Orthonormal Matrix Codebook Design for Adaptive Transform Coding

Data Compression Conference
March 22, 2022

Rashmi Boragolla, Pradeepa Yahampath

Department of Electrical and Computer Engineering

University of Manitoba

Contents

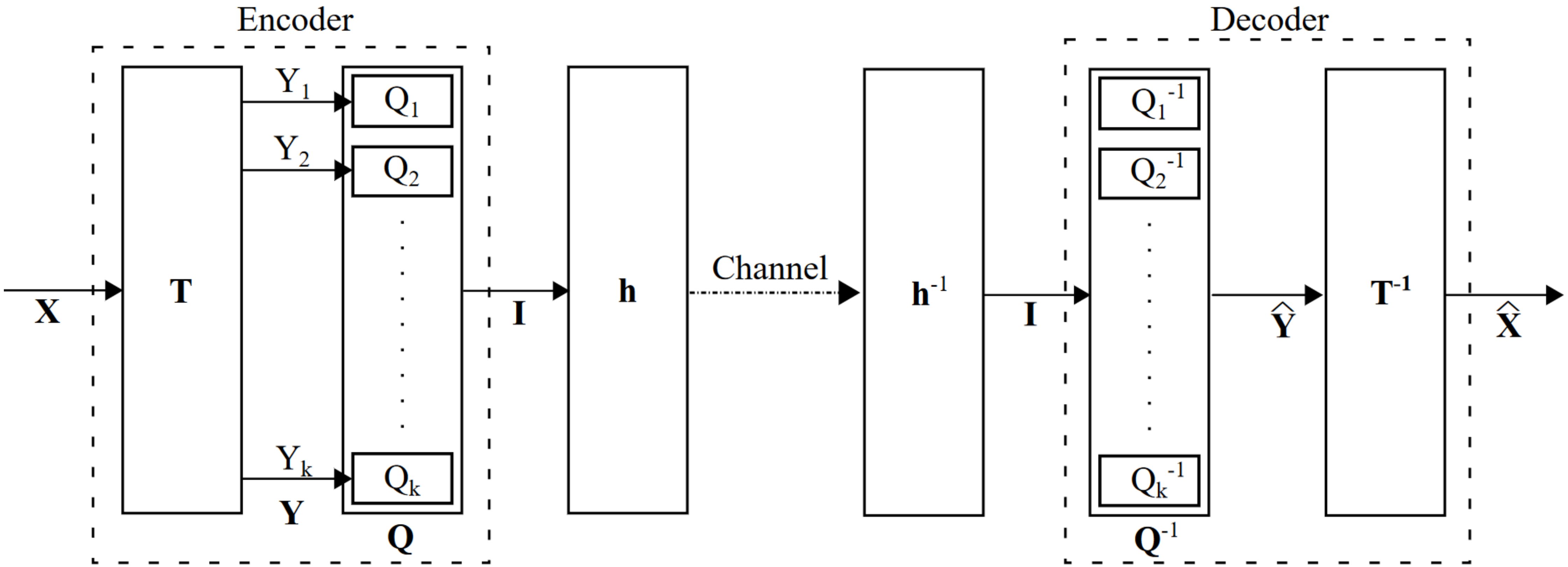
University of Manitoba

Background and Motivation

- Transform coding : Linearly transform random vectors to obtain transform coefficients that would be scaler quantized and entropy coded to achieve compression.
- The optimal transform for coding stationary Gaussian sources based on the mean squared error (MSE) is the Karhunen-Loève transform (KLT).
- In real-world applications, the source data is highly non-stationary, however the tendency has been to use generic fixed-transforms such as DCT.
- Research on adaptive transform coding has shown to yield significant gains over DCT.
- Adaptive transform coding : Linear transform is adapted to signal statistics.
- Previous work on adaptive transform coding
  - specialized to signals with certain characteristics, such as images containing various directional properties and video residuals produced by intra-prediction or motion-compensation
  - more general transforms are designed using a sparsity constraint as a surrogate for a rate constraint
  - do not consider the effect of quantization errors due to finite rate compression.

