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 $\ell^2(\mathbb{Z}/N\mathbb{Z})$

Victor DeBrunner, *Senior Member, IEEE*, Joseph P. Havlicek, *Senior Member, IEEE*, Tomasz Przebinda, and Murad Özaydın

***Abstract*—The traditional Heisenberg–Weyl measure quantifies the joint localization, uncertainty, or concentration of a signal in the phase plane based on a product of energies expressed as signal variances in time and in frequency. In the image processing literature, the term *compactness* also has been used to refer to this same notion of joint localization, in the sense of a signal representation that is efficient simultaneously in time (or space) and frequency. In this paper, we consider Hirschman uncertainty principles based not on energies and variances directly but rather on *entropies* computed with respect to normalized energy densities in time and frequency. Unlike the Heisenberg–Weyl measure, this entropic Hirschman notion of joint uncertainty extends naturally from the case of infinitely supported continuous-time signals to the cases of both finitely and infinitely supported discrete-time signals. For the first time, we consider these three cases together and study them relative to one another. In the case of infinitely supported continuous-time signals, we find that, consistent with the energy-based Heisenberg principle, the optimal time-frequency concentration with respect to the Hirschman uncertainty principle is realized by translated and modulated Gaussian functions. In the two discrete cases, however, the entropy-based measure yields optimizers that may be generated by applying compositions of operators to the Kronecker delta. Study of the discrete cases yields two interesting results. First, in the finitely supported case, the Hirschman-optimal functions coincide with the so-called "picket fence" functions that are also optimal with respect to the joint time-frequency counting measure of Donoho and Stark. Second, the Hirschman optimal functions in the infinitely supported case can be reconciled with continuous-time Gaussians through a certain limiting process. While a different limiting process can be used to reconcile the finitely and infinitely supported discrete cases, there does not appear to be a straightforward limiting process that unifies all three cases: The optimizers from the finitely supported discrete case are decidedly non-Gaussian. We perform a very simple experiment that indicates that the Hirschman optimal transform (HOT) is superior to the discrete Fourier transform (DFT) and discrete cosine transform (DCT) in terms of its ability to separate or resolve two limiting cases of localization in frequency, viz. pure tones and additive white noise. We believe that these differences arise from the use of entropy rather than energy as an optimality criterion and are intimately related to the apparent incongruence between the infinitely supported continuous-time case and the finitely supported discrete-time case.**

***Index Terms*—Denoising, entropy, time-frequency resolution, uncertainty.**

Manuscript received September 25, 2003; revised June 11, 2004. This work was supported in part by the National Science Foundation under Grant DMS-0200724. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jean Pierre Delmas.

V. DeBrunner and J. P. Havlicek are with the School of Electrical and Computer Engineering, The University of Oklahoma, Norman, OK 73019-1023 USA (e-mail: vdebrunn@ou.edu; joebob@ou.edu).

T. Przebinda and M. Özaydın are with the Department of Mathematics, The University of Oklahoma, Norman, OK 73019-1023 USA (e-mail: tprzebinda@ou.edu; mozaydin@ou.edu).

Digital Object Identifier 10.1109/TSP.2005.850329

I. Introduction

In [1], we introduced an entropy-based measure H_p that quantifies the compactness of a discrete-time signal in the sample-frequency phase plane. Use of an entropy-based measure allowed us to overcome the limitations inherent to discretizing the Heisenberg uncertainty. A naïve discretization of the Heisenberg uncertainty leads to a discrete measure that fails to preserve translation invariance and is, therefore, not useful. Our entropy-based measure was used to show that discretized Gaussian pulses may not be the most compact basis with respect to joint time-frequency resolution. At that time, we conjectured a lower limit on the compaction in the phase plane. Later, we became aware that part of this conjectured lower limit was proven in [2] under the moniker "a discrete Hirschman's uncertainty principle." However, that result did not describe the characteristics of the signals that meet the limit, as our conjecture did [3]. We further argued in [4] that this measure indicates two possible "best basis" options for achieving optimal compaction in the time-frequency phase plane:

1) multitransform (nonorthogonal) option;
2) orthogonal discrete-time, discrete-frequency Hirschman uncertainty principle option.

We have discussed many results in the first option (see [1] for pointers to references). The second option was detailed extensively in [5]. In that paper, we found a basis (transform) that is orthogonal and that uniquely minimizes the discrete-time, discrete-frequency Hirschman uncertainty principle. Furthermore, we discussed the relationship between the optimizing time signals in the continuous–continuous and discrete–discrete cases in [6], where we suggested that the notion of sampling needs more investigation. In this paper, we discuss a proof (given in [7]) that covers the minimizers not only in the continuous–continuous and discrete–discrete cases but also in the mixed discrete-time and continuous-frequency case. The proof is valid for one and multiple dimensions in all cases. For brevity, we denote the three cases as follows.

1) We use "continuous Hirschman" when referring to the continuous-time, continuous-frequency case ($L^2(\mathbb{R}^n)$).
2) We use "discrete Hirschman" when referring to the discrete-time, continuous-frequency case ($\ell^2(\mathbb{Z})$).
3) We use "digital Hirschman" when referring to the discrete-time, discrete-frequency case ($\ell^2(\mathbb{Z}/N\mathbb{Z})$).

We also illustrate the implications of these results for the digital Hirschman case through the use of several examples. In particular, we examine experimentally the performance of a transform, called the Hirschman optimal transform (HOT), that is constructed using an orthonormal basis of sequences with optimal digital Hirschman uncertainty. We compare the HOT to the discrete Fourier transform (DFT) and discrete cosine transform (DCT) in separating pure tones from additive white Gaussian noise (AWGN). What we are after is an example that shows the uncertainty of the representation. Analyzing these results, we observe that, when thresholding is used, the HOT is superior to the DFT and DCT over a range of signal-to-noise ratio (SNR) values of greater practical importance. While our examples use the same experimental setup as the most classical of all frequency estimation tests, we want to make it clear that our intent here is not to develop a new frequency estimation technique but rather to explore the time-frequency properties of the HOT.

