

RESEARCH

Open Access

Distributed transform coding via source-splitting

Pradeepa Yahampath

Abstract

Transform coding (TC) is one of the best known practical methods for quantizing high-dimensional vectors. In this article, a practical approach to distributed TC of jointly Gaussian vectors is presented. This approach, referred to as *source-split distributed transform coding* (SP-DTC), can be used to easily implement two terminal transform codes for any given rate-pair. The main idea is to apply *source-splitting* using orthogonal-transforms, so that only Wyner-Ziv (WZ) quantizers are required for compression of transform coefficients. This approach however requires optimizing the bit allocation among dependent sets of WZ quantizers. In order to solve this problem, a low-complexity tree-search algorithm based on analytical models for transform coefficient quantization is developed. A rate-distortion (RD) analysis of SP-DTCs for jointly Gaussian sources is presented, which indicates that these codes can significantly outperform the practical alternative of independent TC of each source, whenever there is a strong correlation between the sources. For practical implementation of SP-DTCs, the idea of using *conditional entropy constrained* (CEC) quantizers followed by Slepian-Wolf coding is explored. Experimental results obtained with SP-DTC designs based on both CEC scalar quantizers and CEC trellis-coded quantizers demonstrate that actual implementations of SP-DTCs can achieve RD performance close to the analytically predicted limits.

Keywords: distributed transform coding, Wyner-Ziv quantization, multi-terminal quantization, Karhunen-Loève transform (KLT), optimal bit-allocation

1 Introduction

Many new applications such as multi-camera imaging systems rely on networks of distributed wireless sensors to acquire signals in the form of high-dimensional vectors [1]. In such situations, 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 which jointly decodes all the sources. The strong statistical dependencies among the signals observed by different sensors can be exploited in the decoder to reduce the transmission bit-rate of each sensor. This problem, in general, is referred to as distributed (or multiterminal) vector quantization (VQ). The design of a distributed VQ for a large number of source variables is a difficult task. A practically simpler, yet very effective approach to quantizing a large number of correlated variables by using a bank of single variable quantizers is transform coding (TC) [2-4]. Clearly, TC can be used for distributed VQ when separately observed vectors have both inter-vector

and intra-vector statistical dependencies, a situation typical in applications such as camera networks. Most of the previous work [5-7] studies Wyner-Ziv (WZ) transform coding (WZ-TC), which is a special case of more general multiterminal transform coding (MT-TC) [8]. In WZ-TC, a single source is quantized given that the decoder has access to side information about the source.

Information-theoretic studies of distributed transform coding (DTC) can be found in [5,8]. In [8], the optimal linear transform for Gaussian WZ-TC under the mean square-error (MSE) criterion is shown to be the conditional Karhunen-Loève transform (CKLT), which is a natural extension of the result in [2]. This result is based on the assumption that each transform coefficient is compressed by a rate-distortion (RD) optimal WZ quantizer and hence describes the optimal performance theoretically attainable (OPTA) in Gaussian WZ-TC. However, the optimal solution to the more general MT-TC problem remains unsolved, even for the Gaussian case. In [8], an iterative descent algorithm for determining the OPTA of Gaussian MT-TC problem is given. It is shown that, while this algorithm [referred to as the distributed KLT (DKLT)] always converges to a solution, the final solution is not necessarily the global optimum. In any case, the practical

Correspondence: pradeepa@ee.umanitoba.ca
Department of Electrical and Computer Engineering, University of Manitoba,
Winnipeg, Manitoba R3T 5V6, Canada

implementation of distributed quantizers implied by the DKLT remains an open problem. In [5] WZ-TC based on high-rate scalar quantization and ideal Slepian-Wolf (SW) coding [9] is studied. In particular, it is shown that, for jointly Gaussian vectors, CKLT followed by uniform scalar quantization is asymptotically optimal, a natural extension of the result in [10] for entropy-coded quantization at high-rate. More importantly, the bit-allocations and quantizer-step sizes found in [5] can be used for practical design of WZ-transform codes as long as high-rate approximations hold. However, we note that, even when scalar quantizers are used, achieving good performance with this approach still requires the use of a subsequent block-based SW coding method (e.g., Turbo codes or LDPC codes). Other previous studies on WZ-TC can be found in [6,7]. However, they rely on WZ scalar quantization of transform coefficients. Such methods are therefore most suitable for applications requiring low coding delay as their performance is strictly inferior to block-based quantization.

In contrast to WZ-TC, we consider in this article the practical design of two-terminal transform codes for jointly Gaussian vectors in which arbitrary transmission rates can be assigned to each terminal. Our approach is based on the idea of *source-splitting* [11,12] to convert the two-terminal TC problem into two WZ-TC problems. Since transform codes quantize linear projections, we perform source splitting in terms of optimal linear approximations, i.e., a linear approximation of one source is provided as decoder side-information for the other source. The proposed source-split DTC (SP-DTC) approach only requires the design of two-sets of WZ quantizers sequentially, and avoids having to iteratively optimize two sets of WZ quantizers to each other as in [8]. However, this approach requires the solution of a bit allocation problem involving dependent WZ quantizers. To solve this problem for Gaussian sources, we propose an efficient tree-search algorithm, which can be used to find the a good SP-DTC under different models for quantization of transform coefficients. When used with the RD-optimal WZ quantization model [8], this algorithm can potentially locate the optimal SP-DTC for Gaussian sources. In practice, with constraints imposed on tree-search complexity, the algorithm yields a near-optimal solution. We refer to the optimal solution to the Gaussian problem as the source-split DKLT (SP-DKLT). Using this algorithm, we numerically compute the rate-region achievable with a SP-DKLT code for two examples of jointly Gaussian vector sources. This study shows that, when there is sufficient inter-source correlation, optimal SP-DKLT codes can achieve substantially better performance than independent transform codes for the two sources. However, we find

that the rates achievable with SP-DKLT codes are strictly inside the optimal achievable rate-region predicted by the DKLT algorithm of [8]. In order to approach the performance predicted by the optimal SP-DKLT in practice, block WZ quantization of transform coefficients is required. For implementation of block WZ quantizers, we consider the use of trellis-coded quantization (TCQ) followed by SW coding. This two stage approach is known to achieve the RD function of Gaussian WZ coding [13]. In order practically implement this approach, we introduce the idea of designing *conditional entropy constrained* TCQ (CEC-TCQ) based on analytically found bit-allocations. We present experimental results to demonstrate that practical implementations of SP-DTCs for Gaussian sources can closely approach the performance limits indicated by the optimal SP-DKLT. On the other hand when SW coded high-rate scalar quantization model [5] is assumed for encoding transform coefficients, the tree-search algorithm proposed in this article can also be used to find asymptotically good SP-DTC codes for scalar quantization based implementations. These codes can be readily implemented using CEC scalar quantizers (CEC-SQ) as demonstrated by experimental studies presented in this article. In our experimental study, we also investigate the design of good SP-DTCs based on widely used discrete cosine transform (DCT).

This article is organized as follows. Section 2 presents a review of WZ-TC of Gaussian vectors and motivates the particular approach introduced in this article. Section 3 presents the idea of SP-DTC and develops the tree-search algorithm for finding the optimal transforms and the bit-allocation for SP-DKLT codes. Section 4 computes the achievable rate region of SP-DKLT codes for two example Gaussian source models, and presents experimental results obtained by designing SP-DTCs based on both KLT and DCT. Finally, some concluding remarks are given in Section 5.

Notation: As usual, bold letters denote vectors and matrices, upper case denotes random variables, and lower case denotes realizations. Σ_X denotes the auto-covariance matrix of the vector \mathbf{X} . Σ_{XY} and $\Sigma_{X|Y}$, respectively denote the joint covariance matrix of (\mathbf{X}, \mathbf{Y}) and the conditional covariance matrix of \mathbf{X} given \mathbf{Y} . The eigenvalues $\lambda_1, \dots, \lambda_M$ of a $M \times M$ covariance matrix are always indexed such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$, and the corresponding KLT matrix has the structure $\mathbf{T} = (\mathbf{u}_1^T, \dots, \mathbf{u}_M^T)$, where \mathbf{u}_m is the eigenvector associated λ_m .

2 WZ-TC of Gaussian vectors

Consider encoding of a Gaussian vector $\mathbf{X} \in \mathbb{R}^{M_1}$ using B bits per vector, given that the decoder has access to a jointly Gaussian vector $\mathbf{Y} \in \mathbb{R}^{M_2}$. Assume that both

vectors have zero mean, and let the auto-covariance matrix of \mathbf{X} be $\Sigma_{\mathbf{X}} = E\{\mathbf{X}\mathbf{X}^T\}$. In WZ-TC, a linear transform is first applied to \mathbf{X} and each component of the transform coefficient vector $\mathbf{U} = \mathbf{T}^T\mathbf{X}$ is separately compressed by a WZ quantizer, considering \mathbf{Y} as decoder side-information, where \mathbf{T} is a $M_1 \times M_1$ unitary matrix. Let $\hat{\mathbf{U}}$ be the quantized value of \mathbf{U} . The decoder then estimates the source vector based on $\hat{\mathbf{U}}$ and \mathbf{Y} . We wish to find the optimal transform and the allocation of B bits among M_1 transform coefficients, which minimize the quantization MSE $E\left\{\|\mathbf{X} - \hat{\mathbf{X}}\|^2\right\}$, where

$\hat{\mathbf{X}} = E\{\mathbf{X}|\hat{\mathbf{U}}, \mathbf{Y}\}$ is the optimal estimate (at the decoder) of the source vector. The solution of this problem requires an analytical model for coefficient quantization. To this end, [8] considers RD optimal WZ quantization (RD-WZQ) model, the solution based on which is appropriate for practical block quantization techniques such as TCQ. On the other hand [5] considers SW-coded high-rate scalar quantization (SWCHRSQ) model.

Let the eigenvalues of the conditional covariance matrix $\Sigma_{\mathbf{X}|\mathbf{Y}} = E\{\mathbf{X}\mathbf{X}^T|\mathbf{Y}\}$ be $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{M_1})^T$. The CKLT of \mathbf{X} given \mathbf{Y} is defined as $\boldsymbol{\theta} = \mathbf{T}^T\mathbf{X}$, where \mathbf{T} is $M_1 \times M_1$ unitary matrix such that $\Sigma_{\mathbf{X}|\mathbf{Y}} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{M_1})$ [8]. It is easy to verify that $E\{\boldsymbol{\theta}\boldsymbol{\theta}^T|\mathbf{Y}\} = \boldsymbol{\Lambda}$, i.e., the components of $\boldsymbol{\theta}$ are conditionally uncorrelated, given \mathbf{Y} . For convenience, define

$$d^*(\boldsymbol{\lambda}, B, N) \triangleq \left(\prod_{m=1}^N \lambda_m\right)^{\frac{1}{N}} 2^{-\frac{2B}{N}},$$

where $N \leq M_1$ is a positive integer.

