation, and leveraging conditional perception metric [18].

References

[1] 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.

[2] 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

[3] 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

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

[5] 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.

[6] S. Zhang, M. Mrak, L. Herranz, M. G. Blanch, S. Wan, and F. Yang, “Dvc-p: Deep video compression with perceptual optimizations,” in *2021 International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2021, pp. 1–5.

[7] 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.

[8] R. Yang, L. Van Gool, and R. Timofte, “Perceptual learned video compression with recurrent conditional gan,” *arXiv preprint arXiv:2109.03082*, vol. 1, 2021.

[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] Y. Wang, P. Bilinski, F. Bremond, and A. Dantcheva, “G^{3}an: Disentangling appearance and motion for video generation,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

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

[12] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool, “Generative adversarial networks for extreme learned image compression,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 221–231.

[13] J. Ballé, V. Laparra, and E. P. Simoncelli, “End-to-end optimized image compression,” in *5th International Conference on Learning Representations*, 2017.

[14] L. Theis, W. Shi, A. Cunningham, and F. Huszár, “Lossy image compression with compressive autoencoders,” in *5th International Conference on Learning Representations*, 2017.

[15] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, “Conditional probability models for deep image compression,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[16] 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.

[17] C. Gao, T. Xu, D. He, Y. Wang, and H. Qin, “Flexible neural image compression via code editing,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 184–12 196, 2022.

[18] F. Mentzer, G. Toderici, M. Tschannen, and E. Agustsson, “High-fidelity generative image compression,” in *Advances in Neural Information Processing Systems*, 2020.

[19] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, “Generative adversarial networks for extreme learned image compression,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 221–231.

[20] G. Lu, C. Cai, X. Zhang, L. Chen, W. Ouyang, D. Xu, and Z. Gao, “Content adaptive and error propagation aware deep video compression,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 456–472.

https://doi.org/10.52202/075280-2093

[21] Y. Blau and T. Michaeli, “The perception-distortion tradeoff,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6228–6237.

[22] D. Freirich, T. Michaeli, and R. Meir, “A theory of the distortion-perception tradeoff in wasserstein space,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 25 661–25 672, 2021.

[23] N. Saldi, T. Linder, and S. Yüksel, “Randomized quantization and optimal design with a marginal constraint,” in *2013 IEEE International Symposium on Information Theory*. IEEE, 2013, pp. 2349–2353.

[24] Z. Yan, F. Wen, R. Ying, C. Ma, and P. Liu, “On perceptual lossy compression: The cost of perceptual reconstruction and an optimal training framework,” in *International Conference on Machine Learning*. PMLR, 2021, pp. 11 682–11 692.

[25] E. Agustsson, D. Minnen, G. Toderici, and F. Mentzer, “Multi-realism image compression with a conditional generator,” *arXiv preprint arXiv:2212.13824*, 2022.

[26] V. M. Panaretos and Y. Zemel, *An invitation to statistics in Wasserstein space*. Springer, 2020.

[27] A. Makur, *Information contraction and decomposition*. PhD Thesis, MIT, 2019.

[28] N. Ma and P. Ishwar, “On delayed sequential coding of correlated sources,” *IEEE Trans. on Info. Theory*, vol. 57, no. 6, pp. 3763–3782, 2011.

[29] 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.

[30] 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.

[31] C. Schuldt, I. Laptev, and B. Caputo, “Recognizing human actions: a local svm approach,” in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3. IEEE, 2004, pp. 32–36.

[32] 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.

[33] J. Li, B. Li, and Y. Lu, “Deep contextual video compression,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 18 114–18 125, 2021.

[34] E. Pergament, P. Tandon, O. Rippel, L. Bourdev, A. G. Anderson, B. Olshausen, T. Weissman, S. Katti, and K. Tatwawadi, “Pim: Video coding using perceptual importance maps,” *arXiv preprint arXiv:2212.10674*, 2022.

[35] 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.

[36] 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.

[37] 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.

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

[39] E. Denton and R. Fergus, “Stochastic video generation with a learned prior,” in *International conference on machine learning*. PMLR, 2018, pp. 1174–1183.

[40] Y.-H. Kwon and M.-G. Park, “Predicting future frames using retrospective cycle gan,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1811–1820.

[41] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

https://doi.org/10.52202/075280-2093

[42] S. Hong, D. Yang, Y. Jang, T. Zhao, and H. Lee, “Diversity-sensitive conditional generative adversarial networks,” in *7th International Conference on Learning Representations, ICLR 2019*. International Conference on Learning Representations, ICLR, 2019.

[43] R. Yang and S. Mandt, “Lossy image compression with conditional diffusion models,” *arXiv preprint arXiv:2209.06950*, 2022.

https://doi.org/10.52202/075280-2093

A Distortion Analysis for 0-PLF-FMD

Recall the definition of Wasserstein-2 distance [26] as follows. For given distributions P_{X_j} and $P_{\tilde{X}_j}$, let

$$W_2^2(P_{\tilde{X}_j}, P_{X_j}) := \inf \mathbb{E}[\|X_j - \tilde{X}_j\|^2], \tag{18}$$

where the infimum is over all joint distributions of $(X_j, \tilde{X}_j)$ with marginals P_{X_j} and $P_{\tilde{X}_j}$.

Theorem 1 *The set $\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$ is characterized as follows:*

$$\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K}) = \{\mathsf{D} : D_j \geq \mathbb{E}_P[\|X_j - \tilde{X}_j\|^2] + W_2^2(P_{\tilde{X}_j}, P_{X_j}),\ j = 1, 2, 3\}, \tag{19}$$

Furthermore, we also have that:

$$\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K}) \supseteq \{\mathsf{D} : D_j \geq 2\mathbb{E}_P[\|X_j - \tilde{X}_j\|^2], \quad j = 1, 2, 3\}, \tag{20}$$

i.e., minimum achievable distortion with 0-PLF-FMD is at most twice the MMSE distortion.

Proof: Define

$$\mathcal{D}^0 := \{\mathsf{D} : D_j \geq \mathbb{E}[\|X_j - \tilde{X}_j\|^2] + W_2^2(P_{\tilde{X}_j}, P_{X_j}), \ \ j = 1, 2, 3\}. \tag{21}$$

First, we show that $\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K}) \subseteq \mathcal{D}^0$. For any $\mathsf{D} \in \Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$, there exists $\hat{\mathsf{X}}_{\mathsf{D}^0} = (\hat{X}_{D_1^0}, \hat{X}_{D_2^0}, \hat{X}_{D_3^0})$ jointly distributed with $(\mathsf{M}, \mathsf{X}, K)$ such that

$$\mathbb{E}[\|X_j - \hat{X}_{D_j^0}\|^2] \leq D_j, \qquad j = 1, 2, 3, \tag{22}$$

$$P_{X_j} = P_{\hat{X}_{D_j^0}}. \tag{23}$$

Then, for example, the analysis for the second frame is as follows

$$\begin{aligned}
D_2 &\geq \mathbb{E}[\|X_2 - \hat{X}_{D_2^0}\|^2] && (24)\\
&= \mathbb{E}[\|(X_2 - \tilde{X}_2) - (\hat{X}_{D_2^0} - \tilde{X}_2)\|^2] && (25)\\
&= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \mathbb{E}[\|\tilde{X}_2 - \hat{X}_{D_2^0}\|^2] && (26)\\
&\geq \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + W_2^2(P_{\tilde{X}_2}, P_{\hat{X}_{D_2^0}}) && (27)\\
&= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + W_2^2(P_{\tilde{X}_2}, P_{X_2}), && (28)
\end{aligned}$$

where (26) holds because both $\tilde{X}_2$ and $\hat{X}_{D_2^0}$ are functions of (M_1, M_2, K) and thus the MMSE $(X_2 - \tilde{X}_2)$ is uncorrelated with $(\hat{X}_{D_2^0} - \tilde{X}_2)$; (28) follows because the 0-PLF-FMD implies that $P_{\hat{X}_{D_2^0}} = P_{X_2}$. Following similar steps for other frames, we get $\Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K}) \subseteq \mathcal{D}^0$.

Next, we show that $\mathcal{D}^0 \subseteq \Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$. Assume that $\mathsf{D} \in \mathcal{D}^0$. Let $\hat{X}_1^*$ be an auxiliary random variable jointly distributed with (M_1, K) such that it satisfies the following conditions

$$P_{\hat{X}_1^*} = P_{X_1}, \tag{29}$$

and

$$P_{\tilde{X}_1 \hat{X}_1^*} = \arg \inf_{\substack{\bar{P}_{\tilde{X}_1 \hat{X}_1^*}: \\ \bar{P}_{\tilde{X}_1} = P_{\tilde{X}_1} \\ \bar{P}_{\hat{X}_1^*} = P_{\hat{X}_1^*}}} \mathbb{E}_{\bar{P}}[\|\tilde{X}_1 - \hat{X}_1^*\|^2]. \tag{30}$$

Moreover, let $\hat{X}_2^*$ be an auxiliary random variable jointly distributed with (M_1, M_2, K) such that the following two conditions are satisfied

$$P_{\hat{X}_2^*} = P_{X_2}, \tag{31}$$

and

$$P_{\tilde{X}_2 \hat{X}_2^*} = \arg \inf_{\substack{\bar{P}_{\tilde{X}_2 \hat{X}_2^*}: \\ \bar{P}_{\tilde{X}_2} = P_{\tilde{X}_2} \\ \bar{P}_{\hat{X}_2^*} = P_{\hat{X}_2^*}}} \mathbb{E}_{\bar{P}}[\|\tilde{X}_2 - \hat{X}_2^*\|^2]. \tag{32}$$

Similarly, we define $\hat{X}_3^*$. Now, notice that since $\mathsf{D} \in \mathcal{D}^0$, we have:

$$D_2 \geq \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + W_2^2(P_{\tilde{X}_2}, P_{X_2}). \tag{33}$$

It then directly follows that

$$\begin{aligned}
\mathbb{E}[\|X_2 - \hat{X}_2^*\|^2] &= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \mathbb{E}[\|\tilde{X}_2 - \hat{X}_2^*\|^2] && (34)\\
&= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + W_2^2(P_{\tilde{X}_2}, P_{\hat{X}_2^*}) && (35)\\
&= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + W_2^2(P_{\tilde{X}_2}, P_{X_2}) && (36)\\
&\leq D_2, && (37)
\end{aligned}$$

where

- (34) follows because $\tilde{X}_2$ and $\hat{X}_2^*$ are functions of (M_1, M_2, K) and thus the MMSE $(X_2 - \tilde{X}_2)$ is uncorrelated with $(\hat{X}_2^* - \tilde{X}_2)$;
- (35) follows from (32);
- (36) follows because $P_{\hat{X}_2^*} = P_{X_2}$.

Following similar steps for other frames, we get $\mathsf{D} \in \Phi_{\mathsf{D}^0}(P_{\mathsf{X}_r|\mathsf{X}})$.

Now, notice that $W_2^2(P_{\tilde{X}_2}, P_{X_2}) \leq \mathbb{E}[\|X_2 - \tilde{X}_2\|^2]$ since the Wasserstein-2 distance takes the infimum over all possible joint distributions $(X_2, \tilde{X}_2)$, but the expectation in $\mathbb{E}[\|X_2 - \tilde{X}_2\|^2]$ is taken over the given $P_{X_2\tilde{X}_2}$. Thus, we get

$$\mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + W_2^2(P_{\tilde{X}_2}, P_{X_2}) \leq 2\mathbb{E}[\|X_2 - \tilde{X}_2\|^2]. \tag{38}$$

This concludes the proof. ■

B Distortion Analysis for 0-PLF-JD

Let $\hat{X}_1^*$ be defined as in (29)–(30). Moreover, let $\hat{X}_2^*$ be an auxiliary random variable jointly distributed with (M_1, M_2, K) such that the following conditions are satisfied

$$P_{\hat{X}_2^*|\hat{X}_1^*=x_1} = P_{X_2|X_1=x_1}, \qquad \forall x_1 \in \mathcal{X}_1, \tag{39}$$

and

$$P_{\tilde{X}_2\hat{X}_2^*|\hat{X}_1^*=x_1} = \arg \inf_{\substack{\bar{P}_{\tilde{X}_2\hat{X}_2^*|\hat{X}_1^*=x_1}: \\ \bar{P}_{\tilde{X}_2|\hat{X}_1^*=x_1} = P_{\tilde{X}_2|\hat{X}_1^*=x_1} \\ \bar{P}_{\hat{X}_2^*|\hat{X}_1^*=x_1} = P_{\hat{X}_2^*|\hat{X}_1^*=x_1}}} \mathbb{E}_{\bar{P}}[\|\tilde{X}_2 - \hat{X}_2^*\|^2|\hat{X}_1^* = x_1], \qquad \forall x_1 \in \mathcal{X}_1. \tag{40}$$

Then, the following result holds.

Theorem 2 *We have*

$$\begin{aligned}
\Phi_{\mathsf{D}^0}^{\textit{joint}}(P_{\mathsf{M}|\mathsf{X}K}) \supseteq \{\mathsf{D} : D_1 &\geq \mathbb{E}[\|X_1 - \tilde{X}_1\|^2] + W_2^2(P_{\tilde{X}_1}, P_{X_1}),\\
D_2 &\geq \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \sum_{x_1} P_{X_1}(x_1) W_2^2(P_{\tilde{X}_2|\hat{X}_1^*=x_1}, P_{X_2|X_1=x_1}),\\
D_3 &\geq \mathbb{E}[\|X_3 - \tilde{X}_3\|^2] + \sum_{x_1,x_2} P_{X_1X_2}(x_1,x_2) W_2^2(P_{\tilde{X}_3|\hat{X}_1^*=x_1,\hat{X}_2^*=x_2}, P_{X_3|X_1=x_1,X_2=x_2})\}.
\end{aligned} \tag{41}$$

Proof: Define

$$\begin{aligned}
\mathcal{D}^0_{\text{joint}} := \{\mathsf{D} : D_1 &\geq \mathbb{E}[\|X_1 - \tilde{X}_1\|^2] + W_2^2(P_{\tilde{X}_1}, P_{X_1}),\\
D_2 &\geq \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \sum_{x_1} P_{X_1}(x_1) W_2^2(P_{\tilde{X}_2|\hat{X}_1^*=x_1}, P_{X_2|X_1=x_1}),\\
D_3 &\geq \mathbb{E}[\|X_3 - \tilde{X}_3\|^2] + \sum_{x_1,x_2} P_{X_1X_2}(x_1,x_2) W_2^2(P_{\tilde{X}_3|\hat{X}_1^*=x_1,\hat{X}_2^*=x_2}, P_{X_3|X_1=x_1,X_2=x_2})\}.
\end{aligned} \tag{42}$$

Now, assume that $\mathsf{D} \in \mathcal{D}^0_{\text{joint}}$. For the first frame, recall that $\hat{X}_1^*$ is an auxiliary random variable jointly distributed with (M_1, K) such that it satisfies (29)–(30). From similar steps to (34)–(36), it then follows that

$$\begin{aligned} \mathbb{E}[\|X_1 - \hat{X}_1^*\|^2] &= \mathbb{E}[\|X_1 - \tilde{X}_1\|^2] + W_2^2(P_{\tilde{X}_1}, P_{X_1}) && (43) \\ &\leq D_1. && (44) \end{aligned}$$

For the second frame, since $\mathsf{D} \in \mathcal{D}^0_{\text{joint}}$, we have:

$$D_2 \geq \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \sum_{x_1} P_{X_1}(x_1) W_2^2(P_{\tilde{X}_2|X_1=x_1}, P_{X_2|X_1=x_1}). \tag{45}$$

Recall that $\hat{X}_2^*$ is an auxiliary random variable jointly distributed with (M_1, M_2, K) such that (39)–(40) hold. It then directly follows that

$$\begin{aligned} \mathbb{E}[\|X_2 - \hat{X}_2^*\|^2] &= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \mathbb{E}[\|\tilde{X}_2 - \hat{X}_2^*\|^2] && (46) \\ &= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \sum_{x_1} P_{\hat{X}_1^*}(x_1) \mathbb{E}[\|\tilde{X}_2 - \hat{X}_2^*\|^2 | \hat{X}_1^* = x_1] && (47) \\ &= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \sum_{x_1} P_{\hat{X}_1^*}(x_1) W_2^2(P_{\tilde{X}_2|\hat{X}_1^*=x_1}, P_{\hat{X}_2^*|\hat{X}_1^*=x_1}) && (48) \\ &= \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \sum_{x_1} P_{X_1}(x_1) W_2^2(P_{\tilde{X}_2|\hat{X}_1^*=x_1}, P_{X_2|X_1=x_1}), && (49) \end{aligned}$$

where

- (46) follows because $\tilde{X}_2$ and $\hat{X}_2^*$ are functions of (M_1, M_2, K) and thus the MMSE $(X_2 - \tilde{X}_2)$ is uncorrelated with $(\hat{X}_2^* - \tilde{X}_2)$,
- (48) follows from (40),
- (49) follows because $P_{\hat{X}_1^* \hat{X}_2^*} = P_{X_1 X_2}$.

Following similar steps for the third frame, we get $\mathsf{D} \in \Phi_{\mathsf{D}^0}(P_{\mathsf{M}|\mathsf{X}K})$. This concludes the proof. ■

B.1 A Counterexample for Factor-Two Bound in Case of 0-PLF-JD

Assume that we have only two frames, i.e., $D_3 \to \infty$. Let M_1 be independent of X_1 and $M_2 = X_2$. Then, we have $\tilde{X}_1 = \emptyset$ and $\tilde{X}_2 = X_2$. Consider the achievable distortion region of Theorem 2. The distortion of the first step is given by the following

$$\mathbb{E}[\|X_1 - \tilde{X}_1\|^2] + W_2^2(P_{\tilde{X}_1}, P_{X_1}) = 2\mathbb{E}[X_1^2]. \tag{50}$$

For the second frame, we have

$$\begin{aligned} &\mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \sum_{x_1} P_{X_1}(x_1) W_2^2(P_{\tilde{X}_2|\hat{X}_1^*=x_1}, P_{X_2|X_1=x_1}) \\ &\quad = \sum_{x_1} P_{X_1}(x_1) W_2^2(P_{X_2|\hat{X}_1^*=x_1}, P_{X_2|X_1=x_1}) && (51) \\ &\quad = \sum_{x_1} P_{X_1}(x_1) W_2^2(P_{X_2}, P_{X_2|X_1=x_1}), && (52) \end{aligned}$$

where (51) follows because $\tilde{X}_2 = X_2$ and (52) follows because X_2 is independent of $\hat{X}_1^*$ (M_1 is independent of X_1, then $\hat{X}_1^*$, which is a function of (M_1, K), would be independent of X_1 and hence independent of X_2).

Now, notice that the MMSE distortion of the second step is zero since $\tilde{X}_2 = X_2$. However, the achievable distortion of the second step for the reconstruction satisfying 0-PLF JD is given in (52) which clearly does not satisfy the factor-two bound.

C Fixed Encoders Operating at Low rate regime

We consider the class of noisy encoders where the encoder distribution can be written as follows

$$P^{\text{noisy}}_{X_j|M_1\dots M_j K} = (1-\mu)P_{X_j} + \mu Q^{\text{noisy}}_{X_j|M_1\dots M_j K}, \qquad j = 1,2,3. \tag{53}$$

where μ is a sufficiently small constant and the distribution $Q^{\text{noisy}}(\cdot)$ could be arbitrary conditional distribution with same marginal as P_{X_j}.

Theorem 3 *For the class of encoders given by* (53)*, we have*

$$\Phi^{\textit{joint}}_{\mathsf{D}^0}(P^{\textit{noisy}}_{\mathsf{M}|\mathsf{X}K}) \supseteq \{\mathsf{D} : D_j \geq 2\mathbb{E}_{P^{noisy}}[\|X_j - \tilde{X}_j\|^2] + O(\mu), \quad j = 2,\dots,3\}. \tag{54}$$

Proof: We analyze the distortion for the second frame. A similar argument holds for other frames.

