DISTRIBUTED TRANSFORM CODING VIA SOURCE-SPLITTING

Pradeepa Yahampath

Department of Electrical and Computer Engineering
University of Manitoba, Winnipeg, Canada
email: pradeepa@ee.umanitoba.ca

ABSTRACT

A practical approach to designing distributed transform codes for high-dimensional correlated Gaussian vectors is presented. In this approach, source-splitting based on linear approximations is used to achieve arbitrary rate-pairs, by using only Wyner-Ziv (WZ) quantizers. The optimal bit-allocation among a dependent set of WZ quantizers is found by using a tree-search algorithm. Experimental results obtained with actual designs, which use conditional entropy constrained trellis coded quantizers (CEC-TCQ) and Slepian-Wolf (SW) codes, are presented.

***Index Terms*—** Distributed transform codes, Wyner-Ziv quantization, Karhunen-Loéve transform, bit allocation.

1. INTRODUCTION

Many new applications rely on networks of distributed wireless sensors to acquire information in the form of high-dimensional vectors (e.g., multi-camera imaging systems and microphone arrays). In such applications, an encoder in each sensor quantizes a vector of observation variables (without exchanging any information with other sensors) and transmits its output to a central processor. A special case of this more general distributed *vector quantization* (VQ) problem is Wyner-Ziv (WZ) quantization, in which a single source is quantized given that the decoder has access to side information about the source, see [1] and references therein.

The design of a distributed VQ for a large number of source variables is in general a difficult task. A practically simpler, yet very effective approach to quantization of a large number of correlated variables by using a bank of single variable quantizers is transform coding (TC) [2]. In TC, the correlation among the variables in a vector is exploited by applying a "de-correlating" linear transform to the vector prior to quantization. Recently, distributed TC (DTC) has been considered in both WZ setting [3–5], as well as in the multi-terminal setting [6]. In [6], the optimal performance theoretically attainable (OPTA) in the Gaussian WZ-TC problem under minimum mean square error (MMSE) quantization is derived. In particular, it is shown that the optimal transform is the conditional Karhunen-Loéve transform (CKLT). The optimal solution to the more general multi-terminal WZ-TC problem remains unsolved, even for the Gaussian case. An iterative descent algorithm which alternately optimizes the WZ-TC of one terminal, while keeping the other terminals fixed, is used in [6] for obtaining a locally optimal solution to the Gaussian multi-terminal problem. The resulting transform is refereed to as the distributed KLT (DKLT). However, the implementation of quantizers implied by the DKLT is not tractable, since it is not practical to actually design a set of near-optimal WZ quantizers in each iteration of the aforementioned algorithm.

In this paper, we present a practical approach to designing a DTC with arbitrary bit-rates for Gaussian two-terminal case. The main idea is to use *source-splitting* based on optimal linear approximations to convert the two-terminal DTC problem into two WZ-TC problems. For each WZ-TC, the optimal transform as well as the bit allocations can be obtained in closed-form, based on which practical WZ quantizers approaching the optimal performance can be designed. However, source-splitting requires optimally allocating bits among a dependent set of WZ quantizers. A low-complexity tree-search algorithm is proposed for solving this problem. We present experimental results obtained with a practical implementation of the proposed source-split DKLT (SP-DKLT) code for a 16 dimensional Gaussian source pair, which uses SW-coded CEC-TCQ to realize block WZ quantizers. These results demonstrate that SP-DKLT can significantly outperform non-distributed TCs.

2. SP-DKLT CODES AND OPTIMIZATION

A block diagram of the proposed SP-DKLT system is shown in Fig. 1. Let the total number of bits available for encoding two jointly Gaussian vectors $\mathbf{X}_1 \in \mathbb{R}^{M_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{M_2}$ be B bits. The terminal 1 provides decoder side-information for WZ coding of the terminal 2 at the rate B_1' $(< B)$ bits/vector. In the transform coding framework, the goal is to provide the best (in MMSE sense) linear approximation of $\mathbf{X}_1$ using B_1' bits/vector as the decoder side-information. Given the quantized version $\mathbf{Y}_1' \in \mathbb{R}^{N_1'}$ of the transform coefficient vector $\mathbf{U}_1' = \mathbf{T}_1'^{T}\mathbf{X}_1 \in \mathbb{R}^{M_1}$ at the decoder, where $N_1' \leq M_1$, the

This work has been supported by a NSERC Discovery Grant.