Transform Coding Framework

- $\boldsymbol{X} \in \mathbb{R}^k$ is a stationary random vector, $\boldsymbol{T} \in \mathbb{R}^{k \times k}$ is an orthonormal matrix, $\boldsymbol{Y} \in \mathbb{R}^k$ is the transform coefficients, $\boldsymbol{Q}$ is the scalar quantizer with quantization step sizes $\Delta = [\Delta_1, \dots, \Delta_k]^T$, $\boldsymbol{h}$ is the entropy coder, $\widehat{\boldsymbol{Y}} \in \mathbb{R}^k$ is the quantized version of $\boldsymbol{Y}$ and $\widehat{\boldsymbol{X}}$ is the reconstructed source vector.



Transform Coding Error

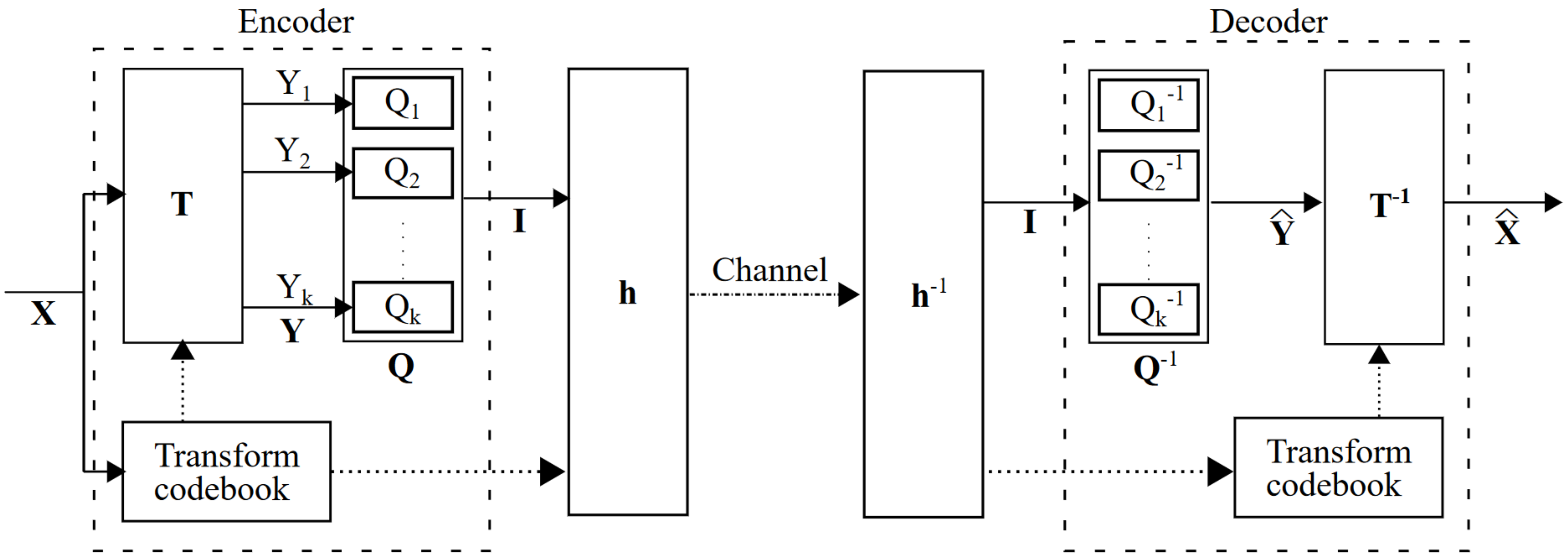
- Assume that each transform coefficient is zero mean and that the quantization MSE is a function of the quantizer's input variance $\sigma_{Y_i}^2 = E[Y_i^2]$.
- The quantization MSE of the coefficient Y_i
$$\theta\left(\sigma_{Y_i}^2, \Delta_i\right) = E\left(Y_i - \hat{Y}_i\right)^2$$
- MSE of transform coding $\boldsymbol{X}$ is given by
$$\Theta(\boldsymbol{T}, \boldsymbol{C_X}) = E\left(\boldsymbol{X} - \widehat{\boldsymbol{X}}\right)^2 = \sum_{i=1}^{k} \theta\left(\sigma_{Y_i}^2, \Delta_i\right)$$

where $\boldsymbol{C_X}$ is the covariance matrix of $\boldsymbol{X}$.