2.1 RD-WZQ model

When RD-WZQ model is used, the quantization MSE for U_m is given by [14]

$$d_m^{\text{RD-WZQ}} = \lambda_m 2^{-2r_m}, \quad (1)$$

where $\lambda_m = \text{var}(U_m|\mathbf{Y})$ and $\text{var}(\cdot|\cdot)$ denotes the conditional variance. The optimal solution to the WZ-TC problem under RD-WZQ model is given by the following theorem.

Theorem 1 *Given jointly Gaussian \mathbf{X} and \mathbf{Y} as defined above, and a total bit budget of B bits, if each transform coefficient U_m , $m = 1, \dots, M_1$, where $\mathbf{U} = \mathbf{T}^T\mathbf{X}$, is quantized by an RD optimal WZ quantizer which uses \mathbf{Y} as decoder side-information, then the transform \mathbf{T} which minimizes $E\|\mathbf{X} - \hat{\mathbf{X}}\|^2$ is the CKLT of \mathbf{X} given \mathbf{Y} , and the number of bits allocated to quantizing U_m is*

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

where $N \leq M$ is the largest integer for which $\lambda_m \geq d^*(\boldsymbol{\lambda}, B, N)$, $m = 1, \dots, N$. The quantization MSE of the m th coefficient is

$$d_m = \lambda_m 2^{-2\rho_m} = \begin{cases} d^*(\boldsymbol{\lambda}, B, N), & m = 1, \dots, N \\ \lambda_m, & m = N + 1, \dots, M \end{cases} \quad (3)$$

and the overall MSE is

$$E\left\{\|\mathbf{X} - \hat{\mathbf{X}}\|^2\right\} = Nd^*(\boldsymbol{\lambda}, B, N) + \sum_{m=N+1}^{M_1} \lambda_m. \quad (4)$$

Proof 1 *Directly follows from* [[8], Section III-B].

Note that RD-WZQ model implies infinite-dimensional VQ of each coefficient and hence the above MSE is the OPTA in the Gaussian WZ-TC problem.

2.2 SWC-HRSQ model

The asymptotically (in rate) optimal solution to the WZ-TC problem under SWC-HRSQ model is given by the following theorem.

Theorem 2 *Let \mathbf{X} , \mathbf{Y} , and B be as in Theorem 1. If each transform coefficient U_m , $m = 1, \dots, M_1$, where $\mathbf{U} = \mathbf{T}^T\mathbf{X}$, is quantized by a high-rate scalar quantizer and the quantizer output is encoded by a SW code which uses \mathbf{Y} as decoder side-information, then the transform \mathbf{T} which asymptotically minimizes $E\|\mathbf{X} - \hat{\mathbf{X}}\|^2$ is the CKLT of \mathbf{X} given \mathbf{Y} and the bit allocation is given by (2). Furthermore, the asymptotically optimal quantizer for U_m , $m = 1, \dots, N$ is a uniform quantizer with step-size $\Delta = \sqrt{(2\pi e)d^*(\boldsymbol{\lambda}, B, N)}$. The resulting quantization MSE is*

$$E\left\{\|\mathbf{X} - \hat{\mathbf{X}}\|^2\right\} = \left(\frac{\pi e}{6}\right) Nd^*(\boldsymbol{\lambda}, B, N) + \sum_{m=N+1}^{M_1} \lambda_m. \quad (5)$$

Proof 2 *See Section "Proof of Theorem 2" in Appendix 1.*

2.3 Sufficiency of scalar side-information

The WZ quantizers with vector-valued decoder side-information as considered in Theorem 1 are difficult to design in practice. However, the following theorem establishes that when CKLT is used and RD-WZQ model applies for quantization of coefficients, a linear transformation of the side-information vector can be used to convert the vector side-information problem into an equivalent scalar side-information problem.

Furthermore, [[5], Section 6] shows that this result applies in asymptotic sense to the SWC-HRSQ model as well.

Theorem 3 *Let the mean-zero vectors $\mathbf{X} \in \mathbb{R}^{M_1}$ and $\mathbf{Y} \in \mathbb{R}^{M_2}$ be jointly Gaussian, and let \mathbf{T} be the CKLT of \mathbf{X} given \mathbf{Y} . Suppose that transform coefficients U_m , $m = 1, \dots, M_1$, where $\mathbf{U} = \mathbf{T}^T \mathbf{X}$, are each compressed by an RD optimal WZ quantizer relative to decoder side-information \mathbf{Y} . Then, the minimum MSE (MMSE) estimate $\tilde{u}_m(\mathbf{y}) = E\{U_m|\mathbf{y}\}$ of U_m given $\mathbf{Y} = \mathbf{y}$ is a sufficient statistic for decoder side-information for quantizing U_m .*

Proof 3 See Section "Proof of Theorem 3" in Appendix 1.

Wyner-Ziv transform coding is a special case of more general MT-TC where two or more terminals apply TC to their respective inputs and transmit the quantized outputs to a single decoder which exploits the inter-source correlation to jointly reconstruct all the sources. In this case, the problem is to optimally allocate a given bit budget among all the terminals such that the total MSE is minimized. However, the closed-form solution to this problem appears difficult, due to the inter-dependence of the encoders in different terminals. An iterative descent algorithm is given in [8] for solving the Gaussian MT-TC problem. Given a total bit-budget, the bit-rate of the system is incremented by a small amount in each iteration, and the optimal WZ-TC for each terminal is determined by fixing the encoders of all other terminals and considering their outputs as decoder-side information. The solution that gives the MMSE is accepted and the iterations are repeated until the total bit-budget is exhausted. While this algorithm, referred to as the DKLT algorithm, is guaranteed to converge to at least a locally optimal solution, there is no tractable way to implement the quantizers implied by the final solution since it is not practical to optimize a set of near-optimal WZ quantizers in each iteration of this algorithm. Note also that, DKLT requires joint decoding of two vector sources.

3 Source-splitting based distributed TC

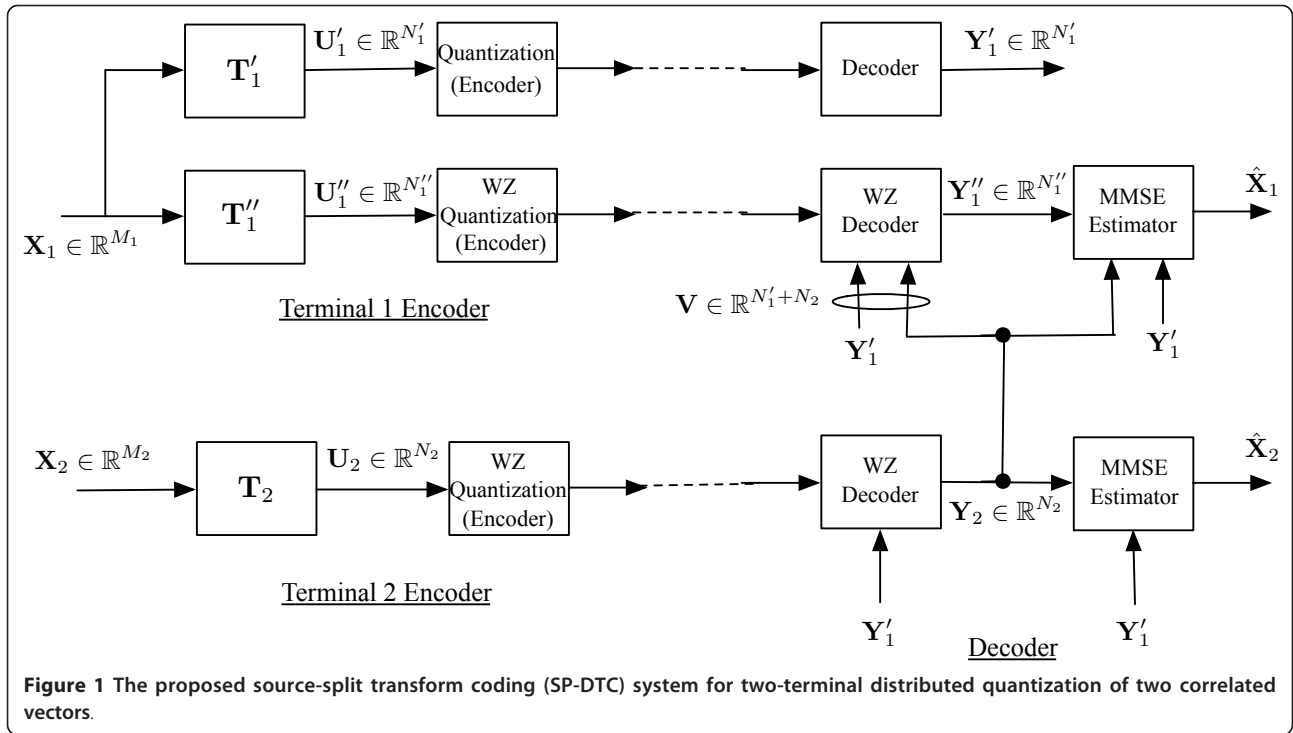
In general, designing a multi-terminal VQ is more difficult than designing a WZ-VQ, due to the mutual dependence among the encoders. However, one could use WZ-VQs to realize a multi-terminal VQ by using source-splitting [12]. It is known that in the quadratic-Gaussian case, source-splitting can be used to realize any rate-pair in the achievable rate-region by only using ideal WZ-VQs which correspond to the corner-points of the achievable rate-region [[12], Section V-C], [15]. While the same optimality properties cannot be claimed for source-splitting by linear transforms, the aforementioned observation still provides us the motivation to take a similar approach in practically realizing the DTCs

which can operate at arbitrary rates, by using only WZ quantizers.