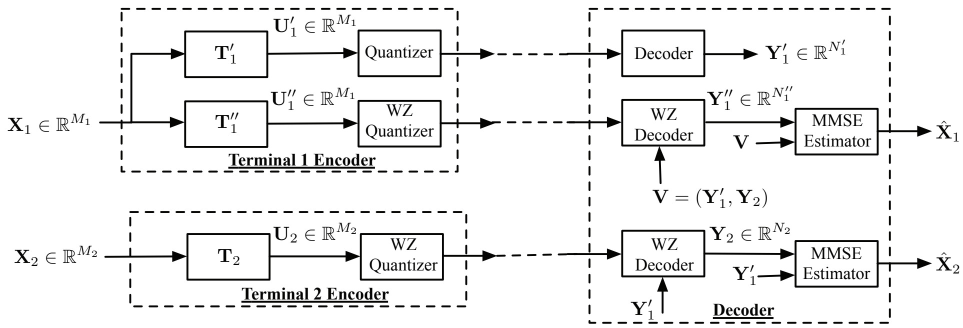

Fig. 1. The proposed source-split transform coding system for distributed quantization of correlated vectors.

terminal 2 then quantizes the best linear approximation of $\mathbf{X}_2$ using B_2 ($< B - B_1'$) bits/vector. Finally, given the linear approximations of both $\mathbf{X}_1$ and $\mathbf{X}_2$ available at the decoder, the terminal 1 WZ quantizes the best linear approximation using B_1'' ($= B - B_1' - B_2$) bits/vector. The total transmission rate for source $\mathbf{X}_1$ is thus $B_1 = B_1' + B_1''$ bits/vector. Let the quantized version of the $\mathbf{U}_2 = \mathbf{T}_2^T \mathbf{X}_2 \in \mathbb{R}^{M_2}$ be $\mathbf{Y}_2 \in \mathbb{R}^{N_2}$, $N_2 \leq M_2$ and that of $\mathbf{U}_1'' = \mathbf{T}_1''^T \mathbf{X}_1 \in \mathbb{R}^{M_1}$ be $\mathbf{Y}_1'' \in \mathbb{R}^{N_1''}$, $N_1'' \leq M_1$. Also let $\mathbf{V} = (\mathbf{Y}_1', \mathbf{Y}_2)^T$. Given a total of B bits for encoding both $\mathbf{X}_1$ and $\mathbf{X}_2$, the design of this system involves determining the optimal transforms $\mathbf{T}_1'$, $\mathbf{T}_1''$, and $\mathbf{T}_2''$, and the corresponding bit allocation among the transform coefficients $\mathbf{U} = (\mathbf{U}_1', \mathbf{U}_1'', \mathbf{U}_2)^T$ such that the total MSE

$$D = E\{\|\mathbf{X}_1 - \hat{\mathbf{X}}_1\|^2 + \|\mathbf{X}_2 - \hat{\mathbf{X}}_2\|^2\} \tag{1}$$

is minimized, where $\hat{\mathbf{X}}_1 = E\{\mathbf{X}_1 | \mathbf{V}, \mathbf{Y}_1''\}$ and $\hat{\mathbf{X}}_2 = E\{\mathbf{X}_2 | \mathbf{V}\}$. Let $\mathbf{U}_1' = (U_{1,1}', \ldots, U_{1,M_1}')^T$, $\mathbf{U}_1'' = (U_{1,1}'', \ldots, U_{1,M_1}'')^T$ and $\mathbf{U}_2 = (U_{2,1}, \ldots, U_{2,M_2})^T$. Let the bit-rates allocated to quantizing these transform coefficients be $\mathbf{r}_1' = (r_{1,1}', \ldots, r_{1,M_1}')^T$, $\mathbf{r}_1'' = (r_{1,1}'', \ldots, r_{1,M_1}'')^T$, and $\mathbf{r}_2 = (r_{2,1}, \ldots, r_{2,M_2})^T$ respectively. Define $\mathbf{r} = (\mathbf{r}_1', \mathbf{r}_1'', \mathbf{r}_2)^T$. The bit allocation problem can now be stated as follows: given a total bit-budget of B bits, minimize $D(\mathbf{r})$ subject to

