MEASURING TIME-FREQUENCY INFORMATION CONTENT USING THE RÉNYI ENTROPIES

Richard G. Baraniuk,[◇] *Patrick Flandrin,*[□] *Augustus J. E. M. Janssen,*[○] *and Olivier Michel* [•] [*]

[◇] Department of Electrical and Computer Engineering
Rice University
6100 Main Street, Houston, TX 77005–1892, USA
E-mail: richb@rice.edu, Fax: 713.348.6196

[□] Laboratoire de Physique (URA 1325 CNRS)
Ecole Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, France
E-mail: flandrin@ens-lyon.fr

[○] Philips Research Laboratories
Prof. Holstlaan 4, WY–81
5656 AA Eindhoven, The Netherlands
E-mail: a.j.e.m.janssen@philips.com

[•] Laboratoire d'Astrophysique
Parc Valrose
06108 Nice Cedex 2, France
E-mail: olivier.michel@unice.fr

Submitted to *IEEE Transactions on Information Theory*, August 1998
Revised September 2000
Index Terms: Time-frequency analysis, Wigner distribution, Rényi entropy, complexity

Abstract

The generalized entropies of Rényi inspire new measures for estimating signal information and complexity in the time-frequency plane. When applied to a time-frequency representation from Cohen's class or the affine class, the Rényi entropies conform closely to the notion of complexity that we use when visually inspecting time-frequency images. These measures possess several additional interesting and useful properties, such as accounting and cross-component and transformation invariances, that make them natural for time-frequency analysis. This paper comprises a detailed study of the properties and several potential applications of the Rényi entropies, with emphasis on the mathematical foundations for quadratic TFRs. In particular, for the Wigner distribution we establish that there exist signals for which the measures are not well defined.

[*]This work was supported by the National Science Foundation, grants MIP–9457438 and CCR–997318, the Office of Naval Research, grants N00014–95–1–0849 and N00014–99–1–0813, the NATO Collaborative Research Programme, grant GRG–950202, and URA 1325 CNRS.

Portions of this work have been presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing in Adelaide, Australia, May 1994, and at the IEEE International Symposium on Information Theory in Whistler, Canada, September 1995.

1 Introduction

The term *component* is ubiquitous in the signal processing literature. Intuitively, a component is a concentration of energy in some domain, but this notion is difficult to translate into a quantitative concept [1–3]. In fact, the concept of a signal component may never be clearly defined.

The use and abuse of this term is particularly severe in the literature on time-frequency analysis. Time-frequency representations (TFRs) generalize the concept of the time and frequency domains to a joint time-frequency function $C_s(t,f)$ that indicates how the frequency content of a signal s changes over time [4–6]. Common themes in the literature include the suppression of TFR cross-components, the concentration and resolution of auto-components, and the property that the time-varying spectral analysis of TFRs separates signal components such as parallel chirps that overlap in both time and frequency. Moreover, the quality of particular TFRs is very often judged based on subjective criteria related to the components of the signal being analyzed.

In this paper, rather than address the question "what is a component?" directly, we will investigate a class of quantitative measures of deterministic signal *complexity* and *information content*.[1] While they do not yield direct answers regarding the locations and shapes of components, these measures are intimately related to the concept of a signal component, the connection being the intuitively reasonable supposition that signals of high complexity (and therefore high information content) must be constructed from large numbers of elementary components.

Moment-based measures, such as the time-bandwidth product and its generalizations to second-order time-frequency moments [4–6, 8, 9] have found wide application, but unfortunately measure neither signal complexity nor information content [1,2]. To demonstrate, consider a signal comprised of two components of compact support, and note that while the time-bandwidth product increases without bound with separation, signal complexity clearly does not increase once the components become disjoint.

A more promising approach to complexity based on entropy functionals exploits the analogy between signal energy densities and probability densities [1]. Just as the instantaneous and spectral amplitudes $|s(t)|^2$ and $|S(f)|^2$ behave as unidimensional densities of signal energy in time and frequency, TFRs try very hard to act as bidimensional energy densities in time-frequency. In particular, there exist TFRs whose marginal properties parallel those of probability densities:

$$\int C_s(t,f)\,df \;=\; |s(t)|^2, \qquad \int C_s(t,f)\,dt \;=\; |S(f)|^2, \tag{1}$$

$$\iint C_s(t,f)\,dt\,df \;=\; \int |s(t)|^2\,dt \;=:\; \|s\|_2^2. \tag{2}$$

The quadratic TFRs of the large and useful Cohen's class can be obtained as the convolution [4–6]

$$C_s(t,f) \;:=\; \iint W_s(u,v)\;\Phi(t-u,\;f-v)\,du\,dv \;:=\; (W_s * \Phi)(t,f) \tag{3}$$

[1] Many alternative measures of complexity exist (such as Kolmogorov's [7]). These measures lie beyond the scope of this paper, however, since they are typically applied to *signals*, whereas our analysis is based on *TFRs*, considered as pseudo probability densities.

of a kernel function Φ with the Wigner distribution W_s of the signal

$$W_s(t,f) \;:=\; \int s\left(t+\frac{\tau}{2}\right) s^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi\tau f}\, d\tau. \tag{4}$$

The probabilistic analogy evoked by (1) and (2) suggests the classical Shannon entropy [10] (given here for unit-energy signals)

$$H(C_s) \;:=\; -\iint C_s(t,f)\, \log_2 C_s(t,f)\, dt\, df \tag{5}$$

as a natural candidate for measuring the complexity of a signal through its TFR:[2] The peaky TFRs of signals comprised of small numbers of elementary components would yield small entropy values, while the diffuse TFRs of more complicated signals would yield large entropy values. Unfortunately, however, the negative values taken on by most TFRs (including all fixed-kernel Cohen's class TFRs satisfying (1)) prohibit the application of the Shannon entropy due to the logarithm in (5).

In [1], Williams, Brown, and Hero sidestepped the negativity issue by employing the generalized entropies of Rényi [11] (again for unit-energy signals)

$$H_\alpha(C_s) \;=\; \frac{1}{1-\alpha}\log_2 \iint C_s^\alpha(t,f)\, dt\, df. \tag{6}$$

Parameterized by $\alpha > 0$, this class of information measures is obtained simply by relaxing the mean value property of the Shannon entropy from an arithmetic to an exponential mean. (Shannon entropy appears as $\alpha \to 1$.) In several empirical studies, Williams, Brown, and Hero found that in addition to appearing immune to the negative TFR values that invalidate the Shannon approach, the 3rd-order Rényi entropy seemed to measure signal complexity. Figure 1 repeats the principal experiment of [1]: The 3rd-order Rényi entropy $H_3(W_s)$ of the Wigner distribution of the sum $s(t) := g(t-\Delta t)+g(t+\Delta t)$ of two Gaussian pulses is plotted versus the separation distance $2\Delta t$. (At $\Delta t = 0$, the two pulses coincide and therefore, because of the assumed energy renormalization, have the same information content as a solitary pulse. The time-bandwidth product of s is also plotted. It is clear from the Figure that, unlike the time-bandwidth product, which grows without bound with Δt, the Rényi entropy saturates exactly one bit above the value $H_3(W_g) = -\frac{1}{2}\log_2\frac{4}{3} \approx -0.208$.[3] Similar results hold for N separated copies of $g(t)$ ($\log_2 N$ bits information gain). To summarize, independent of the definition of signal component, the Rényi entropy indicates a "doubling of complexity" in s as the separation a moves from 0 to ∞.

This paper comprises a detailed study of the properties and some potential applications of the Rényi time-frequency information measures (6), with emphasis on the mathematical foundations for quadratic TFRs. In Section 2, after reviewing the development of these measures, we examine their existence and show that for each odd $\alpha \geq 3$ there exist signals for which (6) is *not* defined (due to $\iint C_s^\alpha(t,f)\, dt\, df < 0$). This unprecedented result surprises, for it indicates that the Rényi

[2]Note that the complexity measure (Shannon entropy here) is applied not to a signal or process, but to a TFR that plays a rôle analogous to a probability density function.

[3]Readers should not be alarmed by negative Rényi entropy values. Even the Shannon entropy takes on negative values for certain distributions in the continuous-variable case.