A block diagram of the SP-DTC system is shown in Figure 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 performs source splitting by providing an approximation $\mathbf{Y}'_1 \in \mathbb{R}^{N'_1}$ of \mathbf{X}_1 at the rate $B'_1 (< B)$ bits/vector as decoder side-information for WZ coding of terminal 2, where $N'_1 \leq M_1$. In a TC framework, the goal is to provide the best (in MMSE sense) linear approximation of \mathbf{X}_1 as the decoder side-information. Therefore, \mathbf{Y}'_1 is the B'_1 -bit approximation of a linear projection $\mathbf{U}'_1 = \mathbf{T}'_1{}^T \mathbf{X}_1$, where \mathbf{T}'_1 is a $M_1 \times M_1$ unitary matrix. Given the side-information \mathbf{Y}'_1 at the decoder, the terminal 2 quantizes a linear projection $\mathbf{U}_2 = \mathbf{T}_2{}^T \mathbf{X}_2$ of \mathbf{X}_2 using $B_2 (< B - B'_1)$ bits/vector, where \mathbf{T}_2 is a $M_2 \times M_2$ unitary matrix. Let $\mathbf{Y}_2 \in \mathbb{R}^{N_2}$, be the quantized value of \mathbf{U}_2 , where $N_2 \leq M_2$. Then, given the quantized linear projections of both \mathbf{X}_1 and \mathbf{X}_2 available at the decoder, the terminal 1 quantizes the a linear projection $\mathbf{U}''_1 = \mathbf{T}''_1{}^T \mathbf{X}_1$ of \mathbf{X}_1 using $B''_1 = B - B'_1 - B_2$ bits/vector, where \mathbf{T}''_1 is a $M_1 \times M_1$ unitary matrix. Let $\mathbf{Y}''_1 \in \mathbb{R}^{N''_1}$ be the quantized value of \mathbf{U}''_1 , where $N''_1 \leq M_1$. In the receiver, each source vector is reconstructed by a WZ decoder. The MMSE optimal reconstructions for \mathbf{X}_1 and \mathbf{X}_2 are, respectively given by $\hat{\mathbf{X}}_1 = E\{\mathbf{X}_1|\mathbf{Y}''_1, \mathbf{V}\}$ and $\hat{\mathbf{X}}_2 = E\{\mathbf{X}_2|\mathbf{V}\}$, where $\mathbf{V} = (\mathbf{Y}'_1 \mathbf{Y}_2)^T$. The total transmission rate for source \mathbf{X}_1 is thus $B_1 = B'_1 + B''_1$ bits/vector. The rates used by terminals 1 and 2 in bits/sample are given by $R_1 = B_1/M_1$ and $R_2 = B_2/M_2$, respectively.

Let $\mathbf{U}'_1 = (U'_{1,1}, \dots, U'_{1,M_1})^T$, $\mathbf{U}''_1 = (U''_{1,1}, \dots, U''_{1,M_1})^T$ and $\mathbf{U}_2 = (U_{2,1}, \dots, U_{2,M_2})^T$. Also, let the bit-rates allocated to quantizing these transform coefficients be $\mathbf{r}''_1 = (r''_{1,1}, \dots, r''_{1,M_1})^T$, $\mathbf{r}'_1 = (r'_{1,1}, \dots, r'_{1,M_1})^T$, and $\mathbf{r}_2 = (r_{2,1}, \dots, r_{2,M_2})^T$ respectively, and define $\mathbf{r} = (\mathbf{r}'_1, \mathbf{r}''_1, \mathbf{r}_2)^T$. Given a total of B bits for encoding both \mathbf{X}_1 and \mathbf{X}_2 , the design of a SP-DTC involves determining the values of the transforms \mathbf{T}'_1 , \mathbf{T}''_1 , and \mathbf{T}_2 , and a bit allocation among the transform coefficients $\mathbf{U} = (\mathbf{U}'_1, \mathbf{U}''_1, \mathbf{U}_2)^T$ such that the total MSE

$$D(\mathbf{r}) = E \left\{ \left\| \mathbf{X}_1 - \hat{\mathbf{X}}_1 \right\|^2 + \left\| \mathbf{X}_2 - \hat{\mathbf{X}}_2 \right\|^2 \right\} \quad (6)$$



is minimized. By writing the quantization MSEs of $U_{1,i}'$ and $U_{2,j}$ as $d_{1,i}'(r_{1,i}', \mathbf{r}_1', \mathbf{r}_2)$, and $d_{2,j}(r_{2,j}, \mathbf{r}_1')$, respectively, $i = 1, \dots, M_1, j = 1, \dots, M_2$, the total MSE can be expressed as

$$D(\mathbf{r}) = \sum_{i=1}^{M_1} d_{1,i}'(r_{1,i}', \mathbf{r}_1', \mathbf{r}_2) + \sum_{j=1}^{M_2} d_{2,j}(r_{2,j}, \mathbf{r}_1'). \quad (7)$$

The bit allocation problem can now be stated as follows:

Given a total bit-budget of B bits

$$\min_{\mathbf{r}} D(\mathbf{r}) \quad (8)$$

subject to

$$\sum_{m=1}^{2M_1+M_2} r_m \leq B, \\ r_m \geq 0, \quad m = 1, \dots, 2M_1 + M_2,$$

where $\mathbf{r} = (r_1, \dots, r_{2M_1+M_2})^T$. The explicit solution of this minimization problem is unfortunately intractable due to the inter-dependence of the three transform codes involved.

However, an explicit solution can be found for a variant of this problem obtained by fixing B_1' , B_1'' and B_2 , so that the number of bits allocated to each transform code is fixed and it is only required to optimize the bit allocation

among the quantizers within each transform code. For simplicity, we refer to this problem as the *constrained bit-allocation problem*. In the following, an explicit solution to this problem is derived. Based on the result, we then present a tree-search algorithm to solve the *unconstrained problem* (8). Under both RD-WZQ and SWC-HRSQ models, the optimal transforms for Gaussian sources are CKLTs. Therefore, we refer to the solution to problem (8) as the SP-DKLT.

3.1 Solution to the constrained bit-allocation problem

3.1.1 RD optimal quantization

Let B_1' , B_1'' , and B_2 be fixed in Figure 1 and let the coefficient quantization be represented by the RD-WZQ model (for components of \mathbf{U}_1' , this reduces to the non-distributed RD-optimal quantization). From *Theorem 1*, it follows that the MMSE optimal transform (in the sense of providing the best linear approximation as decoder side information for terminal 2) \mathbf{T}_1' is the KLT of \mathbf{X}_1 . Let the eigenvalues of $\sum_{\mathbf{X}_1}$ be $\lambda_1' = (\lambda_{1,1}', \dots, \lambda_{1,M_1}')^T$. Then, using (2), the optimal bit allocation for $U_{1,m}'$ can be given by

$$r_{1,m}' = \rho_m(\lambda_1', B_1', N_1'), \quad m = 1, \dots, M_1. \quad (9)$$

for some $N_1' \leq M_1$. Let the quantized value of $U_{1,m}'$ be $\hat{U}_{1,m}'$. We note that $E U_{1,m}'^2 = \lambda_{1,m}'$ for $m = 1, \dots, M_1$ and $E(U_{1,m}' - \hat{U}_{1,m}')^2 = d^*(\lambda_1', B_1', N_1')$ for $m = 1, \dots, N_1'$. In

RD theoretic sense, the quantized value (up to a scaling factor) of the mean-zero Gaussian variable $U'_{1,m}$ can be given by $Y'_{1,m} = U'_{1,m} + Z'_{1,m}$, $m = 1, \dots, N'_1$, where $Z'_{1,m}$ is a mean-zero Gaussian random-variable independent of $U'_{1,m}$ such that

$$EZ'^2_{1,m} = \frac{\lambda'_{1,m} d^* (\lambda'_1, B'_1, N'_1)}{\lambda'_{1,m} - d^* (\lambda'_1, B'_1, N'_1)}, \quad (10)$$

see [[16], Section 10.3.2]. Therefore, we can write

$$\mathbf{Y}'_1 = \mathbf{K}'_1{}^{(N'_1)T} \mathbf{X}_1 + \mathbf{Z}'_1, \quad (11)$$

where $\mathbf{K}'_1{}^{(N'_1)}$ denotes the $M_1 \times N'_1$ matrix consisting of first N'_1 columns of \mathbf{K}'_1 . The covariance matrix of “quantization noise” vector $\mathbf{Z}'_1 = (Z'_{1,1}, \dots, Z'_{1,N'_1})^T$ is given by $\sum_{Z'_1} = \text{diag}(Z'_{1,1}, \dots, Z'_{1,N'_1})$. According to (11), \mathbf{Y}'_1 and \mathbf{X}_1 are jointly Gaussian, and it follows that

$$\sum_{Y'_1} = \mathbf{K}'_1{}^{(N'_1)T} \sum_{X_1} \mathbf{K}'_1{}^{(N'_1)} + \sum_{Z'_1}. \quad (12)$$

Furthermore, \mathbf{X}_2 and \mathbf{Y}'_1 are jointly Gaussian with the conditional covariance matrix

$$\sum_{X_2|Y'_1} = \sum_{X_2} - \sum_{X_1 X_2}^T \mathbf{K}'_1{}^{(N'_1)} \sum_{Y'_1}^{-1} \mathbf{K}'_1{}^{(N'_1)T} \sum_{X_1 X_2}. \quad (13)$$

Next consider TC \mathbf{X}_2 given \mathbf{Y}'_1 as decoder side-information. From *Theorem 1*, it follows that the $E\|\mathbf{X}_2 - \hat{\mathbf{X}}_2\|^2$ is minimized by choosing \mathbf{T}_2 as the CKLT of \mathbf{X}_2 given \mathbf{Y}'_1 and by applying RD optimal WZ quantization to each element of $\mathbf{U}_2 = \mathbf{T}_2^T \mathbf{X}_2$ given decoder side information \mathbf{Y}'_1 based on a bit-allocation specified by the eigenvalues of $\sum_{X_2|Y'_1}$, $\lambda_2 = (\lambda_{2,1}, \dots, \lambda_{2,M_2})$. The optimal bit allocation for $U_{2,m}$ is given by

$$r_{2,m} = \rho_m (\lambda_2, B_2, N_2), \quad m = 1, \dots, M_2, \quad (14)$$

for some $N_2 \leq M_2$. The resulting MSE is given by

$$D_2(B'_1, B_2) = E\|\mathbf{X}_2 - \hat{\mathbf{X}}_2\|^2 = N_2 d^* (\lambda_2, B_2, N_2) + \sum_{m=N_2+1}^{M_2} \lambda_{2,m}. \quad (15)$$

As before, the quantized value of \mathbf{U}_2 up to a scaling factor, can be represented by [[8], Theorem 3],

$$\mathbf{Y}_2 = \mathbf{K}_2{}^{(N_2)T} \mathbf{X}_2 + \mathbf{Z}_2, \quad (16)$$

where $\mathbf{K}_2{}^{(N_2)}$ denotes the $M_2 \times N_2$ matrix consisting of first N_2 columns of \mathbf{K}_2 and the quantization noise $\sum_{Z_2} = \text{diag}(Z_{2,1}, \dots, Z_{2,N_2})$ is a mean zero iid Gaussian

vector independent of \mathbf{X}_2 . The covariance matrix of \mathbf{Z}_2 is given by $\sum_{Z_2} = \text{diag}(Z_{2,1}, \dots, Z_{2,N_2})$, where $EZ_{2,m}^2$, $m = 1, \dots, N_2$ are given by

$$EZ_{2,m}^2 = \frac{\lambda_{2,m} d^* (\lambda_2, B_2, N_2)}{\lambda_{2,m} - d^* (\lambda_2, B_2, N_2)}. \quad (17)$$

The covariance matrix of \mathbf{Y}_2 is

$$\sum_{Y_2} = \mathbf{K}_2{}^{(N_2)T} \sum_{X_2} \mathbf{K}_2{}^{(N_2)} + \sum_{Z_2}. \quad (18)$$

Therefore \mathbf{X}_1 and $\mathbf{V} = (\mathbf{Y}'_1 \mathbf{Y}_2)^T$ are jointly Gaussian with the cross-covariance matrix

$$\sum_{X_1|V} = \sum_{X_1} - \sum_{V X_1}^T \sum_V^{-1} \sum_{V X_1}, \quad (19)$$

where

$$\sum_{V X_1} = \begin{pmatrix} \mathbf{K}'_1{}^{(N'_1)} \sum_{X_1} \\ \sum_{X_2} \sum_{X_2 X_1} \end{pmatrix}$$

and

$$\sum_V = \begin{pmatrix} \sum_{Y'_1} & \mathbf{K}'_1{}^{(N'_1)} \sum_{X_1 X_2} \sum_{X_2}^T \\ \sum_{X_2} \sum_{X_2 X_1} \mathbf{K}'_1{}^{(N'_1)T} & \sum_{Y_2} \end{pmatrix}.$$

Finally, consider quantizing \mathbf{X}_1 , given $\mathbf{V} = (\mathbf{Y}'_1 \mathbf{Y}_2)^T \in \mathbb{R}^{N'_1+N_2}$ as the decoder side-information.

