Delayed Sequential Coding of Correlated Sources

Nan Ma, Ye Wang, and Prakash Ishwar

Department of Electrical and Computer Engineering
Boston University, Boston, MA 02215
{nanma, yw, pi}@bu.edu

Abstract— Motivated by video coding applications, we study the problem of sequential coding of correlated sources with (noncausal) encoding and/or decoding frame-delays. The fundamental tradeoffs between individual frame rates, individual frame distortions, and encoding/decoding frame-delays are derived in terms of a single-letter information-theoretic characterization of the rate-distortion region for general inter-frame source correlations and certain types of (potentially frame-specific and coupled) single-letter fidelity criteria. For video sources which are spatially stationary memoryless and temporally Gauss–Markov, MSE frame distortions, and a sum-rate constraint, our results expose the optimality of differential predictive coding among all causal sequential coders. Somewhat surprisingly, causal sequential encoding with one-step delayed noncausal sequential decoding can *exactly match* the sum-rate-MSE performance of *joint coding* for all nontrivial MSE-tuples satisfying certain positive semi-definiteness conditions. Thus, even a single frame delay holds potential for yielding huge performance improvements. A rate-distortion performance equivalence of, causal sequential encoding with delayed noncausal sequential decoding, and, delayed non-causal sequential encoding with causal sequential decoding, is also established.

I. Introduction

Differential predictive coded modulation (DPCM) is a popular and well-established sequential predictive source compression method with a long history of development (see [1]–[5] and the references therein). DPCM has had wide impact on the evolution of compression standards for speech, image, audio, and video coding. The classical DPCM system consists of a causal sequential predictive encoder and a causal sequential decoder. This is aligned with applications having low delay tolerance at both encoder and decoder. However, there are many interesting scenarios where these constraints can be relaxed. There are three additional sequential source coding systems possible when limited delays are allowed at the encoder and/or the decoder: (i) causal (C) encoder and noncausal (NC) decoder; (ii) NC-encoder and C-decoder; and (iii) NC-encoder and NC-decoder. Application examples of these include, respectively, non-real-time display of live video for C–NC, zero-delay display of non-real-time encoded video for NC–C, and non-real-time display of non-real-time video for NC–NC (see Figs. 1, 2, and 3). Of special interest, for performance comparison, is joint coding (JC) which may be interpreted as an extreme special case of the C–NC, NC–C, and the NC–NC systems where all frames are jointly processed and jointly reconstructed (Fig. 3(c)).

The goal of this work is to provide a computable (single-letter) characterization of the fundamental information-

theoretic rate-distortion performance limits for the different scenarios and to quantify and compare the potential value of systems with limited encoding and decoding delays in different rate-distortion regimes. The primary motivational application of our study is video coding (see Section II-B) with encoding and decoding *frame* delays.[1] However, this work also has implications for certain sensor-network applications in which the communication constraints impose a sequential order for coding the blocks of observations at different sensors.

To characterize the fundamental tradeoffs between individual frame-rates, individual expected frame-distortions, encoding and decoding frame-delays, and source inter-frame correlation, we build upon the information-theoretic framework of sequential coding of correlated sources. This mathematical framework was first introduced in [6] (and independently studied in [7], [8] under a nonasymptotic stochastic control framework involving dynamic programming) within the context of the *purely C–C*[2] sequential source coding system. As noted in [6], the results for the well-known successive-refinement source coding problem can be derived from those for the C–C sequential source coding problem by setting all sources to be identically equal to the same source. The complete (single-letter) rate region for two sources (with a remark regarding generalization to multiple sources) and certain types of perceptually motivated coupled (and uncoupled) single-letter distortion criteria was derived in [6]. Our results cover not only the C–C problem studied [6] but also the C–NC, the NC–C, and the JC cases[3] for arbitrary number of sources and for coupled single-letter distortion criteria similar to those in [6]. We have also been able to simplify some of the key derivations in [6] (the C–C case).

The benefits of decoding delay on the rate versus MSE performance was investigated in [4] for a, spatially independent-vector-Gaussian, temporally Gaussian first-order-autoregressive model for video, with a DPCM structure imposed on both the encoder and the decoder. In contrast to conventional rate-distortion studies of *scalar* DPCM systems based on *scalar quantization* and *high-rate* asymptotics (see [1]–[3] and references therein), [4] studied DPCM systems with vector-valued sources and large spatial (as opposed to high rate) asymptotics similar in spirit to [6]–[8] but with decoding frame-delays. The main findings of [4] were that (i)

---

[1] Accordingly, terms like frame-delay and "causal" and "noncausal" encoding and/or decoding should be interpreted within this application context.

[2] The terminology is ours.

[3] The general NC–NC case is currently under investigation.

NC-decoders offer a significant *relative* improvement in the MSE at medium to low rates for video sources with strong temporal correlation, (ii) most of this improvement can be attained with a modest decoding frame-delay, and (iii) the gains vanish at very high and very low rates.

In contrast to the insistence on DPCM encoders and decoders in [4], here we consider arbitrary rate-constrained coding structures as in conventional rate-distortion studies. When specialized to spatially stationary memoryless, temporally Gauss–Markov video sources, with MSE as the fidelity metric and a sum-rate constraint, our results reveal the information-theoretic optimality of DPCM encoders and decoders for the C–C sequential coding system (Corollary 1.3). A second, somewhat surprising, finding is that for the just mentioned Gauss–Markov video sources with a sum-rate constraint, a C-encoder with a one-step-delayed NC-decoder (Fig. 3(a)) can *exactly match* the sum-rate-MSE performance of the *joint coding system* (Fig. 3(c)) which can wait to collect *all* frames of the video segment before jointly processing and jointly reconstructing them[4] (Corollary 3.2). Interestingly, this performance equivalence does not hold for all MSE-tuples. It holds for a *non-trivial* subset which satisfies certain positive semi-definiteness conditions. The performance-matching region expands with increasing frame-delays allowed at the decoder until it completely coincides with the set of all reachable tuples of the JC system. In simple words, the benefit of even a single frame-delay can be huge. These two specific architectural results constitute the main high-level take-away messages of this work.

For clarity of exposition, our discussion in this paper is limited to the exemplary case of three discrete, memoryless, (*spatially*) stationary (DMS) correlated sources taking values in finite alphabets and standard uncoupled single-letter fidelity criteria. The results presented in this work in fact hold for an arbitrary number of correlated sources and for more general coupled fidelity criteria which are similar to those considered in [6] and will be reported elsewhere. Analogous results can be established for continuous alphabets (e.g., Gaussian sources) and unbounded distortion criteria (e.g., MSE) using the techniques in [9] but are not discussed here. To keep the exposition clutter-free, all proof-sketches involving technical derivations are presented in the Appendices. Detailed technical proofs of all results will be provided in a technical report under preparation.[5]

The rest of this paper is organized as follows. Four delayed sequential coding systems and their associated operational rate-distortion regions are formulated in Section II. All the technical findings with discussions of the underlying intuition and implications for the C–C, JC, C–NC, and NC–C systems are presented in Sections III, IV, V, and VI respectively. We conclude in Section VII with remarks on ongoing work.

Notation: The nonnegative cone of real numbers is de-

---

[4]This is similar to the coding of *correlated* parallel vector Gaussian sources but with an *individual* MSE constraint on each source component.

