

QUANTIZATION PERFORMANCE OF 4×4 BLOCK HIRSCHMAN TRANSFORMS

Joseph P. Havlicek¹, Tristan N. Arian¹, Hoda Soltani², Tomasz Przebinda³, and Murad Özaydin³

¹School of Electrical and Computer Engineering, University of Oklahoma, USA

²Center for Intelligent Transportation Systems, University of Oklahoma, USA

³Department of Mathematics, University of Oklahoma, USA

ABSTRACT

We introduce two new 2D 4×4 block Hirschman transforms, one separable and one nonseparable, that are inherently real and integer-valued and also admit highly desirable numerical properties. The new separable transform represents the first construction of a real-valued true Hirschman optimal transform. We apply these two new transforms to perform block-based video coding and investigate their quantization performance in terms of PSNR as a function of quantization parameter QP relative to the DCT and the well-known H.264 4×4 integer transform. Our results show that the separable Hirschman optimal transform achieves superior PSNR performance at all QP against all four of our well-known video test sequences. The results also suggest that the advantages of true Hirschman optimality outweigh any potential advantages associated with the nonseparability of the second new transform. Moreover, both new transforms outperform the H.264 integer transform across a wide range of QP values.

Index Terms— video compression, quantization, DCT, Hirschman transform, H.264

1. INTRODUCTION

Hirschman transforms admit optimal joint spatio-spectral localization and have been previously considered in several image analysis applications [1–3]. However, like the DFT (to which they bear a strong relationship, see e.g. [4]), Hirschman transforms are complex-valued in general; consequently, their use in video coding has not been widely contemplated. Recently, we demonstrated construction of an 8×8 2D block Hirschman transform that, for the first time, is inherently both real and integer-valued and also admits highly desirable numerical properties [5]. In our preliminary experiments presented in [5], this 8×8 Hirschman transform achieved significantly superior PSNR performance across a wide range of quantization parameter (QP) values relative to both the 2D DCT [6] and the H.264 8×8 integer transform [7, 8].

In this paper we introduce two new real and integer-valued Hirschman transforms applicable to 4×4 image blocks, one separable and the other nonseparable, thereby demonstrating for the first time basic techniques which may be used to construct both separable and nonseparable real and integer Hirschman transforms for arbitrary even block sizes. The sep-

arable 4×4 Hirschman transform we develop in Section 3 is a true *Hirschman Optimal Transform* (HOT – see the discussion in Section 2); thus, in this paper we demonstrate construction of a real-valued true HOT for the first time. In Section 4, we argue without proof that such real-valued HOTs can only exist on $\mathbb{Z}/4\mathbb{Z}$ and products thereof. Accordingly, the nonseparable 4×4 Hirschman transform we develop in Section 4 is $H_{1/2}$ -optimal but is *not* Hirschman optimal.

In Section 5 we compare quantization performance, quantified in terms of PSNR measured over a range of $0 \leq \text{QP} \leq 51$, of the two new 4×4 block Hirschman transforms developed here relative to the directly comparable H.264 4×4 integer transform process [7, 8]. While our experiments are admittedly limited in scope (run on the luminance components of only four well-known test sequences), the results are consistent and show a clear PSNR advantage at all QP for the new 4×4 separable block Hirschman transform. In addition, we find that the new 4×4 nonseparable block Hirschman transform outperforms the H.264 integer transform over a wide range of QP including all but the largest values.

These results are encouraging and strongly suggest that Hirschman transforms may have the potential to deliver substantial coding gains in future video CODECs. However, our investigation in this paper is limited to the quantization performance of the studied transforms and there is another significant open question that must be answered before Hirschman transforms can be considered for deployment in practical CODECs. Specifically, the efficacy of the ubiquitous block DCT derives primarily from its energy compaction properties whereby the magnitudes of the transform coefficients decay towards zero with increasing frequency, leading to runs of zeros in the quantized and 1D-ordered DCT coefficients which may then be efficiently run-length and entropy coded (see, e.g., [8–12]). We demonstrated in [5] that, unlike the DCT, the quantized Hirschman transform coefficients do not decay towards zero, but rather tend to cluster around a small number of integer attractor values. This suggests that m -ary run-length coding may be very efficiently applied to the quantized Hirschman transform coefficients – but only if advantageous scanning patterns can be developed to produce significant runs at the attractor values. The development of such scanning patterns is an important open research problem that, while beyond the scope of this paper, is a central theme

of our ongoing work in this area.