- Optimal transform
$$\boldsymbol{T}^* = \arg\min_{\boldsymbol{T}} \Theta(\boldsymbol{T}, \boldsymbol{C_X})$$

such that $\boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{I_k}$, where $\boldsymbol{I_k}$ is the identity matrix.

Adaptive Transform Coding Framework



Problem Statement

- Let
  - $B \in \mathbb{S}_B$ be a locally stationary block of vectors in a non-stationary process, where $\mathbb{S}_B$ is the set of all possible blocks. $\boldsymbol{C}_B$ be the covariance matrix for $\boldsymbol{X} \in B$, and $\mathbb{S}_C$ be the ensemble of covariance matrices
- If $\boldsymbol{X}$ is a Gaussian vector, then the optimal transform for stationary block B is the KLT of $\boldsymbol{C}_B$.
  - The main difficulty is KLT undo the advantage gained by requiring an additional bit-rate overhead to encode data-dependent transform matrices.
- Adaptive transform coding : Use a codebook of orthonormal transform matrices $\mathcal{T} = \{\widetilde{\boldsymbol{T}}_1, \ldots, \widetilde{\boldsymbol{T}}_N\}$ and select the optimal transform matrix for B as

$$\boldsymbol{T}^* = \underset{\boldsymbol{T}\in\mathcal{T}}{\arg\min} \frac{1}{|B|} \sum_{\boldsymbol{X}\in B} \left\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\right\|^2 = \underset{\boldsymbol{T}\in\mathcal{T}}{\arg\min}\, \Theta(\boldsymbol{T}, \boldsymbol{C}_B)$$

where $\widehat{\boldsymbol{X}}$ is the reconstructed version of source vector $\boldsymbol{X}$ using the transform $\boldsymbol{T}$ and $|B|$ is the number of vectors in B.

University of Manitoba

Transform Matrix Codebook Optimization

- Model the non-stationary process by a block-wise stationary vector process and encode each stationary block of vectors using a single transform optimized to local statistics.
- Optimal partitioning of $\mathbb{S}_B$, $\boldsymbol{\Omega}^* = \{\Omega_1^*, \ldots, \Omega_N^*\}$ for a fixed $\mathcal{T}$
$$\Omega_i^* = \left\{\boldsymbol{C}_B \in \mathbb{S}_C : \Theta\left(\widetilde{\boldsymbol{T}}_i, \boldsymbol{C}_B\right) < \Theta\left(\widetilde{\boldsymbol{T}}_j, \boldsymbol{C}_B\right) \; \forall \; i \neq j\right\} \qquad (1)$$
- Optimal codebook $\mathcal{T}^*$ for fixed partition $\boldsymbol{\Omega} = \{\Omega_1, \ldots, \Omega_N\}$ of $\mathbb{S}_B$
$$\boldsymbol{T}^* = \underset{\boldsymbol{T} \in \mathbb{R}^{k \times k}}{\arg\min} \, \mathrm{E}[\Theta(\boldsymbol{T}, \boldsymbol{C}_B) | \boldsymbol{C}_B \in \mathbb{S}_C], \qquad \textit{subject to} \quad \boldsymbol{T}^T \boldsymbol{T} = \boldsymbol{I}_{\boldsymbol{k}} \qquad (2)$$
- $\boldsymbol{T}^*$ should be orthonormal and hence the solution space is the set of all $k \times k$ orthogonal matrices $O(k)$.
- Rather than solving a constrained minimization problem in (2) on the Euclidian space $\mathbb{R}^{k \times k}$, we use low-complexity, modified steepest descent algorithm on $O(k)$ [1] to solve (2) as an unconstrained minimization problem on $O(k)$.
- This algorithm requires that the objective function in (2) be differentiable
- We propose two alternatives

University of Manitoba

[1] J. H. Manton, "Optimization algorithms exploiting unitary constraints," IEEE Trans. Signal Process., vol. 50, no. 3, pp. 635–650, Aug. 2002.