[5]This will be posted on `<arXiv.org>` and submitted to the IEEE Transactions on Information Theory.

noted by $\mathbb{R}^+$ and 'iid' denotes independent and identically distributed. Vectors are denoted in boldface (e.g., $\mathbf{x}$, $\mathbf{X}$). With the exception of T denoting the size of a group of pictures (GOP) in a video segment, random quantities are denoted in upper case (e.g., X, $\mathbf{X}$), and their specific instantiations in lower case (e.g., $X = x$, $\mathbf{X} = \mathbf{x}$). A^n denotes the ordered tuple $(A_1, \ldots, A_n)$ and A_m^n denotes $(A_m, \ldots, A_n)$.

II. Problem formulation

A. Statistical model for $T = 3$ correlated sources

Three correlated DMSs taking values in finite alphabets are defined by $(X_1(i), X_2(i), X_3(i))_{i=1}^n \in (\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)^n$ with $(X_1(i), X_2(i), X_3(i)) \sim$ iid $p_{X_1 X_2 X_3}(x_1, x_2, x_3)$. Potentially, the (spatially) iid assumption can be relaxed to spatially stationary ergodic sources by a general AEP argument, but is not treated in this paper for simplicity. Motivated by the video coding application, these sources have the following temporal structure. Instead of being available simultaneously for encoding, initially, only $\{X_1(i)\}_{i=1}^n$ is available. Then, $\{X_2(i)\}_{i=1}^n$ "appears" and finally, $\{X_3(i)\}_{i=1}^n$ "arrives". In the following video coding application, this assumption captures the temporal order of frames shown in Fig. 1.

B. Motivating application contexts

(*i*) *Video coding:* Here (Fig. 1), $\mathbf{X}_j = \{X_j(i)\}_{i=1}^n, j = 1, \ldots, T$ represent $T = 3$ *video frames* with i = discrete index of the spatial location of a picture element (pixel) relative to a certain spatial scan order (e.g., zig-zag or raster scan), and $X_j(i)$ = discrete pixel intensity level at spatial location i in frame number j (Fig. 1). The statistical structure implies that the sources are *spatially independent but temporally dependent*.
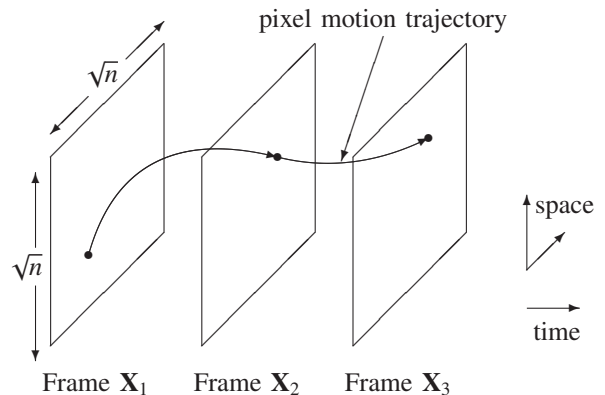

Fig. 1. *Illustrating motion-compensated video coding for $T = 3$ frames.*

While this is rarely an accurate statistical model for the *unprocessed* frames of a stationary video segment (usually corresponding to the GOP in video coding standards), it is a reasonable approximation for the evolution of the video innovations process along optical-flow motion trajectories for groups of adjacent pixels (see [4] and references therein). This model assumes arbitrary temporal correlation but iid spatial correlation. The statistical law $p_{X_1 X_2 X_3}$ is assumed

to be known here. In practice, this may be learnt from pre-operational training using clips from video databases used by video-codec standardization groups such as H.26x and MPEG-x quite similar in spirit to the offline optimization of quantizer tables in commercial video codecs. Single-letter information-theoretic coding results need asymptotics along some problem dimension to exploit some version of the law of large numbers. Here, the asymptotics are in the *spatial dimension* and is appropriate for video coding applications where it is quite typical to have frames of size $n = 352 \times 288$ pixels at 30 frames per second (full CIF[6]). It is also fairly common to code video in groups of $T = 15$ pictures.

(*ii*) *Sensor networks:* Here, $\mathbf{X}_j = \{X_j(i)\}_{i=1}^n, j = 1, \ldots, T$ represent $T = 3$ statistically correlated sources observed at spatially separated sensor locations with i = discrete time index (corresponding to the temporal sampling frequency of the sensors) and j = discrete spatial index (sensor location). In this application, the asymptotics are in time corresponding to a large block of samples gathered over a certain time duration. The sensor observations can have arbitrary spatial correlation which is assumed to be known. In practice, the statistics can be ascertained from the physical laws governing the spatio-temporal phenomenon being sensed. See [10]–[12] for related sensor network problems having sequential coding constraints.

The primary application focus of this work is video coding. Accordingly, terms such as (frame) delay "causal" and "non-causal" encoding and/or decoding should be interpreted within this application context.

C. Delayed sequential coding systems

• *C–C systems:* Causal (zero-delay) sequential encoding with (zero-delay) causal sequential decoding is illustrated in Fig. 2. Implicitly, the encoders and decoders have enough memory to store all previous frames and messages.
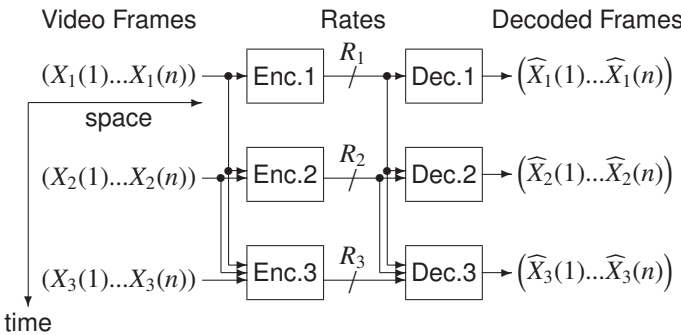

Fig. 2. *C–C: Causal (zero-delay) sequential encoding with causal sequential decoding.* Sum-rate $= R_{sum}^{C-C} = R_1 + R_2 + R_3$.

• *C–NC systems:* Causal sequential encoding with one-step delayed noncausal sequential decoding is illustrated in Fig. 3(a).
• *NC–C systems:* One-step delayed noncausal sequential

[6]Progressively scanned HDTV is typically $n = 1280 \times 720 \approx$ one million pixels at 60 frames per second.

encoding with causal sequential decoding is illustrated in Fig. 3(b).
• *The JC system:* Of special interest is *joint* (noncausal) encoding and decoding illustrated in Fig. 3(c). Note that here the encoding frame delay is $(T - 1)$.

The C–C blocklength-n encoders and decoders are formally defined by the maps

(Enc.j) $f_j : \ \mathcal{X}_1^n \times \ldots \times \mathcal{X}_j^n \rightarrow \{1, \ldots, M_j\},$

(Dec.j) $g_j : \ \{1, \ldots, M_1\} \times \ldots \times \{1, \ldots, M_j\} \rightarrow \widehat{\mathcal{X}}_j^n$

for $j = 1, \ldots, T$, where $(\log_2 M_j)/n$ is the j-th frame coding rate in bits per pixel (bpp) and $\widehat{\mathcal{X}}_j$ is the j-th (finite cardinality) reproduction alphabet.