$$\begin{aligned} \sum_{m=1}^{2M_1+M_2} r_m &\leq B, \\ r_m &\geq 0,\ m = 1, \ldots, 2M_1 + M_2, \end{aligned} \tag{2}$$

where $\mathbf{r} = (r_1, \ldots, r_{2M_1+M_2})^T$. The explicit solution of this problem is unfortunately intractable due to the interdependence of the three transform codes involved. However, an explicit solution can be found to a variant of this problem in which B_1', B_1'', and B_2 are fixed. This constraint converts the *main bit-allocation problem* in (2) into three sub-problems which only require the optimal bit allocation among the coefficients within a transform code. We refer to this as the *constrained bit-allocation problem*. In the following, we first derive an explicit solution to this problem based on the rate-distortion optimal WZ quantization model of [6] for compressing the transform coefficients. Based on this result, we present a tree-search algorithm to obtain a solution to the main bit-allocation problem.

Theorem 1: Suppose $\mathbf{X} \in \mathbb{R}^M$ and $\mathbf{Y} \in \mathbb{R}^{M_2}$ are mean-zero and jointly Gaussian vectors. Let the *conditional* covariance matrix of $\mathbf{X}$ given $\mathbf{Y}$ be *diagonal* and have diagonal elements $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)^T$, where $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_M$. Then, given a total bit budget of B bits, the optimal (in MMSE sense) bit allocation $(\rho_1, \ldots, \rho_M)^T$ for separate R-D optimal WZ quantization of each component of $\mathbf{X}$, given the decoder side-information $\mathbf{Y}$, is given by

$$\rho_m(\boldsymbol{\lambda}, B, N) = \begin{cases} \frac{1}{2}\log_2\left[\frac{\lambda_m}{d^*(\boldsymbol{\lambda}, B, N)}\right] & m = 1, \ldots, N \\ 0 & m = N+1, \ldots, M, \end{cases}$$

where $N < M$ is the largest integer for which $\lambda_m \geq d^*(\boldsymbol{\lambda}, B, N) \triangleq \left(\prod_{i=1}^{N} \lambda_i\right)^{\frac{1}{N}} 2^{-\frac{2B}{N}}$, $m = 1, \ldots, N$. The resulting minimum quantization MSE is

$$J(\boldsymbol{\lambda}, B) = N d^*(\boldsymbol{\lambda}, B, N) + \sum_{m=N+1}^{M} \lambda_m. \tag{3}$$

Proof: Directly follows from [6, Theorem 3].

2.1. Solution to the constrained bit-allocation problem

First consider the non-distributed transform code for $\mathbf{X}_1$ based on $\mathbf{T}_1'$. It is well known that the MMSE optimal $\mathbf{T}_1'$ is the KLT of $\mathbf{X}_1$. Let the eigenvalues of the covariance matrix of $\mathbf{X}_1$ be $\boldsymbol{\lambda}_1' = (\lambda_{1,1}', \ldots, \lambda_{1,M_1}')^T$. Then, it directly follows from *Theorem 1*, that the optimal bit allocation among $U_{1,1}', \ldots, U_{1,M_1}'$ is given by $r_{1,m}' = \rho_m(\boldsymbol{\lambda}_1', B_1', N_1')$ bits/sample, $m = 1, \ldots, M_1'$. With RD-optimal quantization of Gaussian variables, we can represent the quantized value of $\mathbf{U}_1' = K_1'^T \mathbf{X}_1$ (up to a scaling factor) by [6, (29)].

$$\mathbf{Y}_1' = \mathbf{K}_1'^{(N_1')T} \mathbf{X}_1 + \mathbf{Z}_1', \tag{4}$$

where $\mathbf{K}_1'^{(N_1')}$ denotes the $M_1 \times N_1'$ matrix consisting of first N_1' columns of $\mathbf{K}_1'$ and the quantization noise $\mathbf{Z}_1' \in \mathbb{R}^{N_1'}$

is a mean zero iid Gaussian vector independent of $\mathbf{X}_1$. The variance of $Z'_{1,m}$, $m = 1, \ldots, N'_1$ is [6, (30)]

$$E\{Z^2_{1,m}\} = \lambda'_{1,m} d^* \left(\boldsymbol{\lambda}'_1, B'_1, N'_1\right) / \left[\lambda'_{1,m} - d^* \left(\boldsymbol{\lambda}'_1, B'_1, N'_1\right)\right].$$