Denote the reconstruction of the second step by $\hat{X}_2^*$ and consider the expected distortion. From a similar justification starting from (24) and leading to (26), we can write the distortion as follows

$$\mathbb{E}[\|X_2 - \hat{X}_2^*\|^2] = \mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + \mathbb{E}[\|\tilde{X}_2 - \hat{X}_2^*\|^2]. \tag{55}$$

Now, we study the expected term $\mathbb{E}[\|\tilde{X}_2 - \hat{X}_2^*\|^2]$ as follows

$$\mathbb{E}[\|\tilde{X}_2 - \hat{X}_2^*\|^2] = \sum_{x_1} P_{\hat{X}_1^*}(x_1)\mathbb{E}[\|\tilde{X}_2 - \hat{X}_2^*\|^2|\hat{X}_1^* = x_1]. \tag{56}$$

In order to analyze the above expression, we first approximate the MMSE reconstruction $\tilde{X}_2$ as follows

$$\begin{aligned}
\tilde{X}_2 &= \mathbb{E}_{P^{\text{noisy}}}[X_2|M_1, M_2, K] \quad (57)\\
&= (1-\mu)\mathbb{E}_P[X_2] + \mu\mathbb{E}_{Q^{\text{noisy}}}[X_2|M_1, M_2, K] \quad (58)\\
&= \mathbb{E}[X_2] + O(\mu), \quad (59)
\end{aligned}$$

where (58) follows from (53). Moreover, notice that (59) implies that

$$\begin{aligned}
\mathbb{E}[\|X_2 - \tilde{X}_2\|^2] &= \mathbb{E}[\|X_2 - \mathbb{E}[X_2] + \mu(\mathbb{E}_{Q^{\text{noisy}}}[X_2|M_1, M_2, K] - \mathbb{E}[X_2])\|^2] \quad (60)\\
&= \mathbb{E}[\|X_2 - \mathbb{E}[X_2]\|^2] + O(\mu). \quad (61)
\end{aligned}$$

Next, consider the expected term in (56) as follows

$$\begin{aligned}
\sum_{x_1} P_{\hat{X}_1^*}(x_1)\mathbb{E}[\|\tilde{X}_2 - \hat{X}_2^*\|^2|\hat{X}_1^* = x_1] &= \sum_{x_1} P_{\hat{X}_1^*}(x_1)\mathbb{E}[\|\mathbb{E}[X_2] - \hat{X}_2^*\|^2|\hat{X}_1^* = x_1] + O(\mu) \quad (62)\\
&= \sum_{x_1} P_{\hat{X}_1^*}(x_1)\mathbb{E}[\|\mathbb{E}[X_2] - X_2\|^2|X_1 = x_1] + O(\mu) \quad (63)\\
&= \sum_{x_1} P_{X_1}(x_1)\mathbb{E}[\|\mathbb{E}[X_2] - X_2\|^2|X_1 = x_1] + O(\mu) \quad (64)\\
&= \mathbb{E}[\|\mathbb{E}[X_2] - X_2\|^2] + O(\mu) \quad (65)\\
&= \mathbb{E}[\|\tilde{X}_2 - X_2\|^2] + O(\mu), \quad (66)
\end{aligned}$$

where

- (62) follows from (59);
- (63) follows because the 0-PLF-JD implies that $P_{\hat{X}_2^*|\hat{X}_1^*} = P_{X_2|X_1}$ and $\mathbb{E}[X_2]$ is just a constant;
- (64) follows from 0-PLF-JD where $P_{\hat{X}_1^*} = P_{X_1}$;
- (66) follows from (61).

Considering (55) and (66), we get

$$\mathbb{E}[\|X_2 - \hat{X}_2^*\|^2] = 2\mathbb{E}[\|X_2 - \tilde{X}_2\|^2] + O(\mu). \tag{67}$$

The proof for the third frame follows similar steps. ■

https://doi.org/10.52202/075280-2093

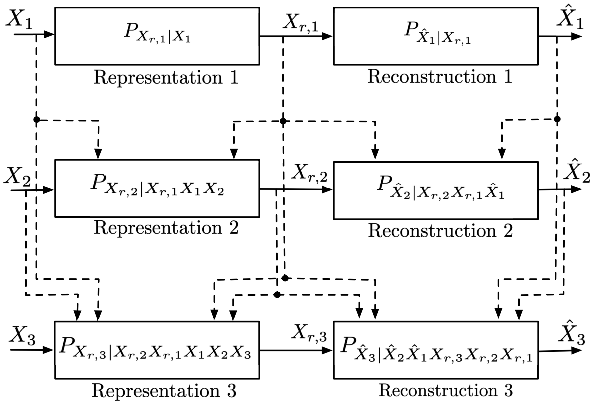


Figure 5: Encoded representations and reconstructions of the iRDP region $\mathcal{C}_{\mathsf{RDP}}$.

D Operational RDP Region

Recall the definition of iRDP region $\mathcal{C}_{\mathsf{RDP}}$ for first-order Markov sources (Definition 4) as follows. It is the set of all tuples $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ satisfying

$$R_1 \geq I(X_1; X_{r,1}), \tag{68}$$

$$R_2 \geq I(X_2; X_{r,2}|X_{r,1}), \tag{69}$$

$$R_3 \geq I(X_3; X_{r,3}|X_{r,1}, X_{r,2}), \tag{70}$$

$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j\|^2], \qquad j = 1, 2, 3, \tag{71}$$

$$P_j \geq \phi_j(P_{X_1 \ldots X_j}, P_{\hat{X}_1 \ldots \hat{X}_j}), \qquad j = 1, 2, 3, \tag{72}$$

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

$$\hat{X}_1 = \eta_1(X_{r,1}), \;\; \hat{X}_2 = \eta_2(X_{r,1}, X_{r,2}), \;\; \hat{X}_3 = X_{r,3}, \tag{73}$$

$$X_{r,1} \to X_1 \to (X_2, X_3), \tag{74}$$

$$X_{r,2} \to (X_2, X_{r,1}) \to (X_1, X_3), \tag{75}$$

$$X_{r,3} \to (X_3, X_{r,1}, X_{r,2}) \to (X_1, X_2), \tag{76}$$

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

Theorem 4 *For first-order Markov sources, a given* (D, P) *and* $\mathsf{R} \in \mathcal{R}(\mathsf{D}, \mathsf{P})$*, we have*

$$\mathsf{R} + \log(\mathsf{R} + 1) + 5 \in \mathcal{R}^o(\mathsf{D}, \mathsf{P}). \tag{77}$$

Moreover, the following holds:

$$\mathcal{R}^o(\mathsf{D}, \mathsf{P}) \subseteq \mathcal{R}(\mathsf{D}, \mathsf{P}). \tag{78}$$

Proof: Before stating the achievable scheme, we first discuss the strong functional representation lemma [35]. 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

$$H(Y|U) \leq I(X; Y) + \log(I(X; Y) + 1) + 4. \tag{79}$$

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

$$H(Y|W, U) \leq I(X; Y|W) + \log(I(X; Y|W) + 1) + 4. \tag{80}$$

Proof of (77) *(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 (74)–(76) hold and the rate constraints in (68)–(70)

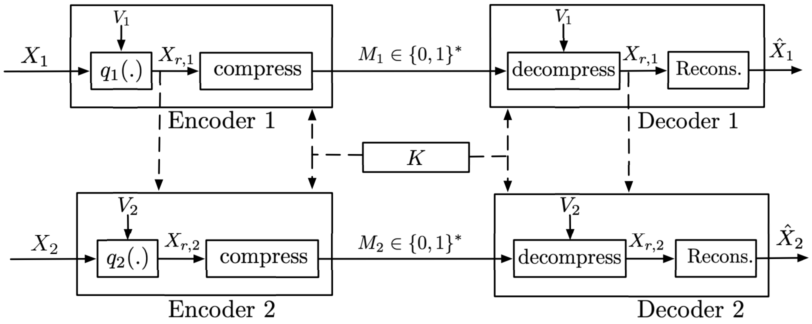


Figure 6: Strong functional representation lemma for $T = 2$ frames.

are satisfied such that there exist $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ for which distortion-perception constraints (71)–(72) 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 (74)–(76), it factorizes as the following

$$\begin{aligned} P_{\mathsf{X}\mathsf{X}_r\hat{\mathsf{X}}} &= P_{X_1X_2X_3} \cdot P_{X_{r,1}|X_1} \cdot P_{X_{r,2}|X_{r,1}X_2} \cdot P_{X_{r,3}|X_{r,2}X_{r,1}X_3} \\ &\cdot \mathbb{1}\{\hat{X}_1 = g_1(X_{r,1})\} \cdot \mathbb{1}\{\hat{X}_2 = g_2(X_{r,1}, X_{r,3})\} \cdot \mathbb{1}\{\hat{X}_3 = X_{r,3}\}. \end{aligned} \tag{81}$$

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. 5.

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 (81) which constructs the codebook, given by

$$P_{\mathsf{X}_r} = P_{X_{r,1}} P_{X_{r,2}|X_{r,1}} P_{X_{r,3}|X_{r,2}X_{r,1}}. \tag{82}$$

From the strong functional representation lemma [35], 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

  $$H(X_{r,1}|V_1) \leq I(X_1; X_{r,1}) + \log(I(X_1; X_{r,1}) + 1) + 4, \tag{83}$$

  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 (82) (see Fig. 6);
- 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

  $$H(X_{r,2}|X_{r,1}, V_2) \leq I(X_2; X_{r,2}|X_{r,1}) + \log(I(X_2; X_{r,2}|X_{r,1}) + 1) + 4. \tag{84}$$

  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 (82) (see Fig. 6);
- 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

  $$H(X_{r,3}|X_{r,1}, X_{r,2}, V_3) \leq I(X_3; X_{r,3}|X_{r,1}, X_{r,2}) + \log(I(X_3; X_{r,3}|X_{r,1}, X_{r,2}) + 1) + 4. \tag{85}$$

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. 6).
- 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. 6).

- 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 (78) *(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

$$R_j \geq \mathbb{E}[\ell(M_j)], \qquad j = 1, 2, 3, \tag{86}$$

and

$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j\|^2], \qquad j = 1, 2, 3, \tag{87}$$

$$P_j \geq \phi_j(P_{X_1 \ldots X_j}, P_{\hat{X}_1 \ldots \hat{X}_j}), \qquad j = 1, 2, 3, \tag{88}$$

we lower bound the expected length of the messages. Define

$$X_{r,1} := (M_1, K), \tag{89}$$

$$X_{r,2} := (M_1, M_2, K), \tag{90}$$

and recall that according to the decoding functions, we have

$$\hat{X}_j = g_j(M_1, \ldots, M_j, K), \qquad j = 1, 2, 3. \tag{91}$$

We can write

$$\begin{aligned} R_1 \geq \mathbb{E}[\ell(M_1)] &\geq H(M_1|K) && (92)\\ &= I(X_1; M_1|K) && (93)\\ &= I(X_1; M_1, K) && (94)\\ &= I(X_1; X_{r,1}). && (95) \end{aligned}$$

Now, consider the following set of inequalities

$$\begin{aligned} R_2 \geq \mathbb{E}[\ell(M_2)] &\geq H(M_2|M_1, K) && (96)\\ &= I(X_1, X_2; M_2|M_1, K) && (97)\\ &= I(X_1, X_2; X_{2,r}|X_{r,1}). && (98) \end{aligned}$$

Similarly, we have

$$\begin{aligned} R_3 \geq \mathbb{E}[\ell(M_3)] &\geq H(M_3|M_1, M_2, K) && (99)\\ &= I(X_1, X_2, X_3; M_3|M_1, M_2, K) && (100)\\ &\geq I(X_1, X_2, X_3; \hat{X}_3|X_{r,1}, X_{r,2}). && (101) \end{aligned}$$

Notice that the definitions in (89)–(90) imply the following Markov chains

$$X_{r,1} \to X_1 \to (X_2, X_3), \tag{102}$$

$$X_{r,2} \to (X_1, X_2, X_{r,1}) \to X_3. \tag{103}$$

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

$$\hat{X}_1 = g_1(M_1, K), \tag{104}$$

$$\hat{X}_2 = g_2(M_1, M_2, K), \tag{105}$$

where together with definitions in (89) and (90), we can write

$$\hat{X}_1 = g_1(M_1, K) := \eta_1(X_{r,1}), \tag{106}$$

$$\hat{X}_2 = g_2(M_1, M_2, K) := \eta_2(X_{r,1}, X_{r,2}), \tag{107}$$

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

https://doi.org/10.52202/075280-2093

Now, consider the fact that the set of constraints in (87)–(88), (95), (98), (101) with Markov chains in (102)–(103) and deterministic functions in (106)–(107) constitute an iRDP region, denoted by $\bar{\mathcal{C}}_{\mathsf{RDP}}$, which is the set of all tuples $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ such that

$$R_1 \geq I(X_1; X_{r,1}), \tag{108}$$
$$R_2 \geq I(X_1, X_2; X_{r,2}|X_{r,1}), \tag{109}$$
$$R_3 \geq I(X_1, X_2, X_3; \hat{X}_3|X_{r,1}, X_{r,2}), \tag{110}$$
$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j\|^2], \qquad j = 1, 2, 3, \tag{111}$$
$$P_j \geq \phi_j(P_{X_1 \dots X_j}, P_{\hat{X}_1 \dots \hat{X}_j}), \qquad j = 1, 2, 3, \tag{112}$$

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

$$\hat{X}_1 = \eta_1(X_{r,1}), \;\; \hat{X}_2 = \eta_2(X_{r,1}, X_{r,2}) \tag{113}$$
$$X_{r,1} \to X_1 \to (X_2, X_3), \tag{114}$$
$$X_{r,2} \to (X_1, X_2, X_{r,1}) \to X_3. \tag{115}$$

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

Comparing the two regions $\bar{\mathcal{C}}_{\mathsf{RDP}}$ and $\mathcal{C}_{\mathsf{RDP}}$, we identify the following differences. The Markov chain in (74) is more restricted comparing to (115). Moreover, the Markov chain (75) does not exist in $\bar{\mathcal{C}}_{\mathsf{RDP}}$. The following lemma states that $\bar{\mathcal{C}}_{\mathsf{RDP}} = \mathcal{C}_{\mathsf{RDP}}$. Now, for a given (D, P), let $\bar{\mathcal{R}}(\mathsf{D}, \mathsf{P})$ denote the set of rate tuples R such $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \bar{\mathcal{C}}_{\mathsf{RDP}}$, then this lemma implies that $\bar{\mathcal{R}}(\mathsf{D}, \mathsf{P}) = \mathcal{R}(\mathsf{D}, \mathsf{P})$ which completes the proof of the outer bound. Moreover, notice that the above proof only deals with the statistics of the representations and reconstructions and does not depend on the choice of the PLF. So, it holds for both PLF-FMD and PLF-JD. This concludes the proof.

We conclude this section by the following lemma.

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

$$\mathcal{C}_{\mathsf{RDP}} = \bar{\mathcal{C}}_{\mathsf{RDP}}. \tag{116}$$

Proof: This result for the scenario without perception constraint has been similarly observed in [36, Eq. (12)]. The proof in this section is provided for completeness.

First, notice that the set of Markov chains in (74)–(76) is more restricted than the ones in (114)–(115), hence $\mathcal{C}_{\mathsf{RDP}} \subseteq \bar{\mathcal{C}}_{\mathsf{RDP}}$. Now, it remains to prove that $\bar{\mathcal{C}}_{\mathsf{RDP}} \subseteq \mathcal{C}_{\mathsf{RDP}}$. Consider the following facts

1. The distortion constraints in (111) depend only on the joint distribution of $(X_j, \hat{X}_j)$, and thus on the joint distribution of $(X_j, X_{r,1}, \dots, X_{r,j})$. 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 (112) depend on the joint distributions $P_{X_1 \dots X_j}$ and $P_{\hat{X}_1, \dots, \hat{X}_j}$ (hence on $P_{X_{r,1} \dots X_{r,j}}$). Thus, imposing $X_{r,2} \to (X_2, X_{r,1}) \to X_1$ does not affect $\phi_2(P_{X_1 X_2}, P_{\hat{X}_1 \hat{X}_2})$ 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 (109) and (110) would be further lower bounded by
$$R_2 \geq I(X_1, X_2; X_{r,2}|X_{r,1}) \geq I(X_2; X_{r,2}|X_{r,1}), \tag{117}$$
$$R_3 \geq I(X_1, X_2, X_3; \hat{X}_3|X_{r,1}, X_{r,2}) \geq I(X_3; \hat{X}_3|X_{r,1}, X_{r,2}). \tag{118}$$
Thus, the set of rate constraints is optimized by the set of Markov chains (74)–(76).
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 (73)–(74). Now, since we have first-order Markov sources (see Definition 3), 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

$$X_{r,1} \to X_1 \to (X_2, X_3), \tag{119}$$

$$X_{r,2} \to (X_2, X_{r,1}) \to (X_1, X_3), \tag{120}$$

$$\hat{X}_3 \to (X_3, X_{r,1}, X_{r,2}) \to (X_1, X_2). \tag{121}$$

This concludes the proof for the PLF-JD. For the PLF-FMD, notice that the only difference is the second fact stated above. But, this also holds since the perception constraints depend only on P_{X_j} and $P_{\hat{X}_j}$ (hence on $P_{X_{r,1}\dots,X_{r,j}}$). ■

■

E Gauss-Markov Source Model

We first remark that the Wasserstein-2 distance can also be replaced by the KL-divergence in most of the following analysis. The common properties between these two measures are convexity and the fact that they both depend on only second-order statistics when restricted to Gaussian source model.

Theorem 5 *For the Gauss-Markov source model, any tuple* $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \mathcal{C}_{\mathsf{RDP}}$ *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 4.*

Proof: First, notice that a proof for the setting without perception constraint is provided in [37]. The following proof is different from [37] in some steps and also involves the perception constraint.

For a given tuple $(\mathsf{R}, \mathsf{D}, \mathsf{P}) \in \mathcal{C}_{\mathsf{RDP}}$, 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 (73)–(75). 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.

$$\text{cov}(\hat{X}^G_1, X_1) = \text{cov}(\hat{X}^*_1, X_1), \tag{122}$$

$$\text{cov}(\hat{X}^G_1, \hat{X}^G_2, X_2) = \text{cov}(\hat{X}^*_1, \hat{X}^*_2, X_2), \tag{123}$$

$$\text{cov}(\hat{X}^G_1, \hat{X}^G_2, \hat{X}^G_3, X_3) = \text{cov}(\hat{X}^*_1, \hat{X}^*_2, \hat{X}^*_3, X_3), \tag{124}$$

In general, the Gaussian random variables which satisfy the constraints in (122)–(124) can be written in the following format

$$X_1 = \nu \hat{X}^G_1 + Z_1, \tag{125}$$

$$\hat{X}^G_2 = \omega_1 \hat{X}^G_1 + \omega_2 X_2 + Z_2, \tag{126}$$

$$\hat{X}^G_3 = \tau_1 \hat{X}^G_1 + \tau_2 \hat{X}^G_2 + \tau_3 X_3 + Z_3, \tag{127}$$

for some real ν, ω_1, ω_2, τ_1, τ_2, τ_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 $\alpha^2_1, \alpha^2_2, \alpha^2_3$, 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 (125) by $\hat{X}^G_1$ and taking an expectation, we get

$$\mathbb{E}[X_1 \hat{X}^G_1] = \nu \sigma^2_{\hat{X}^G_1}. \tag{128}$$

According to (122), the above equation can be written as follows

$$\mathbb{E}[X_1 \hat{X}^*_1] = \nu \mathbb{E}[\hat{X}^{*2}_1]. \tag{129}$$

Multiplying both sides of (126) by the vector $[\hat{X}^G_1 \;\; X_2]$ and taking an expectation, we have

$$[\mathbb{E}[\hat{X}^G_1 \hat{X}^G_2] \;\; \mathbb{E}[X_2 \hat{X}^G_2]] = [\omega_1 \;\; \omega_2] \begin{pmatrix} \sigma^2_{\hat{X}^G_1} & \mathbb{E}[X_2 \hat{X}^G_1] \\ \mathbb{E}[X_2 \hat{X}^G_1] & \sigma^2_2 \end{pmatrix} \tag{130}$$

Considering the fact that $\mathbb{E}[X_2 \hat{X}^G_1] = \rho_1 \mathbb{E}[X_1 \hat{X}^G_1]$ and according to (123), the above equation can be written as follows

$$[\mathbb{E}[\hat{X}^*_1 \hat{X}^*_2] \;\; \mathbb{E}[X_2 \hat{X}^*_2]] = [\omega_1 \;\; \omega_2] \begin{pmatrix} \mathbb{E}[\hat{X}^{*2}_1] & \rho_1 \mathbb{E}[X_1 \hat{X}^*_1] \\ \rho_1 \mathbb{E}[X_1 \hat{X}^*_1] & \sigma^2_2 \end{pmatrix}. \tag{131}$$

Similarly, multiplying both sides of (127) by the vector $[\hat{X}_1^G \;\; \hat{X}_2^G \;\; X_3]$, taking an expectation and considering (124), we get

$$[\mathbb{E}[\hat{X}_1^*\hat{X}_3^*] \;\; \mathbb{E}[\hat{X}_2^*\hat{X}_3^*] \;\; \mathbb{E}[X_3\hat{X}_3^*]] = [\tau_1 \;\; \tau_2 \;\; \tau_3] \begin{pmatrix} \mathbb{E}[\hat{X}_1^{*2}] & \mathbb{E}[\hat{X}_1^*\hat{X}_2^*] & \rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] \\ \mathbb{E}[\hat{X}_1^*\hat{X}_2^*] & \mathbb{E}[\hat{X}_2^{*2}] & \rho_2\mathbb{E}[X_2\hat{X}_2^*] \\ \rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] & \rho_2\mathbb{E}[X_2\hat{X}_2^*] & \mathbb{E}[\hat{X}_3^{*2}] \end{pmatrix}. \tag{132}$$