As before, $E\|\mathbf{X}_1 - \hat{\mathbf{X}}_1\|^2$ is minimized by choosing \mathbf{T}'_1 as the CKLT of \mathbf{X}_2 given \mathbf{V} , and RD optimal WZ quantization of each element of $\mathbf{U}'_1 = \mathbf{T}'_1{}^T \mathbf{X}_1$ given \mathbf{V} , based on a bit-allocation specified by the eigenvalues of $\sum_{X_1|V}$, $\lambda''_1 = (\lambda''_{1,m}, \dots, \lambda''_{1,M_1})^T$. The bit rate allocated to quantizing $U'_{1,m}$ given by

$$r''_{1,m} = \rho_m (\lambda''_1, B''_1, N''_1), \quad m = 1, \dots, M_1, \quad (20)$$

for some $N''_1 \leq M_1$ and the resulting MSE is

$$D_1(B'_1, B''_1, B_2) = E\|\mathbf{X}_1 - \hat{\mathbf{X}}_1\|^2 = N''_1 d^* (\lambda''_1, B''_1, N''_1) + \sum_{m=N''_1+1}^{M_1} \lambda''_{1,m}. \quad (21)$$

Given a rate tuple (B'_1, B''_1, B_2) , the MMSE achievable with a SP-DKLT code for two jointly Gaussian vectors is given by

$$D(B'_1, B''_1, B_2) = D_1(B'_1, B''_1, B_2) + D_2(B'_1, B_2). \quad (22)$$

3.1.2 High-resolution scalar quantization and SW coding

Due to *Theorem 2*, the expressions for bit-allocations given by (9), (14) and (20) applies to SWC-HRSQ

model as well. However, the resulting MSE and hence the quantization noise variances are different to those of the RD-WZQ model. More specifically, since the decoder side-information in a SP-DTC depends on quantization noise of the other terminals, the optimal transforms \mathbf{T}_2 and \mathbf{T}'_1 and the associated bit allocations obtained with the SWC-HRSQ model are different to those obtained with the RD-WZQ model. In order to make the problem tractable, we assume that the quantization noise of SWC-HRSQ model also follows (11) and (16) (for a discussion on the validity of this assumption, see [17]). This assumption essentially allows us to compute the conditional covariance matrices $\sum_{X_2|Y'_1}$ and $\sum_{X_1|V}$ as in the previous case, and then apply (9), (14) and (20) to find the optimal transforms and the bit-allocations. However, due to (27), the quantization noise variance in (10) in this case is given by

$$EZ_{1,m}^2 = \frac{(\pi e/6) \lambda'_{1,m} d^*(\lambda'_1, B'_1, N'_1)}{\lambda'_{1,m} - (\pi e/6) d^*(\lambda'_1, B'_1, N'_1)}. \quad (23)$$

A similar expression exists for the quantization noise variance in (17).

3.2 A tree-search solution to the unconstrained bit-allocation problem

We note that the optimal solution to the unconstrained problem defined in (8) corresponds to the MMSE solution of the constrained problem over the set of rate-tuples $\mathcal{S} = \{(B'_1, B''_1, B_2) : B'_1 \in (0, B), B''_1 \in (0, B), B_2 \in (0, B), B'_1 + B''_1 + B_2 \leq B\}$. This set is shown in Figure 2. One approach to locating the MMSE solution is to search over an appropriately discretized grid of points inside \mathcal{S} . As we will see, even though an exhaustive search on a fine grid can be prohibitively complex, a much simpler constrained tree-search algorithm exists which can be used to locate the required solution with a very high probability.

The proposed 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 [[4], Section 8.4]. Unfortunately, this type of a greedy search cannot guarantee that the final solution is overall optimal and can yield poor results in our problem where the bit allocation among three sets of dependent quantizers must be achieved. On the other hand, if the increment ΔB is chosen small enough, a near-optimal solution can be found by resorting to a tree-search.

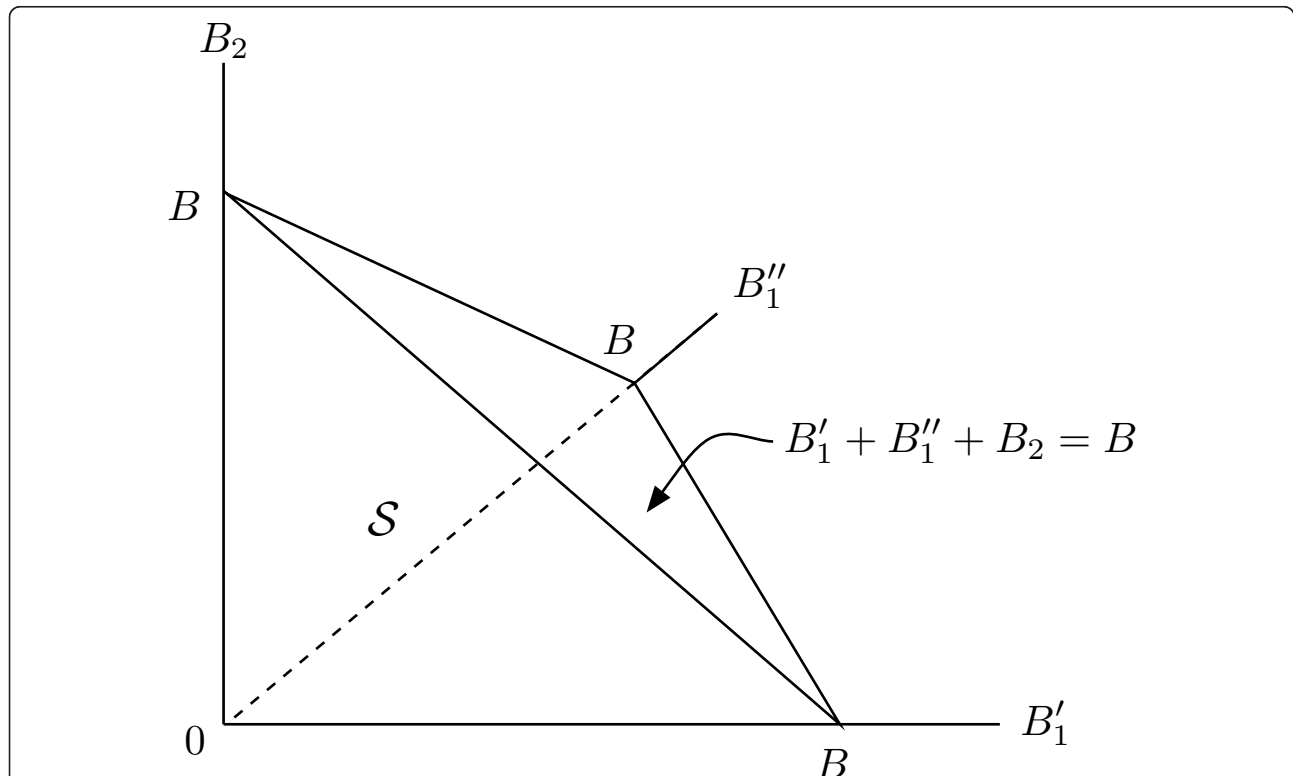


Figure 2 The solution space \mathcal{S} for the unconstrained bit-allocation problem. The tree-search algorithm uses a search-grid of regularly spaced points (i.e., a cubic lattice) with a separation of ΔB inside \mathcal{S} .

Even though a full tree-search is intractable, a simple algorithm referred to as the (M, L) -algorithm [18] exists for detecting the minimum cost path in the tree with a high probability. We use this insight to formulate a tree-search algorithm for solving the unconstrained bit allocation problem, in which a set of constrained bit allocation problems are solved in each iteration.

In order to describe the proposed tree-search algorithm in detail, let ΔB be the incremental amount of bits to be allocated in each step of the search, where $0 < \Delta B \ll B$. The algorithm is initialized by setting $(B'_1, B''_1, B_2) = (0, 0, 0)$, i.e., the origin in Figure 2. Now if we are to allocate 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 as described in the previous section and find the MMSE solution. Each of these candidate solutions can be viewed as a node in a tree as shown in Figure 3. The root node of the tree corresponds to a SP-DTC of rate 0, and a node in the first level of nodes obtained in the first iteration of the algorithm corresponds to a SP-DTC of rate of ΔB bits per source pair $(\mathbf{X}_1, \mathbf{X}_2)$. In the second iteration of the algorithm, we allocate ΔB more bits to each of the three candidate SP-DTCs (but to one of the SP-DTCs at a time) in the first level of nodes. Note that, for each SP-DTC we can allocate ΔB bits in three different ways, i.e., ΔB bits can be added to either B'_1, B''_1 , or B_2 . This requires the solution

of three constrained bit allocations problems for each of the 3 nodes in the first-level. As a result, the tree will be extended to a second level of 3^2 nodes, in which each node corresponds to a SP-DTC of $2\Delta B$ bits per source-pair, as shown in Figure 3. We can repeat this procedure, allocating ΔB bits to each of terminal node of the tree in a given iteration, until all B bits are exhausted. After the final iteration, the tree would consist of $L = \lceil B/\Delta B \rceil$ levels with 3^L nodes in the last level (terminal nodes). Each terminal node corresponds to a candidate SP-DTC of rate B , and rate-tuples of these SP-DTCs lie on the plane $B'_1 + B''_1 + B_2 = B$ in Figure 2. The MMSE terminal node of the tree is the optimal solution to the unconstrained bit allocation problem, provided that the latter solution is on the search-grid. If ΔB is chosen small enough, then we can ensure that the optimal solution is nearly on the search-grid. Suppose that, in each iteration, the algorithm saves the MSE of the solution to the constrained bit-allocation problem associated with each node. In theory, the optimal solution can be found by an exhaustive tree-search, using the MSE of a node [given by (22)] as the path-cost. In order to practically implement the tree-search, we use the (M, L) -algorithm, 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 the (M, L) algorithm [[18], p. 216] for a tree of depth L , one only retains the M best (lowest MSE) nodes in each iteration. When $M = 1$ we have a completely greedy search. On the other hand when $M_n = 3^n$ in