II. Continuous Case Hirschman Uncertainty Principle

The Hirschman uncertainty principle was developed for the continuous case [8]. We provide some definitions first and then give the Hirschman measure. Let $S(\mathbb{R}^n)$ be the Schwartz space of functions on the Euclidean space $\mathbb{R}^n$ [9]. Recall the (multidimensional) Fourier transform

$$X(f) = \int_{\mathbb{R}^n} x(t) e^{-j2\pi f \cdot t} dt, \quad (t, f \in \mathbb{R}^n, x \in S(\mathbb{R}^n))$$

where we have used $f \cdot t$ for the Euclidean inner product $t_1 f_1 + t_2 f_2 + \cdots t_n f_n$. The L^2-norm on $S(\mathbb{R}^n)$ is given by

$$||x||_2 = \left(\int_{\mathbb{R}^n} |x(t)|^2 \, dt \right)^{\frac{1}{2}}.$$

Consider a function $x \in S(\mathbb{R}^n)$ with $||x||_2 = 1$. Then, $|x(t)|^2$ may be interpreted as the distribution of signal energy in time in the usual sense, i.e., as a probability distribution on $\mathbb{R}^n$. Hence, the notion of entropy, introduced by Shannon [10], applies to $|x(t)|^2$. We use $H(x)$ to denote the entropy of $|x(t)|^2$

$$H(x) = -\int_{\mathbb{R}^n} |x(t)|^2 \ln\left(|x(t)|^2\right) dt.$$

Since the Fourier transform is unitary, we have $||x||_2 = ||X||_2$. Define an entropy-based joint uncertainty measure according to

$$H_p(x) = pH(x) + (1-p)H(X).$$

In the special case, $p = 1/2$, $H_{1/2}(x)$ is called the Hirschman uncertainty of x. The following theorem, conjectured by Hirschman [8], has been proven by Beckner [11], [12].

Theorem 1: Let $x \in S(\mathbb{R}^n)$ with $||x||_2 = 1$. Then

$$H_{\frac{1}{2}}(x) \geq \frac{n}{2} \ln\left(\frac{e}{2}\right). \tag{1}$$

It is easy to check that equality occurs in (1) if x is obtained by arbitrary translations, dilations, and/or modulations of the Gaussian $x(t) = e^{-(\pi/2)x^2}$. In [8], Hirschman conjectured that $H_{1/2}(x)$ is minimal [i.e., that $H_{1/2}(x) = (n/2)\ln(e/2)$ after the work of Beckner] only for these functions—a result that we proved recently in [7]. We observe that these minimizers are identically the functions shown to minimize the Heisenberg uncertainty in one dimension (1–D) by Gabor [13] and in two dimensions (2–D) by Daugman [14].

III. Discrete-Time Hirschman Uncertainty Principles

In this section, we consider the digital and discrete cases for sampled-data systems. In many applications, the most important case is the digital Hirschman case because of its relative ease of use in computer-based system implementation. This case is considered in Section III-A. We develop in Section III-B an orthogonal transform that is analogous to the DFT but uses Hirschman optimal basis functions. In Section III-C, we also determine the optimizing functions for the discrete Hirschman case.

A. Digital Hirschman Uncertainty Principle

In this section, we consider discrete 1–D signals on a finite domain. Fix a finite set of nonnegative integers $\mathcal{D} = \{0, 1, 2, \ldots, N-1\}$. Let $\mathcal{H}_N$ denote the Hilbert space of sequences $x : \mathcal{D} \to \mathbb{C}$ with squared-norm

$$||x||_2^2 = \sum_{n=0}^{N-1} |x[n]|^2.$$

With the standard twiddle factor notation $W_N = e^{-j(2\pi/N)}$, the DFT is

$$X(k) = Fx[n] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] W_N^{nk}. \tag{2}$$

This defines an isometry on $\mathcal{H}_N$ with inverse given by

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X(k) W_N^{-nk}. \tag{3}$$

By the *digital phase plane*, we mean the set of all points $(n, k) \in \mathcal{D} \times \mathcal{D}$. The operators that allow us to view the digital phase plane (i.e., to move a signal from point to point in the digital phase plane) are the translation and modulation operators. For $x \in \mathcal{H}_N$ with $0 \leq n, D, a \leq N-1$, these operators are defined by

- Translation: $T_D x[n] = x[n \dot{-} D]$
- Modulation: $M_a x[n] = W_N^{\dot{-}an} x[n]$

where the binary operator "$\dot{-}$" in the translation definition means subtraction modulo N. Both of these operators are well defined and admit interpretations related to the corresponding continuous case operators. However, in the digital case, there is no geometric analog of the continuous case

dilation operator in the sense of "zooming in and out." This is because of the inherently discrete (countable) nature of the domain. In the continuous case dilation formula $D_a x(t) = a^{-(1/2)} x(a^{-1}t)$, the role of the constant $a^{-(1/2)}$ is to maintain $\|x\| = \|D_a x\|$. In our digital case, a^{-1} makes sense only when a is relatively prime to N, and multiplication (modulo N) by a^{-1} simply permutes $\{0, 1, \ldots, N-1\}$.[1] So, $\|x\|^2 = \sum_{n=0}^{N-1} |x[n]|^2 = \sum_{n=0}^{N-1} |x[a^{-1}n]|^2$, and consequently, we cannot multiply by $a^{-(1/2)}$ if we wish to preserve the norm. So, discrete dilation should be defined by $D_a x[n] = x[a^{-1}n]$. However, $n \mapsto a^{-1}n$ is a pseudo-random permutation without a physical interpretation, so unlike translations, we choose not to highlight dilations.

We next recall some simple facts. If we view x as a column vector in $\mathbb{C}^N$ with entries $x[0], x[1], \ldots, x[N-1]$, the DFT is accomplished by premultiplying with the matrix

$$F = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)^2} \end{bmatrix}. \quad (4)$$

Hence

$$F^2 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \quad (5)$$

and consequently, $F^4 = I_N$ [the $(N \times N)$ identity matrix].

We are not aware of any straightforward analog of the Heisenberg inequality for signals in $\mathcal{H}_N$. One problem is that the "position operator" $x(t) \mapsto t x(t)$ is well defined on $L^2(\mathbb{R})$ but not on $\mathcal{H}_N$, where t takes values in the group $\mathcal{D}$ (equipped with addition modulo N). For $x : \mathcal{D} \to \mathbb{C}$, with $\|x\|_2 = 1$, consider the following seemingly obvious analog for the mean: $\overline{n} = \sum_{n=0}^{N-1} n|x(n)|^2$. The immediate problem is that $\overline{x}$ need not belong to the set $\mathcal{D}$. A more serious problem is that addition modulo N, with respect to which the DFT is defined (unlike the continuous Fourier transform), is not the same as addition of real numbers. In particular, for signals in $\mathcal{H}_N$, the mean of the translated signal will only rarely be the translate of the mean. The implied periodicity inherent in both the signal and its DFT are also problematic with respect to translation of the independent variable: The element 0 in $\mathcal{D}$ is associated with N, but these two representatives yield different values for the mean $\bar{x}$.

Because of these problems, simple discretization of the continuous case Heisenberg measure leads to an uncertainty measure for the digital case that fails to be translation invariant. This is clearly undesirable: Time shifting or modulating (frequency shifting) a signal x should not change its uncertainty. These flaws appear to be inherent to energy-based (variance-based) uncertainty measures in the digital case.

[1]Let $N = 7$ and $a = 3$. Then, under multiplication modulo N, $a^{-1} = 5$ since $(3 \cdot 5) \bmod 7 = 1$.