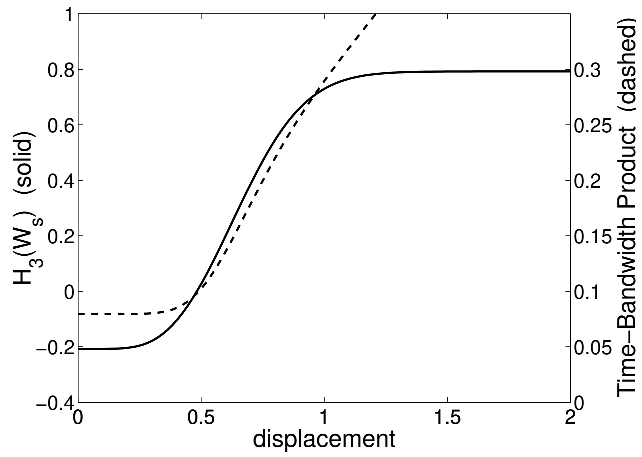


Figure 1: *Solid curve and axis on left: The 3rd-order Rényi entropy* $H_3(W_s)$ *of the Wigner distribution of the sum* $s(t) := g(t - \Delta t) + g(t + \Delta t)$ *of two Gaussian components plotted versus the displacement parameter* Δt *(see (32) in Section 3.2 for the exact signal definition). (The asymptotic* $H_3(W_s)$ *levels are* $-\frac{1}{2}\log_2 \frac{4}{3} \approx -0.208$ *and* $\frac{1}{2}\log_2 3 \approx 0.792$ *bits.) Dotted curve and axis on right: Time-bandwidth product of* s*.*

formalism is not universally applicable to time-frequency analysis. Counterexample signals are easily constructed for large odd α; however, our $\alpha = 3$ counterexamples are quite contrived. This is consistent with the ample numerical evidence [1,8,9,12,13] indicating that the 3rd-order entropy is defined for a broad class of signals and TFRs, including even those distributions taking locally negative values. When defined, these measures have some striking properties that we investigate in Section 3:

1. $H_\alpha(C_s)$ counts the "number of components" in a multicomponent signal.

2. For odd orders $\alpha > 1$, $H_\alpha(C_s)$ is asymptotically invariant to TFR cross-components and therefore does not count them.

3. $H_\alpha(C_s)$ exhibits extreme sensitivity to phase differences between closely spaced components. This sensitivity can be reduced through smoothing in time-frequency. We provide analytical results for the sum of two Gaussian signals.

4. The range of $H_\alpha(C_s)$ values is bounded from below. For the Wigner distribution, a single Gaussian pulse attains the lower bound.

5. The values of $H_\alpha(C_s)$ are invariant to time and frequency shifts of the signal. Certain TFRs provide an additional invariance to scale changes, while the Wigner distribution boasts complete invariance to symplectic transformations on the time-frequency plane. For more general invariances, the Rényi theory extends easily to encompass not only the TFRs of the affine class [14] but also the generalized representations of the unitarily equivalent Cohen's and affine classes [15–17].

In Section 4, we discuss the application of these measures as objective functions in optimized time-frequency analysis and introduce the notion of Rényi dimension. We close with a discussion and conclusions. Proofs of the various results are contained in a number of Appendices.

2 The Rényi Entropies

2.1 Rényi entropy of a probability density

In [11] Rényi introduced an alternative axiomatic derivation of entropy based on *incomplete* probability mass distributions $p = \{p_1, p_2, \dots p_n\}$ whose total probabilities sum to $w(p) := \sum_i p_i \leq 1$. He observed that the Shannon entropy $H(p) := \frac{-1}{w(p)} \sum_i p_i \log_2 p_i$ uniquely satisfies the axioms of symmetry, continuity, normalization, additivity, and, in addition, the mean value condition

$$H(p \cup q) \;=\; \frac{w(p)\, H(p) \;+\; w(q)\, H(q)}{w(p) \;+\; w(q)}\,. \tag{7}$$

Here p and q are any two incomplete densities such that $w(p) + w(q) \leq 1$, and $p \cup q$ signifies the composite density $\{p_1, p_2, \dots, p_n, q_1, q_2, \dots, q_m\}$.

Extending the arithmetic mean in (7) to a generalized mean yields generalized entropies closely resembling Shannon's. Considering the generalized mean value condition

$$H^{\mathrm{R}}(p \cup q) \;=\; m^{-1}\left(\frac{w(p)\, m[H^{\mathrm{R}}(p)] \;+\; w(q)\, m[H^{\mathrm{R}}(q)]}{w(p) \;+\; w(q)}\right), \tag{8}$$

with m a continuous monotone function, Rényi demonstrated that just two types of functions m are compatible with the other four axioms. The first, $m_1(x) \;=\; ax + b$, yields the arithmetic mean (7) and the Shannon entropy. The second,

$$m_2(x) \;=\; 2^{(\alpha-1)x}, \qquad \alpha > 0, \quad \alpha \neq 1, \tag{9}$$

yields the functional

$$H_\alpha^{\mathrm{R}}(p) \;:=\; \frac{1}{1-\alpha} \log_2 \frac{\sum_i p_i^\alpha}{\sum_i p_i} \tag{10}$$

now known as the Rényi entropy of order α. The Shannon entropy can be recovered as $\lim_{\alpha\to 1} H_\alpha^{\mathrm{R}} = H$. Extension of $H_\alpha^{\mathrm{R}}(p)$ to continuous-valued bivariate densities $P(x,y)$ is straightforward:

$$H_\alpha^{\mathrm{R}}(P) \;:=\; \frac{1}{1-\alpha} \log_2 \frac{\iint P^\alpha(x,y)\, dx\, dy}{\iint P(x,y)\, dx\, dy}\,. \tag{11}$$

We emphasize that since the passage from the Shannon entropy to the class of Rényi entropies involves only the relaxation of the mean value property from an arithmetic to an exponential mean, H_α^{R} behaves much like H [11].

2.2 Rényi entropy of a time-frequency representation

The central theme of this paper is the application of entropy measures to TFRs to measure the complexity and information content of nonstationary signals indirectly via the time-frequency plane. Our primary TFR tools of choice lie in Cohen's class [4–6], which can be expressed as in (3) as the convolution between the Wigner distribution and a real-valued kernel Φ.[4] The kernel and its inverse Fourier transform ϕ completely determine the properties of the corresponding TFR. For example, a fixed-kernel TFR possesses the energy preservation property (2) provided $\phi(0,0) = 1$ and the marginal properties (1) provided $\phi(\theta,0) = \phi(0,\tau) = 1 \ \forall\, \theta,\tau$. Besides the Wigner distribution, examples of Cohen's class TFRs include the spectrogram (ϕ = ambiguity function of the time-reversed window function) and the smoothed pseudo-Wigner distributions ($\phi(\theta,\tau) = h_1(\theta)\, h_2(\tau)$) [4–6]

The analogy between TFRs and bidimensional probability densities discussed in the Introduction breaks down at at least two key points. First, because of the freedom of choice of kernel function, the TFR of a given signal is nonunique, with many different distributions "explaining" the same data. Second, and more pertinent to the present discussion, most Cohen's class TFRs are nonpositive and, therefore, cannot be interpreted strictly as densities of signal energy.[5] These locally negative values will clearly play havoc with the logarithm in the Shannon entropy (5).

While the Rényi entropies (6) appear intriguing and encouraging for time-frequency application [1, 8, 9, 12, 13], it has remained an open question whether in general these measures can cope with the locally negative values of Cohen's class TFRs. In order for (6) to be defined for a signal s, we clearly need $C_s^\alpha(t,f)$ to be real and such that

$$\iint C_s^\alpha(t,f)\, dt\, df \ > \ 0. \tag{12}$$

Noninteger orders α yield complex $C_s^\alpha(t,f)$ values and so appear of limited utility. Integer orders α that are even pose no such hazards, since the integral of the positive function $C_s^\alpha(t,f)$ remains positive.