2. BRIEF REVIEW OF HIRSCHMAN TRANSFORMS

The notion of joint uncertainty for a signal defined on a continuous domain may be rigorously quantified in terms of the well-known Heisenberg-Weyl inequality [13–15]. However, as we discussed in [16], this traditional notion of joint uncertainty does not generalize well to discrete signals because the resulting measure fails to preserve translation invariance. Therefore, it has been argued in [4, 17–20] that joint uncertainty for discrete signals is more appropriately quantified in terms of *entropy*.

Proceeding as in [4, 16, 17, 19], we consider the ring $A = \mathbb{Z}/N\mathbb{Z}$ with addition and multiplication modulo N . Let j denote the imaginary unit and let $W_N = \exp(-j2\pi/N)$. Let $u \in \ell^2(A)$ have unit norm ($\|u\|_{\ell^2} = 1$) and discrete Fourier transform given by

$$(\mathcal{F}u)[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] W_N^{nk}. \quad (1)$$

We quantify the time uncertainty or localization of u by the entropy

$$H(u) = - \sum_{n=0}^{N-1} |u[n]|^2 \ln(|u[n]|^2). \quad (2)$$

The joint time-frequency localization of u may then be rigorously quantified by the measure

$$H_p(u) = pH(u) + (1-p)H(\mathcal{F}u), \quad (3)$$

where parameter $0 \leq p \leq 1$ balances the relative contributions of time and frequency localization to the joint measure H_p .

Orthogonal transforms on $\ell^2(A)$ where every basis signal minimizes H_p for one or more p are known generally as *discrete Hirschman transforms*. Constructions of such Hirschman transforms with $p = 1/2$ were demonstrated in [2, 4, 16] but are unknown for other individual values of $p \neq 1/2$. Remarkably however, techniques for constructing Hirschman transforms that optimize H_p for *all* $0 \leq p \leq 1$ are known and were given in [4, 16]; we refer to these very special transforms as *Hirschman optimal transforms* (HOTs).

To construct a discrete Hirschman transform, we begin by selecting an additive subgroup B of $A = \mathbb{Z}/N\mathbb{Z}$. Let $\mathbb{I}_B[n]$ be the indicator function of B and let $v[n] = \mathbb{I}_B[n]/\|\mathbb{I}_B\|_{\ell^2}$. Then an orthonormal basis for $\ell^2(A)$ wherein every basis signal minimizes $H_{1/2}$ may be generated by applying compositions of the following operators to $v[n]$ [4, 16]:

1. translation modulo N : $T_D v[n] = v[(n-D) \bmod N]$,
2. modulation: $M_a v[n] = W_N^{-an} v[n]$.

Provided that a sufficient number of such signals can be obtained to form a complete basis, they may then be used as the rows (equivalently, the columns) of a transform matrix to yield an $H_{1/2}$ -optimal discrete Hirschman transform. In the special case that N is a perfect square *and* $|B| = \sqrt{|A|}$ (where $|\cdot|$ denotes cardinality) the resulting transform is also a HOT – minimizing $H_p \forall 0 \leq p \leq 1$ [4, 16]. Note that appearance of the term W_N^{-an} in the modulation operator M_a implies that these transforms are complex-valued in general.

3. SEPARABLE REAL-VALUED 4×4 BLOCK HOT

Proceeding analogously to our development of the 8×8 block Hirschman transform given in [5], we let $N = 4$ and take the additive subgroup to be $B = \{0, 2\}$, giving

$$v[n] = \frac{1}{\sqrt{2}} [1 \ 0 \ 1 \ 0]^T. \quad (4)$$

Applying the operators T_1 , M_1 , and $(T_1 \circ M_1)$ to $v[n]$ we obtain the symmetric, involutory orthogonal transform matrix

$$\mathcal{B} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad (5)$$

defining a 1D Hirschman optimal (HOT) transform on $\ell^2(\mathbb{Z}/4\mathbb{Z})$ given by

$$y = \mathcal{B}x, \quad (6)$$

$$x = \mathcal{B}^{-1}y = \mathcal{B}y. \quad (7)$$

We then obtain a separable 2D HOT \mathcal{H}_2 applicable to 4×4 image blocks \mathbf{X} according to

$$\mathbf{Y} = \mathcal{H}_2 \mathbf{X} = \mathcal{B}_2 \mathbf{X} \mathcal{B}_2, \quad (8)$$

$$\mathbf{X} = \mathcal{H}_2^{-1} \mathbf{Y} = (\mathcal{B}_2 \mathbf{Y} \mathcal{B}_2) / 4, \quad (9)$$

where $\mathcal{B}_2 = \sqrt{2}\mathcal{B}$.