Mean Square Error Modeling - Models for $\Theta(\boldsymbol{T}, \boldsymbol{C}_B)$

1. **High-rate Gaussian model**
   - Given a target rate R_0bits/vector, the minimum MSE of transform coding a Gaussian vector

$$\Theta(\boldsymbol{T}, \boldsymbol{C}_B) = \frac{k\pi e}{6}\left(\prod_{i=1}^{k} \sigma_{Y_i}^2\right)^{\frac{1}{\mathrm{k}}} 2^{-\frac{2R_o}{k}}$$

2. **Laplacian model**
   - MSE of quantizing a mean-zero Laplace variable with variance σ^2 using a uniform quantizer with a dead zone $(-{}^{z}/_{2}, {}^{z}/_{2})$, quantization step-size Δ and $b = \sqrt{\sigma^2/2}$

$$\theta(\sigma^2, \Delta, z) = 2b^2 - e^{z/2b}\left(\frac{z^2 - \Delta^2}{4} + zb + \Delta b \frac{(e^{\Delta/b} + 1)}{(e^{\Delta/b} - 1)}\right)$$

   - Therefore

$$\Theta(\boldsymbol{T}, \boldsymbol{C}_B) = \sum_{i=1}^{k} \theta\left(\sigma_{Y_i}^2, \Delta_{\mathrm{i}}, z_i\right)$$

University of Manitoba

Algorithm for Transform Codebook Design

Input	:	A training set of covariance matrices $\tilde{S}$
Parameters	:	A tolerance parameter $\varepsilon > 0$ Maximum allowed iterations M
Initialize	:	An initial codebook of orthonormal matrices $\mathcal{T}^{(0)} = \{\tilde{T}_1^{(0)}, \ldots, \tilde{T}_N^{(0)}\}$ Iteration index $t \leftarrow 1$

while $\dfrac{\overline{\Theta}^{(t-1)} - \overline{\Theta}^{(t)}}{\overline{\Theta}^{(t-1)}} \geq \varepsilon$ or $t < M$

Step 1 : Given $\mathcal{T}^{(t-1)}$ partition $\tilde{S}$ into N subsets $\{\Omega_1^{(t)}, \ldots, \Omega_N^{(t)}\}$

Step 2 : Given $\{\Omega_1^{(t)}, \ldots, \Omega_N^{(t)}\}$ find the optimal transform codebook $\mathcal{T}^{(t)}$

Step 3 : Estimate by sample averaging $\overline{\Theta}^{(t)}$

end

University of Manitoba

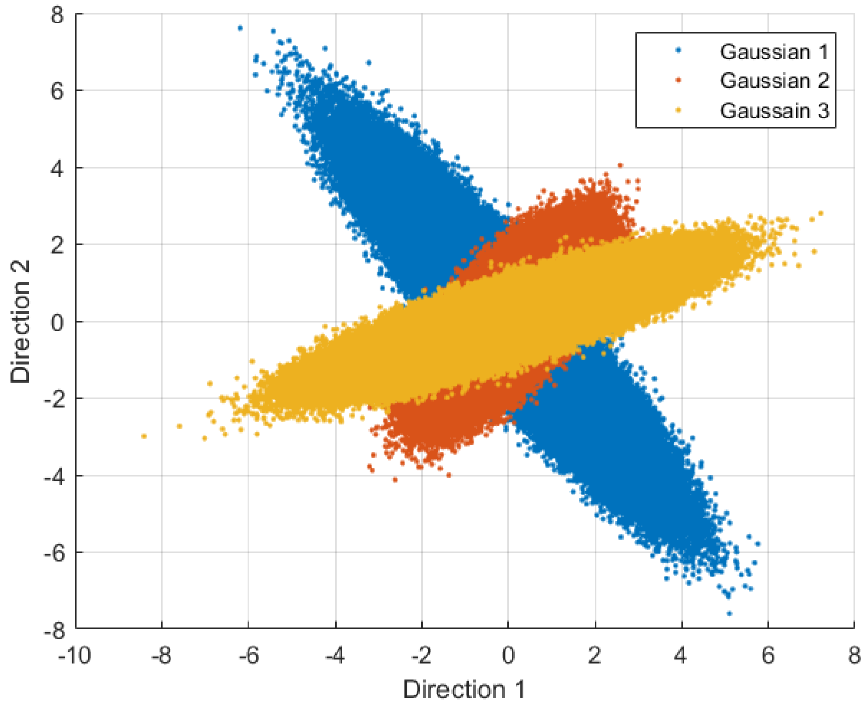
A Toy Example

- Optimize a single transform matrix for 2-D vectors drawn from a Gaussian mixture with 3 mean-zero components whose covariance matrices are

$$C_1 \begin{bmatrix} 1.54 & -1.84 \\ -1.84 & 2.62 \end{bmatrix}, C_2 = \begin{bmatrix} 0.46 & 0.40 \\ 0.40 & 0.70 \end{bmatrix},$$

$$C_3 = \begin{bmatrix} 2.22 & 0.77 \\ 0.77 & 0.38 \end{bmatrix}$$

- Our design algorithm finds a transform matrix noticeably better than the KLT and DCT.