Next consider WZ transform coding of $\mathbf{X}_2$ given $\mathbf{Y}'_1$ at the decoder. Invoking [6, Theorem 3], we find that $\mathbf{T}_2$ which minimizes $E\|\mathbf{X}_2 - \hat{\mathbf{X}}_2\|^2$ is the CKLT computed from the contrivance matrix of $\mathbf{X}_2$ given $\mathbf{Y}'_1$, $\boldsymbol{\Sigma}_{X_2|Y'_1}$ [which can be determined from (4)]. According to *Theorem 1*, the bit allocation which achieves the MMSE is $r_{2,m} = \rho_m(\boldsymbol{\lambda}_2, M_2, N_2)$, $m = 1, \ldots, M_2$, where $\boldsymbol{\lambda}_2$ is the vector of eigenvalues of $\boldsymbol{\Sigma}_{X_2|Y'_1}$ and the resulting MSE is given by [see (3)]

$$E\|\mathbf{X}_2 - \hat{\mathbf{X}}_2\|^2 = N_2 d^*_2(\boldsymbol{\lambda}_2, B_2, N_2) + \sum_{m=N_2+1}^{M_2} \lambda_{2,m}. \quad (5)$$

As before, let the quantized value of the transform coefficients be given by

$$\mathbf{Y}_2 = \mathbf{K}_2^{(N_2)T}\mathbf{X}_2 + \mathbf{Z}_2, \quad (6)$$

where $\mathbf{Z}_2$ is a zero-mean iid Gaussian vector independent of $\mathbf{X}_2$. Finally, $\mathbf{T}''_1$ which minimizes $E\|\mathbf{X}_1 - \hat{\mathbf{X}}_1\|^2$ is the CKLT computed from the covariance matrix $\boldsymbol{\Sigma}_{X_1|V}$ of $\mathbf{X}_1$ given $\mathbf{V}$, the corresponding bit allocations is $r''_{1,m} = \rho_m\left(\boldsymbol{\lambda}''_1, B''_1, N''_1\right)$, $m = 1, \ldots, M_1$, where $\boldsymbol{\lambda}''_1$ is the vector of eigenvalues of $\boldsymbol{\Sigma}_{X_1|V}$, and the resulting MSE is

$$E\|\mathbf{X}_1 - \hat{\mathbf{X}}_1\|^2 = N''_1 d^*\left(\boldsymbol{\lambda}''_1, B''_1, N''_1\right) + \sum_{m=N''_1+1}^{M_1} \lambda''_{1,m}. \quad (7)$$

2.2. Solution to the main bit-allocation problem

The solution to the main bit-allocation problem defined in (2) corresponds to the minimum MSE solution of the constrained problem over the set $\mathcal{S} = \{(B'_1, B''_1, B_2) : B'_1 \in (0, B), B''_1 \in (0, B), B_2 \in (0, B), B'_1 + B''_1 + B_2 = B\}$. One approach to locating this minimum is to search over an appropriately discretized grid of points inside $\mathcal{S}$. The proposed search algorithm is a generalization of a class of bit-allocation algorithms in which a small fraction ΔB of the total bit-budget B is allocated to the "most deserving" quantizer among a set of quantizers in an incremental fashion, until the entire bit-budget is exhausted [2, Sec. 8.4]. However, this type of greedy search cannot guarantee that the final solution is overall optimal. In fact, locating the optimal solution (provided that it is on the grid) requires a tree-search, where each candidate solution corresponds to a path in the tree. Even though a full tree-search is intractable, a simple algorithm exists for detecting the minimum cost path in a tree with a high probability.