The formal definitions of C–NC encoders are identical to that for the C–C encoders. However, the C–NC decoders with *one-step frame delay* are formally defined by the maps

(Dec.j) $g_j : \ \{1, \ldots, M_1\} \times \ldots \times \{1, \ldots, M_{\min\{j+1,T\}}\} \rightarrow \widehat{\mathcal{X}}_j^n,$

for $j = 1, \ldots, T$. Similarly, the NC–C decoder definitions are identical to those for the C–C decoders and the NC–C encoders are formally defined by the maps

(Enc.j) $f_j : \ \mathcal{X}_1^n \times \ldots \times \mathcal{X}_{\min\{j+1,T\}}^n \rightarrow \{1, \ldots, M_j\},$

for $j = 1, \ldots, T$. Finally the JC encoder and decoder are defined by the maps

(Enc.) $f : \ \mathcal{X}_1^n \times \ldots \times \mathcal{X}_T^n \rightarrow \{1, \ldots, M\},$

(Dec.) $g : \ \{1, \ldots, M\} \rightarrow \widehat{\mathcal{X}}_1^n \times \ldots \times \widehat{\mathcal{X}}_T^n.$

Note that for $T = 2$, a one-step delayed sequential coding system (C–NC or NC–C) is operationally *equivalent* to a JC system. The first nontrivial delayed sequential coding problem arises for $T = 3$. Therefore we consider $T = 3$ frames in this paper. As will become clear, it is straightforward to generalize all results of this paper to more than three frames (see [13]).

For a frame-delay k, there are boundary effects associated with the decoders (resp. encoders) of the last $(k+1)$ frames for the C–NC (resp. NC–C) systems. For example, the last two decoders in Fig. 3(a) are operationally equivalent to a single decoder since both use the same set of encoded messages. Similarly, the last two encoders in Fig. 3(b) are operationally equivalent to a single encoder since both use the same set of source frames. However, it should be noted that whereas combining the last two decoders in Fig. 3(a) into a single decoder does not change the rate-distortion region, combining the last two encoders in Fig. 3(b) into a single encoder reduces the dimension of the rate-tuples by one. Hence, although redundant, we will retain the distinction of the boundary encoders/decoders for clarity and to aid comparison.

D. Operational rate-distortion regions

For each $j = 1, \ldots, T$, the pixel reproduction quality is measured by a single-letter distortion (fidelity) criterion $d_j : \mathcal{X}_j \times \widehat{\mathcal{X}}_j \rightarrow \mathbf{R}^+$. The frame reproduction quality is in terms of the average pixel distortion $d_j(x_j^n, \hat{x}_j^n) = \sum_{i=1}^n d_j(x_j(i), \hat{x}_j(i))/n$. Of interest are the expected frame distortions $E[d_j(\mathbf{X}_j, \widehat{\mathbf{X}}_j)]$. It
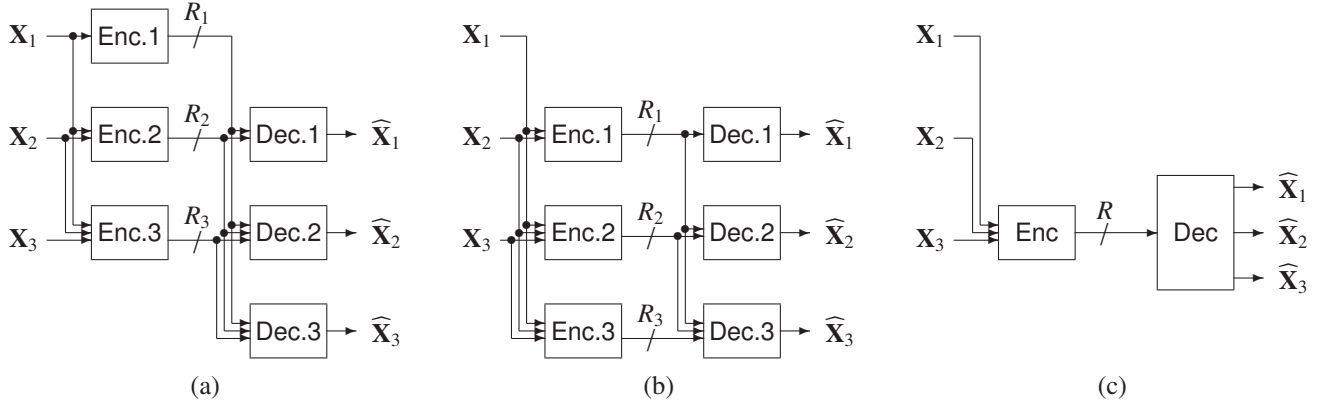
Fig. 3. *(a) C–NC: Causal sequential encoding with one-step delayed noncausal sequential decoding; (b) NC–C: one-step delayed noncausal sequential encoding with causal sequential decoding; (c) JC: $(T-1)$-step delayed joint (noncausal) encoding with joint (noncausal) decoding.*

is important to notice that these are *frame-specific* distortions as opposed to an average distortion across all frames. This makes the JC problem distinctly different from a standard parallel vector source coding problem. The results presented here also apply to certain perceptually-motivated coupled fidelity criteria reflecting dependencies on previous frame reproductions as in [6] (e.g., $d_2 : \mathcal{X}_2 \times \widehat{\mathcal{X}}_2 \times \widehat{\mathcal{X}}_1 \to \mathbb{R}^+$) but are not entered into here for clarity.

For a distortion-tuple $\mathbf{D} = (D_1, \ldots, D_T)$, a rate-tuple $\mathbf{R} = (R_1, \ldots, R_T)$ is said to be $\mathbf{D}$-admissible for a given delayed sequential coding system if, for every $\epsilon > 0$, and all sufficiently large n, there exist block encoders and decoders satisfying

$$\frac{1}{n} \log M_j \le R_j + \epsilon,$$
$$E[d_j(\mathbf{X}_j, \widehat{\mathbf{X}}_j)] \le D_j + \epsilon,$$

simultaneously for all $j = 1, \ldots, T$. For system $A \in \{\text{C–C, C–NC, NC–C, JC}\}$ and a distortion tuple $\mathbf{D}$, the operational rate region $\mathcal{R}_{op}^A(\mathbf{D})$ is the set of all $\mathbf{D}$-admissible rate tuples. and the sum-rate region $R_{sum}^A(\mathbf{D})$ is the set of all the $\mathbf{D}$-admissible sum-rates $\sum_{j=1}^{T} R_j$.

III. RESULTS FOR C–C SYSTEMS

The C–C rate region can be formulated as a single-letter mutual information optimization problem subject to distortion constraints and natural Markov chains involving auxiliary and reproduction random variables and deterministic functions.

Theorem 1 *(C–C rate-distortion region)* The single-letter rate-distortion region for a $T = 3$ frame C–C system is given by

$$\mathcal{R}^{C-C}(\mathbf{D}) = \{\mathbf{R} \mid \exists\, U^2, \widehat{X}^3, g_1(\cdot), g_2(\cdot, \cdot), s.t.$$
$$R_1 \ge I(X_1; U_1),$$
$$R_2 \ge I(X^2; U_2|U_1),$$
$$R_3 \ge I(X^3; \widehat{X}_3|U^2),$$
$$E[d_j(X_j, \widehat{X}_j)] \le D_j, \quad j = 1, 2, 3,$$
$$\widehat{X}_1 = g_1(U_1), \ \widehat{X}_2 = g_2(U_1, U_2),$$
$$U_1 - X_1 - X_2^3, \ U_2 - (X^2, U_1) - X_3\}$$

where $\{U_1, U_2, \widehat{X}_1, \widehat{X}_2, \widehat{X}_3\}$ are auxiliary and reconstruction random variables and $\{g_1(\cdot), g_2(\cdot, \cdot)\}$ are deterministic functions. Cardinality bounds on the alphabets of auxiliary random variables can be derived using the Carathéodory theorem and the support lemma as in [6] but are omitted.

Note that this C–C rate region, as described here, differs from the direct extension of the result in [6]: the formulation in Theorem 1 has different rate inequalities and fewer Markov chain conditions than in [6]. However, as discussed below, this form of the rate region offers a somewhat easier natural interpretation and therefore an easier generalization to the case of multiple frames.