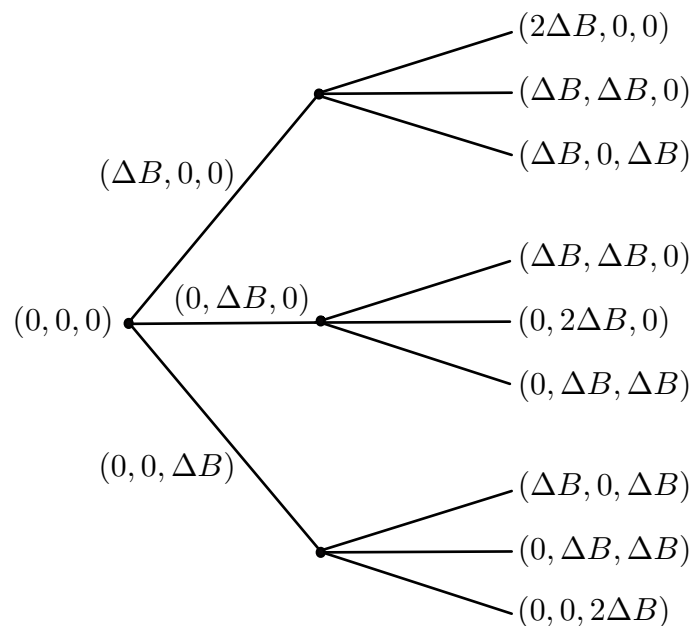


Figure 3 Bit allocation tree with the values of the rate-tuple (B'_1, B''_1, B_2) after two iterations of the tree-search algorithm.

the n th iteration, we have a full-tree search which has a complexity that grows exponentially with the iteration number. When M ($1 \leq M \leq 3^L$) is a prescribed constant, the complexity is linear in M , independent of the number of iterations. In obtaining the simulation results presented in Section 4, $M = 27$ and $L = 135$ ($\Delta B = 0.2$) were found to be sufficient to obtain near optimal results. For example, it was observed that even for $M = 81$ and $L = 405$, nearly the same result was obtained.

4 Numerical results and discussion

Source model A: Let the components of \mathbf{X}_1 be M_1 consecutive samples of a first-order Gauss-Markov process with a unit-variance and the correlation coefficient $|\rho| < 1$, i.e., $X_{1,m} = \rho X_{1,(m-1)} + Z_m$, $m = 2, \dots, M_1$, where Z_m , $m = 1, \dots, M_1$ are mean-zero iid Gaussian variables such that $EZ_m^2 = 1 - \rho^2$. The auto-covariance matrix \sum_{X_1} is a Teoplitz matrix with the first row $(1, \rho, \rho^2, \dots, \rho^{M_1-1})$. Now define the components of \mathbf{X}_2 to be noisy observations of the components of \mathbf{X}_1 , i.e., $X_{2,m} = \gamma X_{1,m} + W_m$, where $|\gamma| < 1$ and W_m is a mean-zero, iid Gaussian variable with $EW_m^2 = 1 - \gamma^2$, $m = 1, \dots, M_1$ ($M_2 = M_1$). It follows that, \sum_{X_2} is a Teoplitz matrix with the first row $(1, \gamma^2 \rho, \gamma^2 \rho^2, \dots, \gamma^2 \rho^{M_1-1})$. Furthermore, the cross-covariance matrix $\sum_{X_1 X_2} = \gamma^2 \sum_{X_1}$. Note that \mathbf{X}_1 and \mathbf{X}_2 are not statistically similar and the components of \mathbf{X}_1 are more correlated than those of \mathbf{X}_2 .

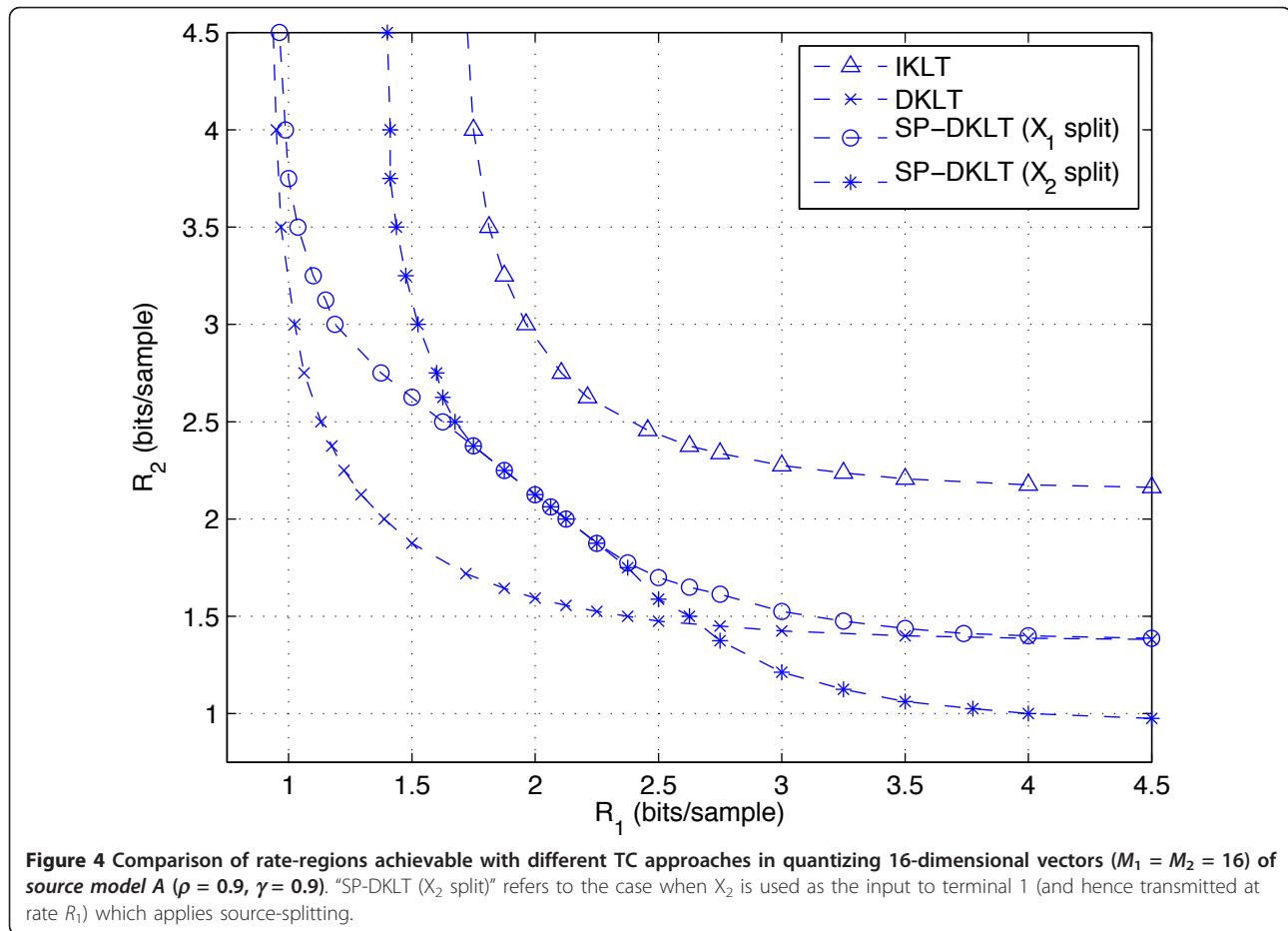
Source model B: Consider a spatial Gaussian random field in which the correlation function decays with distance d according to the squared exponential model [19]. We define the random vectors \mathbf{X}_1 and \mathbf{X}_2 to be observations picked-up by a pair of sensor arrays placed in this random field. In this case, the auto-covariance matrix of \mathbf{X}_1 is given by $[\sum_{X_1}]_{ij} = \exp\{-(\alpha d_{ij})^2\}$, where α is a constant and d_{ij} is the distance between $X_{1,i}$ and $X_{1,j}$. The auto-covariance matrix of \mathbf{X}_2 also has a similar form. For simplicity assume that the sensors in each array are placed on a $M \times M$ square grid of unit spacing (i.e., $M_1 = M_2 = M^2$), the two arrays are on parallel planes separated by a distance r , and the two grids are aligned so that the distance between $X_{1,i}$ and $X_{2,i}$ is r for all i . With this setup, the distance between $X_{1,i}$ and $X_{2,j}$ is $\sqrt{d_{ij}^2 + r^2}$, and the cross-covariance matrix is given by $[\sum_{X_1 X_2}]_{ij} = \theta \exp\{-(\alpha d_{ij})^2\}$, where $\theta = \exp\{-(\alpha r)^2\}$. This sensor structure ensures that \mathbf{X}_1 and \mathbf{X}_2 are statistically similar. However, $\sum_{X_1 X_2}$ can be chosen

independently (by choosing array separation r) of \sum_{X_1} and \sum_{X_2} .

4.1 RD performance

We compute the rate-pairs (R_1, R_2) achievable with a SP-DKLT code 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 [given by (6)]. The rate-pairs achievable for $D = 0.01$ with SP-DKLT coding of vectors from *source model A* with $\rho = 0.9$, $\gamma = 0.9$, are plotted in Figure 4. These values of ρ and γ result in a source cross-covariance matrix with the largest element 0.9. In Figure 4, the curve “SP-DKLT (\mathbf{X}_1 split)” corresponds to a system in which the input to terminal 1 (which applies source splitting) is \mathbf{X}_1 as shown in Figure 1. The curve “SP-DKLT (\mathbf{X}_2 split)” corresponds to a system in which the input to the terminal 1 is \mathbf{X}_2 . Note that the two curves are not symmetric in rates and they coincide if R_1 and R_2 are inter-changed in one of the curves. This is because \mathbf{X}_1 and \mathbf{X}_2 have different auto-covariance matrices, and hence inter-changing the rates is equivalent to inter-changing the terminals. Importantly, this result indicates that when the two sources are not statistically identical, which source is chosen for splitting does not affect the SP-DKLT performance. Table 1 lists the best bit allocations found by the tree-search algorithm for SP-DKLT codes shown in Figure 4. Note that, for the same (B_1, B_2) , the rate-split between B'_1 and B''_2 when \mathbf{X}_1 is applied to the terminal 1 is not identical to that when \mathbf{X}_2 is applied to the terminal 1.