Unfortunately, odd integer orders are not so robust: For each odd $\alpha \geq 3$, there exist signals in $L^2(\mathbb{R})$ and TFRs such that (12) fails, leaving (6) undefined. In Appendix A, we develop the following counterexamples for the Wigner distribution W_s:[6]

1. For sufficiently large odd α, (12) fails for any smooth, rapidly decaying, odd signal.
2. For odd integers $\alpha \geq 5$, (12) fails for the 1st-order Hermite function.
3. For $\alpha = 3$, (12) fails for a particular linear combination of the 3rd and 9th-order Hermite functions (see Figure 2).

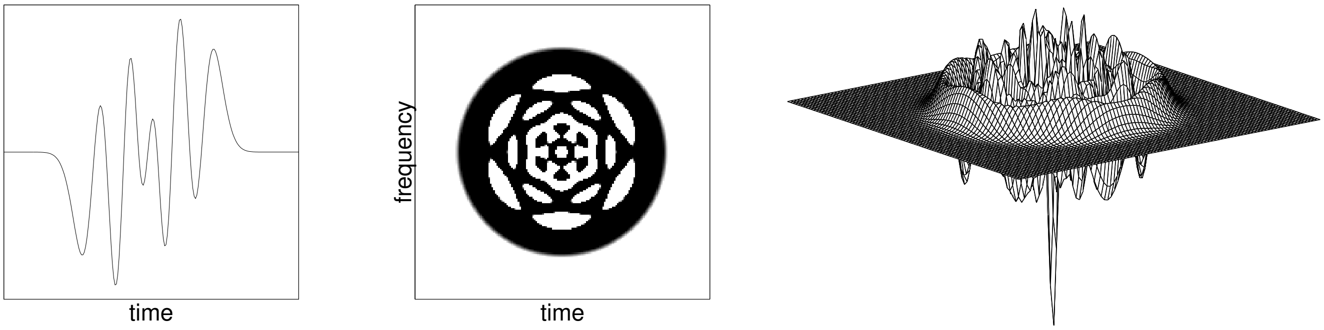
As noted in the Introduction, these surprising results suggest that we must proceed with caution when applying TFR-based Rényi entropies.

Negative results aside, a preponderance of numerical evidence [1, 8, 9, 12] indicates that the 3rd order entropies are well-defined for large classes of signals and TFRs (our counterexamples only

[4] To avoid cumbersome machinations, we will restrict our attention to the Wigner distribution and all TFRs obtained from (3) with $\Phi \in L^1(\mathbb{R}^2)$. Since we will be interested in odd powers of TFRs (see (6)), we furthermore assume that that kernel is a real-valued function.

[5] While there do exist nonquadratic classes of positive TFRs that satisfy (2) and (1) [4], we will consider only quadratic TFRs in this paper.

[6] Clearly these counterexamples invalidate the "proofs" of existence sketched in [8, 9].



(a) (b) (c)

Figure 2: *Example of a signal for which the 3rd-order Rényi entropy $H_3(W_s)$ of the Wigner distribution is not defined. (a) The signal s consists of a special combination of two odd-order Hermite functions (see Appendix A.3). Its Wigner distribution (b) as an image and (c) in 3-d (note the large negative peak).*

apply to the Wigner distribution). In Appendix A, we spend a considerable effort to find a signal for which (12) fails for $\alpha = 3$. Also, a small amount of Gaussian smoothing of W_s is generally enough for (12) to hold. This indicates that the examples for which (12) fails for $\alpha = 3$ are rather exceptional.

Throughout the balance of this paper, we will assume that all signals under consideration are such that the formula (6) is well-defined.

We close this section with some important notes on normalization. In their experiments, Williams, Brown, and Hero actually employed not H_α^{R} from (11), but a pre-normalized version equivalent to normalizing the signal energy *before* raising the TFR to the α power:

$$H_\alpha(C_s) \ := \ \frac{1}{1-\alpha} \log_2 \iint \left(\frac{C_s(t,f)}{\iint C_s(u,v)\, du\, dv} \right)^\alpha dt\, df. \tag{13}$$

The two measures are related by

$$H_\alpha^{\mathrm{R}}(C_s) \ = \ H_\alpha(C_s) \ - \ \log_2 \|s\|_2^2, \tag{14}$$

and thus $H_\alpha^{\mathrm{R}}(C_s)$ varies with the signal energy. Since an information measure should be invariant to the energy of the signal being analyzed, we will adhere strictly to the definition (13) for the duration of this paper. Discretization of this measure (by setting $t = n\delta_t, f = k\delta_f$ with $n, k \in \mathbb{Z}$) for use with computer generated, discrete TFRs yields

$$H_\alpha(C_s[n,k]) \ := \ \frac{1}{1-\alpha} \log_2 \sum_n \sum_k \left(\frac{C_s[n,k]}{\sum_{n'} \sum_{k'} C_s[n',k']} \right)^\alpha \ + \ \log_2 \delta_t \delta_f. \tag{15}$$

The frequency step constant is computed as $\delta_f = \frac{F}{K}$, given K uniform frequency samples spanning the frequency range of F Hz/sample. For both continuous and discrete TFRs, operation in (t,ω) coordinates, with radial frequency $\omega = 2\pi f$ rad/s, introduces an offset: $H_\alpha(C_s(t,\omega)) = H_\alpha(C_s(t,f)) + \log_2 2\pi$. Sang and Williams explore an alternative magnitude normalization of the Rényi entropy in [13].

3 Properties of the Rényi Time-Frequency Information Measure

We now conduct a detailed analysis of the properties of the Rényi entropy (when it exists) that make it a fascinating and useful tool for studying the information content of time-varying deterministic signals.

3.1 Component counting and cross-component invariance

If TFRs were "quasi-linear" — such that each signal component contributed essentially separately to the overall time-frequency representation with no intervening cross-components — then the analogy between TFRs and probability density functions would predict an additive or counting behavior from the Rényi entropy. This is reasonable, since a (non-overlapping) combination of basic signal components is more "complex" than the individual components.

To gain more intuition into this most fundamental property of H_α, imagine applying this measure first to a compactly supported signal s using an ideal, quasi-linear TFR $I_s(t,f)$. Denote the length of the supporting interval J_σ of s by σ and assume that $I_s(t,f)=0$ for all $t \notin J_\sigma$ and for all f.

Form the two-component signal $s+\mathcal{T}s$, where $(\mathcal{T}s)(t) := s(t-\Delta t)$ represents translation by time Δt. Assuming that $\Delta t > \sigma$, the distribution is given by

$$I_{s+\mathcal{T}s}(t,f) \;=\; I_s(t,f) \;+\; I_s(t-\Delta t, f). \tag{16}$$

Since $I_s(t,f)$ is compactly supported in the time direction, we can appeal to the analogy between the right hand side of (16) and the composite probability distribution $p \cup q$ in (8) to compute $H_\alpha(I_{s+\mathcal{T}s})$. In particular, substituting (14) into (8) with $m(x) = 2^{(\alpha-1)x}$ and employing the facts $H_\alpha(I_{\mathcal{T}s}) = H_\alpha(I_s)$ and $\|s+\mathcal{T}s\|_2^2 = 2\,\|s\|_2^2$, some simple algebra yields

$$H_\alpha(I_{s+\mathcal{T}s}) \;=\; H_\alpha(I_s) \;+\; 1. \tag{17}$$

In words, the two-component signal $s+\mathcal{T}s$ contains exactly one bit more information than the one-component signal s.[7] The saturation levels of the entropy curve in Figure 1 display precisely this behavior.

While this simple analysis provides considerable insight into the counting behavior of H_α, it does not take into account the nonideal, nonlinear behavior of the quadratic TFRs of Cohen's class. In particular, we have ignored the presence of cross-components in these distributions [4–6], which violate the linearity assumption underlying (16). We will broaden our analysis to encompass actual TFRs in two stages.

First, consider the Wigner distribution (4) of the compactly supported, two-component signal $s+\mathcal{T}s$:

$$W_{s+\mathcal{T}s}(t,f) \;=\; W_s(t,f) \;+\; X_{s,\mathcal{T}s}(t,f) \;+\; W_{\mathcal{T}s}(t,f). \tag{18}$$

[7] Note that the post-normalized entropy H_α^{R} from (11) exhibits the invariance $H_\alpha^{\mathrm{R}}(I_{s+\mathcal{T}s}) = H_\alpha^{\mathrm{R}}(I_s)$, since the energies of s and $s+\mathcal{T}s$ differ by a factor of two.