Note that, like the separable 8×8 Hirschman transform (i.e., *not* HOT) developed in [5], the 4×4 2D separable HOT (8), (9) is inherently real and integer-valued. Computation of the forward transform (8) requires additions only, while computation of the inverse transform (9) requires only additions and bit shift operations.

4. NONSEPARABLE REAL-VALUED 4×4 BLOCK HIRSCHMAN TRANSFORM

The use of multiple transforms in a competition scheme for video coding was recently studied in [21] where it was pointed out that nonseparable transforms can potentially exploit inter-pixel correlations more effectively than separable transforms. In this section, we develop a nonseparable discrete Hirschman transform applicable to 4×4 image blocks.

To this end, we reshape the 4×4 input block \mathbf{X} by stacking the columns to obtain a 16×1 column vector $\mathcal{X} = \text{vec}(\mathbf{X})$.

Let $N = 16$ so that $A = \mathbb{Z}/16\mathbb{Z}$. To obtain a true HOT, we would have to choose the subgroup B such that $|B| = 4$, meaning that $v[n]$ would necessarily contain four nonzero entries. However, this would preclude the existence of any integer a for which $M_a v[n]$ could still be real-valued. In fact, reasoning similarly we believe that real-valued HOTs exist only for $N = 4$. Therefore, for the present case of $N = 16$ we strive instead for a real $H_{1/2}$ -optimal transform and choose $B = \{0 \ 8\}$. There results

$$v[n] = \frac{1}{\sqrt{2}}[1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T \quad (10)$$

and

$$M_1 v[n] = \frac{1}{\sqrt{2}}[1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T. \quad (11)$$

Applying the operator T_D to (10) and (11) with $1 \leq D \leq 7$, we obtain the symmetric, involutory orthogonal transform matrix

$$\mathfrak{B} = \begin{bmatrix} v[n]^T \\ T_1 v[n]^T \\ \vdots \\ T_7 v[n]^T \\ M_1 v[n]^T \\ (T_1 \circ M_1) v[n]^T \\ \vdots \\ (T_7 \circ M_1) v[n]^T \end{bmatrix}. \quad (12)$$

Letting $\mathfrak{B}_2 = \sqrt{2}\mathfrak{B}$, we then have a nonseparable block Hirschman transform \mathfrak{H}_2 applicable to 4×4 image blocks \mathbf{X} and given by

$$\mathcal{X} = \text{vec}(\mathbf{X}), \quad (13)$$

$$\mathcal{Y} = \mathfrak{B}_2 \mathcal{X}, \quad (14)$$

$$\mathbf{Y} = \text{vec}^{-1}(\mathcal{Y}) \quad (15)$$

with inverse

$$\mathcal{Y} = \text{vec}(\mathbf{Y}), \quad (16)$$

$$\mathcal{X} = \mathfrak{B}_2 \mathcal{Y} / 2, \quad (17)$$

$$\mathbf{X} = \text{vec}^{-1}(\mathcal{X}). \quad (18)$$

Like the separable \mathcal{H}_2 transform developed in Section 3, the nonseparable \mathfrak{H}_2 transform (13)-(18) is inherently real and integer-valued, the forward transform $\mathbf{Y} = \mathfrak{H}_2 \mathbf{X}$ requires only additions, and the inverse transform $\mathbf{X} = \mathfrak{H}_2^{-1} \mathbf{Y}$ requires only additions and bit shift operations.