Figure 4 also shows the rate region achievable if each source is independently compressed using the KLT (i.e., only intra-vector correlation is utilized), labeled IKLT, and the OPTA lower bound for distributed TC predicted by the iterative DKLT algorithm [8]. The performance of both distributed and non-distributed TC of *source model A* degrades as ρ decreases, since both auto- and cross-covariance matrices of \mathbf{X}_1 and \mathbf{X}_2 are functions of ρ . Next consider the *source model B* for which the lowest achievable rate-pairs corresponding to $D = 0.005$ are plotted in Figure 5. The source parameter $\alpha = 0.32$ results in auto-covariance matrices whose largest off-diagonal element is 0.9. Also recall that θ is the largest element in the source cross-covariance matrix. Note that changing θ only affects the cross-covariance matrix, and hence has no effect on the best achievable rates for independent coding of the two sources. On the other hand, as θ increases, the rates achievable with distributed coding do improve. Since in *source model B*, the two sources are statistically similar, the curves in Figure 5 are symmetric in rates and the optimal bit



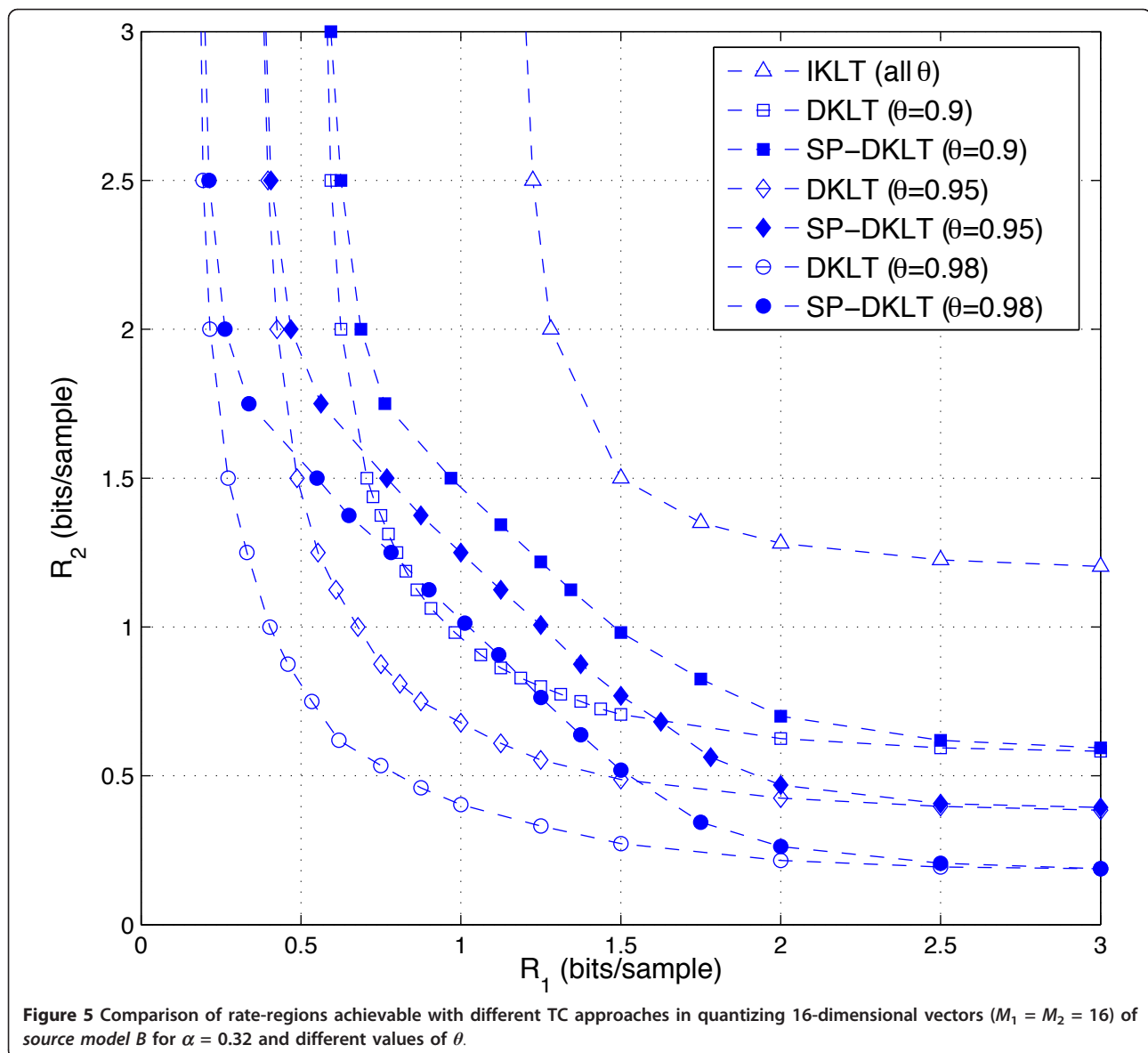
allocation does not depend on which source is chosen for splitting.

The RD performance in Figures 4 and 5 indicate that SP-DKLT codes can significantly outperform IKLT codes at all rates when there is sufficient correlation between the two distributed sources. The performance of SP-DKLT coding necessarily approaches OPTA (DKLT) bound when either R_1 or R_2 is sufficiently high. That is, the terminal with the higher bit rate can independently transform code its input with negligible distortion, and the other terminal can then apply WZ-TC at the minimum rate achievable with "almost unquantized" decoder side-information. However, for both source-models the rate-region achievable with source-splitting is strictly inside that of DKLT. In other words, there are some rate-pairs inside the DKLT rate-region for a given MSE D , which cannot to be achieved by a SP-DKLT code. A closely related issue is that, for a range of values of (R_1, R_2) , the sumrate $R_1 + R_2$ of SP-DKLT codes remains constant and reaches its minimum. For example, it can be seen from Figure 4 and Table 1 that the sum-rate is about 4.125 bits when the rate of X_1 is in the range 1.375 - 2.5 bits/sample. From

Table 1, it can be seen that when the sum-rate is greater than its minimum value, the optimal SP-DKLT code approaches a WZ transform code, i.e., no source-splitting occurs. This situation, which also exists in Figure 5, suggests that optimal SP-DKLT codes for the sum-rate at which a given D can be achieved, are equivalent to time-sharing [12] of two "corner points". Figure 6 illustrates this situation for optimal SP-DKLT codes at $D = 0.005$ for source model B ($\theta = 0.9$ in Figure 5). It should however be noted that, unlike source-splitting, time-sharing between the two terminals requires synchronization of their encoders [11].

4.2 Design examples

In this section, we focus on the practical design of SP-DKLT codes for a given pair of rates (R_1, R_2) based on both scalar and block-quantization. RD-WZQ model used in Section 3.1.1 implies infinite block-length WZ-VQ of each coefficient. A practically realizable approach to block WZ quantization is SWC-TCQ [20]. Experimental results obtained with LDPC codes of block length up to 10^6 bits and TCQs up to 8,192 states are presented in [20] for quadratic Gaussian WZ



quantization, which indicate that performance very close to the theoretical limit can be achieved with SWC-TCQ. Motivated by these results, we aim to implement SP-DTCs which can approach theoretical performance predicted in Section 3.1.1 using TCQ and SW coding for encoding transform coefficients. However, the SWC-TCQ design procedure followed in [20] is to first design a TCQ whose MSE satisfies a constraint (by choosing a sufficiently high rate) and then to estimate the output conditional entropy (which is the target rate of the SW code) of the resulting TCQ. This is sufficient for verifying the achievable rate pairs for a given MSE which is the goal of [20]. Our problem is different in that the rate of the SW code is specified by the solution to the bit-allocation problem and our goal is to design a TCQ

which minimizes the MSE, subject to a constraint on the output conditional entropy. This requires an alternative formulation of the design procedure, which we refer to as CEC-TCQ. In previous work on non-distributed quantization, entropy constrained TCQ (EC-TCQ) has been investigated in [21-24]. CEC-TCQ is a modification of EC-TCQ in [21,22] to accommodate block SW-coding of the TCQ output relative to a decoder side-information sequence. Our formulation of CEC-TCQ follows the *supersentropy* formulation of EC-TCQ in [22].

Suppose that a sequence of source samples $\{U_n \in \mathbb{R}\}$ has to be quantized, given that the sequence $\{Y_n \in \mathbb{R}\}$, is available at the decoder as side-information, where $n = 1, 2, \dots$ denotes the discrete-time. Similar to an ordinary TCQ [25], a CEC-TCQ uses a size $2^{R_{TCQ}+1}$ scalar

Table 1 Bit allocations found by the tree-search algorithm for 16-dimensional SP-DKLT coding of source model A ($\rho = 0.9, \gamma = 0.9$)

	X ₁ Split				B	X ₂ Split				B
	B' ₁	B'' ₁	B ₁	B ₂		B' ₁	B'' ₁	B ₁	B ₂	
0	14.8	14.8	96	110.8	0	22	22	96	118	
0	15	15	80	95	0	22.4	22.4	80	102.4	
0	15.8	15.8	64	79.8	0	22.6	22.6	64	86.6	
0	19	19	48	67	0	24.4	24.4	48	72.4	
0	22	22	44	66	0	25.6	25.6	44	69.6	
3.2	22.8	26	40	66	0	26.8	26.8	40	66.8	
10	20	30	36	66	0	30	30	36	66	
16.6	16.4	33	33	66	2.8	30.2	33	33	66	
18.6	15.4	34	32	66	4.2	29.8	34	32	66	
23.2	12.8	36	30	66	7.6	28.4	36	30	66	
34.2	5.8	40	27.2	67.2	17.4	22.6	40	26	66	
44	0	44	25.8	69.8	32	12	44	22	66	
48	0	48	24.4	72.4	43.8	4.2	48	19.4	68.4	
64	0	64	22.6	86.6	64	0	64	16	90	
80	0	80	22.2	102.2	80	0	80	15.2	95.2	
96	0	96	22	118	96	0	96	14.8	110.8	

codebook to quantize the input sequence U_1, U_2, \dots , into a R_{TCQ} bits/sample output sequence $\hat{U}_1, \hat{U}_2, \dots$. However, CEC-TCQ output satisfies the additional property that the conditional entropy $H(\hat{U}_n|Y_n) = E\{-\log_2 P(\hat{U}_n|Y_n)\} \leq R$ for some given R . It follows from [9] that if the CEC-TCQ output is SW-coded relative to the decoder side-information sequence $\{Y_n\}$ then LR bits are sufficient to (almost) losslessly transmit a sequence of L source samples as $L \rightarrow \infty$. The optimal CEC-TCQ minimizes the MSE $E\{(U_n - \hat{U}_n)^2\}$, subject to the constraint $E\{-\log_2 P(\hat{U}_n|Y_n)\} \leq R$, or equivalently, minimizes the Lagrangian

$$J = \lim_{L \rightarrow \infty} \sum_{n=1}^L \left[E\{(U_n - \hat{U}_n)^2\} + \beta E\{-\log_2 P(\hat{U}_n|Y_n)\} \right] \quad (24)$$

where $\beta > 0$ is the Lagrange multiplier. This implies that, given a specific sequence of input samples u_1, u_2, \dots , the CEC-TCQ encoder should use the Viterbi algorithm based on the path-cost function

$$\sum_n \left[(u_n - \hat{u}_n)^2 + \beta E\{-\log_2 P(\hat{u}_n|Y_n)\} \right]. \quad (25)$$