Transform	SNR (dB)	Entropy (bits/sample)
KLT	3.21	0.71
DCT	3.69	0.63
High-Rate	4.0	0.59
Laplace	4.0	0.59

[High-rate and Laplace respectively refer to transforms optimized using the high-rate Gaussian model and the Laplacian model dataset.]

University of Manitoba

Preliminary Results for Motion-compensated (MC) Video Residuals

- Considered a set of 9 standard CIF resolution 30 fps gray-scale video sequences (Bus, Coastguard, Crew, Football, Foreman, Mobile, Soccer, Stefan, Tennis), to generate MC residuals from HM test model.
- Transform coding of 4×4 pixel blocks have been considered to keep the computational complexity low as we are dealing with non-separable transforms.
- To estimated a single covariance matrix, residual frames are divided into 16×16 non-overlapping blocks and a set of time-aligned spatially stationary blocks in 8 adjacent frames have been considered as a spatio-temporal stationary block.
- Various codebook designs were tested using a separate set of 7 video sequences.
- Since the DCT will be good for some stationary blocks, we included the DCT as an additional codeword, after designing a codebook.

University of Manitoba

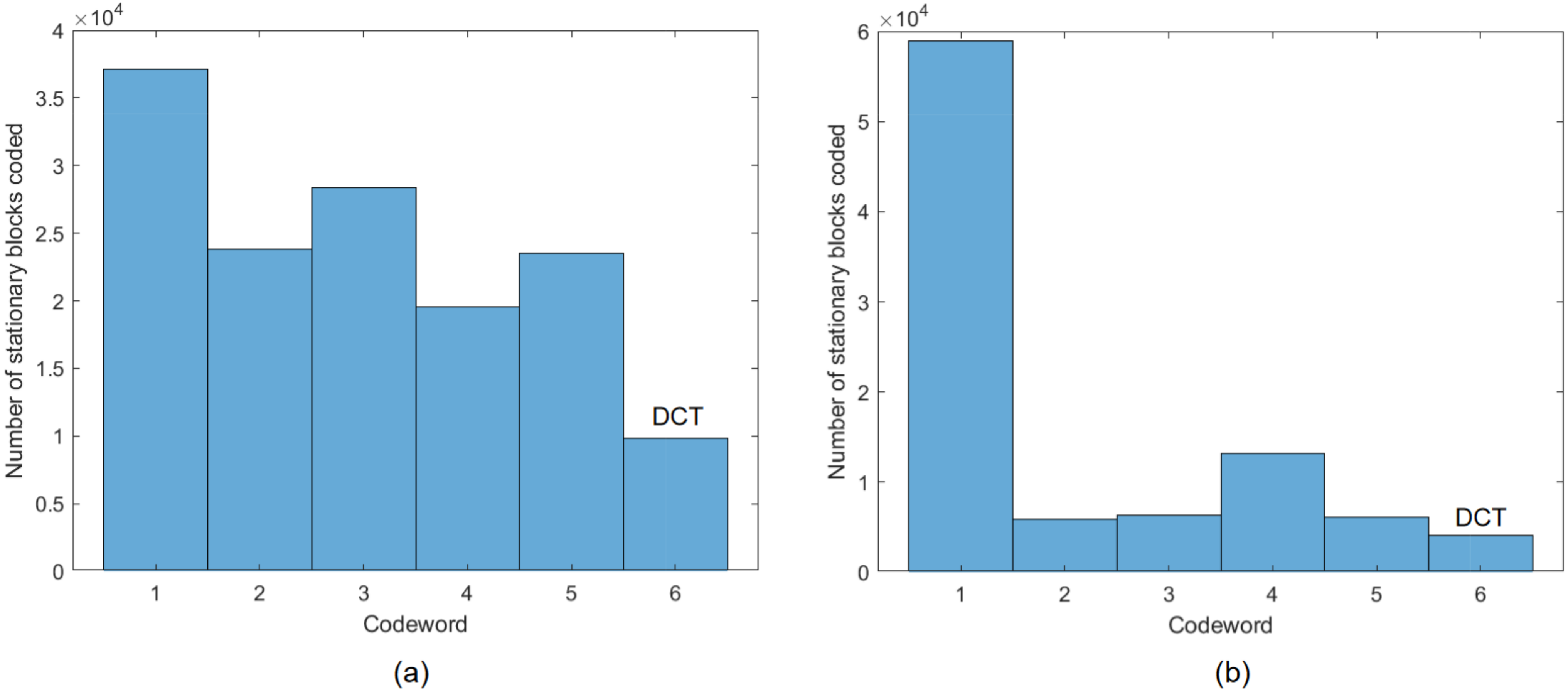
BD-PSNR and BD-Rate Gains

Sequence	Codebook size N											
	3				6				9			
	BD-PSNR (dB)		BD-Rate (%)		BD-PSNR (dB)		BD-Rate (%)		BD-PSNR (dB)		BD-Rate (%)	
	High-Rate	Laplace	High-Rate	Laplace	High-Rate	Laplace	High-Rate	Laplace	High-Rate	Laplace	High-Rate	Laplace
Akiyo	0.096	0.148	-2.86	-5.10	0.122	0.185	-3.60	-5.39	0.135	0.192	-3.94	-5.43
City	0.096	0.233	-1.44	-3.80	0.180	0.332	-2.75	-5.14	0.180	0.343	-2.74	-5.35
Flower	-0.009	0.152	0.15	-1.44	0.048	0.221	-0.41	-2.04	0.048	0.211	-0.39	-1.88
Hall Monitor	0.137	0.250	-3.44	-6.49	0.187	0.313	-4.46	-7.46	0.185	0.323	-4.46	-7.65
Ice	0.157	0.393	-2.91	-7.31	0.180	0.467	-3.27	-8.27	0.179	0.465	-3.29	-8.28
Mother daughter	0.224	0.290	-5.83	-8.66	0.259	0.367	-6.61	-9.08	0.264	0.379	-6.74	-9.58
Waterfall	0.152	0.239	-1.86	-3.53	0.281	0.356	-4.10	-5.00	0.292	0.354	-4.30	-5.17
Average	0.122	0.244	-2.60	-5.19	0.180	0.320	-3.60	-6.05	0.183	0.324	-3.69	-6.19

BD-PSNR and BD-Rate gains achieved by transform matrix codebooks over the standard DCT.