The proof of achievability follows standard random coding and random binning arguments. This region has the following natural interpretation. First, $\mathbf{X}_1$ is quantized into $\mathbf{U}_1$ using a random codebook-1 for encoder-1 without access to $\mathbf{X}_2^3$. Decoder-1 recovers $\mathbf{U}_1$ and reproduces $\mathbf{X}_1$ as $\widehat{\mathbf{X}}_1 = g_1^n(\mathbf{U}_1)$. Next, the tuple $\{\mathbf{X}^2, \mathbf{U}_1\}$ is (jointly) quantized into $\mathbf{U}_2$ without access to $\mathbf{X}_3$ using a random codebook-2 for encoder-2. The codewords are further randomly distributed into bins and the bin index of $\mathbf{U}_2$ is sent to the decoder. Decoder-2 identifies $\mathbf{U}_2$ from the bin with the help of $\mathbf{U}_1$ as side-information (available from decoder-1) and reproduces $\mathbf{X}_2$ as $\widehat{\mathbf{X}}_2 = g_2^n(\mathbf{U}_1, \mathbf{U}_2)$. Finally, encoder-3 (jointly) quantizes $\{\mathbf{X}^3, \mathbf{U}^2\}$ into $\widehat{\mathbf{X}}_3$ using encoder-3's random codebook, bins the codewords and sends the bin index of $\widehat{\mathbf{X}}_3$ such that decoder-3 can identify $\widehat{\mathbf{X}}_3$ with the help of $\mathbf{U}^2$ as side-information available from decoders 1 and 2. The constraints on the rates and Markov chains ensure that with high probability (for all large enough n) both encoding (quantization) and decoding (recovery) succeed and the recovered words are jointly strongly typical with the source words to meet the target distortions. Notice that the conditioning random variables that appear in the conditional mutual information expressions at each stage correspond to quantities that are known to both the encoding and decoding sides at that stage due to the previous stages. By formalizing this observation one can write down an achievable rate-distortion region for general delayed sequential coding

systems by inspection. Of course, the converses will have to be established to claim these regions to be the entire rate-distortion region. As discussed later in this paper, it turns out that the converses can indeed be established.

The (weak) converse part of Theorem 1, is proved following [6] using standard information inequalities by defining auxiliary random variables $U_j(i) = \{S_j, X_j^{(i-1)}\}, j = 1, 2$, where S_j denotes the message sent by the j-th encoder satisfying all Markov-chain and distortion constraints, and a convexification (timesharing) argument as in [14, p.397]. The important steps in the derivation are sketched in Appendix B.

Corollary 1.1 *(C–C Sum-rate region)* The sum-rate region for the C–C system is $\mathcal{R}_{sum}^{C-C}(\mathbf{D}) = [R_{sum}^{C-C}(\mathbf{D}), \infty)$ where the minimum sum-rate is

$$R_{sum}^{C-C}(\mathbf{D}) = \min_{\substack{E[d_j(X_j, \widehat{X}_j)] \leq D_j, j=1,2,3, \\ \widehat{X}_1 - X_1 - X_2^3 \\ \widehat{X}_2 - (X^2, \widehat{X}_1) - X_3}} I(X^3; \widehat{X}^3). \quad (3.1)$$

The proof is sketched in Appendix B. The proof follows from the rate region of Theorem 1 and the Markov chain constraints. The main simplification is the *absence* of the auxiliary random variables U^2.

As will become clear in the sequel, the minimum sum-rate for any type of delayed sequential coding system is given by the minimization of the mutual information between the source random variables X_1^T and the reproduction random variables $\widehat{X}_1^T$ subject to several expected distortion and Markov-chain constraints involving these random variables of a form similar to 3.1. In the case of Gaussian sources and MSE distortion criteria, using the following lemma it will then follow that the minimum sum-rate of any delayed sequential coding system can be achieved by reproduction random variables which are jointly Gaussian with the source random variables.

Lemma If $(X_1, \ldots, X_T)$ are jointly Gaussian, the minimizer of the optimization problem

$$\min_{E[(X_j - \widehat{X}_j)^2] \leq D_j, j=1, \ldots, T} I(X_1^T; \widehat{X}_1^T)$$

with some additional Markov chain constraints involving X_1^T and $\widehat{X}_1^T$ is achieved by reproduction random variables $(\widehat{X}_1, \ldots, \widehat{X}_T)$ which are jointly Gaussian with $(X_1, \ldots, X_T)$.

The proof which is based on the Shannon lower bound is sketched in Appendix C. Since Gaussian vectors are characterized by covariance matrices, the minimum sum-rate computation reduces to a determinant optimization problem involving second-order and Markov matrix constraints.

For Gauss–Markov sources, $p_{X_1 X_2 X_3} = \mathcal{N}(\mathbf{0}, \Sigma_X)(x_1, x_2, x_3)$ where the covariance matrix

$$\Sigma_X = \begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \rho_1\rho_2\sigma_1\sigma_3 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 & \rho_2\sigma_2\sigma_3 \\ \rho_1\rho_2\sigma_1\sigma_3 & \rho_2\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$$

has a structure which is consistent with the Markov chain $X_1 - X_2 - X_3$ associated with the Gauss–Markov assumption. Define a distortion region $\mathcal{D}^{C-C} = \{\mathbf{D} \mid D_1 \leq \sigma_1^2, D_2 \leq \sigma_{e_2}^2, D_3 \leq \sigma_{e_3}^2\}$ where

$$\sigma_{e_j}^2 = \rho_{j-1}^2 \frac{\sigma_j^2}{\sigma_{j-1}^2} D_{j-1} + (1 - \rho_{j-1}^2)\sigma_j^2, \quad j = 2, 3 \quad (3.2)$$

whose significance will be discussed below. The C–C minimum sum-rate evaluated for any MSE tuple $\mathbf{D}$ in this region is given by the following corollary.

Corollary 1.2 *(C–C minimum sum-rate for Gauss–Markov sources and MSE)* In the distortion region $\mathcal{D}^{C-C}$, the C–C minimum sum-rate for Gauss–Markov sources and MSE is

$$R_{sum}^{CCGM}(\mathbf{D}) = \frac{1}{2}\log\left(\frac{\sigma_1^2}{D_1}\right) + \frac{1}{2}\log\left(\frac{\sigma_{e_2}^2}{D_2}\right) + \frac{1}{2}\log\left(\frac{\sigma_{e_3}^2}{D_3}\right). \quad (3.3)$$