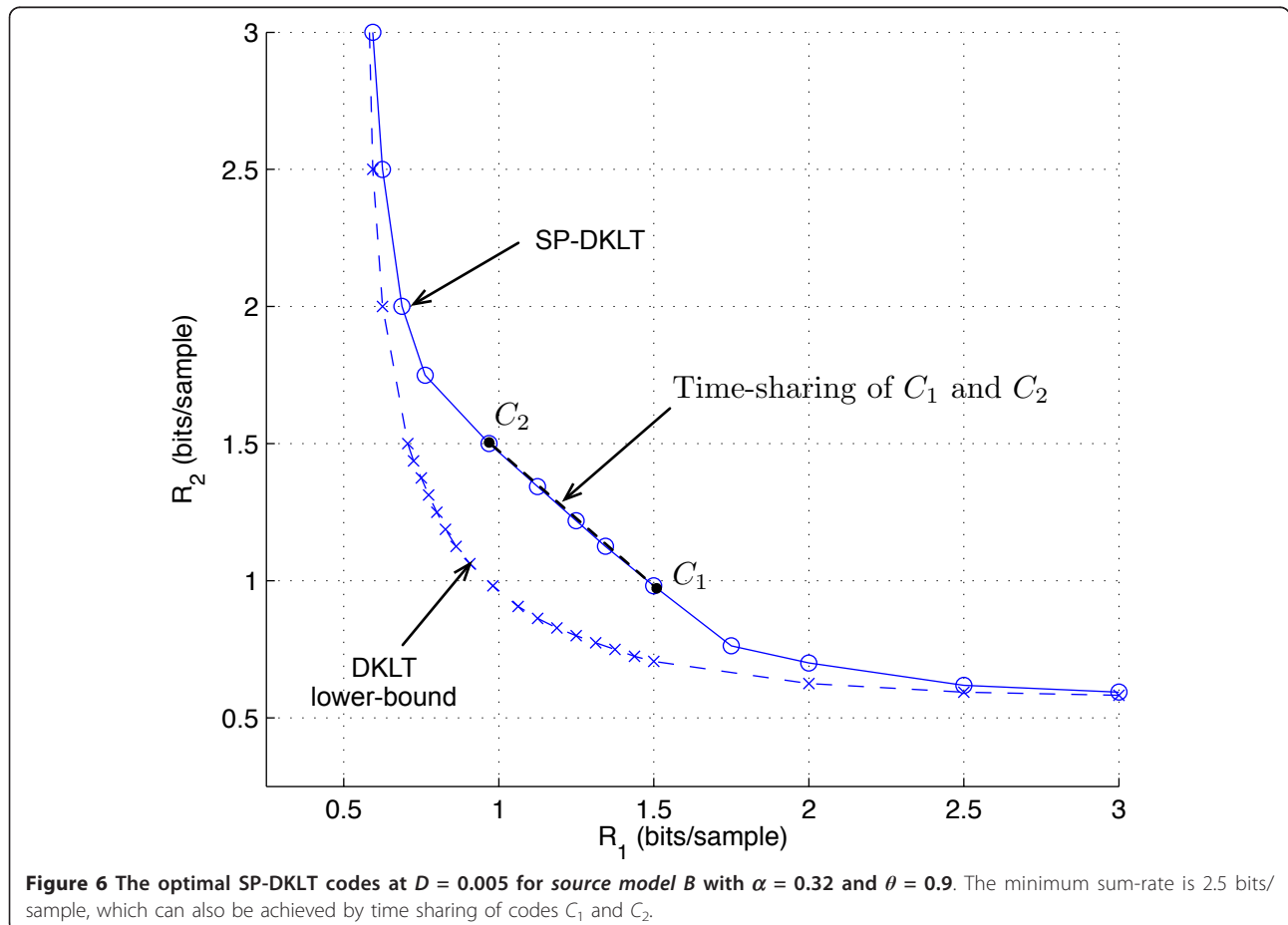


Figure 6 The optimal SP-DKLT codes at $D = 0.005$ for source model B with $\alpha = 0.32$ and $\theta = 0.9$. The minimum sum-rate is 2.5 bits/sample, which can also be achieved by time sharing of codes C_1 and C_2 .

In a rate R_{TCQ} TCQ, each codeword c_k , $k = 1, \dots, 2^{R_{\text{TCQ}}+1}$, in the codebook is labeled with R_{TCQ} -bit binary string [25]. Let b_i , $i = 1, \dots, 2^{R_{\text{TCQ}}}$ be these binary labels. Then, to compute (25), the cost $\beta E\{-\log_2 P(b_i|Y)\}$ must also be stored for each binary label, where Y is the random variable representing the decoder side-information. For a fixed value of β , we can use a slight modification of the algorithm in [21] for optimizing the CEC-TCQ codebook, by replacing the codeword entropies $E\{-\log_2 P(b_i)\}$ by the conditional entropies $E\{-\log_2 P(b_i|Y)\}$, and by using a training sequence of (U_n, Y_n) pairs. In order to approximate the expectations by sample averages computed from training data, the side-information variable Y is discretized to $\hat{Y} \in \{\eta_1, \dots, \eta_{\mathcal{Y}}\}$, where \mathcal{Y} is a large enough positive integer. Then $E\{-\log_2 P(B = b_i | Y)\}$ can be approximated by

$$\hat{H}(b_i) = - \sum_{k=1}^{\mathcal{Y}} \log_2 [P(B = b_i | \hat{Y} = \eta_k)] P(\hat{Y} = \eta_k | B = b_i) \quad (26)$$

where $B \in \{b_1, \dots, b_{2^{R_{\text{TCQ}}}}\}$ is the binary-labeled output of the TCQ. Given a TCQ code-book, the probabilities $P(B = b_i | \hat{Y} = \eta_k)$ and $P(\hat{Y} = \eta_k | B = b_i)$ can be estimated using the training set [it is sufficient to estimate $p_{i,k} = P(B = b_i, \hat{Y} = \eta_k)$, $k = 1, \dots, \mathcal{Y}, i = 1, \dots, N$]. To complete the design, it is necessary to search for the value of β for which $E\{-\log_2 P(\hat{U}_n | Y_n)\} \approx R$ by repeating the codebook optimization for an appropriately chosen sequence of β values.

For block WZ-code designs, the transforms and the bit allocations are found by RD-WZQ model (Section 3.1.1) and WZ quantizers are implemented using CEC-TCQ followed by binary SW coding. More specifically, the rate found by the bit allocation algorithm for each transform coefficient is used as the conditional entropy constraint in the design of a CEC-TCQ for that coefficient. As described in Section 2.3, the CEC-TCQ designs are based on scalar-side information obtained by a linear transform of the vector side-information at the decoder, see *Theorem 3*. All CEC-TCQ designs are based on the 8-state trellis used in JPEG2000 [[26], Figure 3.16]. For trellis encoding and decoding, a sequence length of 256 source samples has been used. For design and testing quantizers, sample sequences of length 5×10^5 have been used. Since, the main focus this paper is the design of transforms and the quantizers, we assume ideal SW coding of the binary output of each CEC-TCQ, so that our results do not depend on any particular SW coding method. In a practical implementation (e.g., [20]), near optimal performance can be obtained by employing a sufficiently long SW code (note that sequence length for SW-coding can be chosen arbitrary larger than the sequence length used for TCQ encoding). This type of coding is well suited for applications such as distributed

image compression, where the coding is inherently block-based.

We also consider SP-DKLT code designs based on scalar quantization. In this case, the transforms and bit allocations are found by using the SWC-HRSQ model (Section 3.1.2). While it is possible to use the step-size predicted by SWC-HRSQ model to design uniform quantizers, we found that such quantizers in reality do not satisfy the required entropy constraint at lower rates. We instead use conditional entropy constrained scalar quantizers (CEC-SQ), designed by modifying the algorithm in [27] to accommodate a conditional entropy constraint similar to CEC-TCQ approach above.

The reconstruction signal-to-noise ratio (RSNR) [with MSE as given by (6)] of SP-DKLT code designs for *source model B* is shown in the rows labeled *Design* in Table 2 where SP-DKLT/CECSQ and SP-DKLT/CECTCQ refer to scalar quantization and TCQ based designs respectively. The rows labeled *Analytical* show the performance predicted by the SWC-HRSQ and RD-WZQ models upon which the transforms and bit-allocations are based (note however that the performance predicted by SWC-HRSQ model is not necessarily an upper-bound for CEC-SQ designs which are not constrained to be uniform quantizers). We compare the performance of our SP-DKLT code designs with IKLT codes for which the bit-allocations are obtained by using either entropy coded high-rate quantization model (for scalar quantizer design) [[4], Section 9.9] or RD-optimal quantization model (for block quantizer design) [[16], Section 10.3.3] for Gaussian variables. The IKLT codes with scalar quantization have been implemented by using entropy constrained scalar quantizers (EC-SQ) while those with block quantizers have been implemented by using EC-TCQ [23], where we assume ideal entropy coding of the quantizer outputs. In Table 2,

Table 2 RSNR (in dB) of KLT-based transform code designs for quantizing 16-dimensional vectors ($M_1 = M_2 = 16$) of source model B ($\alpha = 0.32$, $\theta = 0.9$) at $R_1 = R_2 = R$ bits/sample

R (bits/sample)	IKLT		SP-DKLT		
	EC-SQ	EC-TCQ	CEC-SQ	CEC-TCQ	
0.5	10.3	11.7	12.4	14.0	(Analytical)
	10.3	11.0	11.6	12.7	(design)
1	16.6	18.0	19.1	20.6	
	16.4	17.3	18.6	19.6	
1.5	21.6	23.0	23.8	25.6	
	21.7	22.5	23.3	24.4	
2	25.7	27.1	28.2	29.8	
	25.2	26.6	27.8	28.6	

Analytical values refer to SNR of the quantization model assumed for determining the optimal transforms and the bit-allocation

IKLT/ECSQ and IKLT/ECTCQ respectively refer to these designs.

From a practical view point, the use of DCT instead of KLT is interesting [26]. We therefore consider the design of SP-DTCs based on the DCT, referred to as source-split distributed DCT (SP-DDCT) codes. Since DCT is a fixed transform, we only need to optimize the bit allocations. To do this, we assume that DCT is approximately a decorrelating transform for Gaussian vectors [4]. Then, the bit-allocations given by (9), (14), and (20) are still valid provided that the eigenvalues in these expressions are replaced by the variances of the corresponding DCT coefficients. The rest of the design procedure is the same as that with SP-DKLT. The RSNR of DCT based designs are presented in Table 3 (again, the performance predicted by SWC-HRSQ model is not an upper bound for corresponding practical codes). The results show that, for this particular source model, even the scalar quantization based SP-DDCT codes outperform TCQ-based IDCT codes.