These problems do not appear when the uncertainty measure is based on entropy instead of energy. For $x \in \mathcal{H}_N$ with $\|x\|_2 = 1$, the entropy is defined as

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

Using entropy, we define a general class of digital uncertainty measures for $0 \le p \le 1$ according to

$$H_p(x) = pH(x) + (1-p)H(Fx), \quad (x \in \mathcal{H}_N, \|x\|_2 = 1). \quad (6)$$

In the special case where $p = 1/2$, the measure (6) is termed the *digital Hirschman uncertainty*. In the general case, the parameter p allows for a tradeoff between concentration in time and in frequency. In the extreme case where $p = 1$, the measure (6) ignores frequency localization, and the minimizing signals are those concentrated at single points. Similarly, if $p = 0$, the minimizing signals are those for which all the sample magnitudes $|x(n)|$ are equal. Intermediate values of p give a weighted measure of joint time-frequency localization of the signal x. Before describing the minimizers of (6), we define periodization.

Definition 1: For $N = KL$, the periodization of $v \in \mathbb{C}^K$ is $x \in \mathbb{C}^N$ defined as $x(sK + n) = (1/\sqrt{L})v(n)$ for $0 \le s \le L-1$ and $0 \le n \le K-1$.

We refer to the sequence $v \in \mathbb{C}^K$ given by $v(0) = 1, v(1) = 0, \ldots v(K-1) = 0$, as the Kronecker delta or impulse (unit sample) sequence, without specifying the signal length K. We proved the following theorem in [5].

Theorem 2: The only sequences $x \in \mathbb{C}^N$, with $\|x\|_2 = 1$, for which $H_{1/2}(x)$ is minimal are obtained from the Kronecker delta sequence by applying any composition of periodization, translation, modulation, the DFT, and multiplication by a complex number of unit magnitude.

Proof: A detailed and comprehensive proof of a more general result that includes Theorem 2 as two parts is given in [5] and [7]. ■

We have shown that the signals minimizing the continuous Hirschman uncertainty measure are Gaussians [7]. In contrast, for any N $(= K^2)$, the HOT basis vectors that minimize the digital Hirschman uncertainty measure are strikingly dissimilar to discretized Gaussians. Furthermore, no discretization of a Gaussian will lead to a discrete signal minimizing the digital Hirschman uncertainty measure [5]. This rather surprising fact contradicts the intuitive expectation that the limiting case of digital minimizers as $N \to \infty$ should be Gaussian.

The discrete case may be obtained from the continuous case by a limiting process on the variance of Gaussian minimizers. Alternatively, the discrete case can be viewed as deriving from the digital case by the application of an appropriate limiting process, where the time support becomes bidirectionally infinite, and the minimizers are periodic trains of Kronecker deltas (the Kronecker comb function) with a fundamental period that tends toward infinity. Thus, we arrive at a deep and fundamental incongruity.

The question of how to unify all three cases could not even be asked in the context of energy-based uncertainty measures. Intuitively, one would hope to be able to show that the digital case

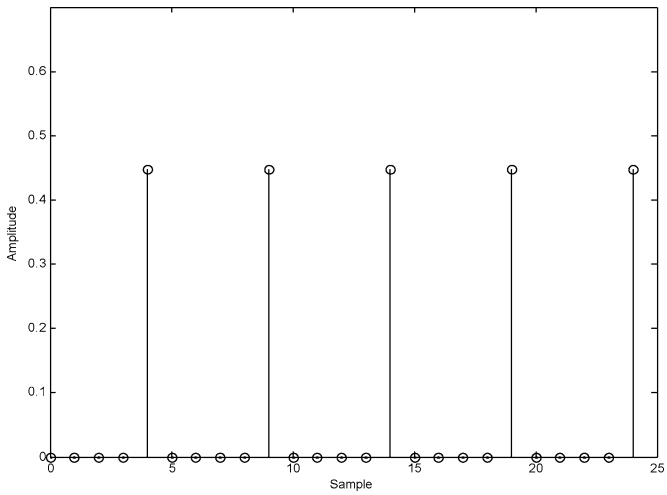


Fig. 1. $N = K^2 = 25$-point optimizer—the picket fence signal.

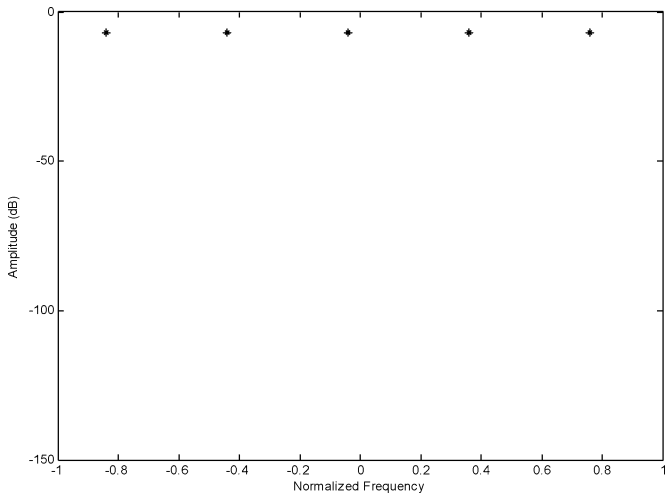


Fig. 2. Spectrum of the $N = K^2 = 25$-point optimizing signal given in Fig. 1.

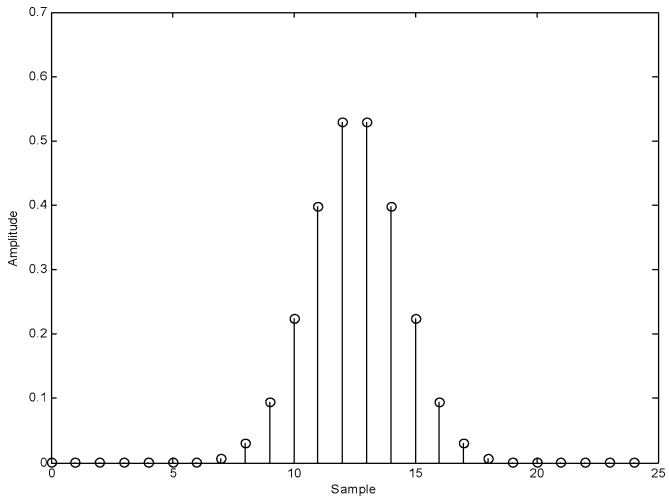


Fig. 3. Twenty-five point discretized Gaussian pulse.

minimizers converge in some sense to the discrete case minimizers as the support (cardinality of $\mathcal{D}$) tends to infinity and that in a closely related sense, the discrete minimizers converge to Gaussians in the limit as the sampling frequency increases without bound. With entropy-based uncertainty measures, the question of unifying all three cases can be addressed, but the answer remains unclear. The discrete case can be obtained from both the continuous and digital cases but through significantly dissimilar limiting processes, as we observed in the preceding paragraph. The result is that we do not yet know of a meaningful way to relate the continuous case minimizers (Gaussians) to the digital case minimizers (picket fence functions). What is needed is a more complete theory regarding the effects of sampling and windowing. In Section IV, we perform a simple test that provides some insight but raises more questions as well.