The proof of the converse part of Corollary 1.2 is sketched in Appendix D. The form of (3.3) suggests the following (achievable) coding scheme (see Fig. 4). Encoder-1 initially quantizes $\mathbf{X}_1$ into $\widehat{\mathbf{X}}_1$ to meet the target MSE D_1 using an ideal Gaussian rate-distortion quantizer and decoder-1 recovers $\widehat{\mathbf{X}}_1$. For notational convenience let $\mathbf{e}_1 = \mathbf{X}_1$ and $\hat{\mathbf{e}}_1 = \widehat{\mathbf{X}}_1$. Next, encoder-2 makes the causal minimum mean squared error (MMSE) prediction of $\mathbf{X}_2$ based on $\widehat{\mathbf{X}}_1$ and quantizes the prediction error $\mathbf{e}_2$ into $\hat{\mathbf{e}}_2$ using an ideal Gaussian rate-distortion quantizer so that decoder-2 can form $\widehat{\mathbf{X}}_2$ to meet the target MSE D_2 with help from $\hat{\mathbf{e}}_1$. The asymptotic per-component variance of $\mathbf{e}_2$ is consistent with (3.2). Specifically, decoder-2 recovers $\hat{\mathbf{e}}_2$ and creates the reproduction $\widehat{\mathbf{X}}_2$ as the causal MMSE estimate of $\mathbf{X}_2$ based on $\hat{\mathbf{e}}^2$. Finally encoder-3 makes the causal MMSE prediction of $\mathbf{X}_3$ based on $\hat{\mathbf{e}}_1^2$ and quantizes the prediction error $\mathbf{e}_3$ into $\hat{\mathbf{e}}_3$ using an ideal Gaussian rate-distortion quantizer so that decoder-3 can form $\widehat{\mathbf{X}}_3$ to meet the target MSE D_3 with help from $\hat{\mathbf{e}}_1^3$. Decoder-3 recovers $\hat{\mathbf{e}}_3$ and makes the reproduction $\widehat{\mathbf{X}}_3$ as the MMSE estimate of $\mathbf{X}_3$ based on $\hat{\mathbf{e}}^3$. The C–C coding scheme just described is called DPCM (see [1]–[5] and references therein). This coding procedure when formalized using random coding arguments leads to the following corollary.
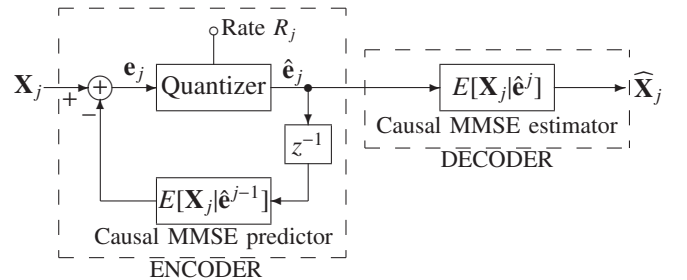

Fig. 4. *Illustrating DPCM.*

Corollary 1.3 *(C–C Optimality of DPCM for Gauss–Markov sources and MSE)* The C–C minimum sum-rate-MSE performance for Gauss–Markov sources is achieved by DPCM for all distortion tuples $\mathbf{D}$ in the distortion region $\mathcal{D}^{C-C}$.

The distortion region $\mathcal{D}^{C-C}$ may be interpreted as the set of distortion levels for which the DPCM encoder uses a positive rate for each frame. It can be shown that $\mathcal{D}^{C-C}$ has a *non-zero volume* for nonsingular sources (see Section V, third para after Corollary 3.2). Hence, the assertion that DPCM is optimal for C–C systems is a nontrivial statement.

IV. RESULTS FOR THE JC SYSTEM

Theorem 2 *(JC rate-distortion function, [15, Problem 14, p.134])* The single-letter rate-distortion function for the joint coding system is given by

$$R^{JC}(\mathbf{D}) = \min_{\substack{E[d_j(X_j,\widehat{X}_j)] \leq D_j \\ j=1,2,3.}} I(X^3; \widehat{X}^3). \tag{4.4}$$

Compared to $R_{sum}^{C-C}(\mathbf{D})$ given by (3.1), the JC rate-distortion function $R^{JC}(\mathbf{D})$ given by (4.4) having no Markov chain constraints is a lower bound on $R_{sum}^{C-C}(\mathbf{D})$. While this follows from a direct comparison of the single-letter rate-distortion functions, from the operational structure of C–C, C–NC, NC–C, and JC systems it is clear that the JC rate-distortion function is in fact a lower bound on the sum-rates for *all* delayed sequential coding systems.

Similar to Corollary 1.2 for C–C systems for Gaussian sources and MSE distortion criteria, for JC systems we have the following corollary.

Corollary 2.1 *(JC rate-MSE function for Gauss–Markov sources)*
(i) For the distortion region $\mathcal{D}^{JC} \stackrel{\text{def}}{=} \{\mathbf{D} \mid (\Sigma_X - \text{diag}(\mathbf{D})) \geq 0\}$, the JC rate-MSE function for jointly Gaussian sources is given by

$$R^{JCGM}(\mathbf{D}) = \frac{1}{2} \log\left(\frac{|\Sigma_X|}{D_1 D_2 D_3}\right). \tag{4.5}$$

(ii) For the distortion region $\mathcal{D}^{JC}$, the JC rate-MSE function for Gauss–Markov sources is given by

$$R^{JCGM}(\mathbf{D}) = \frac{1}{2} \log\left(\frac{\sigma_1^2}{D_1}\right) + \frac{1}{2} \log\left(\frac{\sigma_2^2(1-\rho_1^2)}{D_2}\right) +$$
$$+ \frac{1}{2} \log\left(\frac{\sigma_3^2(1-\rho_2^2)}{D_3}\right). \tag{4.6}$$

Formula (4.5) is the Shannon lower bound [2], [3] of the JC rate-distortion function. It can be achieved in the distortion region $\mathcal{D}^{JC}$ by the test channel

$$\widehat{\mathbf{X}} + \mathbf{Z} = \mathbf{X} \tag{4.7}$$

where $\mathbf{Z} = (Z_1, Z_2, Z_3)$ and $\widehat{\mathbf{X}} = (\widehat{X}_1, \widehat{X}_2, \widehat{X}_3)$ are independent Gaussian vectors with covariance matrices

$$\Sigma_Z = \text{diag}(\mathbf{D}), \quad \Sigma_{\hat{X}} = \Sigma_X - \text{diag}(\mathbf{D}),$$

and $\mathbf{X} = (X_1, X_2, X_3)$. The existence of this channel is guaranteed by the definition of $\mathcal{D}^{JC}$.

Comparing (3.3) and (4.6) for $\mathbf{D} \in \mathcal{D}^{JC} \cap \mathcal{D}^{C-C}$ which generally has a nonempty interior, we find that in general the C–C sum-rate $R_{sum}^{CCGM}(\mathbf{D})$ is *strictly* greater than the JC rate $R^{JCGM}(\mathbf{D})$. However, as $\mathbf{D} \to \mathbf{0}$, the two rates are asymptotically equal. This high-rate asymptotic phenomenon is consistent with the Slepian-Wolf theorem in that the "Slepian-Wolf encoders" can compress the sources individually (without cooperation) with a sum-rate equal to the total (joint) entropy of all sources [14]. However, note that in the classical Slepian-Wolf problem formulation we have spatially correlated sources and temporal asymptotics, whereas in the problem formulation here we have temporally correlated sources and spatial asymptotics. The roles of time and space are exchanged.

V. RESULTS FOR C–NC SYSTEMS

Similar to C–C systems, we can derive the rate-distortion and sum-rate regions for C-NC systems as follows.

Theorem 3 *(C–NC rate-distortion region)* The single-letter rate-distortion region for a C–NC system with one-step decoding frame delay is given by

$$\mathcal{R}^{C-NC}(\mathbf{D}) = \{\mathbf{R} \mid \exists\, U^2, \widehat{X}^3, g_1(\cdot,\cdot), s.t.$$
$$R_1 \geq I(X_1; U_1),$$
$$R_2 \geq I(X^2; U_2|U_1),$$
$$R_3 \geq I(X^3; \widehat{X}_2^3|U^2),$$
$$E[d_j(X_j,\widehat{X}_j)] \leq D_j, \quad j = 1,2,3,$$
$$\widehat{X}_1 = g_1(U_1, U_2),$$
$$U_1 - X_1 - X_2^3, \ U_2 - (X^2, U_1) - X_3\}$$

where $\{U^2, \widehat{X}^3\}$ are auxiliary random variables and $g_1(\cdot, \cdot)$ is a deterministic function.