5. QUANTIZATION PERFORMANCE EXAMPLES

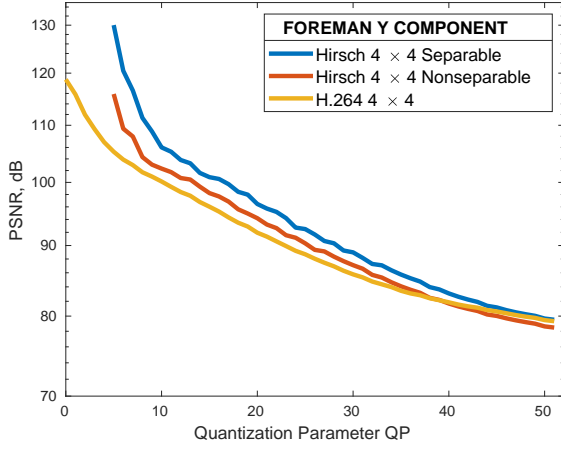
In this section we evaluate the quantization performance of the new \mathcal{H}_2 and \mathfrak{H}_2 transforms developed in Sections 3 and 4

in terms of PSNR as a function of quantization parameter QP on the luminance components of the four well-known CIF video test sequences *Foreman* (300 frames), *News* (300 frames), *Stefan* (90 frames), and *Tempete* (260 frames). For comparison, we also evaluate the H.264 4×4 integer transform, implemented as described in [8, Ch. 7], and the Matlab 4×4 `dct2` discrete cosine transform. While four is a small number of test sequences, our results are extremely consistent – indicating that the trends seen here likely generalize. In all cases, performance of the H.264 4×4 integer transform and the Matlab 4×4 `dct2` were very close. For this reason, results for the Matlab `dct2` transform are omitted from the plots of Fig. 1 in the interest of clarity; at the scale of these graphs, the curves for `dct2` would be visually indistinguishable from those for the H.264 4×4 integer transform.

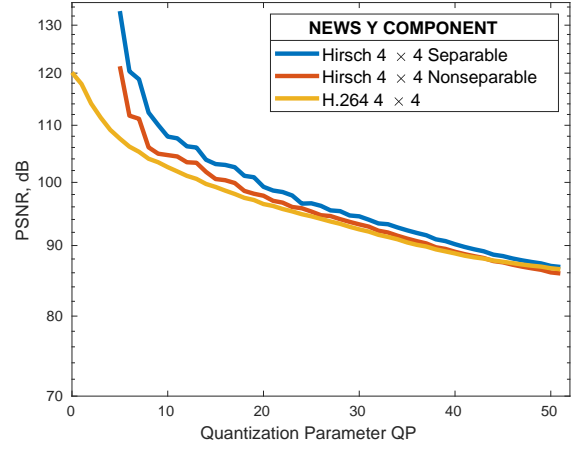
For each sequence, we performed motion estimation (ME) and motion compensation (MC) at a uniform block size of 4×4 using the previous frame for the reference picture and using a search range of $\pm 3 \times \pm 3$ pixels to obtain inter-prediction residual blocks. Sub-pixel ME and MC were not performed. Also, in the interest of simplicity and in view of the fact that we have not implemented a full CODEC for these experiments, we performed the MC and ME using the *original* pixel values from the reference picture rather than decoded pixel values. Consequently, it should be noted that our experiments here do not consider propagation of frame-to-frame decoding errors.

We applied each of the four transforms to the residual blocks using quantization parameter values $0 \leq \text{QP} \leq 51$. Calculations were performed as in [8, Sec. 7.2.3.9]. Quantization is inherent to the H.264 4×4 integer transform. For the \mathcal{H}_2 , \mathfrak{H}_2 , and `dct2` transforms, the quantization step size Q_{step} was back-calculated from QP as described in [8, Sec. 7.2.3.5]. The forward transforms were then computed, divided by Q_{step} , and rounded using floating point arithmetic. To invert, the quantized coefficients were multiplied by Q_{step} , the inverse transforms were applied using floating point arithmetic, and the results were rounded to obtain integer-valued reconstructions of the decoded residual pixels. PSNR was then calculated and averaged over all blocks of all frames (except the first) to obtain the performance curves shown in Fig. 1.