5 Concluding remarks

Rate-distortion analysis and experimental results demonstrate that SP-DTC is a promising practical approach to implementing distributed VQ of high-dimensional correlated vectors. The comparisons shown in Table 2, as well as similar comparisons for *source model A* and the source model in [[8], Example 6], indicate that these codes can substantially outperform the independent transform codes, when there is sufficient inter-vector correlation. This approach has also been demonstrated to be effective for DCT-based systems. Therefore, the proposed approach can be potentially used in applications such as stereo image compression when inter-camera communication is impractical. Our RD analysis however indicates that the achievable rate-region of SP-DKLT codes for jointly Gaussian sources is

Table 3 SNR (in dB) of DCT-based transform code designs for quantizing 16-dimensional vectors ($M_1 = M_2 = 16$) of source model B ($\alpha = 0.32$, $\theta = 0.9$) at $R_1 = R_2 = R$ bits/sample

R (bits/sample)	IDCT		SP-DDCT		
	EC-SQ	EC-TCQ	CEC-SQ	CEC-TCQ	
0.5	7.9	9.1	11.7	12.6	(Analytical)
	8.0	8.4	11.0	12.0	(design)
1	11.9	13.2	15.9	17.8	
	11.9	12.6	16.6	17.0	
1.5	15.3	16.8	19.7	21.9	
	15.3	16.1	20.5	21.3	
2	18.9	20.3	23.6	25.6	
	18.7	19.4	24.5	25.1	

Analytical values refer to SNR of the quantization model assumed for determining the bit-allocation.

strictly inside that predicted by the DKLT of [8]. An interesting avenue of future work is to find implementable distributed transforms codes which can achieve the rate-pairs below the “time-sharing” line in Figure 6. Another issue is the extension of the proposed approach to more than two vector sources. In principle, source-splitting can be easily applied to more than two sources. However, with more than two vector sources, the complexity of the bit-allocation will be significantly higher.

1 Appendix

1.1 Proof of Theorem 2

The optimality of CKLT is proved in [5]. To prove the optimality of the bit allocation, consider high-rate scalar quantization of the transform coefficient U_m and ideal SW coding of the quantizer output \hat{U}_m at the rate $r_m = H(\hat{U}_m | \mathbf{Y})$ bits/sample, $m = 1, \dots, M_1$, where $H(\cdot | \cdot)$ denotes the conditional entropy [16]. In this case, the asymptotically optimal scalar quantizer for each coefficient is known to be uniform [5]. For high-rate uniform quantization, $H(\hat{U}_m | \mathbf{Y}) \approx h(U_m | \mathbf{Y}) - \log_2(\Delta_m)$, where Δ_m is the quantizer step-size and $h(\cdot | \cdot)$ is the conditional differential entropy [[16], Section 8.3]. Since the conditional variance of the transform coefficient U_m , given the side-information \mathbf{Y} is $E\{U_m^2 | \mathbf{Y}\} = \lambda_m$, and (U_m, \mathbf{Y}) are jointly Gaussian, it follows that $h(U_m | \mathbf{Y}) = (1/2) \log_2(2\pi e \lambda_m)$ [16], and hence $\Delta_m = \sqrt{2\pi e \lambda_m} 2^{-r_m}$. Therefore, the MSE of high-rate uniform quantization followed by ideal SW coding of U_m is

$$d_m^{\text{SWC-HRSQ}} = \frac{\Delta_m^2}{12} = \left(\frac{\pi e}{6}\right) \lambda_m 2^{-2r_m}. \quad (27)$$

Since (27) and (1) are the same within a constant factor of $\pi e/6$ (which is identical for all transform coefficients), it is easy to verify that the optimal bit-allocation solution under SWC-HRSQ model is also given by (2). However, the MSE of m th coefficient in this case is

$$d_m = \begin{cases} \left(\frac{\pi e}{6}\right) \lambda_m 2^{-2\rho_m} \\ \left(\frac{\pi e}{6}\right) d^*(\lambda, B, N), & m = 1, \dots, N \\ \lambda_m, & m = N + 1, \dots, M \end{cases} \quad (28)$$

and hence the overall MSE is given by (5). Furthermore,

$$\Delta_m = \Delta = \sqrt{(2\pi e) d^*(\lambda, B, N)}$$

for $m = 1, \dots, N$.

1.2 Proof of Theorem 3

For jointly Gaussian and mean-zero \mathbf{X} and \mathbf{Y} , there exists a matrix \mathbf{A} and a mean-zero Gaussian vector \mathbf{W}_1 independent of \mathbf{Y} such that $\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{W}_1$, where

$\sum_{W_1} = \sum_{X|Y} = \mathbf{TAT}^T$, $\sum_{W_1} = \sum_{X|Y} = \mathbf{TAT}^T$, and Λ is the diagonal matrix of eigenvalues of $\sum_{X|Y}$. Furthermore $\mathbf{U} = \mathbf{T}^T \mathbf{A} \mathbf{Y} + \mathbf{W}_2$, where $\mathbf{W}_2 = \mathbf{T}^T \mathbf{W}_1$ is an uncorrelated Gaussian vector since $\sum_{W_2} = \mathbf{T}^T \sum_{W_1} \mathbf{T} = \Lambda$ (note that for CKLT, $\mathbf{T}^{-1} = \mathbf{T}$). Therefore $\sum_{U|Y} = \Lambda$. The MMSE estimate of \mathbf{U} given \mathbf{Y} is $\hat{\mathbf{U}} = E\{\mathbf{U}|\mathbf{Y}\} = \mathbf{T}^T E\{\mathbf{X}|\mathbf{Y}\} = \mathbf{T}^T \mathbf{A} \mathbf{Y}$. Thus, $\mathbf{U} = \hat{\mathbf{U}} + \mathbf{W}_2$ and $\sum_{U|Y} = \Lambda$, and it follows that U_i is independent of \tilde{U}_j if $j \neq i$ and $\text{var}(U_i | \tilde{U}_i) = \text{var}(U_i | \mathbf{Y})$, where $\text{var}(\cdot | \cdot)$ denotes the conditional variance. Now, since $h(U_i | \mathbf{Y}) = h(U_i | \tilde{U}_i)$, we conclude that \tilde{U}_i is a sufficient statistic [16] for decoder side-information \mathbf{Y} in WZ quantization of U_i .

Competing interests

The author declares that they have no competing interests.

Received: 21 October 2011 Accepted: 5 April 2012

Published: 5 April 2012

References

1. Z Xiong, AD Liveris, S Cheng, Distributed source coding for sensor networks. *IEEE Signal Process Mag.* **21**(5), 80–94 (2004)
2. JY Huang, P Schultheiss, Block quantization of correlated random vectors. *IEEE Trans Commun.* **11**(9), 289–296 (1963)
3. V Goyal, Theoretical foundations of transform coding. *IEEE Signal Process Mag.* **18**(5), 9–21 (2001)
4. A Gersho, RM Gray, *Vector Quantization and Signal Compression*, (Kluwer Academic Publishers, Norwell MA, USA, 1992)
5. D Rebollo-Monedero, S Rane, A Aaron, B Girod, High-rate quantization and transform coding with side information at the decoder. *Signal Process.* **88**(11), 3160–3179 (2006)
6. A Vosoughi, A Scaglione, Precoding and decoding paradigms for distributed vector data compression. *IEEE Trans Signal Process.* **55**(4), 1445–1459 (2007)
7. X Chen, E Tuncel, Low-delay prediction and transform-based Wyner-Ziv coding. *IEEE Trans Signal Process.* **59**(2), 653–666 (2011)
8. M Gastpar, PL Dragotti, M Vetterli, The distributed Karhunen-Loève transform. *IEEE Trans Inf Theory.* **52**, 5177–5196 (2006)
9. D Slepian, JK Wolf, Noiseless coding of correlated information sources. *IEEE Trans Inf Theory.* **19**(4), 471–480 (1973)
10. H Gish, JP Peirce, Asymptotically efficient quantizing. *IEEE Trans Inf Theory.* **14**, 676–683 (1968)
11. B Rimoldi, R Urbanke, Asynchronous Slepian-Wolf coding via source-splitting, in *IEEE Int Symp Inform Theory (ISIT)*, Ulm, Germany, p. 271 (1997)
12. R Zamir, S Shamai, U Erez, Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans Inf Theory.* **48**(6), 1250–1276 (2002)
13. AB Wagner, S Tavildar, P Viswanath, Rate region of the quadratic Gaussian two-encoder source-coding problem. *IEEE Trans Inf Theory.* **54**(5), 1938–1961 (2008)
14. AD Wyner, J Ziv, The rate-distortion function for source coding with side information. *IEEE Trans Inf Theory.* **22**, 1–10 (1976)
15. Y Yang, V Stankovic, Z Xiong, W Zhao, On multiterminal source code design. *IEEE Trans Inf Theory.* **54**(5), 2278–2302 (2008)
16. TM Cover, JA Thomas, *Elements of Information Theory*, (John Wiley, New York, USA, 1991)
17. D Marco, DL Neuhoff, The validity of additive noise model for uniform scalar quantizers. *IEEE Trans Inf Theory.* **51**(5), 1739–1755 (2005)
18. T Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, (Prentice-Hall, Englewood Cliffs NJ, USA, 1971)
19. MC Vuran, OB Akan, IF Akyildiz, Spatio-temporal correlation: theory and applications for wireless networks. *Comput Netw.* **45**, 245–259 (2004)
20. Y Yang, S Cheng, Z Xiong, W Zhao, Wyner-Ziv coding based on TCQ and LDPC codes. *IEEE Trans Commun.* **57**(2), 376–387 (2009)
21. TR Fischer, M Wang, Entropy-constrained trellis-coded quantization. *IEEE Trans Inf Theory.* **38**(2), 415–426 (1992)
22. MW Marcellin, On entropy-constrained trellis-coded quantization. *IEEE Trans Commun.* **42**, 14–16 (1994)
23. MW Marcellin, Transform coding of images using trellis-coded quantization, in *Proc ICASSP*, Albuquerque NM, USA, pp. 2241–2244 (1990)
24. N Farvardin, X Ran, CC Lee, Adaptive DCT coding of images using entropy-constrained trellis-coded quantization, in *Proc ICASSP*, (Minneapolis MN, USA, 1993), pp. 397–400
25. MW Marcellin, TR Fischer, Trellis coded quantization memoryless and Gauss-Markov sources. *IEEE Trans Commun.* **38**, 82–93 (1990)
26. DS Taubman, MW Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, (Kluwer Academic Publishers, Norwell MA, USA, 2004)
27. P Chou, T Lookabaugh, RM Gray, Entropy-constrained vector quantization. *IEEE Trans Acoust Speech Signal Process.* **37**, 31–42 (1989)

doi:10.1186/1687-6180-2012-78

Cite this article as: Yahampath: Distributed transform coding via source-splitting. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:78.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com