The proof of both the achievability and the converse parts of Theorem 3 are similar to that of Theorem 1. Note that $\mathcal{R}^{C-C} \subseteq \mathcal{R}^{C-NC}$ because the encoders and decoders of a C–C system are also permitted in a C–NC system.

Corollary 3.1 *(C–NC sum-rate region)* The sum-rate region for the one-step delayed C–NC system is $\mathcal{R}_{sum}^{C-NC}(\mathbf{D}) = [R_{sum}^{C-NC}(\mathbf{D}), \infty)$ where the minimum sum-rate is

$$R_{sum}^{C-NC}(\mathbf{D}) = \min_{\substack{E[d_j(X_j,\widehat{X}_j)] \leq D_j, j=1,2,3, \\ \widehat{X}_1 - X^2 - X_3}} I(X^3; \widehat{X}^3).$$

The proof is similar to that of Corollary 1.1. As noted earlier, the JC rate-distortion function (4.4) having no Markov

chain constraints is a lower bound on $R_{sum}^{C-NC}(\mathbf{D})$. Remarkably, for Gauss–Markov sources and certain nontrivial MSE tuples $\mathbf{D}$ discussed below, $R_{sum}^{C-NC}(\mathbf{D})$ coincides with the JC rate $R^{JC}(\mathbf{D})$.

Corollary 3.2 *(JC-optimality of one-step delayed C–NC systems for Gauss–Markov sources and MSE)* For all distortion tuples $\mathbf{D}$ belonging to the distortion region $\mathcal{D}^{JC}$ defined in Section IV, Corollary 2.1(i), we have

$$R_{sum}^{CNCGM}(\mathbf{D}) = R^{JCGM}(\mathbf{D}).$$

The proof is sketched in Appendix E.

Corollary 3.2 implies that the JC rate-distortion performance is achievable in terms of sum-rate by only a single frame decoding delay for Gauss–Markov sources and the MSE region $\mathcal{D}^{JC}$. The benefit of one frame delay is so significant that it is equivalent to arbitrary frame delay in this specific situation. The first-order Markov assumption on sources $X_1 - X_2 - X_3$ is essential for this optimality. An interpretation is that $\mathbf{X}_2$ supplies all the help from $\mathbf{X}_3$ to generate the optimum $\widehat{\mathbf{X}}_1$. More generally (for $T > 3$), C–NC encoders need access to only the present and past frames together with *one* future frame to match the rate-distortion function of the JC system in which *all* future frames are simultaneously available for encoding. Thus, the neighboring future frame supplies all the help from the entire future through the Markovian property of sources.

It is of interest to compare Corollary 3.2 with the real-time source coding problem in [16]. In [16] it is shown that for Markov sources the C–C encoder may ignore the previous sources and only use the current source and decoder's memory without loss of performance. This is a purely structural result (no spatial asymptotics and computable single-letter information-theoretic characterizations) exclusively focused on C–C systems. In contrast, Corollary 3.2 is about achieving the JC-system performance with a C–NC system. Additionally, [16] deals with a frame-averaged expected distortion criterion as opposed to frame-specific individual distortion constraints treated here.

The JC-optimality of one-step delayed C–NC systems holds within the distortion region $\mathcal{D}^{JC}$ defined as the set of all distortion tuples $\mathbf{D}$ satisfying the positive semidefiniteness condition $(\Sigma_X - \mathrm{diag}(\mathbf{D})) \geq 0$. For nonsingular sources $\Sigma_X > 0 \Rightarrow \lambda_{\min}(\Sigma_X) > 0$ where $\lambda_{\min}(\Sigma_X)$ is the smallest eigenvalue of the positive definite symmetric (covariance) matrix Σ_X which is strictly positive. Thus $\mathcal{D}^{JC}$ contains the closed hypercube $[0, \lambda_{\min}(\Sigma_X)]^T$ which has a strictly positive volume in $\mathbb{R}^T \Rightarrow \mathcal{D}^{JC}$ has a *non-zero volume*. Hence, the JC-optimality of a C-NC system with one-step decoding delay discussed here is a nontrivial assertion. $\mathcal{D}^{JC}$ includes all distortion tuples with components below certain thresholds corresponding to "sufficiently good" reconstruction qualities. However, it should be noted that this is *not* a high-rate (zero-distortion) asymptotic (a la Slepian-Wolf).

On the contrary, the JC-optimality of a C–NC system with one-step decoding frame delay does not hold for all distortion tuples:

Counter example: Consider Gauss–Markov sources X^3 where $X_1 = X_2$ and MSE tuple $\mathbf{D}$ where $D_1 = D_2 = D$. The JC problem reduces to a *two-stage JC problem* where the encoder jointly quantizes $(\mathbf{X}_1, \mathbf{X}_3)$ into $(\widehat{\mathbf{X}}_1, \widehat{\mathbf{X}}_3)$ and the decoder simply sets $\widehat{\mathbf{X}}_2 = \widehat{\mathbf{X}}_1$. However, the C–NC problem reduces to a *two-stage C–C problem* with sources $(\mathbf{X}_1, \mathbf{X}_3)$ because the first two C–NC encoders are operationally equivalent to the first C–C encoder observing $\mathbf{X}_1$ and the last C–NC encoder is operationally equivalent to the second C–C encoder observing all sources. As mentioned in Section IV, generally speaking, a two-stage C–C system does not match (in sum-rate) the JC-system rate-distortion performance. Therefore the three-stage C–NC system also does not match the JC performance for these specific sources and certain distortion tuples $\mathbf{D}$. Note that these sources are actually singular (Σ_X has a zero eigenvalue) and $\mathcal{D}^{JC}$ only contains trivial points (either $D = 0$ or $D_3 = 0$). So for the nontrivial distortion tuples $\mathbf{D}$ described above (which do not belong to $\mathcal{D}^{JC}$), the JC-optimality of a C–NC system with a one-step decoding delay fails to hold.

To construct a counter example with nonsingular sources, one can slightly perturb Σ_X such that it becomes positive definite. However, the JC rate and C–NC sum-rate only change by limited amounts due to continuity properties of the sum-rate-distortion function with respect to the source distributions (similar to [15, Lemma 2.2, p.124]). Therefore we can find a small enough perturbation such that the rates do not match.

For general C–NC systems with increasing system frame-delays[7], the expressions of the minimum sum-rates contain the same objective function $I(X_1^T; \widehat{X}_1^T)$ and distortion constraints $E[d_j(X_j, \widehat{X}_j)] \leq D_j, j = 1, \ldots, T$ but with a decreasing number of Markov chain constraints. In the limit of maximum possible system frame-delay, we arrive at the JC system with purely distortion (no Markov chain) constraints. For Gauss–Markov sources and MSE, the distortion region for which a C–NC system matches (in sum-rate) the rate-distortion performance of the JC-system, expands with increasing delays until it completely coincides with the set of all reachable tuples of the JC system.