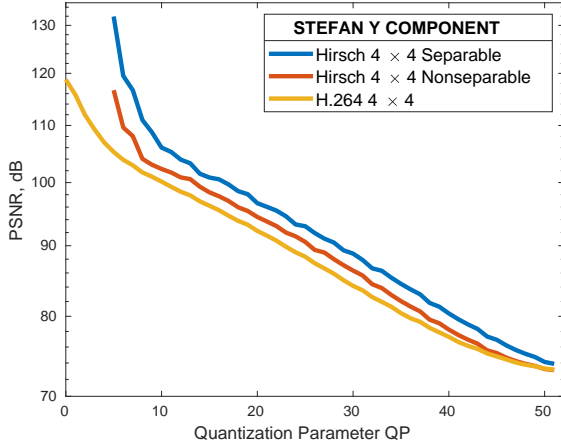
As shown in Fig. 1, the Hirschman optimal 4×4 separable transform \mathcal{H}_2 developed in Section 3 achieved the best PSNR in all cases tested and *significantly* outperformed the H.264 4×4 integer transform over a wide range of QP values. The \mathcal{H}_2 transform also outperformed the nonseparable 4×4 Hirschman transform \mathfrak{H}_2 developed in Section 4 at all QP against all four test sequences. This strongly indicates that the advantage of Hirschman H_p optimality for all $0 \leq p \leq 1$ enjoyed by the \mathcal{H}_2 transform outweighs the potential advantages of nonseparability provided by the \mathfrak{H}_2 transform. Against all four test sequences, the H.264 4×4 integer transform did narrowly outperform the nonseparable \mathfrak{H}_2 Hirschman transform at the largest values of QP. In particular, the crossover where



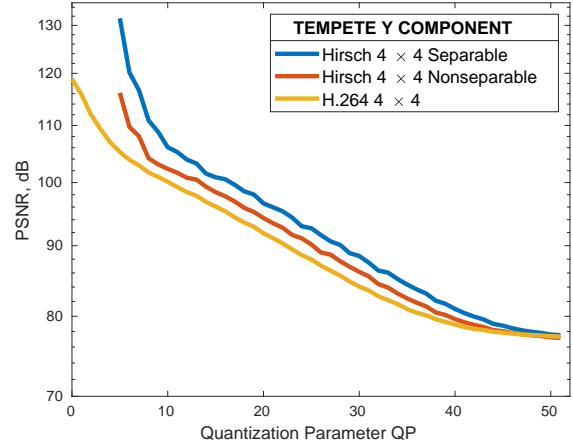
(a) *Foreman* luminance component



(b) *News* luminance component



(c) *Stefan* luminance component



(d) *Tempete* luminance component

Fig. 1. PSNR as a function of quantization parameter QP for $QP \in [0, 51]$. (a) Results for luminance component of *Foreman*. (b) Results for luminance component of *News*. (c) Results for luminance component of *Stefan*. (d) Results for luminance component of *Tempete*. In all four test cases, the separable Hirschman optimal transform \mathcal{H}_2 and the nonseparable Hirschman transform \mathfrak{H}_2 both produced MSE=0 and PSNR= ∞ for $0 \leq QP \leq 4$.

the H.264 transform achieves better PSNR occurred at $QP=37$ for *Foreman*, at $QP=44$ for *News*, at $QP=51$ for *Stefan*, and at $QP=48$ for *Tempete*.

Finally, there is a fine point that deserves clarification regarding the \mathcal{H}_2 results shown in Fig. 1. At small QP, the \mathcal{H}_2 transform delivers $MSE \ll 1$ and we observed a systematic anomaly whereby the PSNR failed to be monotonically decreasing in QP at $QP=10$. For this reason, the results shown for the \mathcal{H}_2 transform in Fig. 1 used rounding towards zero instead of rounding to the nearest integer for $QP=10$ only.

6. DISCUSSION

We introduced two new real and integer-valued 4×4 Hirschman block transforms and studied their quantization performance relative to the well-known H.264 4×4 integer

transform process. The separable \mathcal{H}_2 transform developed in Section 3 is a true Hirschman optimal transform (HOT), minimizing the entropy-based joint spatio-spectral uncertainty measure H_p for all $0 \leq p \leq 1$. The nonseparable \mathfrak{H}_2 transform developed in Section 4 optimizes H_p for $p = 1/2$, but is suboptimal for other p . This is significant in view of the fact that $p = 1/2$ weights the spatial and spectral localizations equally, whereas the DCT, upon which the H.264 4×4 integer transform is based, exhibits decidedly *unequal* localization in space and spatial frequency. Our results show that the separable HOT transform consistently achieves superior PSNR performance at all QP against all four sequences tested relative to the H.264 integer transform. However, the important question of how to order the quantized HOT coefficients in order to obtain efficient run-length coding remains open.