Solving equations (129), (131) and (132), we get

$$\sigma^2_{\hat{X}_1^G} = \mathbb{E}[\hat{X}_1^{*2}], \tag{133}$$

$$\nu = \frac{\mathbb{E}[X_1\hat{X}_1^*]}{\mathbb{E}[\hat{X}_1^{*2}]}, \tag{134}$$

$$\alpha_1^2 = \sigma_1^2 - \frac{\mathbb{E}[X_1\hat{X}_1^*]}{\mathbb{E}[\hat{X}_1^{*2}]}, \tag{135}$$

$$\omega_1 = \frac{\nu\rho_1\mathbb{E}[\hat{X}_1^*\hat{X}_2^*] - \mathbb{E}[X_2\hat{X}_2^*]}{\nu^2\rho_1^2\sigma^2_{\hat{X}_1^G} - \sigma_2^2}, \tag{136}$$

$$\omega_2 = \frac{\nu\rho_1\sigma^2_{\hat{X}_1^G}\mathbb{E}[X_2\hat{X}_2^*] - \sigma_2^2\mathbb{E}[\hat{X}_1^*\hat{X}_2^*]}{\nu^2\rho_1^2\sigma^4_{\hat{X}_1^G} - \sigma_2^2\sigma^2_{\hat{X}_1^G}}, \tag{137}$$

$$\alpha_2^2 = \mathbb{E}[\hat{X}_2^{*2}] - \alpha_2^2\sigma^2_{\hat{X}_1^G} - \omega_2^2\sigma_2^2 - 2\omega_1\omega_2\rho_1\nu\sigma^2_{\hat{X}_1^G}. \tag{138}$$

For the third step, the coefficients and noise variance of (127) are given as follows

$$\begin{aligned} &[\tau_1 \;\; \tau_2 \;\; \tau_3] \\ &\quad= [\mathbb{E}[\hat{X}_1^*\hat{X}_3^*] \;\; \mathbb{E}[\hat{X}_2^*\hat{X}_3^*] \;\; \mathbb{E}[X_3\hat{X}_3^*]] \begin{pmatrix} \mathbb{E}[\hat{X}_1^{*2}] & \mathbb{E}[\hat{X}_1^*\hat{X}_2^*] & \rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] \\ \mathbb{E}[\hat{X}_1^*\hat{X}_2^*] & \mathbb{E}[\hat{X}_2^{*2}] & \rho_2\mathbb{E}[X_2\hat{X}_2^*] \\ \rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] & \rho_2\mathbb{E}[X_2\hat{X}_2^*] & \mathbb{E}[\hat{X}_3^{*2}] \end{pmatrix}^{-1}, \end{aligned} \tag{139}$$

$$\begin{aligned} \alpha_3^2 = {} & \mathbb{E}[\hat{X}_3^{*2}] - \tau_1^2\mathbb{E}[\hat{X}_1^{*2}] - \tau_2^2\mathbb{E}[\hat{X}_2^{*2}] - \tau_3^2\mathbb{E}[X_3^2] \\ & -2\tau_1\tau_2\mathbb{E}[\hat{X}_1^*\hat{X}_2^*] - 2\tau_1\tau_3\rho_1\rho_2\mathbb{E}[X_1\hat{X}_1^*] - 2\tau_2\tau_3\rho_2\mathbb{E}[X_2\hat{X}_2^*], \end{aligned} \tag{140}$$

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

$$\begin{aligned} R_1 &\geq I(X_1; X^*_{r,1}) && (141) \\ &= H(X_1) - H(X_1|X^*_{r,1}) && (142) \\ &\geq H(X_1) - H(X_1|\hat{X}_1^*) && (143) \\ &= H(X_1) - H(X_1 - \mathbb{E}[X_1|\hat{X}_1^*]|\hat{X}_1^*) && (144) \\ &\geq H(X_1) - H(X_1 - \mathbb{E}[X_1|\hat{X}_1^*]) && (145) \\ &\geq H(X_1) - H(X_1 - \mathbb{E}[X_1|\hat{X}_1^G]) && (146) \\ &= H(X_1) - H(X_1 - \mathbb{E}[X_1|\hat{X}_1^G]|\hat{X}_1^G) && (147) \\ &= I(X_1; \hat{X}_1^G) && (148) \end{aligned}$$

where

- (143) follows because $\hat{X}_1^*$ is a function of $X^*_{r,1}$;
- (146) follows because for a given covariance matrix in (122), the Gaussian distribution maximizes the differential entropy;
- (147) follows because the MMSE is uncorrelated from the data and since the random variables are Gaussian, the MMSE would be independent of the data.

https://doi.org/10.52202/075280-2093

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

$$R_2 \geq I(X_2; X_{r,2}^* | X_{r,1}^*) \tag{149}$$

$$= H(X_2 | X_{r,1}^*) - H(X_2 | X_{r,1}^*, X_{r,2}^*) \tag{150}$$

$$\geq H(X_2 | X_{r,1}^*) - H(X_2 | \hat{X}_1^*, \hat{X}_2^*) \tag{151}$$

$$\geq H(X_2 | X_{r,1}^*) - H(X_2 | \hat{X}_1^G, \hat{X}_2^G) \tag{152}$$

$$= H(\rho_1 X_1 + N_1 | X_{r,1}^*) - H(X_2 | \hat{X}_1^G, \hat{X}_2^G) \tag{153}$$

$$\geq \frac{1}{2} \log \left(\rho_1^2 2^{2H(X_1 | X_{r,1}^*)} + 2^{2H(N_1)} \right) - H(X_2 | \hat{X}_1^G, \hat{X}_2^G) \tag{154}$$

$$\geq \frac{1}{2} \log \left(\rho_1^2 2^{-2R_1} 2^{2H(X_1)} + 2^{2H(N_1)} \right) - H(X_2 | \hat{X}_1^G, \hat{X}_2^G), \tag{155}$$

where

- (151) follows because $\hat{X}_1^*$ and $\hat{X}_2^*$ are deterministic functions of $X_{r,1}^*$ and $(X_{r,1}^*, X_{r,2}^*)$, respectively;
- (152) follows because for a given covariance matrix in (123), the Gaussian distribution maximizes the differential entropy;
- (154) follows from entropy power inequality (EPI) [38, pp. 22];
- (155) follows from (142).

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

$$R_3 \geq I(X_3; \hat{X}_3^* | X_{r,1}^*, X_{r,2}^*) \tag{156}$$

$$= H(X_3 | X_{r,1}^*, X_{r,2}^*) - H(X_3 | X_{r,1}^*, X_{r,2}^*, \hat{X}_3^*) \tag{157}$$

$$\geq H(X_3 | X_{r,1}^*, X_{r,2}^*) - H(X_3 | \hat{X}_1^*, \hat{X}_2^*, \hat{X}_3^*) \tag{158}$$

$$\geq H(X_3 | X_{r,1}^*, X_{r,2}^*) - H(X_3 | \hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) \tag{159}$$

$$= H(\rho_2 X_2 + N_2 | X_{r,1}^*, X_{r,2}^*) - H(X_3 | \hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) \tag{160}$$

$$\geq \frac{1}{2} \log \left(\rho_2^2 2^{2H(X_2 | X_{r,1}^*, X_{r,2}^*)} + 2^{2H(N_2)} \right) - H(X_3 | \hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) \tag{161}$$

$$\geq \frac{1}{2} \log \left(\rho_2^2 2^{-2R_2} 2^{2H(X_2 | X_{r,1}^*)} + 2^{2H(N_2)} \right) - H(X_3 | \hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) \tag{162}$$

$$\geq \frac{1}{2} \log \left(\rho_1^2 \rho_2^2 2^{-2R_1 - 2R_2} 2^{2H(X_1)} + \rho_2^2 2^{-2R_2} 2^{2H(N_1)} + 2^{2H(N_2)} \right) - H(X_3 | \hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G) \tag{163}$$

Next, we look at the distortion constraint.

<u>*Distortion Constraint*</u>: The choices in (122)–(124) imply that

$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j^*\|^2] = \mathbb{E}[\|X_j - \hat{X}_j^G\|^2], \qquad j = 1, 2, 3. \tag{164}$$

Finally, we look at the perception constraint

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

Define the following distribution

$$P_{U^* V^*} := \arg \inf_{\substack{\tilde{P}_{UV}: \\ \tilde{P}_U = P_{X_1} \\ \tilde{P}_V = P_{\hat{X}_1^*}}} \mathbb{E}_{\tilde{P}}[\|U - V\|^2]. \tag{165}$$

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

$$\text{cov}(U^G, V^G) = \text{cov}(U^*, V^*). \tag{166}$$

Then, we have the following set of inequalities:

$$
\begin{aligned}
P_1 \geq W_2^2(P_{X_1}, P_{\hat{X}_1^*}) &= \inf_{\substack{\tilde{P}_{UV}: \\ \tilde{P}_U = P_{X_1} \\ \tilde{P}_V = P_{\hat{X}_1^*}}} \mathbb{E}_{\tilde{P}}[\|U - V\|^2] && (167) \\
&= \mathbb{E}[\|U^* - V^*\|^2] && (168) \\
&= \mathbb{E}[\|U^G - V^G\|^2] && (169) \\
&\geq W_2^2(P_{U^G}, P_{V^G}) && (170) \\
&= \inf_{\substack{\hat{P}_{UV}: \\ \hat{P}_U = P_{U^G} \\ \hat{P}_V = P_{V^G}}} \mathbb{E}_{\hat{P}}[\|U - V\|^2] && (171) \\
&= \inf_{\substack{\hat{P}_{UV}: \\ \hat{P}_U = P_{X_1} \\ \hat{P}_V = P_{\hat{X}_1^G}}} \mathbb{E}_{\hat{P}}[\|U - V\|^2] && (172) \\
&= W_2^2(P_{X_1}, P_{\hat{X}_1^G}), && (173)
\end{aligned}
$$

where

- (168) follows from the definition in (165);
- (169) follows from (166) which implies that (U^*, V^*) and (U^G, V^G) have the same second-order statistics;
- (172) 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 $\sigma^2_{V^G}$ and recall that the variance of $\hat{X}_1^G$ is denoted by $\sigma^2_{\hat{X}_1^G}$. According to (166), $\sigma^2_{V^G}$ is equal to the variance of V^*. Also, from (165), 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 (122), we know that the variance of $\hat{X}_1^*$ is equal to $\sigma^2_{\hat{X}_1^G}$. Thus, we conclude that $\sigma^2_{\hat{X}_1^G} = \sigma^2_{V^G}$, 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 (125)–(127) where the coefficients are chosen according to distortion-perception constraints, provides an outer bound to $\mathcal{C}_{\mathsf{RDP}}$ which is the set of all tuples $(\mathsf{R}, \mathsf{D}, \mathsf{P})$ such that

$$
\begin{aligned}
& R_1 \geq I(X_1; \hat{X}_1^G), && (174) \\
& R_2 \geq \frac{1}{2} \log \left(\rho_1^2 2^{-2R_1} 2^{2H(X_1)} + 2^{2H(N_1)} \right) - H(X_2 | \hat{X}_1^G, \hat{X}_2^G), && (175) \\
& R_3 \geq \frac{1}{2} \log \left(\rho_1^2 \rho_2^2 2^{-2R_1 - 2R_2} 2^{2H(X_1)} + \rho_2^2 2^{-2R_2} 2^{2H(N_1)} + 2^{2H(N_2)} \right) - H(X_3 | \hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G), && (176) \\
& D_j \geq \mathbb{E}[\|X_j - \hat{X}_j^G\|^2], \qquad j = 1, 2, 3 && (177) \\
& P_j \geq W_2^2(P_{X_1 \ldots X_j}, P_{\hat{X}_1^G \ldots \hat{X}_j^G}). && (178)
\end{aligned}
$$

Now, we need to show that the above RDP region is also an inner bound to $\mathcal{C}_{\mathsf{RDP}}$. This is simply verified by the following choice. In iRDP region of (68)–(76), choose the following:

$$
X_{r,j} = \hat{X}_j = \hat{X}_j^G, \qquad j = 1, 2, 3, \tag{179}
$$

where $(\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G)$ satisfy (125)–(127) with coefficients chosen according to distortion-perception constraints. The lower bounds on distortion and perception constraints in (177) and (178) 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

$$
R_1 \geq I(X_1; \hat{X}_1^G), \tag{180}
$$

which immediately coincides with (174). The achievable rate of the second step can be written as follows

$$
\begin{aligned}
R_2 &\geq I(X_2; \hat{X}_2^G | \hat{X}_1^G) && (181)\\
&= H(X_2|\hat{X}_1^G) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (182)\\
&= H(\rho_1 X_1 + N_1|\hat{X}_1^G) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (183)\\
&= \frac{1}{2}\log(\rho_1^2 2^{2H(X_1|\hat{X}_1^G)} + 2^{2H(N_1)}) - H(X_2|\hat{X}_1^G, \hat{X}_2^G) && (184)\\
&\geq \frac{1}{2}\log\left(\rho_1^2 2^{-2R_1} 2^{2H(X_1)} + 2^{2H(N_1)}\right) - H(X_2|\hat{X}_1^G, \hat{X}_2^G), && (185)
\end{aligned}
$$

where

- (184) follows because EPI holds with "equality" for jointly Gaussian distributions [38, pp. 22];
- (185) follows from (175).

Thus, the bound in (185) coincides with (155). 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 $\mathcal{C}_{\mathsf{RDP}}$. ■

For a given rate R, the following corollary provides the optimization programs which lead to the characterization of the DP tradeoff $\mathcal{DP}(\mathsf{R})$ for the Gauss-Markov source model.

Corollary 1 *For a given rate tuple* R *and* $T = 2$ *frames, the optimal reconstructions of the DP-tradeoff* $\mathcal{DP}(\mathsf{R})$ *can be written as follows*

$$
\begin{aligned}
\hat{X}_1^G &= \nu X_1 + Z_1, && (186)\\
\hat{X}_2^G &= \omega_1 \hat{X}_1^G + \omega_2 X_2 + Z_2, && (187)
\end{aligned}
$$

where Z_1 *(resp* Z_2*) is a Gaussian random variable independent of* X_1 *(resp* $(\hat{X}_1^G, X_2)$*) and* $\hat{X}_j^G \sim \mathcal{N}(0, \hat{\sigma}_j^2)$ *for* $j = 1, 2$*, and* $\nu, \omega_1, \omega_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2$ *are the solutions of the following optimization program for the first step,*

$$
\begin{aligned}
&\min_{\nu, \hat{\sigma}_1^2} \ \sigma_1^2 + \hat{\sigma}_1^2 - 2\nu\sigma_1^2, && (188a)\\
\textit{s.t.} \quad &\nu^2\sigma_1^2 \leq \hat{\sigma}_1^2(1 - 2^{-2R_1}), && (188b)\\
&(\sigma_1 - \hat{\sigma}_1)^2 \leq P_1, && (188c)
\end{aligned}
$$

and the following minimization problem for the second step and PLF-FMD,

$$
\begin{aligned}
&\min_{\omega_1, \omega_2, \hat{\sigma}_2^2} \ \sigma_2^2 + \hat{\sigma}_2^2 - 2\nu\omega_1\rho_1\sigma_1\sigma_2 - 2\omega_2\sigma_2^2, && (189a)\\
\textit{s.t.} \quad &\omega_2^2\sigma_2^2(1 - 2^{-2R_2}\frac{\nu^2\rho_1^2\sigma_1^2}{\hat{\sigma}_1^2}) \leq (\hat{\sigma}_2^2 - \omega_1^2\hat{\sigma}_1^2 - 2\omega_1\omega_2\nu\rho_1\sigma_1\sigma_2)(1 - 2^{-2R_2}), && (189b)\\
&(\sigma_2 - \hat{\sigma}_2)^2 \leq P_2, && (189c)
\end{aligned}
$$

or the following minimization problem for the second step and PLF-JD,

$$
\begin{aligned}
&\min_{\omega_1, \omega_2, \hat{\sigma}_2^2} \ \sigma_2^2 + \hat{\sigma}_2^2 - 2\nu\omega_1\rho_1\sigma_1\sigma_2 - 2\omega_2\sigma_2^2 && (190a)\\
\textit{s.t.} \quad &\omega_2^2\sigma_2^2(1 - 2^{-2R_2}\frac{\nu^2\rho_1^2\sigma_1^2}{\hat{\sigma}_1^2}) \leq (\hat{\sigma}_2^2 - \omega_1^2\hat{\sigma}_1^2 - 2\omega_1\omega_2\nu\rho_1\sigma_1\sigma_2)(1 - 2^{-2R_2}), && (190b)\\
&tr(\Sigma_{12} + \hat{\Sigma}_{12} - 2(\Sigma_{12}^{1/2}\hat{\Sigma}_{12}\Sigma_{12}^{1/2})^{1/2}) \leq P_2, && (190c)
\end{aligned}
$$

where $tr(.)$ *denotes the trace of a matrix and*

$$
\begin{aligned}
\Sigma_{12} &:= \begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, && (191)\\
\hat{\Sigma}_{12} &:= \begin{pmatrix} \hat{\sigma}_1^2 & \omega_1\hat{\sigma}_1^2 + \nu\omega_2\rho_1\sigma_1\sigma_2 \\ \omega_1\hat{\sigma}_1^2 + \nu\omega_2\rho_1\sigma_1\sigma_2 & \hat{\sigma}_2^2 \end{pmatrix}. && (192)
\end{aligned}
$$

Proof: We obtain the optimization programs for $T = 2$ frames as follows.

For a given rate tuple R, the DP-tradeoff $\mathcal{DP}(\mathsf{R})$ is given by the set of all tuples (D, P) such that there exists $\hat{\mathsf{X}}^G$ satisfying the following Markov chains

$$\hat{X}_1^G \to X_1 \to X_2, \tag{193}$$

$$\hat{X}_2^G \to (\hat{X}_1^G, X_2) \to X_1, \tag{194}$$

and the following conditions,

$$R_1 \geq I(X_1; \hat{X}_1^G), \tag{195}$$

$$R_2 \geq I(X_2; \hat{X}_2^G | \hat{X}_1^G), \tag{196}$$

and

$$D_j \geq \mathbb{E}[\|X_j - \hat{X}_j^G\|^2], \qquad j = 1, 2, \tag{197}$$

$$P_j \geq W_2^2(P_{X_1 \ldots X_j}, P_{\hat{X}_1^G \ldots \hat{X}_j^G}). \tag{198}$$

In general, the set of reconstructions that satisfy (193)–(194) can be written as follows

$$\hat{X}_1^G = \nu X_1 + Z_1, \tag{199}$$

$$\hat{X}_2^G = \omega_1 \hat{X}_1^G + \omega_2 X_2 + Z_2. \tag{200}$$

Plugging the above into (195) and (196) yields the following rate expressions

$$\frac{1}{2} \log \frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 - \nu^2 \sigma_1^2} \leq R_1, \tag{201}$$

$$\frac{1}{2} \log \frac{\hat{\sigma}_2^2 - (\omega_1 \hat{\sigma}_1 + \frac{\omega_2 \nu \rho_1 \sigma_1 \sigma_2}{\hat{\sigma}_1})^2}{\hat{\sigma}_2^2 - \omega_1^2 \hat{\sigma}_1^2 - \omega_2^2 \sigma_2^2 - 2\omega_1 \omega_2 \nu \rho_1 \sigma_1 \sigma_2} \leq R_2. \tag{202}$$

Re-arranging the terms in the above constraints yields the conditions in (188b) and (190b). Considering (197) with (199)–(200) gives the following expressions for distortions

$$\mathbb{E}[\|X_1 - \hat{X}_1^G\|^2] = \sigma_1^2 + \hat{\sigma}_1^2 - 2\mathbb{E}[X_1 \hat{X}_1^G] = \sigma_1^2 + \hat{\sigma}_1^2 - 2\nu\sigma_1^2, \tag{203}$$

$$\mathbb{E}[\|X_2 - \hat{X}_2^G\|^2] = \sigma_2^2 + \hat{\sigma}_2^2 - 2\mathbb{E}[X_2 \hat{X}_2^G] = \sigma_2^2 + \hat{\sigma}_2^2 - 2\omega_1 \nu \rho_1 \sigma_1 \sigma_2 - 2\omega_2 \sigma_2^2, \tag{204}$$

which are the objective functions in (188a) and (190a). Now, we evaluate the perception constraint. Notice that the covariance matrices of (X_1, X_2) and $(\hat{X}_1^G, \hat{X}_2^G)$ are given by Σ_{12} and $\hat{\Sigma}_{12}$ defined in (191) and (192), respectively. The Wasserstein-2 distance between two Gaussian distributions with covariance matrices Σ_{12} and $\hat{\Sigma}_{12}$ is given in (190c) as discussed in [26, pp. 18].