The term $X_{s,\mathcal{T}s}(t,f)$, called the *cross-component* between s and $\mathcal{T}s$, is derived from the cross-Wigner distribution [4–6, 18]

$$X_{s,\mathcal{T}s}(t,f) \;:=\; 2\,\mathrm{Re} W_{s,\mathcal{T}s}(t,f), \tag{19}$$

with the cross-Wigner distribution between signals r and s defined by

$$W_{r,s}(t,f) \;:=\; \int r\left(t+\frac{\tau}{2}\right) s^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi\tau f} d\tau. \tag{20}$$

In general, the Rényi entropy

$$H_\alpha(W_{s+\mathcal{T}s}) \;=\; \frac{1}{1-\alpha} \log_2 \frac{1}{\|s+\mathcal{T}s\|_2^{2\alpha}} \iint \left[W_s(t,f) + X_{s,\mathcal{T}s}(t,f) + W_{\mathcal{T}s}(t,f)\right]^\alpha dt\, df \tag{21}$$

involves a complicated polynomial in W_s, $X_{s,\mathcal{T}s}$, and $W_{\mathcal{T}s}$. However, due to the compact support of s and thus W_s [4–6], for separations $\Delta t > 2\sigma$, these terms lie disjoint in the time-frequency plane, and a tremendous simplification results:

$$\begin{aligned} H_\alpha(W_{s+\mathcal{T}s}) &= \frac{1}{1-\alpha} \log_2 \frac{1}{2^\alpha \|s\|_2^{2\alpha}} \iint \left[W_s^\alpha(t,f) + X_{s,\mathcal{T}s}^\alpha(t,f) + W_{\mathcal{T}s}^\alpha(t,f)\right] dt\, df \\ &= \frac{1}{1-\alpha} \log_2 \frac{1}{2^\alpha \|s\|_2^{2\alpha}} \left[\iint W_s^\alpha(t,f)\, dt\, df \;+\; \iint W_{\mathcal{T}s}^\alpha(t,f)\, dt\, df\right] \\ &= H_\alpha(W_s) \;+\; 1, \end{aligned}$$

provided

$$\iint X_{s,\mathcal{T}s}^\alpha(t,f)\, dt\, df \;=\; 0. \tag{22}$$

While this is obviously not the case for α even, the oscillatory structure of $X_{s,\mathcal{T}s}$ [4–6] cancels under integration with odd powers for Δt sufficiently large. We prove the following in Appendix B.1 as a special case of Theorem 2.

Proposition 1 *Fix odd $\alpha \geq 3$ and let $s \in L^2(\mathbb{R})$ be a signal of compact support such that W_s and $W_{s+\mathcal{T}s}$ obey* (12)*. Denoting the length of the supporting interval by σ, set $\Delta t > \frac{1}{2}(\alpha+1)\sigma$. Then* (22) *holds and thus $H_\alpha(W_{s+\mathcal{T}s}) = H_\alpha(W_s) + 1$.*

The linear growth of the separation condition $\Delta t > \frac{1}{2}(\alpha+1)\sigma$ recommends the first of the odd integers $\alpha \geq 3$, namely $\alpha = 3$, as the best order for information analysis with the Wigner distribution. Problems with (12) and numerical considerations (stability in the face of quantization errors) also justify small α values. Using the symplectic transformation properties of the Wigner distribution (see Appendix C.1 and [5, 6, 19]), Proposition 1 can be easily extended from signals of compact time support to signals whose Wigner distributions are supported on a strip of arbitrary orientation in the time-frequency plane.

We can extend these counting results to include most Cohen's class TFRs and finite energy signals. For noncompactly supported signals, the auto- and cross-components in the Cohen's class analog to (21) will always overlap to some degree, so we should expect only asymptotic expressions. Define the time-frequency displacement operator

$$(\mathcal{D}s)(t) \;:=\; e^{j2\pi t\,\Delta f}\, s(t-\Delta t) \tag{23}$$

that translates signals by the distance $|\mathcal{D}| := \sqrt{(\Delta t/t_0)^2 + (\Delta f/f_0)^2}$ in the time-frequency plane (with $t_0 := 1$ s and $f_0 := 1$ Hz). The following is the key result of this section [8,9].

Theorem 2 (Component counting) *Let $C_s(t,f)$ be either the Wigner distribution or a Cohen's class TFR defined as in* (3) *with $\Phi \in L^1(\mathbb{R}^2)$. Then, for any $s \in L^2(\mathbb{R})$ and odd $\alpha \geq 3$ such that C_s and $C_{s+\mathcal{D}s}$ obey* (12)*, we have*

$$\lim_{|\mathcal{D}|\to\infty} H_\alpha(C_{s+\mathcal{D}s}) \;=\; H_\alpha(C_s) \;+\; 1. \tag{24}$$

Theorem 2 implies also that the "information" in the cross-components of $C_{s+\mathcal{D}s}$ must decay to zero asymptotically.

Corollary 3 (Asymptotic cross-component invariance) *Let $C_s(t,f)$ be either the Wigner distribution or a Cohen's class TFR defined as in* (3) *with $\Phi \in L^1(\mathbb{R}^2)$. Then, for any $s \in L^2(\mathbb{R})$ and odd $\alpha \geq 3$ such that C_s and $C_{s+\mathcal{D}s}$ obey* (12)*, we have*

$$\lim_{|\mathcal{D}|\to\infty} \iint \Big[C^\alpha_{s+\mathcal{D}s}(t,f) \;-\; C^\alpha_s(t,f) \;-\; C^\alpha_{\mathcal{D}s}(t,f)\Big]\, dt\, df \;=\; 0. \tag{25}$$

Proposition 1 extends to n components and $\log_2 n$ bits of information gain, provided that the auto- and cross-components become sufficiently disjoint in the time-frequency plane. The counting property does *not* hold generally when the signal cross-components overlap with the auto-components or other cross-components, however.

A simple example of such a signal is $u(t) := s(t-\Delta t) + s(t) + s(t+\Delta t)$, where the cross-component $W_{s(t-\Delta t),s(t+\Delta t)}$ lies upon the auto-component $W_{s(t)}$. For s supported on $\left[-\frac{1}{2},\frac{1}{2}\right]$ and $\Delta t \geq \frac{1}{2}(1+\alpha)$, we compute

$$\iint W^\alpha_u(t,f)\, dt\, df \;=\; 2\iint W^\alpha_s(t,f)\, dt\, df \;+\; \iint W^\alpha_s(t,f)\left(1 + e^{j4\pi f\Delta t} + e^{-j4\pi f\Delta t}\right)^\alpha dt\, df. \tag{26}$$

The first term on the right side arises from the auto-components of $s(t \pm \Delta t)$; the second term arises from the (overlapping) auto-component of $s(t)$ and cross-component between $s(t-\Delta t)$ and $s(t+\Delta t)$ (see (100) below). Note that

$$\iint W^\alpha_s(t,f)\, e^{j4\pi r f\Delta t}\, dt\, df \;=\; 0 \tag{27}$$

for integer $r \neq 0$, which follows from the conditions on s and Δt. Now using arguments as in Appendix B.1) and expanding the second term on the right side of (26), we find that

$$\iint W_u^\alpha(t,f)\,dt\,df \;=\; (2+N)\iint W_s^\alpha(t,f)\,dt\,df. \tag{28}$$

Here N denotes the coefficient of y^α in $(1+y+y^2)^\alpha$. Thus, we have

$$H_\alpha(W_u) \;=\; H_\alpha(W_s) \;+\; \frac{1}{1-\alpha}\log_2\frac{N+2}{3^\alpha}. \tag{29}$$

For instance, $\alpha = 3$ yields $\frac{1}{2}\log_2 3$ rather than $\log_2 3$ for the second term on the right side of (29).