Let $0 < \Delta B << B$. If we are to assign ΔB bits to only one of the three transform codes $\mathbf{T}'_1$, $\mathbf{T}''_1$, or $\mathbf{T}_2$, then there are three possible choices for the rate-tuple (B'_1, B''_1, B_2), namely $(\Delta B, 0, 0)$, $(0, \Delta B, 0)$, and $(0, 0, \Delta B)$. For each of these choices, we can explicitly solve the constrained bit allocation problem and find the MMSE solution as described in Sec. 2.1. Each of these solutions corresponds to a node in the first level of nodes in the tree. Each node in this level corresponds to DTC of rate ΔB bits per source pair $(\mathbf{X}_1, \mathbf{X}_2)$, where the root node of the tree corresponds to a DTC of rate 0 bits. Now if we are to allocate ΔB more bits to any one of the DTCs in the first level of nodes, we end up with 3^2 possible bit allocation solutions, which constitute the second-level of nodes in the tree. Each node at this level corresponds to a DTC of $2\Delta B$ bits per source-pair. We can repeat this procedure, allocating ΔB bits to each of the nodes, to grow the tree to a depth of $L = B/\Delta B$ levels. This tree would have 3^L terminal nodes each of which corresponds to a DTC of rate B bits. The MMSE terminal node of this tree is the optimal bit-allocation solution to the main bit allocation problem, provided that the latter solution is on the grid. If ΔB is chosen small enough, then we can ensure that the optimal solution is nearly on the grid. In theory, the optimal solution can be found by searching the tree using quantization MSE of each node as the path-cost. In order to practically implement the tree-search, we use the (M, L)-*algorithm* [7, pp. 216], in which the parameter M can be chosen to reduce the complexity at the expense of decreased accuracy (i.e., the probability of detecting the lowest cost path in the tree). In our experiments, we used $\Delta B = 0.2$ and $M = 27$ ($M = 81$ yielded nearly the same result).

3. EXPERIMENTAL RESULTS AND DISCUSSION

For the purpose of testing SP-DKLT codes, we define the random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ to be observations picked-up by a pair of sensor arrays placed in a spatial Gaussian random field in which the correlation function decays with distance d according to the squared exponential model [8]. In this case, the elements of the auto-covariance matrix $\boldsymbol{\Sigma}_{X_1}$ of $\mathbf{X}_1$ are given by $[\boldsymbol{\Sigma}_{X_1}]_{ij} = \exp\{-(\alpha d_{ij})^2\}$, where $\alpha > 0$ is a constant and d_{ij} is the distance between the i-th and j-th components. The auto-covariance matrix $\boldsymbol{\Sigma}_{X_2}$ of $\mathbf{X}_2$ also has a similar form. For simplicity, assume that the sensors in each array are placed on a square grid of unit spacing, and that the two arrays are on parallel planes separated by a distance r. With this setup, the elements of the cross-covariance matrix $\boldsymbol{\Sigma}_{X_1X_2}$ is given by $[\boldsymbol{\Sigma}_{X_1X_2}]_{ij} = \theta\exp\{-(\alpha d_{ij})^2\}$, where $\theta = \exp\{-(\alpha r)^2\}$. This sensor structure ensures that $\boldsymbol{\Sigma}_{X_1X_2}$ can be chosen independently (by choosing array separation r) of $\boldsymbol{\Sigma}_{X_1}$ and $\boldsymbol{\Sigma}_{X_2}$.

Rate-distortion performance- We can compute the rate-pairs (R_1, R_2) achievable with SP-DKLT, where $R_1 = B_1/M_1$ bits/sample and $R_2 = B_2/M_2$ bits/sample, for a given a total MSE D, by fixing R_1 (or R_2) and then searching for minimum R_2 (or R_1) required to achieve the MSE D, by using the the tree-search algorithm presented in Sec. 2.2. The rate pairs achievable with SP-DKLT for $D = 0.005$ is compared with the OPTA bound predicted by the DKLT [6]

Table 1. SNR (in dB) of transform code designs for sensor arrays in a Gaussian random field ($M_1 = M_2 = 16$, $R_1 = R_2 = R$).

	R=0.5 bit/sample		R=1 bit/sample		R=1.5 bits/sample		R=2 bits/sample	
	Analytic.	Design	Analytic.	Design	Analytic.	Design	Analytic.	Design
IKLT	11.7	11.0	18.0	17.3	23.0	22.5	27.1	26.6
SP-DKLT	14.0	12.7	20.6	19.6	25.6	24.4	29.8	28.6

and the independent (non-distributed) KLT (IKLT) for source parameters $\alpha = 0.32$ (for which the highest value of the auto-correlation coefficient between two components in either $\mathbf{X}_1$ or $\mathbf{X}_2$ is 0.9) and several values of θ (which is the highest value of the cross-correlation coefficient between any two components of $\mathbf{X}_1$ and $\mathbf{X}_2$). Since α is fixed, auto-covariance matrices are the same in all cases. However, higher θ implies stronger inter-vector correlation and hence improved performance with distributed coding compared to independent coding. As one would expect, SP-DKLT performance coincides with OPTA (DKLT) lower bound when either R_1 or R_2 is high. The rate-distortion performance seen in Figs. 2 indicates that SP-DKLT codes can significantly outperform independent KLT codes at all rates when there is sufficient correlation between two distributed source. The gap between SP-DKLT performance and OPTA (DKLT) lower bound seen in Fig.2 is due to source splitting in terms of linear approximations, rather than in terms of optimal VQs.