Similarly, the expressions in (189) for the decoder based on PLF-FMD can be obtained. ■

F Gauss-Markov Source Model: Extremal Rates

In this section, we derive the achievable reconstructions for some special cases. We assume that we have only two frames, i.e., $D_3, P_3 \to \infty$. Moreover, let $\sigma_1^2 = \sigma_2^2 := \sigma^2$ for simplicity. In general, the reconstructions can be written as follows

$$\hat{X}_1^G = \nu X_1 + Z_1, \tag{205}$$

$$\hat{X}_2^G = \omega_1 \hat{X}_1^G + \omega_2 X_2 + Z_2, \tag{206}$$

where $\hat{X}_j^G \sim \mathcal{N}(0, \hat{\sigma}_j^2)$ for $j = 1, 2$. Recall the optimization program of the first step in (188) as follows

$$\min_{\nu, \hat{\sigma}_1^2} \ \sigma^2 + \hat{\sigma}_1^2 - 2\nu\sigma^2, \tag{207a}$$

$$\text{s.t.} \qquad \nu^2 \sigma^2 \leq \hat{\sigma}_1^2 (1 - 2^{-2R_1}), \tag{207b}$$

$$(\sigma - \hat{\sigma}_1)^2 \leq P_1, \tag{207c}$$

For a given $\hat{\sigma}_1^2$, the objective function in (207a) is a monotonically deacreasing function of ν, hence one can restrict ν to be nonnegative, without loss of optimality. So, the above optimization program can be written as

$$\min_{\nu,\hat{\sigma}_1^2} \ \sigma^2 + \hat{\sigma}_1^2 - 2\nu\sigma^2, \tag{208a}$$

$$\text{s.t.} \qquad 0 \le \nu \le \frac{\hat{\sigma}_1}{\sigma}\sqrt{1-2^{-2R_1}}, \tag{208b}$$

$$(\sigma - \hat{\sigma}_1)^2 \le P_1, \tag{208c}$$

Optimizing with respect to ν in the above program, we have

$$\nu = \frac{\hat{\sigma}_1}{\sigma}\sqrt{1-2^{-2R_1}}, \tag{209}$$

where the optimization program reduces to

$$\min_{\hat{\sigma}_1^2} \ \sigma^2 + \hat{\sigma}_1^2 - 2\sigma\hat{\sigma}_1\sqrt{1-2^{-2R_1}}, \tag{210a}$$

$$\text{s.t.} \qquad (\sigma - \hat{\sigma}_1)^2 \le P_1. \tag{210b}$$

Next, recall the optimization program of the second step for PLF-FMD in (189) as follows

$$\min_{\omega_1,\omega_2,\hat{\sigma}_2^2} \ \sigma^2 + \hat{\sigma}_2^2 - 2\nu\omega_1\rho_1\sigma^2 - 2\omega_2\sigma^2, \tag{211a}$$

$$\text{s.t.} \qquad \omega_2^2\sigma^2(1 - 2^{-2R_2}\frac{\nu^2\rho_1^2\sigma^2}{\hat{\sigma}_1^2}) \le (\hat{\sigma}_2^2 - \omega_1^2\hat{\sigma}_1^2 - 2\omega_1\omega_2\nu\rho_1\sigma^2)(1-2^{-2R_2}), \tag{211b}$$

$$(\sigma - \hat{\sigma}_2)^2 \le P_2, \tag{211c}$$

Plugging (209) into the above program, we get

$$\min_{\omega_1,\omega_2,\hat{\sigma}_2^2} \ \sigma^2 + \hat{\sigma}_2^2 - 2\omega_1\rho_1\hat{\sigma}_1\sigma\sqrt{1-2^{-2R_1}} - 2\omega_2\sigma^2, \tag{212a}$$

$$\text{s.t.} \qquad \omega_2^2\sigma^2(1 - \rho_1^2 2^{-2R_2}(1-2^{-2R_1})) \le (\hat{\sigma}_2^2 - \omega_1^2\hat{\sigma}_1^2 - 2\omega_1\omega_2\rho_1\hat{\sigma}_1\sigma\sqrt{1-2^{-2R_1}})(1-2^{-2R_2}), \tag{212b}$$

$$(\sigma - \hat{\sigma}_2)^2 \le P_2, \tag{212c}$$

The optimization program for the second step of PLF-JD is similar to the above program (212) when (212c) is replaced by (190c). In this section, we study different rate regimes and obtain the solutions of the above optimization programs. In particular, we are interested in two perception thresholds $P_2 \to \infty$ and $P_2 = 0$ where the former corresponds to the classical rate-distortion region and the latter is the case of 0-PLF. For the 0-PLF-FMD, we have $\hat{\sigma}_1 = \hat{\sigma}_2 = \sigma$. For the 0-PLF-JD, in addition to preserving the marginals, the correlation $\mathbb{E}[\hat{X}_1^G\hat{X}_2^G] = \rho_1\sigma^2$ should be satisfied. For each of these cases, the optimization program in (212) is simplified in the following.

Optimization Program of the Second Step for $P \to \infty$: In this case, there is no perception constraint in the setting and the optimization program in (212) reduces to the following

$$\min_{\hat{\sigma}_2^2,\omega_1,\omega_2} \ \sigma^2 + \hat{\sigma}_2^2 - 2\omega_1\rho_1\hat{\sigma}_1\sigma\sqrt{1-2^{-2R_1}} - 2\omega_2\sigma^2, \tag{213a}$$

$$\text{s.t.} \qquad \omega_2^2\sigma^2(1 - \rho_1^2 2^{-2R_2}(1-2^{-2R_1})) \le (\hat{\sigma}_2^2 - \omega_1^2\hat{\sigma}_1^2 - 2\omega_1\omega_2\rho_1\hat{\sigma}_1\sigma\sqrt{1-2^{-2R_1}})(1-2^{-2R_2}). \tag{213b}$$

This case corresponds to the classical rate-distortion tradeoff where it is shown that for a given rate, the MMSE reconstructions are indeed optimal [28, 37]. The expressions for MMSE reconstructions are given in Appendix H.1.

Optimization Program of the Second Step for 0-PLF-FMD: In this case, we have $\hat{\sigma}_1 = \hat{\sigma}_2 = \sigma$. So, the optimization program in (212) reduces to the following

$$\min_{\omega_1,\omega_2} \ 2\sigma^2 - 2\omega_1\rho_1\sigma^2\sqrt{1-2^{-2R_1}} - 2\omega_2\sigma^2, \tag{214a}$$

$$\text{s.t.} \qquad \omega_2^2(1 - \rho_1^2 2^{-2R_2}(1-2^{-2R_1})) \le (1 - \omega_1^2 - 2\omega_1\omega_2\rho_1\sqrt{1-2^{-2R_1}})(1-2^{-2R_2}). \tag{214b}$$

Here, ω_1 and ω_2 only need to satisfy the rate constraint given in (214b) which represents a larger search space than that of 0-PLF-JD which will be discussed in the following.

Optimization Program of the Second Step for 0*-PLF-JD*: In this case, in addition to preserving marginals $\hat{\sigma}_1 = \hat{\sigma}_2 = \sigma$, we need to satisfy the constraint $\mathbb{E}[\hat{X}_1^G \hat{X}_2^G] = \rho_1\sigma^2$. Thus, the optimization program of this case has an extra condition $\omega_1 + \nu\omega_2\rho_1 = \rho_1$ comparing to (214) and it is given as follows

$$\begin{aligned} &\min_{\omega_1,\omega_2} \;\; 2\sigma^2 - 2\omega_1\rho_1\sigma^2\sqrt{1-2^{-2R_1}} - 2\omega_2\sigma^2, && \text{(215a)} \\ \text{s.t.} \quad & \omega_2^2(1-\rho_1^2 2^{-2R_2}(1-2^{-2R_1})) \leq (1-\omega_1^2 - 2\omega_1\omega_2\rho_1\sqrt{1-2^{-2R_1}})(1-2^{-2R_2}), \\ & \omega_1 + \nu\omega_2\rho_1 = \rho_1. && \text{(215b)} \end{aligned}$$

Comparing (215) with (214), we notice that the search space of the optimization program for 0-PLF-JD is smaller than that of 0-PLF-FMD. Thus, a larger distortion is expected for 0-PLF-JD.

Before studying each case of extremal rates, we introduce another constraint in the optimization program of all above three cases of perception metrics. We restrict to nonnegative $\omega_1\omega_2\rho_1$ and get an upper bound on the programs (213), (214) and (215). So, in further discussion on these programs, the constraint $\omega_1\omega_2\rho_1 \geq 0$ will be also considered.

1) $R_1 = R_2 = \epsilon$ for small ϵ:

In the low-rate regime, notice that we can approximate the rate term as follows

$$1 - 2^{-2\epsilon} = 2\epsilon\ln 2 + O(\epsilon^2). \tag{216}$$

Plugging the above into (209), we have

$$\nu = \frac{\hat{\sigma}_1}{\sigma}\sqrt{2\epsilon\ln 2 + O(\epsilon^2)}. \tag{217}$$

Also, inserting (216) into the rate constraint of the second step (211c) yields the following

$$\omega_2^2\sigma^2(1-\rho_1^2 2\epsilon\ln 2 + O(\epsilon^2)) \leq (\hat{\sigma}_2^2 - \omega_1^2\hat{\sigma}_1^2 - 2\omega_1\omega_2\rho_1\hat{\sigma}_1\sigma\sqrt{2\epsilon\ln 2 + O(\epsilon^2)})(2\epsilon\ln 2 + O(\epsilon^2)).^2 \tag{218}$$

Re-arranging the terms in the above inequality yields the following

$$\begin{aligned} \hat{\sigma}_2^2 &\geq \frac{\omega_2^2\sigma^2(1-\rho_1^2 2\epsilon\ln 2 + O(\epsilon^2))}{2\epsilon\ln 2 + O(\epsilon^2)} + \omega_1^2\hat{\sigma}_1^2 + 2\omega_1\omega_2\rho_1\hat{\sigma}_1\sigma\sqrt{2\epsilon\ln 2 + O(\epsilon^2)} && \text{(219)} \\ &= \omega_2^2\sigma^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2\hat{\sigma}_1^2 + 2\omega_1\omega_2\rho_1\hat{\sigma}_1\sigma\sqrt{2\epsilon\ln 2 + O(\epsilon^2)} && \text{(220)} \end{aligned}$$

So, in all of the optimization programs of the case $R_1 = R_2 = \epsilon$, the above constraint (220) will replace the rate constraint of the second step.

Now, we consider different cases based on the perception measure.

a) Without a perception constraint: In this case, using (216), the optimization program of the first step in (210) simplifies to the following

$$D_1 = \min_{\hat{\sigma}_1^2} \; \sigma^2 + \hat{\sigma}_1^2 - 2\sigma\hat{\sigma}_1\sqrt{2\epsilon\ln 2 + O(\epsilon^2)}, \tag{221}$$

which gives us the following optimal solution

$$\hat{\sigma}_1 = \sqrt{2\epsilon\ln 2 + O(\epsilon^2)}\sigma = \sqrt{2\epsilon\ln 2}\sigma + O(\epsilon). \tag{222}$$

Plugging the above solution into (217) and (221), we get

$$\nu = 2\epsilon\ln 2 + O(\epsilon^2), \tag{223}$$

[2] The inequalities of the form $f(\epsilon) + O(\epsilon^2) \leq g(\epsilon) + O(\epsilon^2)$, where $f(\epsilon), g(\epsilon) = \Omega(\epsilon^2)$, imply that $f(\epsilon) \leq g(\epsilon)$. So, in such inequalities, we work with dominant terms ($f(\epsilon), g(\epsilon)$) and ignore the small terms $O(\epsilon^2)$. A similar argument holds if we have other orders of ϵ and the functions $f(.), g(.)$ approach zero slower than them.

and

$$D_1 = (1-2\epsilon \ln 2)\sigma^2 + O(\epsilon^2). \tag{224}$$

Now, we look at the optimization program of the second step (213). For a given ω_1 and ω_2, the objective function is an increasing function of $\hat{\sigma}_2^2$, so optimizing over $\hat{\sigma}_2^2$ yields the following

$$\hat{\sigma}_2^2 = \omega_2^2\sigma^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2\hat{\sigma}_1^2 + 2\omega_1\omega_2\rho_1\hat{\sigma}_1\sigma\sqrt{2\epsilon\ln 2 + O(\epsilon^2)}. \tag{225}$$

Thus, the optimization program (213) is further upper bounded by the following

$$\min_{\substack{\hat{\sigma}_2^2,\omega_1,\omega_2:\\ \omega_1\omega_2\rho_1\geq 0}} \sigma^2 + \omega_2^2\sigma^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2\hat{\sigma}_1^2 - 2(1-\omega_2)\omega_1\rho_1\hat{\sigma}_1\sigma\sqrt{2\epsilon\ln 2 + O(\epsilon^2)} - 2\omega_2\sigma^2. \tag{226}$$

The optimal solution of the above minimization is given by the following

$$\omega_1 = \rho_1 + O(\epsilon), \tag{227}$$
$$\omega_2 = 2\epsilon\ln 2 + O(\epsilon^2). \tag{228}$$

Thus, considering the dominant terms of (223), (227) and (228), we have

$$\hat{X}_1^G = (2\epsilon\ln 2)X_1 + Z_1, \tag{229}$$
$$\hat{X}_2^G = \rho_1\hat{X}_1^G + (2\epsilon\ln 2)X_2 + Z_2, \tag{230}$$

and $Z_j \sim \mathcal{N}(0, 2\epsilon\sigma^2\ln 2)$ for $j=1,2$. Notice that

$$D_1 = (1-2\epsilon\ln 2)\sigma^2, \tag{231}$$
$$D_2 = (1-(1+\rho_1^2)2\epsilon\ln 2)\sigma^2. \tag{232}$$

b) 0-*PLF-FMD*: In this case, we have $\hat{\sigma}_1 = \hat{\sigma}_2 = \sigma$. For the optimization program of the first step, (209) reduces to the following

$$\nu = \sqrt{2\epsilon\ln 2} + O(\epsilon), \tag{233}$$

and D_1 is given in the following which is derived by (210)

$$D_1 = 2(1-\sqrt{2\epsilon\ln 2})\sigma^2 + O(\epsilon). \tag{234}$$

Now, we study the optimization program of the second step. The optimization program of (214) is further upper bounded by the following

$$\min_{\substack{\omega_1,\omega_2:\\ \omega_1\omega_2\rho_1\geq 0}} \quad 2\sigma^2 - 2\omega_1\rho_1\sigma^2\sqrt{2\epsilon\ln 2 + O(\epsilon^2)} - 2\omega_2\sigma^2, \tag{235a}$$

$$\text{s.t.} \qquad 1 \geq \sqrt{\omega_2^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2 + 2\omega_1\omega_2\rho_1\sqrt{2\epsilon\ln 2 + O(\epsilon^2)}}. \tag{235b}$$

Now, we further simplify the inequality (235b) in the following. Considering the fact that $\omega_1\omega_2\rho_1 \geq 0$, this inequality implies that

$$\omega_1^2 \leq 1, \tag{236}$$
$$\omega_2^2 \leq 2\epsilon\ln 2 + O(\epsilon^2). \tag{237}$$

So, using the above inequalities, the RHS of (235b) can be upper bounded as follows

$$\begin{aligned}&\sqrt{\omega_2^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2 + 2\omega_1\omega_2\rho_1\sqrt{2\epsilon\ln 2 + O(\epsilon^2)}}\\ &\qquad \leq \sqrt{\omega_2^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2 + (\omega_1^2+\omega_2^2)\rho_1\sqrt{2\epsilon\ln 2 + O(\epsilon^2)}}\\ &\qquad \leq \sqrt{\omega_2^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2 + O(\epsilon^{3/2})}.\end{aligned} \tag{238}$$

https://doi.org/10.52202/075280-2093

Now, according to (238), the optimization program in (235) is further upper bounded by the following

$$\min_{\substack{\omega_1,\omega_2:\\ \omega_1\omega_2\rho_1\geq 0}} \quad 2\sigma^2 - 2\omega_1\rho_1\sigma^2\sqrt{2\epsilon\ln 2 + O(\epsilon^2)} - 2\omega_2\sigma^2, \tag{239a}$$

$$\text{s.t.} \qquad 1 \geq \sqrt{\omega_2^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2 + O(\epsilon^{3/2})}. \tag{239b}$$

For a given ω_1 (resp ω_2), the objective function (239a) is a monotonically decreasing function of ω_2 (resp ω_1), so the optimal solution is attained on the boundary, i.e.,

$$1 = \sqrt{\omega_2^2\left(\frac{1}{2\epsilon\ln 2} + O\left(1\right)\right) + \omega_1^2 + O(\epsilon^{3/2})} \tag{240}$$

Thus, the program (239) further simplifies to the following

$$\min_{\substack{\omega_1:\\ \omega_1\rho_1\geq 0}} 2\sigma^2 - 2\omega_1\rho_1\sigma^2\sqrt{2\epsilon\ln 2 + O(\epsilon^2)} - 2\sigma^2\sqrt{(1-\omega_1^2 - O(\epsilon^{3/2}))(2\epsilon\ln 2 + O(\epsilon^2))}. \tag{241}$$

The optimal solution of the above program is given by

$$\omega_1 = \frac{\rho_1}{\sqrt{1+\rho_1^2}} + O\left(\epsilon\right), \tag{242}$$

which together with (240) yields

$$\omega_2 = \sqrt{\frac{2\epsilon\ln 2}{1+\rho_1^2}} + O(\epsilon). \tag{243}$$

Thus, considering dominant terms of (233), (242) and (243), we get

$$\hat{X}_1^G = \sqrt{2\epsilon\ln 2}X_1 + Z_1, \tag{244}$$

$$\hat{X}_2^G = \frac{\rho_1}{\sqrt{1+\rho_1^2}}\hat{X}_1^G + \sqrt{\frac{2\epsilon\ln 2}{1+\rho_1^2}}X_2 + Z_2, \tag{245}$$

where $Z_1 \sim \mathcal{N}(0, (1-2\epsilon\ln 2)\sigma^2)$ and

$$Z_2 \sim \mathcal{N}(0, (1 - \frac{\rho_1^2}{1+\rho_1^2} - \frac{1+2\rho_1^2}{1+\rho_1^2}2\epsilon\ln 2)\sigma^2). \tag{246}$$

Notice that

$$D_1 = 2(1-\sqrt{2\epsilon\ln 2})\sigma^2, \tag{247}$$

$$D_2 = 2(1-\sqrt{(1+\rho_1^2)2\epsilon\ln 2})\sigma^2. \tag{248}$$

For the special case of $\rho_1 = 1$, the expressions in (244) and (245) simplify as follows

$$\hat{X}_1^G = \sqrt{2\epsilon\ln 2}X_1 + Z_1, \tag{249}$$

$$\hat{X}_2^G = \sqrt{2}\sqrt{2\epsilon\ln 2}X_1 + \frac{1}{\sqrt{2}}Z_1 + Z_2. \tag{250}$$

Define $Z_{\text{FMD}} := \frac{1}{\sqrt{2}}Z_1 + Z_2$ and notice that $Z_{\text{FMD}} \sim \mathcal{N}(0, (1-4\epsilon\ln 2)\sigma^2)$. Moreover, we have

$$D_1 = 2(1-\sqrt{2\epsilon\ln 2})\sigma^2, \tag{251}$$

$$D_2 = 2(1-\sqrt{4\epsilon\ln 2})\sigma^2. \tag{252}$$

c) 0-PLF-JD: In this case, the optimization program of the first step is similar to the previous case. The optimization program of the second step is given in (215) where the condition $\omega_1 + \nu\omega_2\rho_1 = \rho_1$ is introduced. According to (233), $\nu = O(\sqrt{\epsilon})$ which suggests the following form for ω_1,

$$\omega_1 = \rho_1 - \delta_\epsilon, \tag{253}$$

https://doi.org/10.52202/075280-2093

for some small δ_ϵ that goes to zero as $\epsilon \to 0$. The parameter δ_ϵ will be determined later. Plugging $\omega_1 = \rho_1 - \delta_\epsilon$ into (240), we find out that only the constant term of ω_1 contributes to a dominant term for ω_2 which yields the following

$$\omega_2 = \sqrt{2\epsilon \ln 2(1-\rho_1^2)} + O(\epsilon). \tag{254}$$

Thus, we have

$$\begin{aligned} \hat{X}_1^G &= \sqrt{2\epsilon \ln 2} X_1 + Z_1, &(255)\\ \hat{X}_2^G &= (\rho_1 - \delta_\epsilon)\hat{X}_1^G + \sqrt{(1-\rho_1^2)2\epsilon \ln 2} X_2 + Z_2, &(256)\end{aligned}$$

Now, applying the constraint $\mathbb{E}[\hat{X}_1^G \hat{X}_2^G] = \rho_1 \sigma^2$, we get

$$\delta_\epsilon = \rho_1 \sqrt{1-\rho_1^2}(2\epsilon \ln 2). \tag{257}$$

However, notice that since $\delta_\epsilon = O(\epsilon)$, it does not contribute to dominant terms of distortion. So, we can simply represent $\hat{X}_1^G$ and $\hat{X}_2^G$ as follows

$$\begin{aligned} \hat{X}_1^G &= \sqrt{2\epsilon \ln 2} X_1 + Z_1, &(258)\\ \hat{X}_2^G &= \rho_1 \hat{X}_1^G + \sqrt{(1-\rho_1^2)2\epsilon \ln 2} X_2 + Z_2, &(259)\end{aligned}$$

where $Z_1 \sim \mathcal{N}(0, (1-2\epsilon \ln 2)\sigma^2)$ and $Z_2 \sim \mathcal{N}(0, (1-\rho_1^2 - (1-\rho_1^2 + 2\rho_1^2\sqrt{1-\rho_1^2})2\epsilon \ln 2)\sigma^2)$. The following distortions are also achievable

$$\begin{aligned} D_1 &= 2(1-\sqrt{2\epsilon \ln 2})\sigma^2, &(260)\\ D_2 &= 2(1-(\rho_1^2 + \sqrt{1-\rho_1^2})\sqrt{2\epsilon \ln 2})\sigma^2. &(261)\end{aligned}$$

For the special case of $\rho = 1$, according to (259) and (261), we have $\hat{X}_2^G = \hat{X}_1^G$ and $D_2 = D_1$.