B. HOT on $\ell^2(\mathbb{Z}/N\mathbb{Z})$

The basis functions that define the HOT are derived according to Theorem 2. For the case where $N = K^2 = 25$, the HOT basis signal generated from the impulse by only applying periodization with $K = 5$ is shown in Fig. 1, where the picket fence characteristic may be clearly seen. A complete HOT basis is generated from this signal by applying modulations and translations. Note that a different initial basis signal could be obtained by applying a different composition of the operations listed in Theorem 2. This would generate a different HOT basis with the same localization characteristics, e.g., optimal Hirschman localization. The DFT magnitude of the basis signal given in Fig. 1 is shown in Fig. 2. Note that the signal and its spectrum have the same functional form, viz. impulse trains. We think that it is very interesting that these are the same functions that are suggested in [15], where uncertainty was quantified by applying the counting measure to the (finite) support of a discrete signal and its discrete spectrum. Similarly, the digital Hirschman uncertainty depends on the number of nonzero samples in the signal and its transform. The Hirschman uncertainty measure $H_{1/2}$ is invariant under translations and modulations, because entropy is clearly invariant under translations and modulations, and the Fourier transform interchanges these operations. While the entropy is in fact invariant under all permutations, the Hirschman uncertainty is not because the entropy of the transformed signal is not, in general, invariant under arbitrary permutations. For example, $H_{1/2}(1/\sqrt{2}, 0, 1/\sqrt{2}, 0) = \ln(2)$ but $H_{1/2}(1/\sqrt{2}, 1/\sqrt{2}, 0, 0) > \ln(2)$. It is in this sense—which is clearly distinct from the traditional variance-based notion of uncertainty—that the HOT basis functions are optimally concentrated.

In continuous time, the translated and/or modulated Gaussians uniquely minimize the joint uncertainty with respect to both the Heisenberg measure and the Hirschman measure. For comparison, therefore, we consider a discretized Gaussian pulse. By choosing the variance appropriately for the number of samples, the spectrum shape can be matched sample to sample to that of the signal, effectively equating the resolution in time to that in frequency. In [1], we showed that a discretized Gaussian pulse with variance chosen this way has a constant Hirschman uncertainty H_p irrespective of p and is, therefore, most comparable to the HOT basis. Of course, the uncertainty H_p realized by the discretized Gaussian pulse is greater than that of the HOT basis functions. A Gaussian pulse of this type is shown in Fig. 3 for $N = K^2 = 25$; its spectrum is given in Fig. 4.

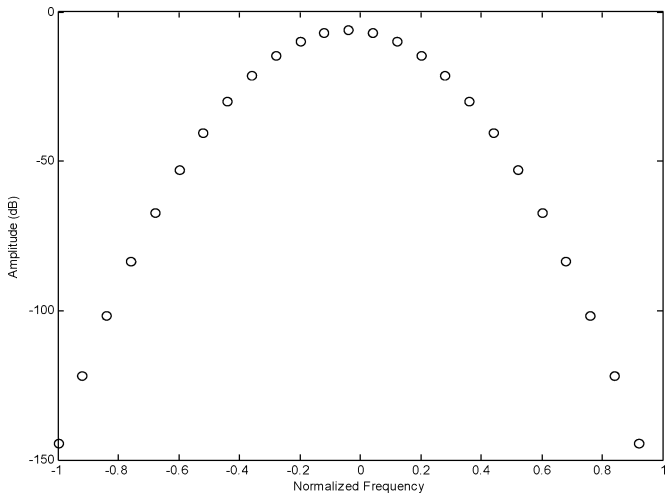


Fig. 4. Spectrum of the 25-point discretized Gaussian pulse.

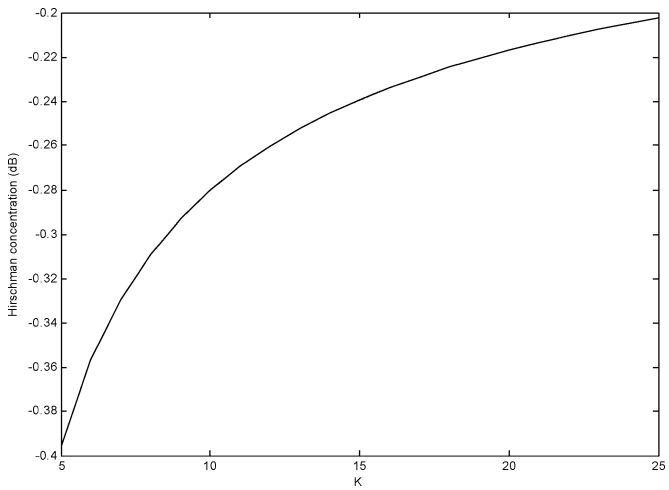


Fig. 5. Improvement of the Hirschman concentration measure for the optimizing pulse over the Gaussian (bottom) pulses. dB versus $K = \sqrt{N}$.

Note that the HOT basis signal (impulse train) and the Gaussian both have a functional form that is invariant under corresponding Fourier transformation, which implies a deep connection between these two signals. Fig. 5 shows the differential uncertainty

$$I = 10\log_{10}\left(H_{\frac{1}{2}}(\text{HOT})\right) - 10\log_{10}\left(H_{\frac{1}{2}}(\text{Gaussian})\right)\ \text{dB} \tag{7}$$

between a HOT basis signal and comparable discretized Gaussian pulse for various lengths $N = K^2$. The uncertainty of the HOT basis signal is the optimal value $H_{1/2}(\text{HOT}) = (1/2)\ln N$. The difference, while not large, can be significant, as we will demonstrate in Section IV. It is highly relevant that the difference vanishes for increasing lengths, again implying a convergence of these two signals in a certain sense—the sense of entropic equivalence—as $N \to \infty$. This is a reflection of the fact that the Hirschman uncertainty principle subsumes the Heisenberg principle in the continuous case. Note that this convergence is indeed slow. The differential uncertainty of (7) drops below 0.1 dB only for $N \geq (2^{10})^2$, a value that exceeds the block length over which any (1-D) real-world signal would exhibit stationarity in practical applications.

We use the K-dimensional DFT kernel as the originator signals for our $N = K^2$-length HOT basis. Each of these basis functions must then be shifted and upsampled to produce the sufficient number of orthogonal basis functions that define the HOT. While other choices are possible, this one leads to an efficient computational structure with a complexity less than that of the N-point DFT. Note that the DFT kernel could also be used in a similar manner to produce transforms for other factorizations $N = KL$, $K \neq L$, but these possess an uncertainty H_p that varies as a function of p and are suboptimal in this sense [5].