7. REFERENCES

- [1] W.-L. Hsue, S.-C. Pei, and J.-J. Ding, "Efficient discrete fractional Hirschman optimal transform and its application," in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 22-27, 2011, pp. 3604–3607.
- [2] R. Thomas, V. DeBrunner, and L. DeBrunner, "Fixed-point implementation of discrete Hirschman transform," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 28-31, 2018, pp. 1507–1511.
- [3] B. Narsimha, E. Suresh, K. P. Chandar, and B. Komuraiah, "Enhancement of color images in HOT domain with quantitative measurements using entropy and relative entropy," in *Proc. Int'l. Conf. Signal, Image Process.*, Chennai, India, Dec. 15-17, 2010, pp. 34–38.
- [4] T. Przebinda, V. DeBrunner, and M. Özaydın, "The optimal transform for the discrete Hirschman uncertainty principle," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 2086–2090, Jul. 2001.
- [5] J.P. Havlicek, T.N. Arian, H. Soltani, T. Przebinda, and M. Özaydın, "A preliminary case for Hirschman transform video coding," in *Proc. IEEE Southwest Symp. Image Anal., Interp.*, Sanga Fe, NM, Mar. 29-31 2020, pp. 104–107.
- [6] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, no. 1, pp. 90–93, Jan. 1974.
- [7] "Advanced video coding for generic audiovisual services," *Recommendation ITU-T H.264 and International Standard ISO/IEC 14496-10*, Jun. 2019.
- [8] I. E. Richardson, *The H.264 Advanced Video Compression Standard*, Wiley, West Sussex, UK, 2nd edition, 2010.
- [9] V. Sze, M. Budagavi, and G. J. Sullivan, Eds., *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, Springer, 2014.
- [10] M. Wien, *High Efficiency Video Coding: Coding Tools and Specification*, Springer, 2015.
- [11] J. Bankoski, *et al.*, "Towards a next generation open-source video codec," in *Vis. Inf. Process. Commun. IV*, Feb. 2013, vol. 8666 of *Proc. SPIE*, pp. 866606–01–866606–13.
- [12] Y. Chen, *et al.*, "An overview of core coding tools in the AV1 video codec," in *Proc. Picture Coding Symp.*, San Francisco, CA, Jun. 24-27, 2018, pp. 41–45.
- [13] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng. London*, vol. 93, no. III, pp. 429–457, 1946.
- [14] J. Ville, "Théorie et applications de la notation de signal analytique," *Cables et Transmission*, vol. 2A, pp. 61–74, 1948, translated from the French in I. Selin, "Theory and applications of the notion of complex signal," Tech. Rept. T-92, The RAND Corporation, Santa Monica, CA, Aug. 1958.
- [15] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am. A*, vol. 2, no. 7, pp. 1160–1169, Jul. 1985.
- [16] V. DeBrunner, J.P. Havlicek, T. Przebinda, and M. Özaydın, "Entropy-based uncertainty measures for $L^2(\mathbb{R}^n)$, $\ell^2(\mathbb{Z})$, and $\ell^2(\mathbb{Z}/N\mathbb{Z})$ with a Hirschman optimal transform for $\ell^2(\mathbb{Z}/N\mathbb{Z})$," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2690–2699, Aug. 2005.
- [17] V. DeBrunner, M. Özaydın, and T. Przebinda, "Resolution in time-Frequency," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 783–788, Mar. 1999.
- [18] A. Dembo, T.M. Cover, and J.A. Thomas, "information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1999.
- [19] V. DeBrunner, M. Özaydın, and T. Przebinda, "Analysis in a finite time-frequency plane," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1831–1832, Jun. 2000.
- [20] V.E. DeBrunner, M. Özaydın, T. Przebinda, and J. Havlicek, "The optimal solutions to the continuous- and discrete-time versions of the Hirschman uncertainty principle," in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Process.*, Istanbul, Turkey, Jun. 5-9, 2000, pp. 81–84.
- [21] T. Biatek, V. Lorcy, and P. Philippe, "Transform competition for temporal prediction in video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 815–826, Mar. 2019.