2) $R_1 \to \infty$, $R_2 = \epsilon$ for small ϵ: In this case, since $R_1 \to \infty$, we have $\hat{X}_1^G = X_1$, $D_1 = 0$, and we only need to solve the optimization program of the second step. Also, we have the following approximation

$$1 - 2^{-2R_2} = 1 - 2^{-2\epsilon} = 2\epsilon \ln 2 + O(\epsilon^2). \tag{262}$$

We consider three different cases based on the perception constraint.

a) Without a perception constraint: In this case, consider the optimization program (213). For a given ω_1 and ω_2, the objective function is an increasing function of $\hat{\sigma}_2^2$, hence optimizing over $\hat{\sigma}_2^2$, we get

$$\hat{\sigma}_2^2 = \frac{\omega_2^2\sigma^2(1-\rho_1^2 + O(\epsilon))}{2\epsilon \ln 2 + O(\epsilon^2)} + \omega_1^2\sigma^2 + 2\omega_1\omega_2\rho_1\sigma^2. \tag{263}$$

The program in (213) is further upper bounded by the following

$$\min_{\substack{\omega_1,\omega_2:\\ \omega_1\omega_2\rho_1 \geq 0}} \sigma^2 + \frac{\omega_2^2\sigma^2(1-\rho_1^2 + O(\epsilon))}{2\epsilon \ln 2 + O(\epsilon^2)} + \omega_1^2\sigma^2 + 2\omega_1\omega_2\rho_1\sigma^2 - 2\omega_1\rho_1\sigma^2 - 2\omega_2\sigma^2, \tag{264}$$

The solution of the above optimization program is given by the following

$$\begin{aligned} \omega_1 &= \rho_1 - \rho_1(2\epsilon \ln 2), &(265)\\ \omega_2 &= 2\epsilon \ln 2. &(266)\end{aligned}$$

Thus, we have

$$\begin{aligned} \hat{X}_1^G &= X_1, &(267)\\ \hat{X}_2^G &= (\rho_1 - \rho_1(2\epsilon \ln 2))X_1 + (2\epsilon \ln 2)X_2 + Z_2, &(268)\end{aligned}$$

where $Z_2 \sim \mathcal{N}(0, (1-\rho_1^2)\sigma^2 2\epsilon \ln 2)$. So, the reconstruction of the second frame closely resembles the first frame. The distortions of the first and second frames are zero and $(1-\rho_1^2-(1-\rho_1^2)2\epsilon \ln 2)\sigma^2$, respectively.

b) 0*-PLF-FMD*: In this case, $\hat{\sigma}_1 = \hat{\sigma}_2 = \sigma$. Thus, the optimization program in (214) is further upper bounded by the following

$$\min_{\substack{\omega_1,\omega_2:\\ \omega_1\omega_2\rho_1\geq 0}} \quad 2\sigma^2 - 2\omega_1\rho_1\sigma^2 - 2\omega_2\sigma^2, \tag{269a}$$

$$\text{s.t.} \qquad \omega_2^2(1-\rho_1^2+O(\epsilon)) \leq (1-\omega_1^2-2\omega_1\omega_2\rho_1)(2\epsilon\ln 2+O(\epsilon^2)). \tag{269b}$$

For a given ω_1 (resp ω_2), the objective function (269a) is a monotonically decreasing function of ω_2 (resp ω_1). So, the optimal solution is attained on the boundary, i.e., (269b) is satisfied with equality given as follows

$$\omega_2^2(1-\rho_1^2+O(\epsilon)) = (1-\omega_1^2-2\omega_1\omega_2\rho_1)(2\epsilon\ln 2+O(\epsilon^2)). \tag{270}$$

It can be easily verified that the first-order terms of ω_1 and ω_2 which optimize the program are 1 and 0, respectively. So, we write ω_1 and ω_2 in the following form

$$\omega_1 = 1 + (2\epsilon\ln 2)\delta_1 + O(\epsilon^2), \tag{271}$$

$$\omega_2 = (2\epsilon\ln 2)\delta_2 + O(\epsilon^2), \tag{272}$$

for some real δ_1 and δ_2. Plugging the above (271) and (272) into (270) and considering the dominant terms, we get

$$\delta_2^2(1-\rho_1^2) = -2\delta_1 - 2\rho_1\delta_2. \tag{273}$$

On the other side, we can write the objective function in (269) as follows

$$\begin{aligned} &2\sigma^2 - 2\omega_1\rho_1\sigma^2 - 2\omega_2\sigma^2 \\ &\quad = 2\sigma^2 - 2\rho_1\omega_1\sigma^2 - 2\omega_2\sigma^2 + O(\epsilon^2) && (274)\\ &\quad = 2\sigma^2 - 2\rho_1\sigma^2 - 2(\rho_1\delta_1\sigma^2 + \delta_2\sigma^2)(2\epsilon\ln 2) + O(\epsilon^2) && (275)\\ &\quad = 2\sigma^2 - 2\rho_1\sigma^2 - (-2\rho_1^2\delta_2\sigma^2 - \rho_1(1-\rho_1^2)\delta_2^2 + 2\delta_2\sigma^2)(2\epsilon\ln 2) + O(\epsilon^2). && (276)\end{aligned}$$

Differentiating the above expression with respect to δ_2 and letting it be zero, we have:

$$\delta_2 = \frac{1}{\rho_1}, \qquad \delta_1 = -\frac{1+\rho_1^2}{2\rho_1^2}. \tag{277}$$

Thus, we have

$$\hat{X}_1^G = X_1, \tag{278}$$

$$\hat{X}_2^G = (1 - \frac{(1+\rho_1^2)2\epsilon\ln 2}{2\rho_1^2})\hat{X}_1^G + \frac{2\epsilon\ln 2}{\rho_1}X_2 + Z_2, \tag{279}$$

where $Z_2 \sim \mathcal{N}(0, (\frac{1-\rho_1^2}{\rho_1^2})2\epsilon\ln 2)$. Again, the reconstruction of the second frame is almost similar to the first frame and the distortion is $2(1-\rho_1-(\frac{1-\rho_1^2}{2\rho_1})2\epsilon\ln 2)\sigma^2$.

c) 0*-PLF-JD*: First consider the case where $\rho_1 \neq 1$. The optimization program is given in (215) where the constraint $\omega_1 + \nu\rho_1\omega_2 = \rho_1$ is introduced. Notice that ω_1 can be written in the following form

$$\omega_1 = \rho_1 + \delta_\epsilon, \tag{280}$$

for some δ_ϵ that goes to zero as $\epsilon \to 0$. The parameter δ_ϵ will be determined later. Plugging $\omega_1 = \rho_1 + \delta_\epsilon$ into (270) yields the following

$$\omega_2 = \sqrt{2\epsilon\ln 2} + O(\epsilon), \tag{281}$$

which is derived only through the first-order term of ω_1 which is ρ_1. Now, considering the fact that $\mathbb{E}[\hat{X}_1^G\hat{X}_2^G] = \rho_1\sigma^2$, we obtain

$$\delta_\epsilon = -\rho_1\sqrt{2\epsilon\ln 2}. \tag{282}$$

Thus, we have

$$\hat{X}_1^G = X_1, \tag{283}$$

$$\hat{X}_2^G = (\rho_1 - \rho_1\sqrt{2\epsilon\ln 2})\hat{X}_1^G + \sqrt{2\epsilon\ln 2}X_2 + Z_2, \tag{284}$$

where $Z_2 \sim \mathcal{N}(0, (1-\rho_1^2)\sigma^2)$. Here, the reconstruction of the second frame closely resembles the first frame. The distortion of the second frame is $2(1-\rho_1^2-(1-\rho_1^2)\sqrt{2\epsilon\ln 2})\sigma^2$.

If $\rho_1 = 1$, we simply have $\hat{X}_2^G = \hat{X}_1^G = X_1 = X_2$ which can be derived from (283)–(284) by letting $X_1 = X_2$.

The analysis for the case of $R_1 = \epsilon$ and $R_2 \to \infty$ is similar and is omitted for brevity. The results of this section are summarized in Table 2.

https://doi.org/10.52202/075280-2093

Table 2: Achievable reconstructions for extremal rates and different PLFs (The first, second and third rows represent reconstructions corresponding to the MMSE, 0-PLF-FMD and 0-PLF-JD, respectively).

	$R_1 = R_2 = \epsilon$	$R_1 \to \infty, R_2 = \epsilon$	$R_1 = \epsilon, R_2 = \infty$
MMSE	$\hat{X}_1^G{=}(2\epsilon \ln 2)X_1 + Z_1$ $\hat{X}_2^G{=}\rho_1 \hat{X}_1^G + (2\epsilon \ln 2)X_2 + Z_2$ $Z_j{\sim}\mathcal{N}(0, 2\epsilon\sigma^2 \ln 2)$ $D_1{=}(1 - 2\epsilon \ln 2)\sigma^2$ $D_2{=}(1 - (1 + \rho_1^2)2\epsilon \ln 2)\sigma^2$	$\hat{X}_1^G{=}X_1$ $\hat{X}_2^G{=}(\rho_1 - \rho_1 2\epsilon \ln 2)\hat{X}_1^G + (2\epsilon \ln 2)X_2 + Z_2$ $Z_2{\sim}\mathcal{N}(0, (1 - \rho_1^2)2\epsilon\sigma^2 \ln 2)$ $D_1{=}0$ $D_2{=}(1 - \rho_1^2 - (1 - \rho_1^2)2\epsilon \ln 2)\sigma^2$	$\hat{X}_1^G{=}(2\epsilon \ln 2)X_1 + Z_1$ $\hat{X}_2^G{=}X_2$ $Z_1{\sim}\mathcal{N}(0, 2\epsilon\sigma^2 \ln 2)$ $D_1{=}(1 - 2\epsilon \ln 2)\sigma^2$ $D_2{=}0$
0-PLF-FMD	$\hat{X}_1^G{=}\sqrt{2\epsilon \ln 2}X_1 + Z_1$ $\hat{X}_2^G{=}\frac{\rho_1}{\sqrt{1+\rho_1^2}}\hat{X}_1^G + \sqrt{\frac{2\epsilon \ln 2}{1+\rho_1^2}}X_2 + Z_2$ $Z_1{\sim}\mathcal{N}(0, (1 - 2\epsilon \ln 2)\sigma^2)$ $Z_2{\sim}\mathcal{N}(0, (1{-}\frac{\rho_1^2}{1+\rho_1^2} - \frac{1+2\rho_1^2}{1+\rho_1^2}2\epsilon \ln 2)\sigma^2)$ $D_1{=}2(1 - \sqrt{2\epsilon \ln 2})\sigma^2$ $D_2{=}2(1 - \sqrt{(1 + \rho_1^2)2\epsilon \ln 2})\sigma^2$	$\hat{X}_1^G{=}X_1$ $\hat{X}_2^G{=}(1 - \frac{(1+\rho_1^2)2\epsilon \ln 2}{2\rho_1^2})\hat{X}_1^G + \frac{2\epsilon \ln 2}{\rho_1}X_2 + Z_2$ $Z_2{\sim}\mathcal{N}(0, (\frac{1-\rho_1^2}{\rho_1^2})2\epsilon \ln 2)$ $D_1{=}0$ $D_2{=}2(1 - \rho_1 - (\frac{1-\rho_1^2}{2\rho_1})2\epsilon \ln 2)\sigma^2$	$\hat{X}_1^G{=}\sqrt{2\epsilon \ln 2}X_1 + Z_1$ $\hat{X}_2^G{=}X_2$ $Z_1{\sim}\mathcal{N}(0, (1 - 2\epsilon \ln 2)\sigma^2)$ $D_1{=}2(1 - \sqrt{2\epsilon \ln 2})\sigma^2$ $D_2{=}0$
0-PLF-JD	$\hat{X}_1^G{=}\sqrt{2\epsilon \ln 2}X_1 + Z_1$ $\hat{X}_2^G{=}\rho_1 \hat{X}_1^G + \sqrt{(1 - \rho_1^2)2\epsilon \ln 2}X_2 + Z_2$[a] $Z_1{\sim}\mathcal{N}(0, (1 - 2\epsilon \ln 2)\sigma^2)$ $Z_2{\sim}\mathcal{N}(0, (1 - \rho_1^2 - (1 - \rho_1^2)2\epsilon \ln 2)\sigma^2)$ $D_1{=}2(1 - \sqrt{2\epsilon \ln 2})\sigma^2$ $D_2{=}2(1 - (\rho_1^2 + \sqrt{1 - \rho_1^2})\sqrt{2\epsilon \ln 2})\sigma^2$	$\hat{X}_1^G{=}X_1$ $\hat{X}_2^G{=}(\rho_1 - \rho_1\sqrt{2\epsilon \ln 2})\hat{X}_1^G + \sqrt{2\epsilon \ln 2}X_2 + Z_2$ $Z_2{\sim}\mathcal{N}(0, (1 - \rho_1^2)\sigma^2)$ $D_1{=}0$ $D_2{=}2(1 - \rho_1^2 - (1 - \rho_1^2)\sqrt{2\epsilon \ln 2})\sigma^2$	$\hat{X}_1^G{=}\sqrt{2\epsilon \ln 2}X_1 + Z_1$ $\hat{X}_2^G{=}\rho_1 \hat{X}_1^G + \sqrt{1 - \rho_1^2}X_2$ $Z_1{\sim}\mathcal{N}(0, (1 - 2\epsilon \ln 2)\sigma^2)$ $D_1{=}2\sigma^2$ $D_2{=}2(1 - \sqrt{1 - \rho_1^2}$ $-\rho_1^2\sqrt{2\epsilon \ln 2})\sigma^2$

[a] As justified in (253)–(259), the coefficient ω_1 (the coefficient of $\hat{X}_1^G$ in $\hat{X}_2^G$) has some correction terms of $O(\epsilon)$ which are ignored in the presentation of $\hat{X}_2^G$ since they do not contribute to dominant terms of distortion.

G Comparison of PLFs in Low-Rate Regime

Theorem 6 *For sufficiently small ϵ, let $R_j = \epsilon$ and suppose that $\rho_j = \rho$ and $\sigma_j = \sigma$, for $j = 1, \ldots, T$. The achievable distortions $D_{FMD,j}$ (for 0-PLF-FMD), and $D_{JD,j}$ (for 0-PLF-JD) are:*

$$D_{FMD,j} = 2(1 - \Delta_{FMD,j}\sqrt{2\epsilon \ln 2})\sigma^2, \quad D_{JD,j} = 2(1 - \Delta_{JD,j}\sqrt{2\epsilon \ln 2})\sigma^2, \tag{285}$$

where $\Delta_{FMD,j} := \sqrt{1 + \rho^2 \frac{(2\rho^2)^{j-1}-1}{2\rho^2-1}}$ and $\Delta_{JD,j} := \rho^{2(j-1)} + \mathbb{1}\{j \geq 2\} \cdot \sqrt{1-\rho^2}(\sum_{i=0}^{j-2} \rho^{2i})$.

Proof: We extend the proof in the previous section for the low-rate regime to T frames.

Distortion Analysis for 0-PLF-FMD:

We follow similar steps to (233)–(248) for optimization problems of the third and fourth frames and then use induction to derive expressions for T frames. For simplicity, we assume that $\rho_j = \rho$ for all j. Notice that in the following proof, $(\hat{X}_1^G, \hat{X}_2^G)$ are as in (205)–(206) where ν, ω_1 and ω_2 are already derived in (233)–(248).

Now, consider the reconstruction of the third frame as follows

$$\hat{X}_3^G = \tau_1 \hat{X}_1^G + \tau_2 \hat{X}_2^G + \tau_3 X_3 + Z_3, \tag{286}$$

for some τ_1, τ_2, τ_3, where $\hat{X}_3^G \sim \mathcal{N}(0, \sigma^2)$ and Z_3 is a Gaussian random variable independent of $(\hat{X}_1^G, \hat{X}_2^G, X_3)$. The rate constraint of the third step is given by

$$R_3 \geq I(X_3; \hat{X}_3^G | \hat{X}_1^G, \hat{X}_2^G). \tag{287}$$

Evaluating the above constraint with the choice of random variables $(\hat{X}_1^G, \hat{X}_2^G, \hat{X}_3^G)$ and re-arranging the terms, we get

$$\begin{aligned}&\tau_3^2\sigma^2(1-2^{-2R_3}(\rho^4 2^{-2R_1-2R_2}+\rho^2(1-\rho^2)2^{-2R_2}-\rho^2)) \leq \\ &\quad(1-2^{-2R_3})(1-\tau_1^2-\tau_2^2-2\tau_1\tau_2\omega_1\nu-2\tau_1\tau_2\omega_2\nu\rho-2\tau_2\tau_3\omega_1\nu\rho^2-2\tau_2\tau_3\omega_2\rho-2\tau_1\tau_3\nu\rho^2)\sigma^2.\end{aligned} \tag{288}$$

Similar to (240), considering the dominant terms of the above rate constraint and the fact that the solution of the optimization problem is attained when the above inequality is satisfied with "equality", we get

$$(1-\tau_1^2-\tau_2^2+O(\epsilon^{3/2}))(2\epsilon\ln 2+O(\epsilon^2)) = \tau_3^2(1+O(\epsilon)). \tag{289}$$

The distortion can be written as follows

$$\mathbb{E}[\|X_3-\hat{X}_3^G\|^2] = 2\sigma^2-2\tau_3\sigma^2-2\tau_2\omega_2\rho\sigma^2-2\tau_2\omega_1\nu\rho^2\sigma^2-2\tau_1\nu\rho^2\sigma^2. \tag{290}$$

So, the goal is to solve the following optimization problem for the third step

$$\min_{\tau_1,\tau_2,\tau_3} \quad 2\sigma^2-2\tau_3\sigma^2-2\tau_2\omega_2\rho\sigma^2-2\tau_2\omega_1\nu\rho^2\sigma^2-2\tau_1\nu\rho^2\sigma^2 \tag{291}$$

$$\text{s.t. :} \qquad (1-\tau_1^2-\tau_2^2+O(\epsilon^{3/2}))(2\epsilon\ln 2+O(\epsilon^2)) = \tau_3^2(1+O(\epsilon)). \tag{292}$$

We restrict the search space to $\tau_1, \tau_2, \tau_3 \geq 0$ and get an upper bound to the above optimization program as follows

$$\min_{\tau_1,\tau_2,\tau_3\geq 0} \quad 2\sigma^2-2\tau_3\sigma^2-2\tau_2\omega_2\rho\sigma^2-2\tau_2\omega_1\nu\rho^2\sigma^2-2\tau_1\nu\rho^2\sigma^2 \tag{293}$$

$$\text{s.t. :} \qquad (1-\tau_1^2-\tau_2^2+O(\epsilon^{3/2}))(2\epsilon\ln 2+O(\epsilon^2)) = \tau_3^2(1+O(\epsilon)). \tag{294}$$

The above optimization problem is equivalent to the following

$$\begin{aligned}\min_{\tau_1,\tau_2\geq 0} \Bigg(2\sigma^2 - 2\sqrt{\frac{(2\epsilon\ln 2+O(\epsilon^2))(1-\tau_1^2-\tau_2^2+O(\epsilon^{3/2}))}{1+O(\epsilon)}}\sigma^2 \\ -2\tau_2\omega_2\rho\sigma^2-2\tau_2\omega_1\nu\rho^2\sigma^2-2\tau_1\nu\rho^2\sigma^2\Bigg).\end{aligned} \tag{295}$$