As a second example, consider $u(t) := s(t-2\Delta t) + s(t-\Delta t) + s(t+\Delta t) + s(t+2\Delta t)$, with the support of s and Δt as above. Here, the cross-component $W_{s(t-2\Delta t),s(t+2\Delta t)}$ overlaps the cross-component $W_{s(t-\Delta t),s(t+\Delta t)}$. As above, we obtain

$$H_\alpha(W_u) \;=\; H_\alpha(W_s) \;+\; \frac{1}{1-\alpha}\log_2\left(\frac{M+4}{4^\alpha}\right), \tag{30}$$

with M the coefficient of $y^{2\alpha}$ in $\left[(1+y)(1+y^3)\right]^\alpha$. For instance, $\alpha = 3$ yields $\frac{1}{2}\log_2\frac{32}{5}$ rather than $\log_2 4$ for the second term on the right side of (30).

These examples show that large inter-component spacing ($\Delta t \to \infty$) is not enough for correct component counting. Notice that the spacings between the signals in these examples are regular; with signals of fixed supports and random spacings that tend to infinity, chances are much better that the component counting property will hold true.

3.2 Amplitude and phase sensitivity

The results of the experiment illustrated in Figure 1 and analyzed in the previous section are very appealing, but are also incomplete, because we introduced no amplitude or phase differences between the two signal components.

First consider amplitude differences. Consider the signal $\sum_i k_i\, s_i$ consisting of weighted components s_i that are time-frequency shifted versions of some basic component s. If the s_i are pulled apart such that the asymptotic overlap of the auto- and cross-components of their TFR decays to zero, then an analysis similar to that of Section 3.1 yields [20]

$$H_\alpha\Big(C_{\sum_i k_i\, s_i}\Big) \;\longrightarrow\; H_\alpha(C_s) \;+\; H_\alpha^{\mathrm{R}}(p), \tag{31}$$

with p a vector with entries $p_i = \frac{|k_i|^2}{\sum_j |k_i|^2}$ and H_α^{R} the discrete Rényi entropy of (10). $H_\alpha^{\mathrm{R}}(p)$ is a continuous function of the k_i bounded by 0 and 1 and maximized with all k_i equal. Thus, amplitude discrepancies alter the asymptotic saturation levels of the Rényi entropy.

We will see that phase offsets induce strong oscillations between the saturation levels. To shed further light on this matter, we will derive an analytic expression for the 3rd-order entropy of the Wigner distribution of the sum of two Gaussian pulses

$$s \;=\; w\, g_\gamma(a,b) \;+\; z\, g_\gamma(c,d), \tag{32}$$

with $g_\gamma(\nu,\xi)$ the Gaussian

$$g_\gamma(\nu,\xi)(t) \;:=\; (2\gamma)^{1/4}\, e^{-\pi\gamma(t-\nu)^2+j2\pi\xi t}, \qquad t\in\mathbb{R} \tag{33}$$

and $\gamma>0$, $a,b,c,d,\xi\in\mathbb{R}$, $z,w\in\mathbf{C}\setminus\{0\}$. We consider

$$H_3(W_s) \;=\; -\frac{1}{2}\log_2 Q_\gamma(a,b,w;c,d,z) \tag{34}$$

with

$$Q_\gamma(a,b,w;c,d,z) \;:=\; \frac{\iint W_s^3(t,f)\,dt\,df}{\left(\iint W_s(t,f)\,dt\,df\right)^3}\,. \tag{35}$$

Using symplectic transformations, we show in Appendix C.1 that the complete $Q_\gamma(a,b,w;c,d,z)$ can be computed from the values of

$$Q(x,y,v) \;:=\; Q_1\Big(a,0,1;-a,0,re^{j\varphi}\Big)\,, \tag{36}$$

with

$$x:=e^{-\frac{2}{3}\pi a^2}\in(0,1], \qquad y:=\cos\varphi\in[-1,1], \qquad v:=\cosh(\log r)\geq 1 \tag{37}$$

related to the time displacement, phase change, and amplitude disparity between the two components, respectively. In Appendix C.2 we derive the following.

Proposition 4 *For the signal* (32) *with* $(a,b,\omega;c,d,z)$ *replaced by* $(a,0,1;-a,0,re^{j\varphi})$ *and with* $H_3(W_s)$ *and* $Q(x,y,v)$ *defined as in* (34)–(37), *we have*

$$H_3(W_s) \;=\; -\frac{1}{2}\,\log_2\!\left[\frac{4}{3} \;-\; (1-x^2)^3\,(1+x^2)\,(v-xy)\,(v+x^3y)^{-3}\right]. \tag{38}$$

This expression is very convenient for studying the effects of component time, phase, and amplitude differences on the 3rd-order Wigner entropy. For example, in Appendix C.3 we find the bounds

$$\frac{1}{2}\log_2 3-1 \;\leq\; H_3(W_s) \;\leq\; \frac{1}{2}\log_2 27-1. \tag{39}$$

From (38), we can compute the effect of component amplitude disparity v on $H_3(W_s)$ by fixing y and varying the separation distance via x. For the asymptotic saturation level, we have

$$\lim_{a\to\infty} H_3(W_s) \;=\; -\frac{1}{2}\,\log_2\!\left(\frac{4}{3}-\frac{1}{v^2}\right), \qquad v\geq 1. \tag{40}$$

The obvious conclusion that equal amplitudes ($v=1$) maximize the complexity of signals composed of multiple identical components appears quite reasonable, for smaller components are dominated by larger ones and therefore carry less information.

In the region between saturation levels (where the TFR components overlap and the assumptions of Section 3.1 fail to hold), the relative phase between components controls the value of the Rényi

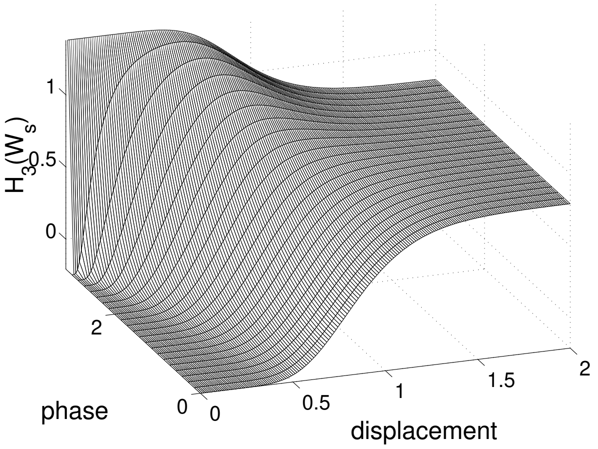


Figure 3: *The 3rd-order Rényi entropy $H_3(W_s)$ of the Wigner distribution of the two-component Gaussian signal (32) plotted versus displacement parameter a and phase φ (in rads). We use (38) with $v = 1$. Comparison with Figure 1 (which coincides with the $\varphi = 0$ slice of the surface) illustrates the sensitivity of $H_3(W_s)$ to relative phase. (The asymptotic levels here are the same as those in Figure 1; the (overestimation) peak value is $\frac{1}{2}\log_2 27 - 1 \approx 1.377$ bits.)*

entropy. Figure 3 extends the experiment of Figure 1 by plotting the $H_3(W_s)$ surface as a function of both inter-component displacement a and phase φ. It is apparent from the curves that while phase changes do not affect the saturation levels of the information measure, they allow many possible trajectories between the two levels, including even trajectories where an "overestimation" (noted numerically in [1] and confirmed in (39)) of information content occurs. Furthermore, if the phase of each component is fixed relative to the center of its envelope (so that the components do not change shape as they are shifted about), then the corresponding $H_3(W_s)$ versus a curve will be a slice of the $H_3(W_s)$ surface along an oblique trajectory in the (a, φ) plane. Curves of this form can be multimodal (see Figure 2 in [8, 9]).

The phase sensitivity of the $H_3(W_s)$ measure for closely spaced components is quite reasonable, given the sensitivity of the signals themselves to relative phase. For example, Figure 4 shows the composite signals and their respective Wigner distributions for a fixed offset a and relative phases $\varphi = \frac{\pi}{2}$ and $\varphi = \frac{3\pi}{2}$. The difference in appearance is striking; clearly the components in the signal in (b) are more separated than those in (a). Accordingly, the $H_3(W_s)$ entropies for the two signals differ widely: from 2.741 to 3.857 bits, respectively.