VI. Results for NC–C systems

We can derive the rate-distortion region for NC–C systems by mimicking the derivations for C–NC systems discussed till this point. However, due to the operational structural relationship between C–NC and NC–C systems, it is not necessary to re-derive the results for the NC–C system at certain operating points, in particular, for the sum-rate region:

Theorem 4 *("Equivalence" between C–NC and NC–C rate-distortion regions)*
(i) The rate-distortion region for the one-step delayed NC–C

---

[7]The general result is presented and discussed in [13].

system is given by

$$
\begin{aligned}
\mathcal{R}^{NC-C}(\mathbf{D}) \quad = \quad &\{\mathbf{R} \mid \exists\, U^2, \widehat{X}^3, g_1(\cdot), g_2(\cdot, \cdot),\ s.t. \\
&R_1 \geq I(X^2; U_1), \\
&R_2 \geq I(X^3; U_2|U_1), \\
&R_3 \geq I(X^3; \widehat{X}_3|U^2), \\
&E[d_j(X_j, \widehat{X}_j)] \leq D_j, \quad j = 1, 2, 3, \\
&\widehat{X}_1 = g_1(U_1), \widehat{X}_2 = g_2(U_1, U_2), \\
&U_1 - X^2 - X^3 \}.
\end{aligned}
$$

(ii) For an arbitrary distortion tuple $\mathbf{D}$, the rate regions $\mathcal{R}^{NC-C}(\mathbf{D})$ and $\mathcal{R}^{C-NC}(\mathbf{D})$ are related in the following manner:

$$
\begin{aligned}
(R_1, R_2, R_3) \in \mathcal{R}^{C-NC}(\mathbf{D}) \quad &\Rightarrow \quad (R_1 + R_2, R_3, 0) \in \mathcal{R}^{NC-C}(\mathbf{D}), \\
(R_1, R_2, R_3) \in \mathcal{R}^{NC-C}(\mathbf{D}) \quad &\Rightarrow \quad (0, R_1, R_2 + R_3) \in \mathcal{R}^{C-NC}(\mathbf{D}).
\end{aligned}
$$

(iii) For an arbitrary distortion tuple $\mathbf{D}$, the minimum sum-rates of one-step delayed C–NC and NC-C systems are equal:

$$
R_{sum}^{C-NC}(\mathbf{D}) = R_{sum}^{NC-C}(\mathbf{D}).
$$

The proof of part (i) is similar to that of Theorem 1. Part (ii) can be proved by either using the definitions of $\mathcal{R}^{C-NC}(\mathbf{D})$ and $\mathcal{R}^{NC-C}(\mathbf{D})$ or more directly from the system structure: the first NC–C encoder is operationally replaceable with the combination of the first two C–NC encoders, and the last C–NC encoder is operationally replaceable with the combination of the last two NC–C encoders (see Figs 3(a) and (b)). Part (iii) follows from part (ii).

In conclusion, the two rate regions are equivalent modulo boundary effects with exact equivalence for sum-rates. The (sum-rate) JC-optimality property of a C–NC system with one-step decoding frame delay given by Corollary 3.2 automatically holds for the NC–C systems with one-step encoding frame delay. This relationship allows one to focus on the performance of only C–NC systems instead of both C–NC and NC–C systems without loss of generality.

This structural principle holds for the general multi-frame problem with multi-step frame delay. Whenever two delayed sequential coding systems have the same *sum* of the encoding plus decoding frame delay, they share the same sum-rate-distortion performance.

VII. CONCLUDING REMARKS

The main message of this study is that even a single frame delay holds potential for yielding huge performance improvements in sequential coding problems, sometimes even matching the joint coding performance. As remarked earlier, the results of this paper hold for an arbitrary number of sources and certain types of coupled fidelity criteria. Our ongoing work includes extensions of Corollaries 1.3 and 3.2 to more general k-th order Gauss–Markov sources and general quadratic criteria [13]. We also expect similar results for other "matched" pairs of source distributions and distortion criteria, example: symmetrically correlated binary sources with Hamming distortion. Finally, it would be of interest to explore the effects of nonergodic "packet erasures", as in the multiple-descriptions coding problem, in future work.

APPENDIX A
SKETCH OF THEOREM 1 CONVERSE PROOF

Denote the messages sent by the three ($T = 3$) encoders respectively by S_1, S_2, and S_3, and define the auxiliary random variables by $U_j(i) = \{S_j, X_j^{(i-1)}\}, j = 1, 2$. For any operationally admissible scheme and $\forall \epsilon > 0$ we have

$$
E[d_j(\mathbf{X}_j, \widehat{\mathbf{X}}_j)] \leq D_j + \epsilon, \quad j = 1, \ldots, T.
$$

For the first coding rate, we have

$$
n(R_1 + \epsilon) \geq H(S_1) = I(S_1; X_1^n) = \sum_{i=1}^{n} I(X_1(i); U_1(i))
$$

In the next stage,

$$
n(R_2 + \epsilon) \geq I(S_2; X_1^n, X_2^n|S_1) = \sum_{i=1}^{n} I(X_1(i), X_2(i); U_2(i)|U_1(i)).
$$

where the last equality is because of the Markov chain $(X_1(i), X_2(i)) - U_1(i) - X_2^{(i-1)}$.

In the final stage,

$$
\begin{aligned}
n(R_3 + \epsilon) \quad &\geq \quad I(S_3; X_1^n, X_2^n, X_3^n|S^2) \\
&\overset{(a)}{=} \quad \sum_{i=1}^{n} I(X^3(i); S_3, X_3^-(i)|U^2(i)) \\
&\geq \quad \sum_{i=1}^{n} I(X^3(i); \hat{X}_3(i)|U^2(i))
\end{aligned}
$$

where step (a) is because of the Markov chain $X^3(i) - U^2(i) - X_3^{(i-1)}$.

The proof can be completed using a timesharing argument as in [14, p.397]. In addition, the cardinality bounds of auxiliary alphabets can be derived using Carathéodory theorem and support lemma [15] as in [6].

APPENDIX B
COROLLARY 1.1 PROOF-SKETCH

The C–C minimum sum-rate directly deduced from the rate region $\mathcal{R}^{C-C}(\mathbf{D})$ is

$$
\begin{aligned}
R_{sum}^{C-C}(\mathbf{D}) \quad = \quad &\min I(X_1; U_1) + I(X^2; U_2|U_1) + I(X^3; \widehat{X}_3|U^2) \\
= \quad &\min I(X^3; U^2, \widehat{X}_3)
\end{aligned}
$$

with the auxiliary random variables and functions satisfying the distortion and Markov chain constraints in the definition of $\mathcal{R}^{C-C}(\mathbf{D})$. It simplifies to the expression in the corollary because

(i) $\{\widehat{X}_1, \widehat{X}_2\}$ are determined by $\{U_1, U_2\}$;
(ii) $\{U_1 = \widehat{X}_1, U_2 = \widehat{X}_2\}$ is a possible choice of $\{U_1, U_2\}$.

APPENDIX C
LEMMA PROOF-SKETCH

Given any reproduction random vector $\widehat{\mathbf{X}} = (\widehat{X}_1, \ldots, \widehat{X}_T)$ satisfying the MSE and Markov chain constraints, we can construct a new random vector $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \ldots, \widetilde{X}_T)$ which is *jointly Gaussian* with $\mathbf{X} = (X_1, \ldots, X_T)$ with the same second-order statistics. Specifically, $cov(\widehat{\mathbf{X}}) = cov(\widetilde{\mathbf{X}})$ and $cov(\mathbf{X}, \widehat{\mathbf{X}}) = cov(\mathbf{X}, \widetilde{\mathbf{X}})$. Since MSEs are fully determined from second-order statistics, $\widetilde{\mathbf{X}}$ automatically satisfies the MSE constraints. The Markov chain relations of $\widehat{\mathbf{X}}$ imply corresponding conditional uncorrelatedness relations, which also hold for $\widetilde{\mathbf{X}}$. Moreover, because $\widetilde{\mathbf{X}}$ is jointly Gaussian, conditional uncorrelatedness is equivalent to conditional independence. Therefore $\widetilde{\mathbf{X}}$ also satisfies the corresponding Markov chain constraints.