We proceed with solving the above optimization program. Taking the derivative of the objective function with respect to η_1 and η_2 yields the following:

$$\frac{\eta_2}{\sqrt{1-\eta_1^2-\eta_2^2}} = \rho\sqrt{1+\rho^2}+O(\epsilon), \tag{296}$$

$$\frac{\eta_1}{\sqrt{1-\eta_1^2-\eta_2^2}} = \rho^2+O(\epsilon). \tag{297}$$

Solving the above set of equations, we get

$$\eta_1 = \frac{\rho^2}{\sqrt{1+\rho^2+2\rho^4}}+O(\epsilon), \tag{298}$$

$$\eta_2 = \frac{\rho\sqrt{1+\rho^2}}{\sqrt{1+\rho^2+2\rho^4}}+O(\epsilon). \tag{299}$$

Thus, considering the dominant terms, we get the following reconstruction for the third frame

$$\hat{X}_3^G = \frac{\rho^2}{\sqrt{1+\rho^2+2\rho^4}}\hat{X}_1^G + \frac{\rho\sqrt{1+\rho^2}}{\sqrt{1+\rho^2+2\rho^4}}\hat{X}_2^G + \frac{\sqrt{2\epsilon\ln 2}}{\sqrt{1+\rho^2+2\rho^4}}X_3 + Z_3. \tag{300}$$

The above reconstruction yields the following distortion for the third frame

$$\mathbb{E}[\|X_3-\hat{X}_3^G\|^2] = 2(1-\sqrt{2\epsilon\ln 2(1+\rho^2+2\rho^4)})\sigma^2. \tag{301}$$

Finally, consider the reconstruction of the fourth frame as follows

$$\hat{X}_4^G = \lambda_1\hat{X}_1^G+\lambda_2\hat{X}_2^G+\lambda_3\hat{X}_3^G+\lambda_4X_4+Z_4, \tag{302}$$

where $\hat{X}_4^G \sim \mathcal{N}(0, \sigma^2)$. The rate constraint of the fourth step implies that

$$(1-\lambda_1^2-\lambda_2^2-\lambda_3^2+O(\epsilon))(2\epsilon \ln 2+O(\epsilon)) = \lambda_4^2(1+O(\epsilon)). \tag{303}$$

The distortion can be written as follows

$$\begin{aligned}
\mathbb{E}[\|X_4-\hat{X}_4^G\|^2] &= 2\sigma^2 - 2\lambda_4\sigma^2 - 2\lambda_3\rho\tau_3\sigma^2 - 2\lambda_3\rho^2\tau_2\omega_2\sigma^2 - 2\lambda_3\rho^3\tau_2\omega_1\nu\sigma^2 \\
&\quad -2\lambda_3\rho^3\tau_1\nu\sigma^2 - 2\lambda_2\rho^3\omega_1\nu\sigma^2 - 2\lambda_2\rho^2\omega_2\sigma^2 - 2\lambda_1\rho^3\nu \qquad (304) \\
&= 2\sigma^2 - 2\sqrt{(2\epsilon\ln 2)(1-\lambda_1^2-\lambda_2^2-\lambda_3^2)}\sigma^2 - 2\lambda_3\rho\tau_3\sigma^2 \\
&\quad -2\lambda_3\rho^2\tau_2\omega_2\sigma^2 - 2\lambda_3\rho^3\tau_2\omega_1\nu\sigma^2 - 2\lambda_3\rho^3\tau_1\nu\sigma^2 \\
&\quad -2\lambda_2\rho^3\omega_1\nu\sigma^2 - 2\lambda_2\rho^2\omega_2\sigma^2 - 2\lambda_1\rho^3\nu + O(\epsilon). \qquad (305)
\end{aligned}$$

We take the derivative of the above expression with respect to λ_1, λ_2 and λ_3 and we get

$$\frac{\lambda_1}{\sqrt{1-\lambda_1^2-\lambda_2^2-\lambda_3^2}} = \rho^3 + O(\epsilon), \tag{306}$$

$$\frac{\lambda_2}{\sqrt{1-\lambda_1^2-\lambda_2^2-\lambda_3^2}} = \rho^2\sqrt{1+\rho^2} + O(\epsilon), \tag{307}$$

$$\frac{\lambda_3}{\sqrt{1-\lambda_1^2-\lambda_2^2-\lambda_3^2}} = \rho\sqrt{1+\rho^2+2\rho^4} + O(\epsilon). \tag{308}$$

Solving the above set of equations yields the following

$$\lambda_1 = \frac{\rho^3}{\sqrt{1+\rho^2+2\rho^4+4\rho^6}} + O(\epsilon), \tag{309}$$

$$\lambda_2 = \frac{\rho^2\sqrt{1+\rho^2}}{\sqrt{1+\rho^2+2\rho^4+4\rho^6}} + O(\epsilon), \tag{310}$$

$$\lambda_3 = \frac{\rho\sqrt{1+\rho^2+2\rho^4}}{\sqrt{1+\rho^2+2\rho^4+4\rho^6}} + O(\epsilon). \tag{311}$$

Thus, considering the dominant terms, we can write

$$\begin{aligned}
\hat{X}_4^G &= \frac{\rho^3}{\sqrt{1+\rho^2+2\rho^4+4\rho^6}}\hat{X}_1^G + \frac{\rho^2\sqrt{1+\rho^2}}{\sqrt{1+\rho^2+2\rho^4+4\rho^6}}\hat{X}_2^G \\
&\quad + \frac{\rho\sqrt{1+\rho^2+2\rho^4}}{\sqrt{1+\rho^2+2\rho^4+4\rho^6}}\hat{X}_3^G + \frac{\sqrt{2\epsilon\ln 2}}{\sqrt{1+\rho^2+2\rho^4+4\rho^6}}X_4 + Z_4. \qquad (312)
\end{aligned}$$

The distortion term then becomes:

$$\mathbb{E}[\|X_4-\hat{X}_4^G\|^2] = 2(1-\sqrt{2\epsilon\ln 2(1+\rho^2+2\rho^4+4\rho^6)})\sigma^2. \tag{313}$$

Now, we use induction to derive the terms for T frames. Define

$$\begin{aligned}
\Delta_{\text{FMD},j} &:= \sqrt{1+\sum_{i=1}^{j-1} 2^{j-1-i}\rho^{2(j-i)}}, \qquad j=2,\dots,T \qquad (314) \\
&= \sqrt{1+\rho^2\frac{(2\rho^2)^{j-1}-1}{2\rho^2-1}}. \qquad (315)
\end{aligned}$$

Thus, we have

$$\hat{X}_j^G = \sum_{i=1}^{j-1}\frac{\Delta_{\text{FMD},i}\rho^{j-i}}{\Delta_{\text{FMD},j}}\hat{X}_i^G + \frac{\sqrt{2\epsilon\ln 2}}{\Delta_{\text{FMD},j}}X_j + Z_j, \qquad j=2,\dots,T, \tag{316}$$

where Z_j is a Gaussian random variable independent of $(\hat{X}_1^G, \dots, \hat{X}_{j-1}^G, X_j)$ and its variance is such that $\mathbb{E}[(\hat{X}_j^G)^2] = \sigma^2$. The distortion is given by the following expression

$$D_{\text{FMD},j} = \mathbb{E}[\|X_j-\hat{X}_j\|^2] = 2(1-\Delta_{\text{FMD},j}\sqrt{2\epsilon\ln 2})\sigma^2, \qquad j=2,\dots,T. \tag{317}$$

For the special case where $\rho = 1$, then the distortion simplifies to the following

$$\mathbb{E}[\|X_j - \hat{X}_j\|^2] = 2(1 - 2^{\frac{j-1}{2}}\sqrt{2\epsilon \ln 2})\sigma^2, \qquad j = 2, \ldots, T, \tag{318}$$

which shows an exponential decrease at each step.

Distortion Analysis for 0*-PLF-JD*:

In this case, the proof for T frames is similar to (254)–(261). Thus, we have

$$\hat{X}^G_j = \rho \hat{X}^G_{j-1} + \sqrt{(1-\rho^2)2\epsilon \ln 2}X_j + Z_j, j = 2, \ldots, T, \tag{319}$$

where Z_j is a Gaussian random variable independent of $(\hat{X}^G_{j-1}, X_j)$ and its variance is such that $\mathbb{E}[(\hat{X}^G_T)^2] = \sigma^2$. It should be mentioned that preserving the correlation coefficients, e.g., $\mathbb{E}[\hat{X}^G_j \hat{X}^G_{j-1}] = \rho$, needs some correction terms of $O(\epsilon)$ as discussed in (257). However, as shown in (261), these correction terms do not contribute to dominant terms of distortion and hence, they can be ignored in the presentation of (319). Now, define

$$\Delta_{\mathrm{JD},j} := \rho^{2(j-1)} + \sqrt{1-\rho^2}(\sum_{i=0}^{j-2} \rho^{2i}), \qquad j = 2, \ldots, T, \tag{320}$$

and notice that

$$\begin{aligned} D_{\mathrm{JD},j} &:= \mathbb{E}[\|X_j - \hat{X}_j\|^2] && (321)\\ &= 2\sigma^2 - 2\mathbb{E}[X_j \hat{X}_j] && (322)\\ &= 2\sigma^2 - 2\mathbb{E}[X_j(\rho \hat{X}^G_{j-1} + \sqrt{(1-\rho^2)2\epsilon \ln 2}X_j)] && (323)\\ &= 2\sigma^2 - 2\mathbb{E}[X_j(\rho^{j-1}X_1 + \sqrt{1-\rho^2}(\rho^{j-2}X_2 + \ldots + X_j))]\sqrt{2\epsilon \ln 2}\sigma^2 && (324)\\ &= 2(1 - \Delta_{\mathrm{JD},j}\sqrt{2\epsilon \ln 2})\sigma^2. && (325) \end{aligned}$$

For the special case of $\rho = 1$, we get $\Delta_{\mathrm{JD},j} = 1$ which remains a constant across different steps. ■

H Universality Statement for Gauss-Markov Source Model

H.1 MMSE Representations for a Given Rate

For a given rate tuple R, the minimum distortions achievable by MMSE representations are derived in [28, 37] and are given by

$$D_1^{\min} = \sigma_1^2 2^{-2R_1}, \tag{326}$$

$$D_2^{\min} = (\rho_1^2 \frac{\sigma_2^2}{\sigma_1^2} D_1^{\min} + \sigma_{N_1}^2)2^{-2R_2}, \tag{327}$$

$$D_3^{\min} = (\rho_2^2 \frac{\sigma_3^2}{\sigma_2^2} D_2^{\min} + \sigma_{N_2}^2)2^{-2R_3}, \tag{328}$$

where

$$\sigma_{N_1}^2 := (1-\rho_1^2)\sigma_2^2, \tag{329}$$

$$\sigma_{N_2}^2 := (1-\rho_2^2)\sigma_3^2. \tag{330}$$

The above distortions are achieved by the following optimal reconstructions $\hat{\mathsf{X}}_r$ given in [28]. Notice that the MMSE representation is $\mathsf{X}_r^{\mathsf{RD}} = \hat{\mathsf{X}}_r$, i.e., the functions $\eta_1(.)$ and $\eta_2(.,.)$ of iRDP region $\mathcal{C}_{\mathsf{RDP}}$ (Definition 4) are identity functions (this statement follows from Theorem 5). Now, we choose the reconstruction $\hat{\mathsf{X}}_r$ in the following.

The reconstruction $\hat{X}_{r,1}$ is chosen such that $\hat{X}_{r,1} \to X_1 \to (X_2, X_3)$ holds a Markov chain and

$$X_1 = \hat{X}_{r,1} + Z_1, \tag{331}$$

where $\hat{X}_{r,1} \sim \mathcal{N}(0, \sigma_1^2 - D_1^{\min})$ and $Z_1 \sim \mathcal{N}(0, D_1^{\min})$ are independent random variables. Then, the reconstruction $\hat{X}_{r,2}$ is chosen as follows. Let

$$W_2 := \rho_1 \frac{\sigma_2}{\sigma_1} Z_1 + N_1, \tag{332}$$

which is the innovation from $\hat{X}_{r,1}$ to X_2. Now, we find the random variables $\hat{W}_2$ and Z_2 such that

$$W_2 = \hat{W}_2 + Z_2, \tag{333}$$

where $\hat{W}_2 \sim \mathcal{N}(0, \rho_1^2 \frac{\sigma_2^2}{\sigma_1^2} D_1^{\min} + \sigma_{N_1}^2 - D_2^{\min})$ and $Z_2 \sim \mathcal{N}(0, D_2^{\min})$ are independent from each other, and the Markov chain $\hat{W}_2 \to (X_2, \hat{X}_{r,1}) \to (X_1, X_3)$ holds. Now, define

$$\hat{X}_{r,2} := \rho_1 \frac{\sigma_2}{\sigma_1} \hat{X}_{r,1} + \hat{W}_2. \tag{334}$$

Finally, we choose the reconstruction $\hat{X}_{r,3}$ as follows. Let

$$W_3 := \rho_2 \frac{\sigma_3}{\sigma_2} Z_2 + N_2, \tag{335}$$

which is the innovation from $\hat{X}_{r,2}$ to X_3. Now, we find random variables $\hat{W}_3$ and Z_3 such that

$$W_3 = \hat{W}_3 + Z_3, \tag{336}$$

where $\hat{W}_3 \sim \mathcal{N}(0, \rho_2^2 \frac{\sigma_3^2}{\sigma_2^2} D_2^{\min} + \sigma_{N_2}^2 - D_3^{\min})$ and $Z_2 \sim \mathcal{N}(0, D_3^{\min})$ are independent from each other, and the Markov chain $\hat{W}_3 \to (X_3, \hat{X}_{r,1}, \hat{X}_{r,2}) \to (X_1, X_2)$ holds. Now, define

$$\hat{X}_{r,3} := \rho_1 \frac{\sigma_3}{\sigma_2} \hat{X}_{r,2} + \hat{W}_3. \tag{337}$$

Thus, the optimal reconstruction $\hat{\mathsf{X}}_r$ is chosen and it satisfies the rate constraint R.

H.2 Universality Statement

Theorem 7 *For a given rate tuple R with strictly positive components, let the MMSE representation be denoted as $\mathsf{X}_r^{RD} = (X_{r,1}^{RD}, X_{r,2}^{RD}, X_{r,3}^{RD})$. Let $(\mathsf{D}, \mathsf{P}) \in \mathcal{DP}(\mathsf{R})$ and let $\hat{\mathsf{X}} = (\hat{X}_1, \hat{X}_2, \hat{X}_3)$ be the corresponding reconstruction achieving it. Then there exist κ_1, θ_1, θ_2, ψ_1, ψ_2 and ψ_3 and noise variables (Z_1, Z_2, Z_3) independent of $(X_{r,1}^{RD}, X_{r,2}^{RD}, X_{r,3}^{RD})$, which satisfy the following*

$$\hat{X}_1 = \kappa_1 X_{r,1}^{RD} + Z_1, \quad \hat{X}_2 = \theta_1 X_{r,1}^{RD} + \theta_2 X_{r,2}^{RD} + Z_2, \quad \hat{X}_3 = \psi_1 X_{r,1}^{RD} + \psi_2 X_{r,2}^{RD} + \psi_3 \hat{X}_{r,3}^{RD} + Z_3.$$

For a given positive rate tuple R, let the MMSE representation X_r^{RD} be in the set $\mathcal{P}^{RD}(\mathsf{R})$. Also, let $(\mathsf{D}, \mathsf{P}) \in \mathcal{DP}(\mathsf{R})$ and X_r, $\hat{\mathsf{X}}$ be the corresponding representation and reconstruction achieving it.

Proof: First, notice that according to the proof of Theorem 5 for the Gauss-Markov source model, one can set $\hat{\mathsf{X}} = \mathsf{X}_r$ in iRDP region of $\mathcal{C}_{\mathsf{RDP}}$, without loss of optimality. So, in the following proof, the reconstruction X_r and representation $\hat{\mathsf{X}}$ are used interchangeably, in some places.

We show the following statement. If

$$R_1 \geq I(X_1; X_{r,1}), \tag{338}$$
$$R_2 \geq I(X_2; X_{r,2} | X_{r,1}), \tag{339}$$
$$R_3 \geq I(X_3; X_{r,3} | X_{r,1}, X_{r,2}), \tag{340}$$

then, there exist κ_1, θ_1, θ_2, ψ_1, ψ_2 and ψ_3 and noise variables Z_1, Z_2, Z_3 independent of $X_{r,1}^{\text{RD}}$, $(X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}})$, $(X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}}, X_{r,3}^{\text{RD}})$, respectively, which satisfy the following

$$\hat{X}_1 = \kappa_1 X_{r,1}^{\text{RD}} + Z_1, \tag{341}$$
$$\hat{X}_2 = \theta_1 X_{r,1}^{\text{RD}} + \theta_2 X_{r,2}^{\text{RD}} + Z_2, \tag{342}$$
$$\hat{X}_3 = \psi_1 X_{r,1}^{\text{RD}} + \psi_2 X_{r,2}^{\text{RD}} + \psi_3 \hat{X}_{r,3}^{\text{RD}} + Z_3. \tag{343}$$

If (338)–(340) are satisfied with equality, then the noise random variables in (341)–(343) do not exist and a linear combination is sufficient for converting $(X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}}, X_{r,3}^{\text{RD}})$ to $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$.

First, we prove the statement when all of inequalities in (338)–(340) hold with "equality". We provide the proof for $T = 2$ frames. The extension to arbitrary number of frames is straightforward. To that end, we first prove the following two lemmas.

Lemma 2 *Without loss of optimality, the reconstruction of the first step $\hat{X}_1$ satisfies the following*

$$\gamma_1 \hat{X}_1 = W_1, \tag{344}$$

where

$$\gamma_1 := \frac{\mathbb{E}[X_1 \hat{X}_1]}{\sigma^2_{\hat{X}_1}}, \tag{345}$$

and W_1 is a Gaussian random variable that its statistics do not depend on the pair (D_1, P_1).

Proof: According to Theorem 5, we know that $(X_1, \hat{X}_1)$ are jointly Gaussian. So, we can write X_1 as follows

$$X_1 = \gamma_1 \hat{X}_1 + T_1, \tag{346}$$

where T_1 is a Gaussian random variable independent of $\hat{X}_1$ with a constant variance $\sigma_1^2 2^{-2R_1}$. Notice that (346) can be written as follows

$$\hat{X}_1 = \alpha_1 (X_1 + Q), \tag{347}$$

where Q is a Gaussian random variable independent of X_1 with a zero-mean and variance $\frac{\sigma_1^2 2^{-2R_1}}{1-2^{-2R_1}}$ and

$$\alpha_1 := \frac{1}{\gamma_1}(1 - 2^{-2R_1}). \tag{348}$$

From (347), we get

$$\gamma_1 \hat{X}_1 = (1 - 2^{-2R_1})(X_1 + Q). \tag{349}$$

Now, defining $W_1 := (1 - 2^{-2R_1})(X_1 + Q)$ yields the desired result. ■

Lemma 3 *Without loss of optimality, the reconstructions of the first and second steps $(\hat{X}_1, \hat{X}_2)$ satisfy the following*

$$\lambda_1 \hat{X}_1 + \lambda_2 \hat{X}_2 = W_2, \tag{350}$$

where

$$\lambda_1 := \frac{\rho_1 \mathbb{E}[X_1 \hat{X}_1] \hat{\sigma}^2_{X_2} - \mathbb{E}[\hat{X}_1 \hat{X}_2] \mathbb{E}[X_2 \hat{X}_2]}{\hat{\sigma}^2_{X_1} \hat{\sigma}^2_{X_2} - \mathbb{E}^2[\hat{X}_1 \hat{X}_2]}, \tag{351}$$

$$\lambda_2 := \frac{\rho_1 \mathbb{E}[X_1 \hat{X}_1] \mathbb{E}[\hat{X}_1 \hat{X}_2] - \hat{\sigma}^2_{X_1} \mathbb{E}[X_2 \hat{X}_2]}{\hat{\sigma}^2_{X_1} \hat{\sigma}^2_{X_2} - \mathbb{E}^2[\hat{X}_1 \hat{X}_2]}, \tag{352}$$

and W_2 is a Gaussian random variable that its statistics do not depend on the pairs (D_1, P_1) and (D_2, P_2).