Design Examples- Since our solution for optimal transforms and bit-allocations assume R-D optimal quantization, we have to use block WZ quantization in practice to approach the predicted performance. To this end, we employ TCQ followed SW coding [1] to design WZ quantizers. We first design a CEC-TCQ for each transform coefficient such that the output conditional entropy of the TCQ, given the decoder side-information is equal to the rate given by the optimal-bit allocation solution. In order to do this, we use the conditional expectation of each transform coefficient, given the decoder side-information vector as the scalar side-information to design the TCQ (it can be shown that this is a sufficient statistic under RD optimal WZ quantization of jointly Gaussian vectors). To design CEC-TCQs, we have used a modified version of the algorithm in [9] (we used a 8-state TCQ with a coding block length of 256). We have assumed ideal SW coding. In Table 1, the overall quantization SNR of practical SP-DKLT designs for 16-dimensional vectors is compared with that of IKLT designs (which use entropy-coded TCQ [9] based on bit allocations for independent coding). While the performance gain relative to IKLT depends on the source correlation model, these results demonstrate that SP-DKLT coding is a relatively simple approach to the practical design of high-dimensional VQ, which only involves the design of a set of WZ quantizers and SW codes. It would however be interesting to seek practical designs which are capable of more closely approaching the performance predicted by the DKLT (see Fig. 2).

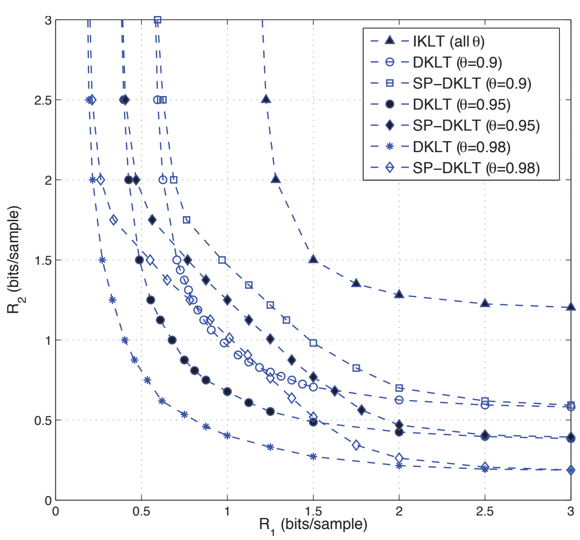

Fig. 2. Comparison of rate-regions achievable with transform coding of sensor arrays in a Gaussian random field.

4. REFERENCES

[1] Z. Xiong, A. D. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” *IEEE Signal Processing Mag.*, vol. 21, no. 5, pp. 80–94, Sept. 2004.

[2] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

[3] D Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, “High-rate quantization and transform coding with side information at the decoder,” *Signal Processing*, vol. 88, pp. 3160–3179, 2006.

[4] A. Vosoughi and A. Scaglione, “Precoding and decoding paradigms for distributed vector data compression,” *IEEE Trans. Signal Processing*, vol. 55, pp. 1445–1459, Apr. 2007.

[5] X. Chen and E. Tuncel, “Low-delay prediction and transform-based Wyner-Ziv coding,” *IEEE Trans. Signal Processing*, vol. 59, no. 2, pp. 653–666, Feb. 2011.

[6] M. Gastpar, P. L. Dragotti, and M. Vetterli, “The distributed Karhunen-Loéve transform,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 5177–5196, Dec. 2006.

[7] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, 1971.

[8] M. C. Vuran, O. B. Akan, and I. F. Akylidz, “Spatio-temporal correlation: theory and applications for wireless networks,” *Computer Networks*, vol. 45, pp. 245–259, 2004.

[9] M. W. Marcellin, “On entropy-constrained trellis-coded quantization,” *IEEE Trans. Commun.*, vol. 42, pp. 14–16, Jan. 1994.