The HOT ($N = K^2$) is unitary to a scale (just like the DFT), and so, the inverse transform can be achieved by taking the conjugate transpose and scaling by $\sqrt{K}$. In general, we have the (unitary) transform relationship [5], [6]

$$H(Kr+l) = \frac{1}{\sqrt{K}}\sum_{n=0}^{K-1} x[Kn+l]e^{-j\frac{2\pi}{K}nr}, \quad 0 \leq r, l \leq K-1$$

and its inverse

$$x[Kn+l] = \frac{1}{\sqrt{K}}\sum_{r=0}^{K-1} H(Kr+l)e^{j\frac{2\pi}{K}nr}, \quad 0 \leq n, l \leq K-1.$$

Given our previous discussions, it should come as no surprise that the HOT shares much structure and appearance with the DFT. In fact, we often think of the HOT as a "1-1/2" dimensional DFT. This notion comes by viewing the $N = K^2$-length 1-D sequence as a $K \times K$ matrix that is transformed either along the rows or along the columns, depending on the ordering of the basis functions. This should be contrasted with the true 2-D DFT, where the action is along both the rows and the columns of the matrix. The 2-D matrix of HOT coefficients must then be reordered onto a 1-D domain. In the case where the action of the K-point DFTs are along the rows (columns), this 1-D reordering consists of appending the rows (columns) of the $K \times K$ matrix of HOT coefficients. In general, the N-point HOT is computationally more efficient than the N-point DFT and increasingly more efficient as $N \to \infty$. We also note the resemblance to the polyphase construct—this relationship is currently under consideration.

C. Discrete Hirschman Uncertainty Principle

One interesting fact regarding the Hirschman uncertainty is its result in the discrete case. Using the discrete-time Fourier transform

$$X(f) = \sum_{n\in\mathbb{Z}} x[n]e^{-j2\pi fn}$$

and its inverse

$$x[n] = \int_{-0.5}^{0.5} X(f)e^{j2\pi fn}df$$

and assuming that $\|x\|_{\ell^2} = 1$, we have shown that the Hirschman uncertainty in this case is [7]

$$H_{\frac{1}{2}}(x) = -\sum_{n\in\mathbb{Z}} |x[n]|^2 \ln\left(|x[n]|^2\right) - \int_{-0.5}^{0.5} |X(f)|^2 \ln\left(|X(f)|^2\right) df \geq 0.$$

The minimizers in this case are the unit sample sequence $\delta[n]$ up to a translation and/or multiplication by a unit-magnitude complex number. As we discussed in Section III-A, in passing from the continuous case through this case to the digital case, the minimizers transition first from Gaussians to the Kronecker delta and second from the Kronecker delta to picket fence sequences.

IV. Examples

We perform a very simple experiment which indicates that the HOT is superior to the DFT and DCT in terms of its ability to separate or resolve two limiting cases of localization in frequency, viz. pure tones and additive white noise. We believe that these differences arise from the use of entropy rather than energy as an optimality criterion and are intimately related to the apparent incongruence between the continuous case and the digital case. As a test signal, we consider the sum of a pure tone and AWGN. This experimental setup is identical to that for the classical problem of estimating the frequency of a pure tone in AWGN, and we maintain that these are, in fact, closely related problems. Here, however, we are not directly interested in developing a frequency estimation algorithm. Rather, we are interested in making fundamental observations about the ability of various transforms to *separate* the coherent and incoherent signal components on the phase plane. Specifically, we are going to use Euclidean distance between a noise-free and a noisy transform domain spectrum to quantify the ability of the transform to separate the pure tone from the noise. Directly or indirectly, this has implications for several important problems, including denoising and frequency estimation. For instance, frequency estimation can be cast as a pattern recognition problem as follows: Suppose that one estimates the frequency by decoding the measured, noisy spectrum as the closest member of a set of noise-free prototype spectra that form a codebook. In this sense, a transform representation that tends to deliver a smaller distance between the noise-free and noisy spectra corresponds to both greater accuracy in estimating the frequency of the pure tone and higher resolution in the sense that a smaller minimum distance implies a codebook that is more dense.

Classical techniques for frequency estimation are referenced in the work of Steinhardt and Bretherton [16]. Newer methods using wavelet transforms may be found in [17]–[19]. In all of these cases, frequency estimation performance is directly impacted by the distance between the thresholded and noise-free transform coefficient vectors. Transform domain thresholding has also been used for the more general denoising problem [20], [21]. The choice of threshold should be determined by the SNR. The binary nature of the threshold is intended to separate the smaller transform coefficients (presumably the noise) from the larger coefficients (presumably the signal). In the context of the counting measure of Donoho and Stark, of course, the binary nature of the threshold is fundamental and required [15], for otherwise, all the transform coefficients would be considered information bearing and relevant. This is unrealistic in the noisy case. Therefore, the goal of thresholding is to separate the noise-dominated coefficients from the transform-domain signal representation. Thresholding can be used to advantage in a number of transform-based signal processing strategies. For example, transform-domain thresholding is a natural process in denoising because the noise masks the support of the transformed signal. In low to moderate SNR situations, thresholding of the HOT coefficients will prove to be very attractive in our problem.

A brief digression at this point is necessary. It has been known for some time that the minimum variance, unbiased estimator for the frequency estimation problem may be found directly from the DFT coefficients [22]. However, this approach requires a computationally intensive nonlinear optimization, and so, alternative methods with reduced computational requirements have been proposed [23]. However, our problem is not that of frequency estimation but rather of signal separability. Nevertheless, for a fixed signal length, the HOT requires fewer computations than the DFT and may, in this regard, be interesting as a foundation for alternative frequency estimation algorithms.

Again, it should be kept in mind that we are not proposing a practical frequency estimation algorithm in this paper. Indeed, spectral estimation using transforms is a well-studied problem, and many excellent techniques applicable to specific application domains have been developed recently [22]–[24]. Instead, we are performing our experiment in the hopes of gaining fundamental insight into the relative joint localization properties of these transforms.