Since the interference pattern generated by cross-components encodes inter-component phase information, signals with low information content ("almost mono-component signals") must exhibit mainly *constructive interference* in the sense of [5, 6]. Relative phase fades from importance after all components become disjoint.

3.3 Effects of smoothing

TFRs based on lowpass kernels lead to more robust Rényi information estimates, since smoothing suppresses the Wigner cross-components that carry the inter-component phase information.

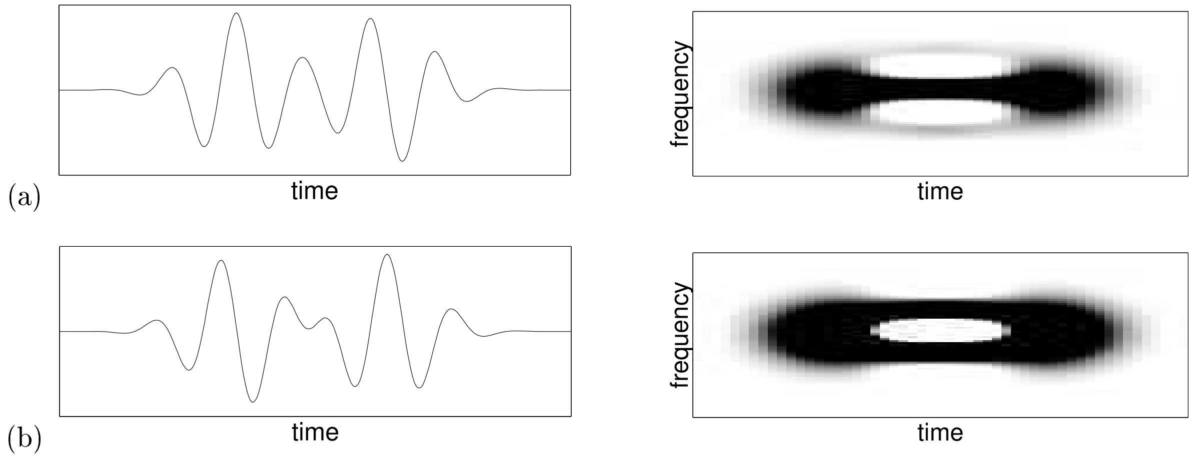


Figure 4: *Signals and positive parts of the Wigner distributions from the experiment illustrated in Figure 3 corresponding to a single fixed separation a and two different phases φ. (a) Signal (left) and Wigner distribution (right) for $\varphi = \frac{\pi}{2}$; $H_3(W_s) = 2.741$ bits. (b) Signal and Wigner distribution for $\varphi = \frac{3\pi}{2}$; $H_3(W_s) = 3.857$ bits.*

In Appendix C.4, we calculate $H_3(C_s)$ for the two-component Gaussian signal (s from (32) with $w = z = \gamma = 1$), with C_s the Wigner distribution smoothed by the Gaussian kernel

$$\Phi_\rho(t,f) \;=\; 2\,\rho^{-2}\,e^{-2\pi\rho^{-2}(t^2+f^2)}. \tag{41}$$

The choice $\rho = 1$ results in the matched-window spectrogram[8] as in Figure 5, and we obtain from (174)

$$H_3(C_s) \;=\; -\frac{1}{2}\,\log_2\left[\frac{1}{3} - \frac{(1-x^2)^2}{4(1+x^3y)^3}\,(2x^2+1+(x^2+2)x^3y)\right], \tag{42}$$

with x and y as in (37). For general $\rho \in (0,1)$, the upper saturation level ($x = 0$) is given by

$$\lim_{x\to 0} H_3(C_s) \;=\; \frac{1}{2}\log_2 3 \;+\; \log_2(1+\rho^2). \tag{43}$$

To illustrate, we repeat in Figure 5 the experiment of Figure 3 using a matched-window spectrogram TFR rather than the Wigner distribution. While the spectrogram information estimate remains somewhat phase sensitive, it climbs more swiftly to the saturation level and with a reduced overshoot compared to the Wigner distribution estimate. In general, the ascent to saturation accelerates with increasing order α once the cross-components are smoothed to the same peak level as the auto-components. (The opposite holds for the Wigner distribution, because Wigner cross-components can tower over Wigner auto-components by up to a factor of two.)

The price paid for the more robust information estimates derived from smoothed TFRs is a signal-dependent bias of entropy levels compared to those derived from the Wigner distribution, with the amount of bias increasing with the amount of smoothing. This bias is difficult to quantify, since the convolution in (3) and the power and integral in (6) do not permute in any simple fashion. In the special case of the matched-window spectrogram applied to a sum of Gaussian

[8] A spectrogram computed using the time-reversed signal as the window.

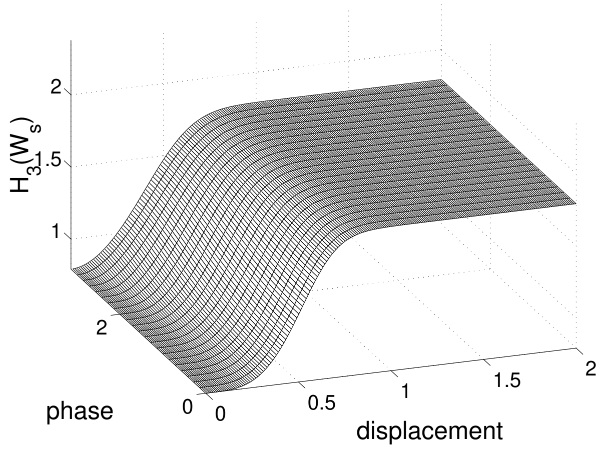


Figure 5: *The 3rd-order Rényi entropy $H_3(C_s)$ of the matched-window spectrogram plotted versus displacement parameter Δt and phase φ (in rads) for the same signal utilized in Figure 3 (see (42)). The reduced sensitivity of the spectrogram to relative phase results in swifter saturation with reduced overshoot. Furthermore, for small φ and Δt, $H_3(C_s)$ actually responds sooner than $H_3(W_s)$ to small increases in Δt. Note also the 1-bit bias in the asymptotic levels of $H_3(C_s)$ versus those of $H_3(W_s)$. (The asymptotic levels here are $\frac{1}{2}\log_2 3 \approx 0.792$ and $\frac{1}{2}\log_2 3 + 1 \approx 1.792$.)*

signal components, a direct computation finds a one-bit bias in asymptotic information compared to that estimated using the Wigner distribution (compare Figure 1 with Figure 5). Despite the introduction of systematic bias, smoothing is essential when measuring entropies for complicated multicomponent signals with overlapping auto- and cross-components.

Although it may be intuitively clear that smoothing attenuates the Wigner distribution's negative values, it is by no means an easy matter to get pertinent results on the existence of $H_\alpha(C_s)$ for odd α. Even for the case of Hermite functions h_n (as considered in Appendix A) and Gaussian smoothing (see (41)) this is a hard problem. It can be shown that the well-definedness of $H_3(W_{h_n} * \Phi_\rho)$ for all $n = 0, 1, \ldots$ and all $0 \leq \rho \leq 1$ requires a Debbi-Gillis type result [21] in which the $\frac{3}{2}$ in the exponential of (73) is replaced by $p = \frac{3}{2}(1 - \rho^2)$. We can summarize our (partial) results on Hermite signals as follows:

$\underline{\rho \geq 1}$: In this case, $(W_{h_n} * \Phi_\rho)(t, f) > 0$ [22] and existence of H_3 is not an issue.

$\underline{\frac{1}{3}\sqrt{3} < \rho < 1}$: It was kindly observed to us by Prof. R. Askey that a Debbi-Gillis result is easily established for $0 < p < 1$, which implies that $H_3(W_{h_n} * \Phi_\rho)$ is well-defined for this range of smoothing.

$\underline{0 < \rho \leq \frac{1}{3}\sqrt{3}}$: We have been unable to prove results for the case $1 \leq p < \frac{3}{2}$.

$\underline{\rho = 0}$: This is the Debbi-Gillis result [21] that shows that $H_3(W_{h_n})$ is well-defined.