Proof: According to Theorem 5, we know that $(X_1, X_2, \hat{X}_1, \hat{X}_2)$ are jointly Gaussian. So, we can write X_2 as follows

$$X_2 = \lambda_1 \hat{X}_1 + \lambda_2 \hat{X}_2 + T_2, \tag{353}$$

where T_2 is a Gaussian random variable independent of $(\hat{X}_1, \hat{X}_2)$ with a constant variance of $\sigma^2_{X_2|\hat{X}_1} 2^{-2R_2}$ where

$$\sigma^2_{X_2|\hat{X}_1} := \frac{1}{2} \log \left(\rho_1^2 \sigma_1^2 2^{-2R_1} + 2^{2H(N_1)} \right). \tag{354}$$

Notice that (353) can be written as follows

$$\lambda_1 \hat{X}_1 + \lambda_2 \hat{X}_2 = (1 - 2^{-2R_2})(X_2 + Q'), \tag{355}$$

where Q' is a Gaussian random variable independent of X_2 with a zero-mean and variance $\frac{\sigma^2_{X_2|\hat{X}_1} 2^{-2R_2}}{1-2^{-2R_2}}$. Defining $W_2 := (1 - 2^{-2R_2})(X_2 + Q')$ yields the desired result. ■

Now, we proceed with the proof of the theorem. According to Lemma 2, there exist real γ_1 and γ_1' such that

$$\gamma_1 \hat{X}_1 = \gamma_1' X_{r,1}^{\text{RD}}. \tag{356}$$

Define

$$\kappa_1 := \frac{\gamma_1'}{\gamma_1}. \tag{357}$$

Then, according to Lemma 3, there exist λ_1, λ_2, λ_1' and λ_2' such that

$$\lambda_1 \hat{X}_1 + \lambda_2 \hat{X}_2 = \lambda_1' X_{r,1}^{\text{RD}} + \lambda_2' X_{r,2}^{\text{RD}}. \tag{358}$$

The above equation can be written as

$$\hat{X}_2 = \frac{\lambda_1' - \lambda_1 \kappa_1}{\lambda_2} X_{r,1}^{\text{RD}} + \frac{\lambda_2'}{\lambda_2} X_{r,2}^{\text{RD}} \tag{359}$$

$$:= \theta_1 X_{r,1}^{\text{RD}} + \theta_2 X_{r,2}^{\text{RD}}. \tag{360}$$

A similar justification holds for the third frame.

Next, we prove the statement when at least one of the rate constraints in (338)–(340) hold with strict inequality. In the following, we construct new reconstructions $(\hat{X}_1', \hat{X}_2')$ based on $(\hat{X}_1, \hat{X}_2)$ such that they satisfy the rate constraints (R_1, R_2) with equality. Then, we will be able to apply the two lemmas we proved to show that $(\hat{X}_1, \hat{X}_2)$ are linearly related to MMSE reconstructions $(X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}})$.

Construction of $\hat{X}_1'$:

Now, let

$$\hat{R}_1 := I(X_1; \hat{X}_1), \tag{361}$$

where $\hat{R}_1 \leq R_1$. Also, recall that

$$R_1 = I(X_1; X_{r,1}^{\text{RD}}). \tag{362}$$

Now, let $\hat{X}_1'$ such that $\hat{X}_1' \to X_{r,1}^{\text{RD}} \to X_1$ holds and

$$\hat{X}_1' = X_{r,1}^{\text{RD}} + W_1, \tag{363}$$

where $W_1 \sim \mathcal{N}(0, \nu_1^2)$ independent of $\hat{X}_1$ and ν_1^2 will be determined in the following. Notice that $I(X_1; \hat{X}_1')$ is a monotonically decreasing function of ν_1^2. So, one choose ν_1^2 such that

$$I(\hat{X}_1'; X_1) = I(X_1; \hat{X}_1) = \hat{R}_1. \tag{364}$$

Now, according to Lemma 2, since $\hat{X}_1'$ and $\hat{X}_1$ have the same rates, there exists a coefficient κ_1' such that

$$\hat{X}_1 = \kappa_1' \hat{X}_1' \tag{365}$$

$$= \kappa_1' X_{r,1}^{\text{RD}} + \kappa_1' W_1. \tag{366}$$

Now, define $Z_1 := \kappa_1' W_1$ and notice that

$$\hat{X}_1 = \kappa_1' X_{r,1}^{\text{RD}} + Z_1. \tag{367}$$

Construction of $\hat{X}_2'$:

Next, consider the second step. Define

$$\hat{R}_2 := I(X_2; \hat{X}_2 | \hat{X}_1), \tag{368}$$

where $\hat{R}_2 \leq R_2$. Also, recall that

$$R_2 = I(X_2; X_{r,2}^{\text{RD}} | X_{r,1}^{\text{RD}}). \tag{369}$$

Define $\tilde{X}_2 := \mathbb{E}[X_2 | X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}}]$ to be the MMSE reconstruction and consider that

$$R_2 = I(X_2; X_{r,2}^{\text{RD}} | X_{r,1}^{\text{RD}}) \tag{370}$$

$$= I(X_2; \tilde{X}_2 | X_{r,1}^{\text{RD}}), \tag{371}$$

where the last equality follows because both Markov chains $X_2 \to (X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}}) \to \tilde{X}_2$ and $X_2 \to \tilde{X}_2 \to (X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}})$ hold where the latter one is satisfied for Gaussian random variables for which we can write $X_2 = \mathbb{E}[X_2|X_{r,1}, X_{r,2}] + W'$ such that W' is independent of $(X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}})$.

Now, we show that $I(X_2; \tilde{X}_2|X_{r,1}^{\text{RD}}) \leq I(X_2; \tilde{X}_2|\hat{X}_1')$. This is justified in the following

$$\begin{aligned}
I(X_2; \tilde{X}_2|\hat{X}_1') &= I(X_2; \tilde{X}_2|X_{r,1}^{\text{RD}} + W_1) && (372)\\
&= H(X_2|X_{r,1}^{\text{RD}} + W_1) - H(X_2|\tilde{X}_2, X_{r,1}^{\text{RD}} + W_1) && (373)\\
&\geq H(X_2|X_{r,1}^{\text{RD}} + W_1, W_1) - H(X_2|\tilde{X}_2, X_{r,1}^{\text{RD}} + W_1) && (374)\\
&= H(X_2|X_{r,1}^{\text{RD}}, W_1) - H(X_2|\tilde{X}_2, X_{r,1}^{\text{RD}} + W_1) && (375)\\
&\geq H(X_2|X_{r,1}^{\text{RD}}, W_1) - H(X_2|\tilde{X}_2) && (376)\\
&= H(X_2|X_{r,1}^{\text{RD}}) - H(X_2|\tilde{X}_2) && (377)\\
&= H(X_2|X_{r,1}^{\text{RD}}) - H(X_2|\tilde{X}_2, X_{r,1}^{\text{RD}}) && (378)\\
&= I(X_2; \tilde{X}_2|X_{r,1}^{\text{RD}}), && (379)
\end{aligned}$$

where (377) follows because W_1 is independent of $(X_2, X_{r,1}^{\text{RD}})$ and (378) follows from the Markov chain $X_2 \to \tilde{X}_2 \to X_{r,1}^{\text{RD}}$.

Define

$$R_2' := I(X_2; \tilde{X}_2|\hat{X}_1'), \tag{380}$$

and consider the fact that $R_2' \geq R_2$. Now, we introduce $\hat{X}_2'$ such that $\hat{X}_2' \to (\tilde{X}_2, \hat{X}_1') \to X_2$ forms a Markov chain and

$$\hat{X}_2' = \tilde{X}_2 + \hat{X}_1' + W_2, \tag{381}$$

where $W_2 \sim \mathcal{N}(0, \nu_2^2)$ independent of $(\tilde{X}_2, \hat{X}_1)$ and ν_2^2 will be determined in the following. Since $I(X_2; \hat{X}_2'|\hat{X}_1')$ is a monotonically decreasing function of ν_2^2, we can choose ν_2^2 such that

$$I(X_2; \hat{X}_2'|\hat{X}_1') = I(X_2; \hat{X}_2|\hat{X}_1) = \hat{R}_2. \tag{382}$$

Then, according to Lemma 3, there exist λ_1', λ_2', $\hat{\lambda}_1$ and $\hat{\lambda}_2$ such that

$$\lambda_1'\hat{X}_1' + \lambda_2'\hat{X}_2' = \hat{\lambda}_1\hat{X}_1 + \hat{\lambda}_2\hat{X}_2. \tag{383}$$

Plugging (363), (367) and (381) into the above expression and letting $\tilde{X}_2 = \alpha X_{r,1}^{\text{RD}} + \beta X_{r,2}^{\text{RD}}$ for some α, β, we get

$$(\lambda_1' + (1+\alpha)\lambda_2' - \hat{\lambda}_1\kappa')X_{r,1}^{\text{RD}} + \lambda_2'\beta X_{r,2}^{\text{RD}} + (\lambda_1' + \lambda_2')W_1 + \lambda_2' W_2 - \hat{\lambda}_1 Z_1 = \hat{\lambda}_2\hat{X}_2. \tag{384}$$

Now define

$$\begin{aligned}
\theta_1 &:= \frac{\lambda_1' + (1+\alpha)\lambda_2' - \hat{\lambda}_1\kappa'}{\hat{\lambda}_2}, && (385)\\
\theta_2 &:= \frac{\lambda_2'\beta}{\hat{\lambda}_2}, && (386)\\
Z_2 &:= \frac{(\lambda_1' + \lambda_2')}{\hat{\lambda}_2}W_1 + \frac{\lambda_2'}{\hat{\lambda}_2}W_2 - \frac{\hat{\lambda}_1}{\hat{\lambda}_2}Z_1. && (387)
\end{aligned}$$

Thus, we have

$$\hat{X}_2 = \theta_1 X_{r,1}^{\text{RD}} + \theta_2 X_{r,2}^{\text{RD}} + Z_2. \tag{388}$$

Notice that the above proof only uses the information about reconstructions of the operating points in DP-tradeoff and it does not depend on the choice of PLF. So, it holds for both PLF-JD and PLF-FMD. This concludes the proof. ■

H.3 Gaussian Example

Assume that the sources are symmetric in the sense that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$, $\rho_1 = \rho_2 = \rho_3 := \rho$ for some $0 < \rho \leq 1$. Also, suppose that the perception thresholds are symmetric, i.e., $P_1 = P_2 = P_3 := P$ for some $0 < P \leq 1$. We choose the rate tuple R such that the minimum distortions $D_j^{\min} = D$ for $j \in \{1, 2, 3\}$. According to Appendix H.1, such rates are given by

$$R_1 = \frac{1}{2} \log \frac{1}{D}, \tag{389}$$

$$R_2 = \frac{1}{2} \log \frac{\rho^2 D + (1-\rho)}{D}, \tag{390}$$

$$R_3 = \frac{1}{2} \log \frac{\rho^2 D + (1-\rho^2)}{D}. \tag{391}$$

The covariance matrix of the MMSE representations $\text{cov}(X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}}, X_{r,3}^{\text{RD}})$ is given by $(1-D)\Sigma$ where

$$\Sigma := \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}. \tag{392}$$

If we introduce the 0-PLF while keeping the rates as those of MMSE reconstructions, it can be shown that the optimal distortions are all equal to $D_1 = D_2 = D_3 = 2 - 2\sqrt{1-D}$. Denote the reconstructions by $(\hat{X}_{D_1}^0, \hat{X}_{D_2}^0, \hat{X}_{D_3}^0)$ and notice that the covariance matrix of the reconstructions is equal to that of the sources and is given by Σ. Thus, the covariance matrix of $(X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}}, X_{r,3}^{\text{RD}})$ is $(1-D)$ times the covariance matrix of $(\hat{X}_{D_1}^0, \hat{X}_{D_2}^0, \hat{X}_{D_3}^0)$. So, the reconstructions $(X_{r,1}^{\text{RD}}, X_{r,2}^{\text{RD}}, X_{r,3}^{\text{RD}})$ and $(\hat{X}_{D_1}^0, \hat{X}_{D_2}^0, \hat{X}_{D_3}^0)$ can be transformed to each other by the scaling factor $\frac{1}{\sqrt{1-D}}$. This inspires the idea that reconstructions corresponding to different tuples (D, P) are linearly related to those of MMSE representations which is the essence of the following Theorem 6. Moreover, both PLFs either based on FMD or JD perform similarly in this example since individually scaling the reconstruction of each frame finally ends up in matching the covariance matrix of all frames.

I Justification of low-rate regime for Moving MNSIT

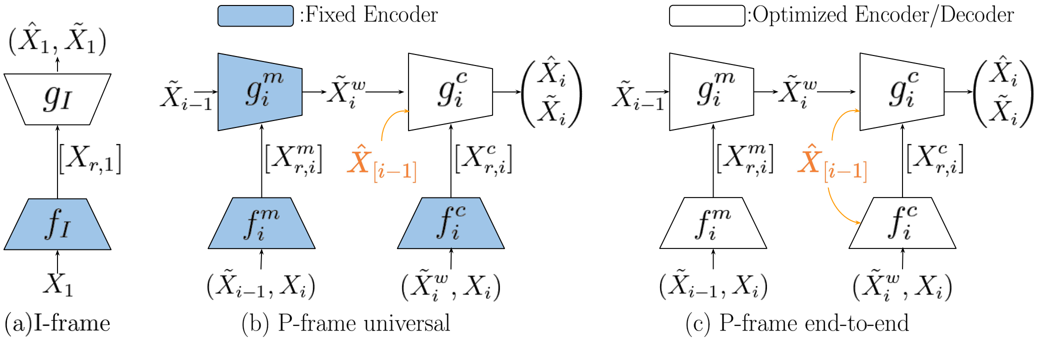
In the MovingMNIST dataset, the digit in I-frame is generated uniformly across the 32×32 center region in a 64×64 image, meaning that $\log(32\times32)$=10 bits are required to localize the digits and any lower rate would result in much less correlated reconstructions. As such, one can consider R_1=12 bits (2 extra bits for content and style) as a low rate. For P-frames, the movement is uniformly constrained within a 10×10 region so any rate $R_2 \leq \log_2(10\times10)$=6.6 bits (excluding residual compensation) can be considered a low rate.

J Experiment Details

J.1 Training Setup and Overview

Our compression architecture is built on the scale-space flow model [32], which allows end-to-end training without relying on pre-trained optical flow estimators. For better compression efficiency, we replace the residual compression module with the conditioning one [33]. In the following, we will interchangeably refer X_1 as the I-frame and subsequent ones as P-frames. The annotation for the encoder, decoder, and critic (discriminator) will be referred to as f, g, and h respectively and their specific functionality (e.g motion compression, joint perception critic) will be described within context through a subscript/superscript.

Distortion and Perception Measurement: We follow the setup in prior works [16, 21] for distortion and perception measurement. Specifically, we use MSE loss $\mathbb{E}[||X - \hat{X}||^2]$ as a distortion metric and Wasserstein-1 distance as a perception metric, which can be estimated through the WGAN critics (following the Kanotorovich-Rubinstein duality). For the marginal perception metric, we optimize



(a)I-frame (b) P-frame universal (c) P-frame end-to-end

Figure 7: Compression diagram for (a) I-frame (b) P-frame with universal representation and (c) P-frame with optimized representation. For simplicity, we do not show the shared randomness K.

our critics h_m to classify between original image X and synthetic ones $\hat{X}$. This will then allow us to measure $W_1(P_X, P_{\hat{X}})$ since:

$$W_1(P_X, P_{\hat{X}}) = \sup_{h_m \in \mathcal{F}} \mathbb{E}[h_m(X)] - \mathbb{E}[h_m(\hat{X})] \tag{393}$$

where $\mathcal{F}$ is a set of all bounded 1-Lipschitz functions. Similarly, the joint perception metric is realized through $W_1(P_{X_1 \dots X_j}, P_{\hat{X}_1 \dots \hat{X}_j})$ by training a critic h_j that classifies between synthetic and authentic sequences:

$$W_1(P_{X_1 \dots X_j}, P_{\hat{X}_1 \dots \hat{X}_j}) = \sup_{h_j \in \mathcal{F}} \mathbb{E}[h_j(X_1, ..., X_i)] - \mathbb{E}[h_j(\hat{X}_1, ..., \hat{X}_i)] \tag{394}$$

In practice, the set of 1-Lipschitz functions is limited by the neural network architecture. Also, although our analysis employs the Wasserstein-2 distance as a perception metric, it is worth noting that the ideal reconstructions (0-PLF) for this metric and the one used in our study should be identical.

I-frame Compressor: We compress I-frames in a similar fashion as previous works [16, 21]. Our encoder f_I and decoder g_I in Figure 7a contain a series of convolution operations and we control the rate R_1 by varying the dimension and quantization level in the bottleneck. The model utilizes common randomness through the dithered quantization operation. For a given rate R_1, we vary the amount of DP tradeoff by controlling the hyper-parameter $\lambda_i^{\text{marginal}}$ in the following minimization objective $\mathcal{L}_1$:

$$\mathcal{L}_1 = \mathbb{E}[||X_1 - \hat{X}_1||^2] + \lambda_i^{\text{marginal}} W_1(P_{X_1}, P_{\hat{X}_1}) \tag{395}$$

Following the results from Zhang et al. [16], we fix the encoder after optimizing the encoder-decoder pair for MSE representations. We then fix the encoder and train another decoder to obtain the optimal reconstruction with perfect perception, i.e, $W_1(P_X, P_{\hat{X}}) \approx 0$. We will leverage these universal representation results to compress P-frames (both end-to-end and universal).

P-frame Compressor: We describe the loss functions before explaining our architectures. Given previous reconstructions $\hat{X}_{[i-1]} := \{\hat{X}_1, \hat{X}_2, ..., \hat{X}_{i-1}\}$, one can adjust the distortion-joint perception tradeoff by controlling the hyper-parameter λ_i^{joint} in the following objective $\mathcal{L}_i$.

$$\mathcal{L}_i^{\text{joint}} = \mathbb{E}[||X_i - \hat{X}_i||^2] + \lambda_i^{\text{joint}} W_1(P_{X_{[i]}}, P_{\hat{X}_{[i]}}) \tag{396}$$

Note that in order to achieve 0-PLF-JD, previous reconstructions $\hat{X}_{[i-1]}$ must also achieve 0-PLF-JD, since it is impossible to reconstruct such $\hat{X}_i$ if the previous $\hat{X}_{[i-1]}$ are not temporally consistent[3]. For the FMD metric, we use the loss function in (395).

In the *universal model* in Figure 7b, the motion encoder f_i^m compresses and sends the quantized flow fields $[X_{r,i}^m]$ between the MMSE reconstruction $\tilde{X}_{i-1}$ and X_i. Given $[X_{r,i}^m]$, the flow decoder and warping module g_i^m will transform $\tilde{X}_{i-1}$ into $\tilde{X}_i^w$ (predicted frame). We use f_i^c to compress the

[3] This follows from the inequality: $W_2^2(P_{X_1,X_2}, P_{\hat{X}_1,\hat{X}_2}) \geq W_2^2(P_{X_1}, P_{\hat{X}_1}) + W_2^2(P_{X_2}, P_{\hat{X}_2})$

residual information $[X^c_{r,i}]$ between X_i and $\tilde{X}^w_i$ [4], which will be decoded by g^c_i. We note that for MMSE representation, g^c_i only requires $\tilde{X}^w_i$ as a conditional input while an additional conditioning input $\hat{X}_{[i-1]}$ is required when perceptual optimization is involved. Together, f^m_i, g^m_i, f^c_i, and g^c_i are optimized for MMSE reconstructions. To train for different DP tradeoffs, we fix f^m_i, g^m_i, f^c_i and adapt the new decoder $\hat{g}^c_i$ (conditioning on $\tilde{X}^w_i, \hat{X}_{[i-1]}$). We note that fixing g^m_i for universal representation is essential since $[X^c_{r,i}]$ is dependent on the outputs $\tilde{X}^w_i$ of g^m_i.

In the *end-to-end model* in Figure 7c, we use an MMSE representation to estimate the motion vector, as in the case of the universal model. The only difference is that the encoder f^c_i also uses previous $\hat{X}_i$ and the encoders will be jointly trained with the decoders.

J.2 Networks Architecture

In this section, we describe the network architecture for universal and end-to-end P-frame compressor models. [5]. In the architecture layout, we denote BN2D and SN for the *Batchnorm2D* and *Spectral Normalization* layers. Convolutional and transposed convolutional layer are denoted as "conv" and "upconv" respectively, which is accompanied by number of filters, kernel size, stride, and padding.

Motion Encoder and Decoder. The universal and optimized end-to-end model shares the same architecture for the motion encoder and decoder. (f^m_i and g^m_i respectively). We follow the original implementations [32] and present the convolutional architecture in Table 3. Different from the original implementation, however, we replace the last layer with dithered quantization layer (as in [16]) in our implementation. The output dimension of the motion encoder is denoted as d_m.

Table 3: Motion Encoder f^m_i and Decoder g^m_i.