To that end, we consider a pure tone $x[n] = A\cos(\omega n + \phi)$, where A, ω, and ϕ are constant parameters, and $n \in \mathcal{D}$. The observed signal is given by $y[n] = x[n] + v[n]$, where $v[n]$ is a zero-mean, AWGN process uncorrelated with the signal $x[n]$, and $n \in \mathcal{D}$. The noise is added at a specific SNR, and the resulting signal $y[n]$ is scaled for unit norm. This scaling operation determines the value of A used for both the noisy signal $y[n]$ and the noise-free signal $x[n]$. Our objective is to compute the distance between the noise-free transform domain representation of $x[n]$ and the noisy transform domain representation of $y[n]$ as well as to study the distance with respect to the selected transform. The three transforms that we consider are the HOT, the DFT, and the DCT. For each trial of the experiment, we use 100 independent realizations. We first compute transform coefficient vectors for the 100 observed noisy signal realizations $y[n]$ and the single noise-free signal $x[n]$. The threshold is set at $\alpha C_{\max}$, where $C_{\max}$ is the largest magnitude coefficient in each transform, and $0 \leq \alpha \leq 1$. The experimental result is then given by the mean Euclidean distance between the 100 thresholded noisy transform coefficient vectors and the single thresholded noise-free transform coefficient vector. This procedure results from the interpretation of the experimental task as a pattern recognition (or feature extraction) problem. This approach

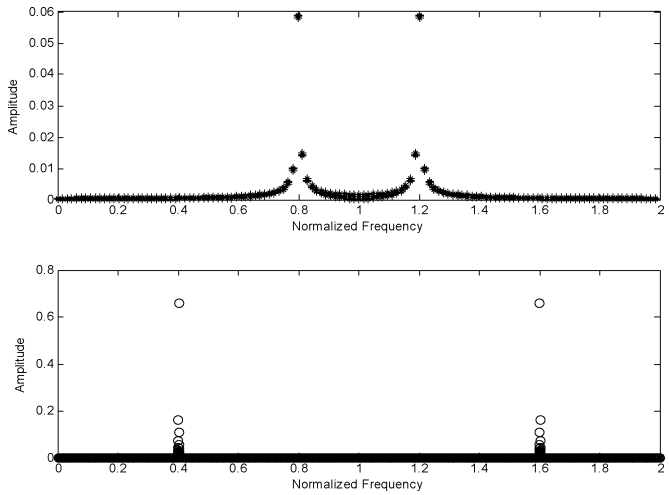


Fig. 6. (Top) HOT and (bottom) DFT coefficients for a pure tone of frequency 0.4 normalized.

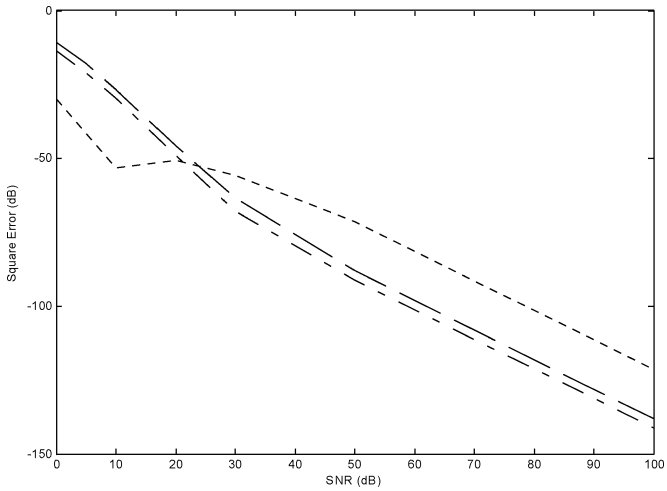


Fig. 7. Squared error versus SNR (DFT is dashed, HOT is dotted, and DCT is the dash-dot).

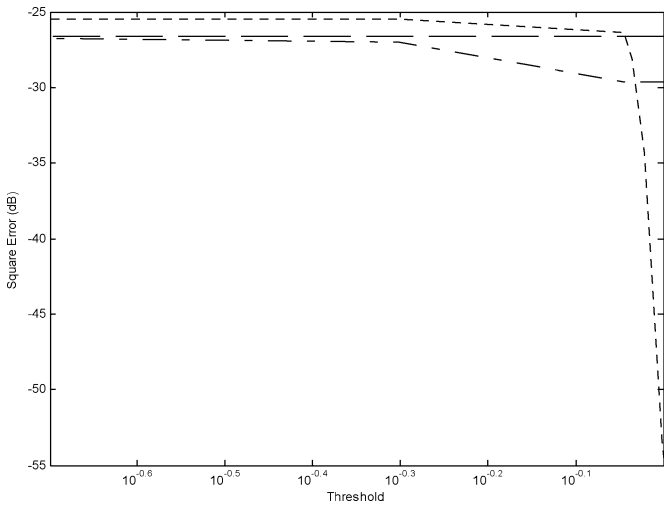


Fig. 8. Squared error versus threshold (DFT is dashed, HOT is dotted, and the DCT is the dash-dot).

places the thresholding as the last step of a two step process: 1) compute the transform, and 2) threshold the transform.

We detail the specific results below. Generally, we find in the high SNR case that the DCT and DFT deliver superior performance in our experiments, as expected. However, the HOT proves superior for the experimental task in the more practically significant low to moderate SNR cases.

We first examine the noise-free case without thresholding. Here, the DFT representation is nearly optimal and exhibits only a few significant (nonzero) coefficients. The HOT representation is also concentrated but results in a larger distance. This may be explained by the 1-1/2 dimensional nature of the HOT. Whereas the DFT uses an N-point inner product with the Fourier kernel, the HOT uses $K = \sqrt{N}$ different K-point inner products. This difference is readily apparent in Fig. 6, where the noise-free signal $x[n] = \cos(0.4\pi n)$ is analyzed for $N = 128^2$ $(K = 128)$ points. As may clearly be seen from the upper graph, the HOT representation of this signal is not as localized as the DFT representation. Note that the largest magnitude HOT coefficients occur at a normalized sequency [24], [25] that is twice the DFT normalized frequency. This doubling results directly from the fact that $N = K^2$.

We now consider the more realistic case, where only noisy observations $y[n]$ are available. The presence of noise introduces uncertainty to the transform coefficients of the observed signal. The presence of the noise means that, in general, none of the transform coefficients will be zero. In the noisy case, we find that, as the SNR increases, the performances of the DFT and the DCT are superior to that of the HOT according to the $N = K^2$ resolution of the basis functions, with the DCT being slightly superior to the DFT, owing to the exact match between the noise-free signal and the DCT basis. This match is clearly evident in Fig. 7, where the DCT curve is uniformly below the DFT curve by a few decibels at all SNR. Observe, however, that the performance of the HOT for low to moderately high SNR (those under 20–25 dB) is superior. This range of SNR is the most relevant for practical applications.

Now, we fix the SNR at the moderate value of 10 dB and observe the experimental performance as it varies with our choices of threshold, transform length, and sinusoidal frequency. We note the following.