3.4 Lower bound on signal information content

Simple to derive from Lieb's inequality [23] (see Appendix D.1), a lower bound on the Rényi entropy corresponds to the "peakiest" Cohen's class TFR.

Theorem 5 (Lower bound on information content for Cohen's class) *For any Cohen's class TFR $C_s(t,f)$ with $\Phi \in L^1(\mathbb{R}^2)$ and $\iint \Phi(t,f)\,dt\,df = 1$, any $\alpha \geq 1$, and any $s \in L^2(\mathbb{R})$*

$$H_\alpha(C_s) \;\geq\; \frac{\log_2 \alpha}{\alpha - 1} \;-\; 1 \;-\; \frac{\alpha}{\alpha - 1} \log_2 \|\Phi\|_1. \tag{44}$$

For the Wigner distribution,

$$H_\alpha(W_s) \;\geq\; \frac{\log_2 \alpha}{\alpha - 1} \;-\; 1, \tag{45}$$

with equality if and only if s is a Gaussian. For the spectrogram C_s^{Gspect} with Gaussian window g,

$$H_\alpha(C_s^{\mathrm{Gspect}}) \;\geq\; \frac{\log_2 \alpha}{\alpha - 1}, \tag{46}$$

with equality if and only if s is a Gaussian of the same form as g.[9]

Compare the lower saturation level in Figure 1 with the theoretical bound $H_3(W_g) = \frac{1}{2}\log_2(\frac{3}{4}) \approx -0.208$ for the Gaussian. Of course, this value also coincides with the lower bound (39).

Theorem 5 can be interpreted as an alternative version of the well-known time-frequency uncertainty principle [4–6,24] that takes the entire time-frequency plane into account rather than just the marginal distributions $|s(t)|^2$ and $|S(f)|^2$. It is also an alternative to the inequality of Hirschman that relates the classical principle to the Shannon entropy of the marginals, as [25]

$$H_1\left(|s|^2\right) \;+\; H_1\left(|S|^2\right) \;\geq\; \log_2 e \;-\; 1 \tag{47}$$

with equality if and only if s is Gaussian. Note that Theorem 5 marks the third breakdown of the analogy between probability density functions and TFRs (the first two being nonpositivity and nonuniqueness), since the Dirac delta function probability density minimizes the Rényi entropies of all orders.

3.5 Information invariant signal transformations and the affine class

An *information invariant* signal transformation $\mathcal{M}$ leaves the Rényi entropy measure unchanged, with $H_\alpha(C_{\mathcal{M}s}) = H_\alpha(C_s)$ [1]. Distributions information invariant to such a transformation $\mathcal{M}$, provided it displaces the center of gravity of the signal in time-frequency, admit a useful generalization of Theorem 2 to $\lim_{|\mathcal{M}|\to\infty} H_\alpha(C_{s+\mathcal{M}s}) = H_\alpha(C_s) + 1$.

The transformations leaving the Rényi entropy invariant correspond to those that do not change the value of the integral in (13). For Cohen's class TFRs, the invariance properties of three nested kernel classes are simple to quantify. All fixed-kernel TFRs are information invariant to time and frequency shifts. Product-kernel TFRs, having kernels of the form $\Phi(t,f) = \kappa(tf)$ with κ a 1-d function, are in addition invariant to scale changes of the form $s(t) \mapsto |k|^{-1/2}\, s(t/k)$. The Wigner distribution is the lone fixed-kernel TFR fully information invariant to time and frequency shifts, scale changes, and the modulation and convolution by linear chirp functions that realize shears in the time-frequency plane (the symplectic transformations of (131)–(132)). It is not coincidental

[9] In this case, C_s^{Gspect} is the matched-window spectrogram of a Gaussian.

that these same five operations leave invariant the form of the (minimum information) Gaussian signal [19].

The *affine class* provides additional TFRs information invariant to time shifts and scale changes [5, 6, 14, 26]. Affine class TFRs are obtained from the affine smoothing

$$\Omega_s(t,f) \;:=\; \iint W_s(u,v)\,\Pi\Big(f(t-u),\,v/f\Big)\,du\,dv \;=:\; (W_s \,@\, \Pi)(t,f) \tag{48}$$

of the Wigner distribution of the signal with a kernel function Π.[10] Given proper normalization of the kernel, we have $\iint \Omega_s(t,f)\,dt\,df \;=\; \|s\|_2^2$. Hence, the Rényi entropy of an affine class TFR can be defined exactly as in (13) (of course with the requirement (12)). The resulting time-scale information measure $H_\alpha(\Omega_s)$ shares all of the properties discussed above in the context of Cohen's class (counting, cross-component invariance, amplitude and phase sensitivity, bounds, etc.), except with time and frequency shifts replaced by time shifts and scale changes. In particular, we have the following.

Theorem 6 (Lower bound on information content for the affine class) *Let $\Omega_s(t,f)$ be an affine class TFR with kernel such that $\Pi \in L^1(\mathbb{R}^2)$ and $\frac{1}{f}\Pi(t,f) \in L^1(\mathbb{R}^2)$. Then for all integers $\alpha \geq 1$ and for all $s \in L^2(\mathbb{R})$*

$$H_\alpha(\Omega_s) \;\geq\; \frac{\log_2 \alpha}{\alpha - 1} \;-\; 1 \;-\; \frac{\alpha}{\alpha-1}\log_2\left\|\frac{1}{f}\,\Pi(t,f)\right\|_1. \tag{49}$$

For the proof, see Appendix D.2. The condition $\frac{1}{f}\Pi(t,f) \in L^1(\mathbb{R}^2)$ implies $\Pi(t,0) = 0\ \forall\, t$ for continuous kernels and ensures that the affine smoothing (48) is defined. Since the kernel generating the scalogram (the squared magnitude of the continuous wavelet transform) corresponds to the Wigner distribution W_ψ of the wavelet function ψ, this condition also generalizes the now classical "wavelet admissibility condition" [5, 6]; in particular, we have

$$\iint \left|\frac{1}{f}\,\Pi(t,f)\right| dt\,df \;=\; \iint \left|\frac{1}{f}\,W_\psi(t,f)\right| dt\,df \;\geq\; \iint W_\psi(t,f)\,\frac{dt\,df}{|f|} \;=\; \int |\Psi(f)|^2\,\frac{df}{|f|}. \tag{50}$$

For information invariances different from time and frequency shifts, scale changes, and chirp modulations and convolutions, we must look beyond Cohen's class and the affine class. Fortunately, all the above results extend easily to the recently developed unitarily equivalent Cohen's and affine classes [15–17]. The TFRs in these new classes are information invariant to generalized time-frequency shifts and time-scale changes.

4 Selected Applications

The foregoing properties of the Rényi entropies (when it is defined) make these new information and complexity measures particularly appropriate for time-frequency analysis. In this section, we briefly discuss two areas of past and potential application.

[10] In order to emphasize the similarity of (48) to (3), we have reparameterized the original time-scale formulation of (48) from [14] in terms of time-frequency coordinates by setting scale $a = f_0/f$, with $f_0 = 1$ Hz.

4.1 Information-based performance measures

The Rényi entropies make excellent measures of the information extraction performance of TFRs. By the analogy to probability density functions, minimizing the complexity or information in a particular TFR is equivalent to maximizing its concentration, peakiness, and, therefore, resolution [27]. Optimization of a TFR (through its kernel) with respect to an information measure yields a high performance "information optimal" TFR that changes its form to best match the signal at hand [13, 28].

Many of the optimal-kernel TFRs in the literature have been based either implicitly or explicitly on information measures. As noted by Williams and Sang [13, 28], the performance index common to the 1/0 [29], radially Gaussian [30], and adaptive optimal kernel [31] optimization formulations can be rewritten using Parseval's theorem as

$$\iint |(W_s * \Phi)(t,f)|^2 \, dt \, df \;=\; 2^{-H_2(W_s * \Phi)}. \tag{51}$$

Since the second-order Rényi entropy squares the TFR, it remains sensitive to cross-components and hence can be considered as a measure of their information content [1]. Thus, maximizing (51) over a class of lowpass smoothing kernels Φ simultaneously minimizes the information in the cross-components of the optimal-kernel TFR. Maximizing the $\|C_s\|_4^4 \,/\, \|C_s\|_2^4$ concentration ratio of [32, 33] can also be viewed in information theoretic terms, since this is equivalent to minimizing the differential entropy $3H_4(C_s) - 2H_2(C_s)$.