(a) Encoder f^m_i

Input-64×64×(2×channels)
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (d_m:4:2:0), BN2D
Quantizer

(b) Decoder g^m_i

Input-d_m
upconv (64:4:1:0), BN2D, l-ReLU
upconv (64:5:2:0), BN2D, l-ReLU
upconv (64:5:2:0), BN2D, l-ReLU
upconv (64:5:2:0), BN2D, l-ReLU
upconv (3:5:2:0), BN2D

Residual Encoder and Decoder. The architecture of the conditional residual encoder is shown in Table 4a, where we stack multiple frames along their channel dimension as an input. As described previously, in the residual encoder, the universal model requires only $X_i, \tilde{X}^w_i$ while the end-to-end model will receive $X_i, \tilde{X}^w_i$ and $\hat{X}_{[i-1]}$. We denote the output dimension of this residual encoder as d_r. In the decoding part, the decoder will first condition all the previous reconstructions $\hat{X}_[i-1]$ by projecting them into an embedding vector of size 192 (conditioning module in Table 4b). Then we concatenate this vector with the output of f^r_i. The concatenated vector will be fed into the decoder (Table 4c) to produce the reconstruction $\hat{X}_i$.

FMD and JD Critics. For the video critics, our PLF-JD critic architecture is inspired by the work of Kwon and Park [40], where we concatenate frames sequentially along their channel dimensions. For both PLF-FMD and PLF-JD critics, we add spectral normalization layers for better convergence. Their architecture is shown in Table 5.

Rate and output dimension The rate R is computed by $\log_2(d_{enc}{\times}L)$, where L is the number of quantization levels and $d_{enc}{=}d_r{+}d_m$. Table 6 provides configurations of the rate, d_m, d_r, and L in the experiment.

Training Details: We use a batch size of 64, RMSProp optimizer with a learning rate of $5{\times}10^{-5}$, and train each model with 360 epochs, where the training set contains 60000 images. To accelerate

[4] Here, we use conditioning [33] instead of sending $X_i - \tilde{X}^w_{i-1}$ as in the original work [32]

[5] For the I-frame compressor, we follow the DCGAN implementation by Denton et al [39], adding the dithered quantization layer in the encoder's last layer(`https://github.com/edenton/svg/blob/master/models/dcgan_64.py`)

Table 4: Residual Encoder, Conditional Module, and Residual Decoder.

(**a**)Encoder f_i^c

Input
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (d_r:4:1:0), BN2D
Quantizer

(**b**)Conditional Module

Input
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (64:5:2:0), BN2D, l-ReLU
conv (192:4:1:0), BN2D

(**c**)Decoder

Input-(d_r+192)
upconv (64:4:1:0) uc4s1, BN2D, l-ReLU
upconv (64:5:2:0), BN2D, l-ReLU
upconv (64:5:2:0), BN2D, l-ReLU
upconv (64:5:2:0), BN2D, l-ReLU
upconv (channels:5:2:0), BN2D

Table 5: PLF-FMD and PLF-JD critic for frame i.

(**a**) PLF-FMD Critic

Input–$64\times64\times$channels
SN, conv (64:4:2:1), l-ReLU
SN, conv (128:4:2:1), l-ReLU
SN, conv (256:4:2:1), l-ReLU
conv (512:4:2:1), l-ReLU
Linear

(**b**) PLF-JD Critic

Input–$64\times64\times(i\times$channels$)$
SN, conv (64:4:2:1), l-ReLU
SN, conv (128:4:2:1), l-ReLU
SN, conv (256:4:2:1), l-ReLU
conv (512:4:2:1), l-ReLU
Linear

training, we pre-train each model for 60 epochs with the MSE objective only. Under WGAN-GP framework [30], we use the gradient penalty of 10 and update the encoders/decoders for every 5 iterations. The parameters λ controlling the tradeoff are in Table.7. Training takes 2 days per model on a single NVIDIA P100 GPU. For the MovingMNIST factor of two bound and permanence of error experiments, we repeat the training 3 times.

Table 6: Rate, embedding dimension d_m, d_r and quantization level L.

(**a**) P-frame encoder, $R_1 = \infty$.

R_2	d_m	d_r	L
1 bit	1	0	2
2 bits	1	1	2
3.17 bits	1	1	3

(**b**) P-frame encoder, $R_1 = \epsilon$ (12 bits).

R_2	d_m	d_r	L
4 bit	2	2	2
8 bits	4	2	2
12 bits	6	2	2

J.3 Permanence of Error on KTH Datasets

The KTH dataset is a widely-used benchmark dataset in computer vision research, consisting of video sequences of human actions performed in various scenarios. We show more examples supporting our argument for the permanence of error on this realistic dataset. We use 16 bits for each frame. In general, the 0-PLF-JD decoder consistently outputs correlated but incorrect reconstructions due to the error induced by the first reconstructions, i.e., the P-frames will follow the wrong direction induced from the I-frame reconstruction. Besides the moving direction, we also notice that the type of actions (i.e. walking, jogging, and running) is also affected. On the other hand, while losing some temporal cohesion, MMSE and 0-PLF FMD decoders manage to fix the movement error.

Finally, we computed LPIPS [41] in Table 8, which is known as a full-reference perceptual metric for images. We compare LPIPS for each reconstruction (MMSE, 0-PLF-FMD and 0-PLF-JD) at each timeframe. This result aligns with our results on distortion loss, as the MMSE and 0-PLF-FMD tend to correct the reconstruction (so the ground truth and reconstruction look more similar over time). On the other hand, due to error permanence, the 0-PLF-JD reconstructions become different from their source sequence, causing the score to go up.

Table 7: Perception loss and their associated λ

Perception Loss	$\lambda \times 10^{-3}$
Joint Distance (JD)	$0.0, 0.7, 1.0, 1.15, 1.2, 1.25, 1.3, 1.5, 1.7$ $2.0, 3.0, 5.0, 8.0, 10.0, 40.0, 80.0$
Frame Marginal Distance (FMD)	$0.0, 0.4, 0.7, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 7.0, 10.0, 40.0$

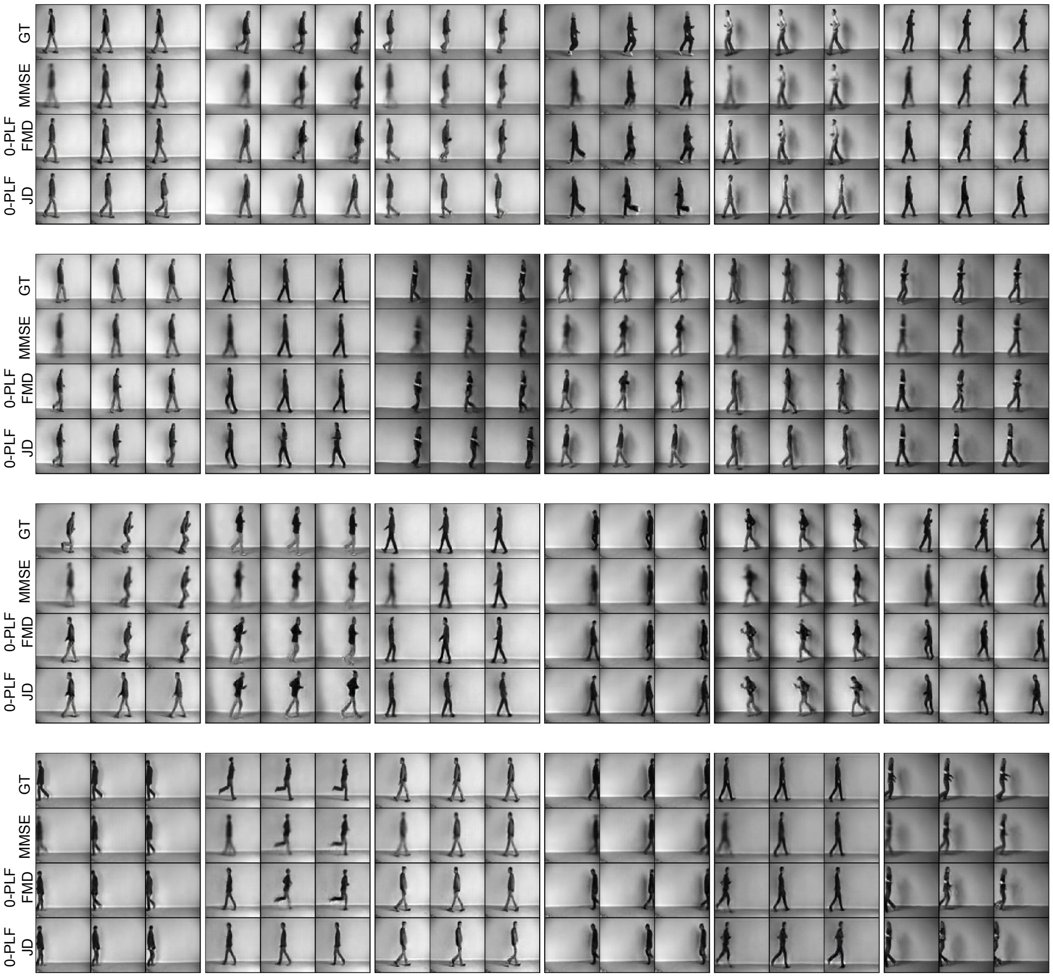


Figure 8: Additional Experimental Results for the Permanence of Error Phenomenon on KTH Dataset.

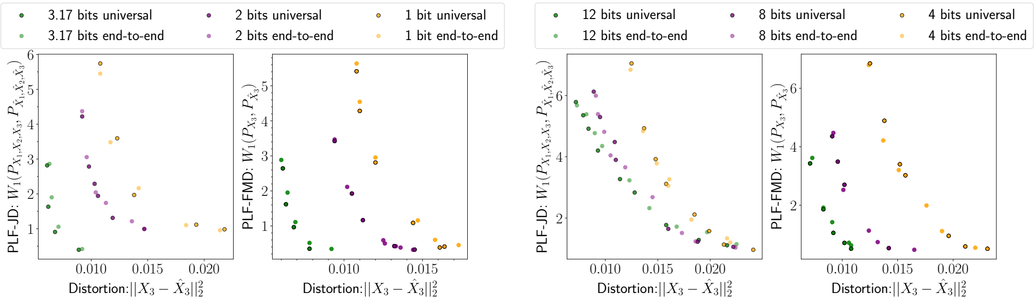
Table 8: LPIPS score on KTH dataset (lower is better).

	MMSE	PLF-FMD	PLF-JD
1st frame	0.1036	0.0584	0.0584
2nd frame	0.0521	0.0313	0.0594
3rd frame	0.0413	0.0232	0.0613

J.4 RDP Tradeoff for 3 frames

We extend our experimental results for the RDP-tradeoff and the principal of universality to the case of GOP size 3. As mentioned in the main paper, while the universal model only requires MMSE representations, the optimal end-to-end model also requires the MMSE reconstructions from previous frames to provide best estimates for motion flow vectors. Practically, this is challenging for our employed architecture since only previous $\hat{X}_1, \hat{X}_2$ are available. As a result, to compare the RDP tradeoff between universal and end-to-end model, we also provide the end-to-end model with the MMSE estimate from previous frames while noting that this is unfeasible in practice. Interestingly, we show in Figure 9 the RDP tradeoff curves for the third frame X_3 and its reconstruction $\hat{X}_3$,

https://doi.org/10.52202/075280-2093



(a) Case 1: $R_1{=}\infty$, $R_2{=}R_3{=}\{1, 2, 3.17\}$ bits

(b) Case 2: $R_1{=}12$, $R_2{=}R_3{=}\{4, 8, 12\}$ bits

Figure 9: RDP tradeoff curves for end-to-end and universal models. We plot the tradeoff for the two regimes: $R_1{=}\infty$ and $R_1{=}\epsilon$ in (a) and (b) respectively. The universal and optimal curves are close to each other.

observing that the universal and optimized curves are still relatively close to each other. When $(R_1, R_2, R_3){=}(\infty, \epsilon, \epsilon)$, we note that the distortion for X_3 is larger than X_2 since the allocated rate is not enough to correct the motion. Finally, for the case $(R_1, R_2, R_3){=}(\epsilon, \epsilon, \epsilon)$, we note that the curves again converge as in the case of $(R_1, R_2){=}(\epsilon, \epsilon)$ due to the incorrect reconstruction in the I-frame.

J.5 Diversity and Correlation

When $(R_1, R_2){=}(\infty, \epsilon)$, our theoretical analysis predicted that the decoder optimized for JD is capable of producing diverse reconstructions. On the other hand, an optimized decoder for FMD will tend to produce reconstructions that are highly correlated with the previous reconstruction $\hat{X}_1$[6]. In our experiment, we also observe such behavior, summarized in Table 9 and show several examples for $R_2 = 2$ bits in Figure 10. We observe that reconstructions from the joint metric deviate more randomly from X_1 than the marginal reconstructions. The marginal reconstructions, on the other hand, stay much closer to their original reconstruction $\hat{X}_1$.

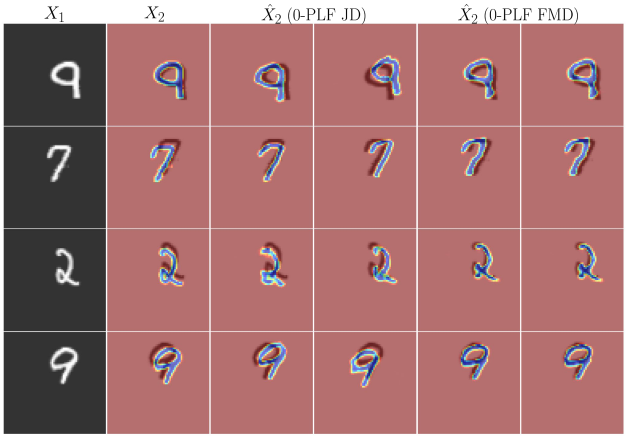


Figure 10: Diversity in reconstruction $\hat{X}_2$ for 0-PLF-JD and correlation with previous frames $\hat{X}_1$ for 0-PLF-JMD. We show X_1 in the first column. From the second column, the light-dark region represents X_1 and the color digit represents $X_2, \hat{X}_2$. For each perception metric, we show two samples.

We measure the diversity in $\hat{X}_2$ reconstruction using $\mathrm{E}[\mathrm{Var}(\hat{X}_2|X_1, X_2)]$ and the correlation with $\hat{X}_1$ by $\mathrm{E}[\mathrm{sim}(\hat{\mathrm{X}}_2, \mathrm{X}_1)]$, where $\mathrm{sim}(u, v)$ is the cosine distance between u, v. Table 9a shows that as we increase the number of bits in R_2, the diversity decreases as the decoder can reconstruct the frame more precisely. In Table 9b, we see that the joint metric keeps the correlation relatively constant, showing that it actually preserves the temporal consistency. On the other hand, as the rate becomes larger, 0-PLF-FMD reconstruction tends to be less correlated with the previous frame X_1. Finally, we note that our architecture innately utilizes common randomness to produce diverse reconstructions and does not suffer from mode-collapse behavior in general conditional GAN settings [42].

[6] $X_1 = \hat{X}_1$ in this regime.

https://doi.org/10.52202/075280-2093

Table 9: Diversity (a) between $\hat{X}_2$ and Correlation Measures (b) between $\hat{X}_2$ and X_1.

(a) Diversity Measures ↑.

R_2	Joint	Marginal
1 bit	0.0096	0.0004
2 bits	0.0082	0.0029
3.17 bits	0.0042	0.0022

(b) Correlation Measures. ↑

R_2	Joint	Marginal
1 bit	0.5218	0.6202
2 bits	0.5190	0.5969
3.17 bits	0.5205	0.5508

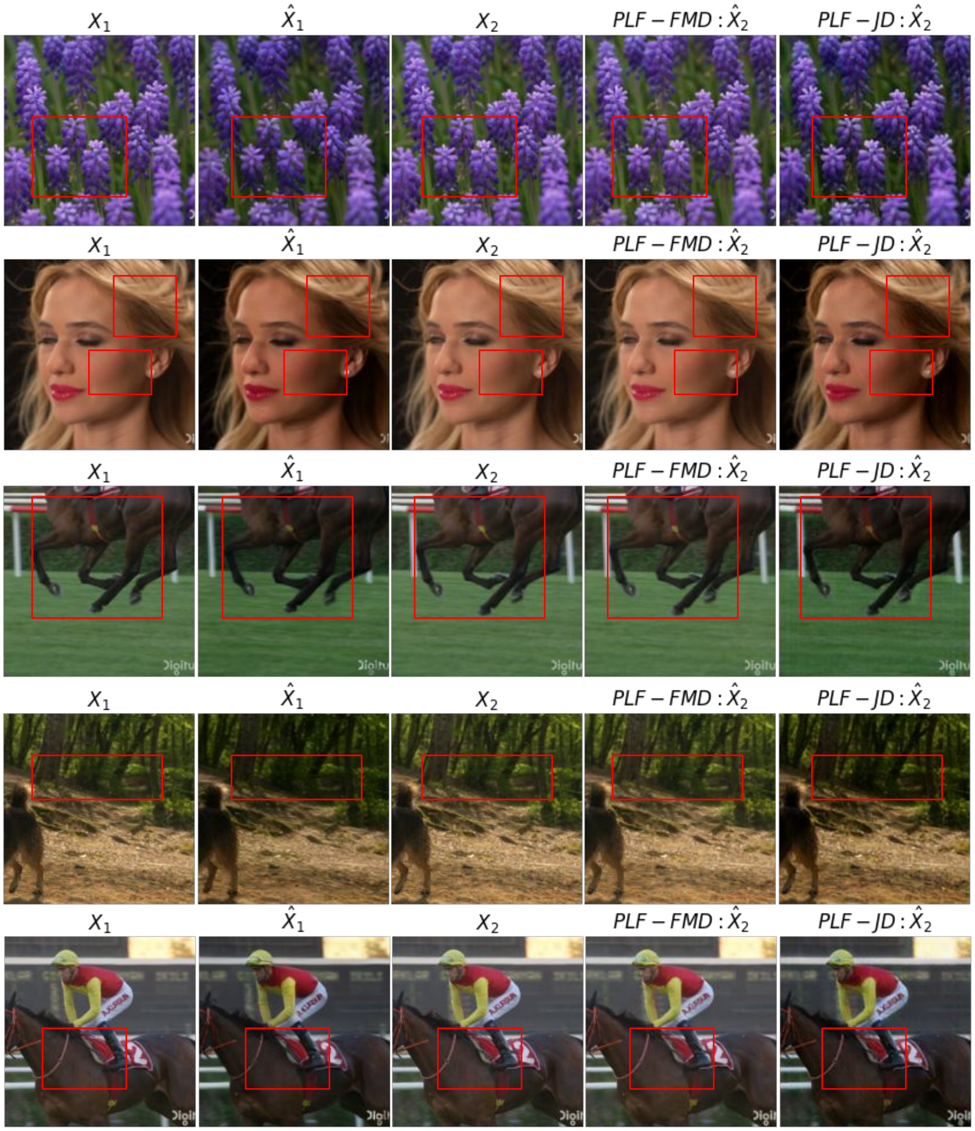


Figure 11: Visualization of error permanence. The PLF-JD reconstructions propagate the flaws in the color tone from the previous I-frame reconstruction while the PLF-FMD is able to fix such error. Compression rate for I-frame and P-frame are 0.144bpp (low rate) and 4.632bpp (high rate) respectively.

K Error Permanence on UVG Dataset

We demonstrate the phenomenon of error permanence in a large-scale scenario using the UVG dataset. Our P-frame compressor is trained on the Vimeo-90k dataset. As illustrated in Figure 11, when reconstructing I-frames, an inaccurate color tone is introduced, which persists when employing PLF-JD. However, PLF-FMD effectively rectifies this issue within P-frames. Numerical results are

Table 10: Distortion $(X_2 - \hat{X}_2)^2$ evaluated across 3900 UVG frames. The PLF-JD reconstructions exhibit notably greater distortion compared to MMSE and PLF-FMD, aligning with our theoretical findings.

	MMSE	PLF-FMD	PLF-JD
Distortion (MSE)	0.0026	0.0032	0.0168

in Table. 10. We note that for the I-frame, we use the pretrained model in [43] that targets high perceptual quality and is also trained on Vimeo-90k dataset.

L Limitations

This work studies the effects of different perception loss functions, namely the PLF-JD and PLF-FMD, on the performance of lossy causal video compression. Our theoretical analysis and experiment reveal the error permanence phenomenon and show the universality principle, suggesting that MMSE representation can be transformed into other points on the DP tradeoffs.

In practice, one might want to combine these two losses, for example, perfect framewise realism (0-PLF FMD) while retaining some degree of temporal cohesion (PLF-JD small), which is not considered in this work. Furthermore, analysis for other types of video compression schemes, such as with B-frame, and scaling the universality compression architecture to high-definition videos are also desired.

https://doi.org/10.52202/075280-2093