Let the linear MMSE estimate of $\mathbf{X}$ based on $\widehat{\mathbf{X}}$ be given by $A\widehat{\mathbf{X}}$ where A is a matrix. Note that by the orthogonality principle and the joint Gaussianity of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ we have $(\mathbf{X} - A\widehat{\mathbf{X}}) \perp \widehat{\mathbf{X}}$, and further $(\mathbf{X} - A\widetilde{\mathbf{X}}) \perp\!\!\!\perp \widetilde{\mathbf{X}}$. Therefore,

$$
\begin{aligned}
I(\mathbf{X}; \widehat{\mathbf{X}}) &= h(\mathbf{X}) - h(\mathbf{X} - A\widehat{\mathbf{X}}|\widehat{\mathbf{X}}) \\
&\geq h(\mathbf{X}) - h(\mathbf{X} - A\widehat{\mathbf{X}}) \\
&\overset{(b)}{\geq} h(\mathbf{X}) - h(\mathbf{X} - A\widetilde{\mathbf{X}}) \\
&\overset{(c)}{=} h(\mathbf{X}) - h(\mathbf{X} - A\widetilde{\mathbf{X}}|\widetilde{\mathbf{X}}) \\
&= I(\mathbf{X}; \widetilde{\mathbf{X}}).
\end{aligned}
$$

Step (b) is because $(\mathbf{X} - A\widetilde{\mathbf{X}})$ has the same second-order statistics as $(\mathbf{X} - A\widehat{\mathbf{X}})$ and it is a *jointly Gaussian* random vector. Step (c) is because $(\mathbf{X} - A\widetilde{\mathbf{X}})$ is independent of $\widetilde{\mathbf{X}}$.

In conclusion, given an *arbitrary* reproduction vector, we can construct a *Gaussian* random vector $\widetilde{\mathbf{X}}$ satisfying the MSE and Markov chain constraints and $I(\mathbf{X}; \widehat{\mathbf{X}}) \geq I(\mathbf{X}; \widetilde{\mathbf{X}})$. Hence the minimum of the optimization problem can be achieved by a reproduction random vector which is jointly Gaussian with $\mathbf{X}$.

APPENDIX D
SKETCH OF COROLLARY 1.2 CONVERSE PROOF

For any choice of reproduction random variables, we have

$$
\begin{aligned}
R_{sum}^{CCGM}(\mathbf{D}) &\geq \frac{1}{2}\log\frac{\sigma_1^2}{D_3} + \min\left\{h(X_2|\widehat{X}_1) - h(X_1|\widehat{X}_1)\right\} \\
&\quad + \min\{h(X_3|\widehat{X^2}) - h(X_2|\widehat{X^2})\}
\end{aligned}
$$

where the minimization is subject to the distortion and Markov chain constraints in the definition of $R_{sum}^{C-C}(\mathbf{D})$. Using [6, Lemma 5], we have the inequality

$$
\min\left\{h(X_2|\widehat{X}_1) - h(X_1|\widehat{X}_1)\right\} \geq \frac{1}{2}\log\left(\frac{\sigma_{e_1}^2}{D_1}\right).
$$

One can also similarly establish the inequality

$$
\min\left\{h(X_3|\widehat{X^2}) - h(X_2|\widehat{X^2})\right\} \geq \frac{1}{2}\log\left(\frac{\sigma_{e_2}^2}{D_2}\right).
$$

The proof follows from these inequalities.

APPENDIX E
COROLLARY 3.2 PROOF-SKETCH

The JC rate-distortion function is achieved by the test channel (4.7) for the distortion region $\mathcal{D}^{JC}$. We need to verify that the Markov chain $\widehat{X}_1 - X^2 - X_3$ holds for this test channel.

Note that because all the variables are jointly Gaussian, they have the property that $A \perp\!\!\!\perp B$ and $A \perp\!\!\!\perp C$ implies $A \perp\!\!\!\perp \{B, C\}$ for any Gaussian vector (A, B, C).

By the Markov chain $X_1 - X_2 - X_3$, the MMSE estimate of X_3 based on X_1 and X_2 is

$$
X_3 = \rho_2 \frac{\sigma_3}{\sigma_2} X_2 + N \tag{E.1}
$$

where N is Gaussian and independent of $\{X_1, X_2\}$.

By the structure of the test channel, $Z_1 \perp\!\!\!\perp \{Z_2, Z_3, \widehat{X}_2, \widehat{X}_3\}$ implies $Z_1 \perp\!\!\!\perp \{X_2, X_3\}$, which further implies $Z_1 \perp\!\!\!\perp N$. Moreover, because $N \perp\!\!\!\perp \{X_1, Z_1\}$, we have $N \perp\!\!\!\perp \widehat{X}_1$. Therefore $N \perp\!\!\!\perp \{X_1, X_2, \widehat{X}_1\}$. So the best estimate of X_3 based on $\{X_1, X_2, \widehat{X}_1\}$ is still formula (E.1). It follows that the Markov chain $X_3 - X_2 - (X_1, \widehat{X}_1)$ holds implying $\widehat{X}_1 - X^2 - X_3$ and completing the proof.

REFERENCES

[1] N. Farvardin and J. W. Modestino, "Rate-distortion performance of DPCM schemes for autoregressive sources," *IEEE Trans. Info. Theory*, vol. IT-31, pp. 402–418, May 1985.

[2] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, NJ: Prentice–Hall, 1971.

[3] T. Berger, J. Gibson, "Lossy Source Coding," *IEEE Trans. Info. Theory*, vol. IT-16, no. 6, pp. 2693–2723, Oct. 1998.

[4] P. Ishwar and K. Ramchandran, "On decoder-latency versus performance tradeoffs in differential predictive coding," *Proc. IEEE International Conference on Image Processing (ICIP)*, Singapore, volume 2, pp. 1097–1100, Oct. 2004.

[5] R. Zamir, Y. Kochman and U. Erez, "Achieving the Gaussian rate-distortion function by prediction," *Proc. IEEE Inttl. Symp. Info. Theory (ISIT)*, Seattle, Jul. 2006.

[6] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *IEEE Trans. Info. Theory*, vol. IT–46, pp. 236–246, Jan. 2000.

[7] S. Tatikonda, "Control Under Communication Constraints," Ph.D. dissertation, MIT, 2000.

[8] V. Borkar, S. Mitter, and S. Tatikonda, "Optimal Sequential Vector Quantization of Markov Sources," *SIAM Journal on Control and Optimization*, vol. 40, no. 1, pp. 135–148, Jan. 2001.

[9] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Tran. Info. Theory*, vol. IT–44, pp. 55–67, May 1998.

[10] S. C. Draper and G. W. Wornell, "Successively Structured CEO Problems," in *Proc. Inttl. Symp. Info. Theory (ISIT)*, Lausanne, Switzerland, Jul. 2002.

[11] S. C. Draper and G. W. Wornell, "Side Information Aware Coding Strategies for Sensor Networks," *J. Select. Areas Commun. (JSAC)*, vol. 22, no. 6, pp. 966–976, Aug. 2004.

[12] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," *Proc. IEEE Inttl. Symp. Info. Theory (ISIT)*, Chicago, IL, USA, 27 Jun. – 2 Jul. 2004, page 119.

[13] N. Ma and P. Ishwar, "The value of frame-delays in the sequential coding of correlated sources," submitted to *IEEE Inttl. Symp. Info. Theory (ISIT)*, Nice, France, Jun. 2007.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[15] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Budapest, Hungary: Akadémiai Kiadó, 1986.

[16] H. S. Witsenhausen, "On the structure of real-time source coders," *Bell Syst. Tech. Jour. (BSTJ)*, vol. 58, no. 6, pp. 1437–1451, 1979.