University of Manitoba

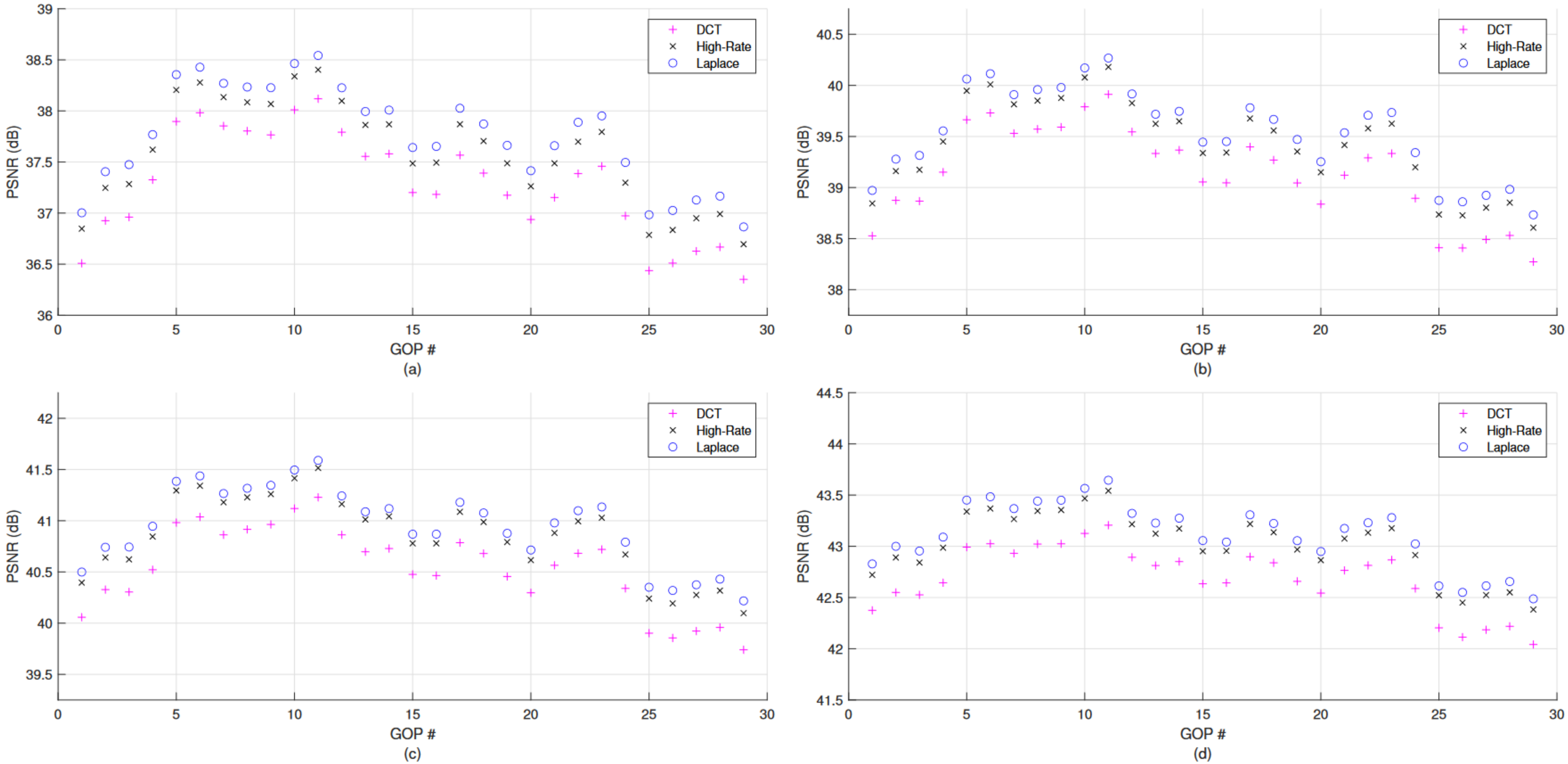
Advantage of Adaptive Transforms over the DCT



Histograms of transform-matrix codeword usage in (a) Football and (b) Ice sequences. The size of the codebook N = 6. The first 5 codewords have been optimized using the proposed algorithm, using the Laplacian MSE model (codeword 6 is the standard DCT.)

University of Manitoba

Adaptive Transform Coding vs DCT



PSNR of coding the Ice sequence using adaptive transforms (codebook size N = 6) and the DCT (non-adaptive). (a) 0.45 bits/pixel, (b) 0.75 bits/pixel, (c) 1.02 bits/pixel and (d) 1.44 bits/pixel. PSNR has been computed for groups of 8 consecutive frames.

Discussion

- The difference between the codebook optimized with the high-rate Gaussian model and the finite-rate Laplacian model diminishes as the rate increases.
- However, in all our experiments with MC residuals, it was observed that that the Laplacian model always yielded a better codebook. As at low rates, the high-rate Gaussian model can be quite inaccurate or even outright invalid.
- The codebooks are designed off-line, hence the slight complexity increase associated with the use of Laplacian model would not affect encoding complexity.
- In terms of the robustness of the codebooks designed with Laplacian model, our experiments showed that single Laplace codebook optimized for QP=34 is nearly as good as the codebooks optimized for each QP value in the entire range.

University of Manitoba

Conclusions

- We presented a novel algorithm for designing orthogonal matrix codebooks for transform coding block-wise stationary vector processes
- In contrast to previous work, the proposed algorithm explicitly minimizes the transform coding MSE with respect to the matrix codebook, and hence produces better transforms
- Experimental results obtained with video inter-prediction residual have shown significant coding gains over the DCT
- So far algorithm is applicable only to non-separable transforms, an extension to separable transforms is being currently developed.

University of Manitoba