- As shown in Fig. 8, we see that, in the limit as the threshold vanishes, the performances of the DFT and the DCT converge. Note that the performance yielded by the HOT basis in this regime is worse because of the $K = \sqrt{N}$ relationship. However, when the threshold is large and lots of noise is actually being removed, the performance of the HOT-based algorithm is superior to both the DFT- and DCT-based methods by an increasingly significant amount. This is a striking demonstration of the superior ability of the HOT basis to separate signal from noise at moderate SNR. This plot also shows that the choice of threshold is significant and that an appropriate threshold must be chosen with the transform in mind.
- In Fig. 9, we study the performance as a function of the signal length N. As N increases, we see the performance advantage afforded by the HOT increases from about 5 dB to more than 26 dB. Block lengths greater

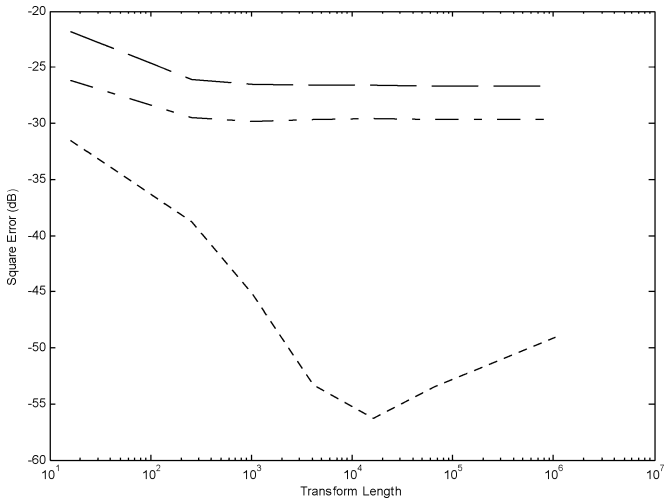


Fig. 9. Squared error versus the transform length, $N = K^2$ (DFT is dashed, HOT is dotted, and the DCT is the dash-dot).

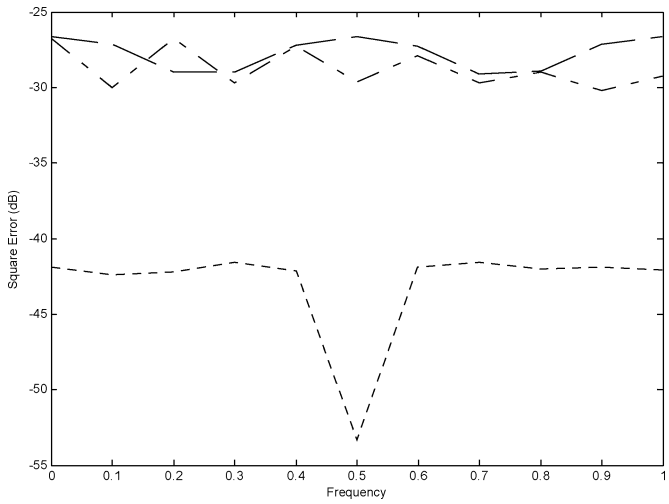


Fig. 10. Squared error versus the frequency (DFT is dashed, HOT is dotted, and the DCT is the dash-dot).

than $N = K^2 = 2^{14}$ are rarely desirable in real-world applications. Over the most important range of typical block lengths, the HOT performance, as measured by the Euclidean distance, is between one and two orders of magnitude superior to both the DFT and the DCT. This performance improvement delivered by the HOT is predicted by the improving frequency resolution (in the sense of the interleaved length-K DFTs sampling the unit circle more densely) of the HOT basis as $N \to \infty$.

Note that the HOT curve is not monotonically decreasing in the data length. This results in part from the choice of threshold and our distance computation method—too many of the coefficients are above the threshold as the number of coefficients increases significantly. Accumulating numerical roundoff errors may also contribute to the observed nonmonotonicity as the length of the required DFTs grows exponentially beyond $N = 10^6$. However, these two factors cannot fully explain the graph. We suspect that additional factors are involved, but we are unable to identify them at present. We have verified that the point where the upturn occurs is sensitive to the threshold value.

- Fig. 10 depicts the performance as a function of the sinusoidal frequency of the pure tone signal. The relative performances between the three approaches are not much affected by the absolute frequency of the sinusoid. For the case of SNR $=$ 10 dB depicted in Fig. 10, the HOT is superior to both the DFT and DCT at all frequencies.

We also considered limited cases where the additive noise is not Gaussian. The results achieved are consistent with those reported above for the Gaussian case. However, we did see that the performance may be somewhat sensitive to the choice of threshold, as would be predicted by our analysis in the first bullet above (see, e.g., Fig. 8). For instance, a larger threshold is required to yield a performance similar to the above Gaussian case when the additive noise $v[n]$ is a sample of a rectified Gaussian process. However, no change of threshold relative to the Gaussian case seems to be required when $v[n]$ is a uniformly distributed zero-mean process.

V. Conclusion

In this paper, we have considered joint uncertainty measures based on entropy as opposed to energy or variance. The classical notion of uncertainty for continuous-time signals imparted by the Heisenberg–Weyl inequality cannot be directly extended to discrete-time signals because of the difficulty in defining a translation-invariant concept of variance in the discrete domain. Using entropy instead of variance, we obtained Hirschman uncertainty measures that generalize the Heisenberg–Weyl notion and that may be extended to the discrete domains in a straightforward manner. For the first time, this enabled us to consider relationships between a single family of measures that comprehensively treats all three cases: $L^2(\mathbb{R}^n)$, $\ell^2(\mathbb{Z})$, and $\ell^2(\mathbb{Z}/N\mathbb{Z})$. We have shown that the minimal time-frequency concentrations using the entropy-based Hirschman uncertainty principles are the expected Gaussians in the continuous-time, continuous-frequency case $(L^2(\mathbb{R}^n))$ but are not in both the discrete-time, continuous-frequency case $(\ell^2(\mathbb{Z}))$ and the discrete-time, discrete-frequency case $(\ell^2(\mathbb{Z}/N\mathbb{Z}))$. In the two discrete-time cases, the optimizers are based on modulations, translations, and periodizations of the unit-sample sequence $\delta[n]$. In particular, the optimizers for $\ell^2(\mathbb{Z}/N\mathbb{Z})$ are the picket fence signals developed in [15], where concentration was quantified by the counting measure of the signal support in the joint time-frequency product space after thresholding. The optimizers for $\ell^2(\mathbb{Z})$ may be obtained from a limiting process on the optimizers from $L^2(\mathbb{R}^n)$. Likewise, the optimizers for $\ell^2(\mathbb{Z})$ may be obtained from a different limiting process on the optimizers from $\ell^2(\mathbb{Z}/N\mathbb{Z})$. At present, however, we know of no direct way to unify the continuous $(L^2(\mathbb{R}^n))$ and digital $(\ell^2(\mathbb{Z}/N\mathbb{Z}))$ cases. We have shown, however, that the entropic view is consistent with both the Heisenberg–Weyl interpretation and the newer interpretations due to Donoho and Stark [15] and Dembo, Cover, and Thomas [2]. These results are summarized in Table I. The arrows in the convergence column indicate the

TABLE I
MINIMIZERS IN THE THREE CASES

Case	Time	Frequency	Convergence
$L^2(\mathbb{R}^n)$	Gaussian	Gaussian	$\downarrow$ as $t \to 0$
$\ell^2(\mathbb{Z})$	Kronecker Delta	Constant	
$\ell^2(\mathbb{Z}/N\mathbb{Z})$	Picket Fence	Picket Fence	$\uparrow$ as $t \to \infty$

direction of connection as well as the kind of limit that must be performed.

We performed a very simple experiment that indicates that the HOT is superior to the DFT and DCT in terms of its ability to separate or resolve two limiting cases of localization in frequency, viz. pure tones and additive white noise. We believe that these differences arise from the use of entropy rather than energy as an optimality criterion and are intimately related to the apparent incongruence between the infinitely supported continuous-time case and the finitely supported discrete-time case. It is of great future interest to investigate the dual problem of time-transient estimation in noise and the intermediate case of gated sinusoids in noise.

REFERENCES

[1] V. DeBrunner, M. Özaydın, and T. Przebinda, "Resolution in time-frequency," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 783–788, Mar. 1999.

[2] 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.

[3] 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.

[4] T. Przebinda, V. E. DeBrunner, and M. Özaydın, "Using a new uncertainty measure to determine optimal bases for signal representations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Phoenix, AZ, Mar. 15–19, 1999, paper 1575, pp. 1365–1368.

[5] ——, "The optimal transform for the discrete Hirschman uncertainty principle," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 2086–2090, Jul. 2001.

[6] 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. Conf. Acoust., Speech, Signal Process.*, vol. 1, Istanbul, Turkey, Jun. 5–9, 2000, paper 1312, pp. 81–84.

[7] M. Özaydın and T. Przebinda, "An entropy-based uncertainty principle for a locally compact abelian group," *J. Functional Anal.*, pp. 241–252, 2004, to be published.

[8] I. I. Hirschman, "A note on entropy," *Amer. J. Math.*, vol. 79, pp. 152–156, 1957.

[9] L. Hörmander, *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis*. New York: Springer-Verlag, 1983.

[10] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, pp. 379–656, 1948.

[11] W. Beckner, "Inequalities in Fourier analysis," *Ann. Math.*, pp. 159–182, 1975.

[12] ——, "Pitt's inequality and uncertainty principle," *Proc. Amer. Math. Soc.*, pp. 1897–1905, 1995.

[13] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng. Lond.*, vol. 93, no. III, pp. 429–457, 1946.

[14] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 2, no. 7, pp. 1160–1169, Jul. 1985.

[15] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, pp. 906–931, 1989.

[16] A. O. Steinhardt and C. Bretherton, "Thresholds in frequency estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 10, Tampa, FL, Apr. 1985, pp. 1273–1276.

[17] P. Moulin, "Wavelet thresholding techniques for power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3126–3136, Nov. 1994.

[18] S. S. Chen and D. L. Donoho, "Application of basis pursuit in spectrum estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Seattle, WA, May 12–15, 1998, pp. 1865–1868.

[19] A. T. Walden, D. B. Percival, and E. J. McCoy, "Spectrum estimation by wavelet thresholding of multitaper estimators," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3153–3165, Dec. 1998.

[20] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[21] M. Jansen, *Noise Reduction by Wavelet Thresholding*, ser. ser. Lecture Notes in Statistics. New York: Springer-Verlag, 2001, vol. 161.

[22] B. G. Quinn, "A fast efficient technique for the estimation of frequency: interpretation and generalization," *Biometrika*, vol. 86, no. 1, pp. 213–220, Mar. 1999.

[23] B. G. Quin and P. J. Kootsookos, "Threshold behavior of the maximum-likelihood estimator of frequency," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3291–3294, Nov. 1994.

[24] C. J. Zarowski, M. Yunik, and G. O. Martens, "DFT spectrum filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 4, pp. 461–470, Apr. 1988.

[25] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, pp. 30–44, 1991.

Victor DeBrunner (S'82–M'90–SM'00) was born in Auburn, AL, on August 21, 1962. He received the B.E.E. degree from Auburn University in 1984 and the M.S. and Ph.D. degrees in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, in 1986 and 1990, respectively.

He has been with the University of Oklahoma, Norman, since 1990, where he is currently a Professor in the School of Electrical and Computer Engineering. His research interests include signal and image processing algorithms and implementations.

Dr. DeBrunner has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and he is currently an Associate Editor for both the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I. He is also a member of the steering committee for the Asilomar Conference on Signals, Systems, and Computers.

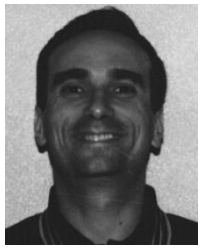
Joseph P. Havlicek (S'84–M'88–SM'98) was born in West Lafayette, IN, on April 7, 1964. He received the B.S. degree in 1986 and the M.S. degree in 1988 from Virginia Polytechnic Institute and State University, Blacksburg, and the Ph.D. degree in 1996 from the University of Texas at Austin, all in electrical engineering.

He was a software developer with Management Systems Laboratories, Blacksburg, from 1984 to 1987 and with IBM, Austin, TX, throughout 1993, where he was affiliated with Ralph Kirkley Associates. From 1987 to 1997, he was with the Naval Research Laboratory, Washington, DC, where he was associated with SFA, Inc., from 1987 to 1989. He joined the School of Electrical and Computer Engineering, University of Oklahoma, Norman, in 1997, where he is currently an Associate Professor. His research interests include signal, image, and video processing and intelligent transportation systems.

Dr. Havlicek was the recipient of the University of Oklahoma IEEE Favorite Instructor Award in 1998 and 2000, the Department of the Navy Award of Merit for Group Achievement in 1990, and the 1992 University of Texas Engineering Foundation Award for Exemplary Engineering Teaching while Pursuing a Graduate Degree. He serves on the organizing committee of the IEEE Southwest Symposium on Image Analysis and Interpretation and is a member of Tau Beta Pi, Phi Kappa Phi, and Eta Kappa Nu.

Tomasz Przebinda received the Ph.D. degree from Yale University, New Haven, CT, in 1987.

He held positions with the University of Utah, Salt Lake City, and Louisiana State University, Baton Rouge. He has been at the University of Oklahoma, Norman, since 1990, where he is currently a Professor of mathematics. His interests include representation theory of real reductive groups, harmonic analysis, wavelets, and signal processing. He also served as a faculty consultant to the Advanced Placement Calculus Reading program in 1997, 1999, and 2001.



Murad Özaydın received the B.Sc. degree from Middle East Technical University, Ankara, Turkey, in 1977 and the Ph.D. degree from Purdue University, West Layayette, IN, in 1984, both in mathematics.

After four years at the University of Wisconsin, Madison, he joined the mathematics department at the University of Oklahoma, Norman, where he is currently a Professor and Director of Graduate Studies. His research interests include algebraic topology, group theory, combinatorics, harmonic analysis, and signal processing.