Differential performance measures formed with odd and even order entropies also prove interesting [28]. For example, the differential measure

$$H_3(C_s) \;-\; \beta H_2(C_s), \qquad 0 \le \beta \le 1 \tag{52}$$

exploits the fact that odd and even order entropies decouple to some degree the information content in the auto- and cross-components in a TFR. Minimizing this measure balances *(i)* maximizing the information in the auto-components (by keeping them peaky through less smoothing) with *(ii)* minimizing the information in the cross-components (by flattening them through more smoothing). For positive TFRs, the special choice $\beta = \frac{3}{4}$, minimizing (52) is equivalent to maximizing the concentration ratio $\|C_s\|_3^6 \,/\, \|C_s\|_2^6$, making it an interesting alternative to the $\|C_s\|_4^4 \,/\, \|C_s\|_2^4$ measure used in [32, 33].

Figure 6 explores the effect of time-frequency smoothing on (52) as a function of the parameter β. Forming a signal s from two well-separated Gaussian pulses, we smooth the Wigner distribution of s with a Gaussian kernel (41) of increasing volume ρ to generate a series of smoothed TFRs C_s. In the Figure, the smoothing parameter ρ corresponds to the normalized degree of smoothing, with $\rho = 0$ leaving the Wigner distribution untouched and $\rho = 1$ generating the matched-window spectrogram. We plot $H_3(C_s) - \beta H_2(C_s)$ versus the smoothing parameter for several values of β ranging between 0 and 1. From the Figure, it is clear that β controls the tradeoff between measuring auto-component concentration and measuring cross-component suppression: Small β favors auto-component concentration, and (52) is minimized by very little smoothing — no smoothing at all

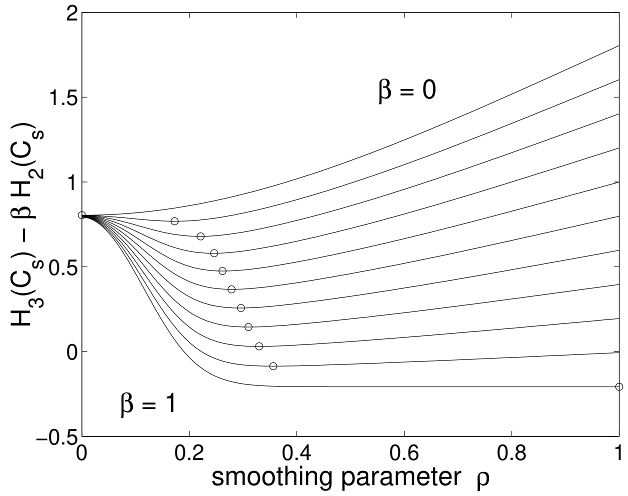


Figure 6: *The effect of time-frequency smoothing on (52) as a function of the parameter β. For a signal consisting of two well-separated Gaussian pulses ($a > 2$ in (37)), we form the TFR $C_s = W_s * \Phi_\rho$, with Φ_ρ from (41). The parameter ρ controls the degree of time-frequency smoothing: $\rho = 0$ generates the Wigner distribution; $\rho = 1$ generates the matched-window spectrogram. We plot $H_3(C_s) - \beta H_2(C_s)$ versus the smoothing parameter ρ for the eleven values $\beta = 0, 0.1, \ldots, 1$. (The upper curve corresponds to $\beta = 0$, while the lower curve corresponds to $\beta = 1$.) The minimum point of each curve, marked by a circle, corresponds to an "information optimal TFR." Data obtained by numerical simulation; results from the analytical approximation (53) correspond closely.*

for the extreme $\beta = 0$ case. On the other hand, large β favors cross-component suppression, and (52) is minimized only after considerable smoothing.

In Appendix C.4, we compute $H_3(C_s)$ and $H_2(C_s)$ for this case analytically. Some additional approximations valid for large component separation a lead us to the simple formula

$$H_3(C_s) - \beta H_2(C_2) \;\approx\; -0.208 \;+\; (1-\beta)\log_2[2(1+\rho^2)] \;+\; \beta \log_2\left[1 + e^{-4\pi a^2 \rho^2 (1+\rho^2)^{-1}}\right]. \quad (53)$$

The second term on the right side increases gently in $\rho \in [0,1]$, while the third term decreases sharply in ρ^2 and is flat for larger values of ρ^2 (but still decreasing). Among other things, it follows that the minimum of (53) decreases when β increases and that the minimum point $\rho_{\min}(\beta)$ shifts very slowly towards 1 as β approaches 1. For β extremely close to 1, we have that (53) decreases in ρ, whence $\rho_{\min}(\beta) = 1$ for these β. More complicated signals will exhibit local minima.

4.2 Rényi dimensions

Based on the counting property of the Rényi entropy (Section 3.1), we can define a *Rényi dimension* $D_\alpha(C_s)$ of a signal s in terms of its TFR C_s and a basic building block function b [8,9]

$$D_\alpha(C_s) \;:=\; 2^{H_\alpha(C_s) \,-\, H_\alpha(C_b)}. \quad (54)$$

This dimension attempts to indicate — in terms of a highly overcomplete set of building blocks obtained from b by all possible translations and modulations — the number of blocks required to "cover" the TFR of s. For the Wigner TFR, a Gaussian is the natural choice for the building block function, since it has minimum intrinsic information (in the sense of Theorem 5) and leads to an always positive dimension. A similar dimension can be defined for affine class TFRs.

By permitting redundant time-frequency building blocks, the Rényi time-frequency dimension generalizes the concepts of the number of "independent degrees of freedom" and number of "independent coherent states" that have proved useful in signal analysis and quantum physics [34, p. 23]. Desirable invariance properties result from this redundancy: Cohen's class Rényi dimension estimates remain invariant under time and frequency shifts in the signal, while affine class estimates remain invariant under time and scale changes. Alternative dimensions that measure signal complexity with respect to an orthonormal basis of (wavelet or Gabor) functions (see [35], for example) cannot share these invariances without carrying out an optimization over all "nice" bases [2].

For the simplest signals, composed of disjoint, equal-amplitude copies of one basic function, the Rényi dimension simply counts the number of components. As the relative amplitudes of these components change, however, the dimension estimate will also change, as some components begin to dominate others.

5 Conclusions

Taking off where Williams, Brown, and Hero left off in [1], this paper has studied a new class of signal analysis tools — the Rényi entropies. Users must proceed with caution, for as we have shown, the higher-order entropies are not defined for large classes of signals. Counterexamples are much harder to find for the 3rd-order entropy, however, especially for suitably smoothed TFRs (we have encountered none). This finding supports the numerous numerical studies [1,8,9,12] that have indicated these measures' general utility.

When well-defined, the accounting, and cross-component and transformation invariance properties of the Rényi entropies make them natural for estimating the complexity of deterministic signals through TFRs. Simple to apply, these measures also provide new insights into the structure of the time-frequency plane. For instance, a lower bound on the entropy of the Wigner distribution yields a new time-frequency uncertainty principle (Theorems 5 and 6) based on the entire time-frequency plane as a whole rather than on the time and frequency domains separately.

The explorations of Section 4 into TFR performance measures and Rényi dimensions merely scratch the surface of potential applications of the Rényi entropies in time-frequency analysis. Worthy of pursuit seems the extension of our results to TFRs outside the quadratic Cohen's and affine classes. The positive TFRs of the Cohen-Posch class [4], for example, would allow the unrestricted use of the Shannon entropy. Moreover, an axiomatic derivation of the "ideal" time-frequency complexity measure along the lines of Rényi's work in probability theory [11] could yield other entropies meriting investigation.

In information theory, entropies form the basis for distance and divergence measures between probability densities. In time-frequency analysis, analogous measures between TFRs would find immediate application in detection and classification problems. Unfortunately, the Rényi entropy complicates the formation of distances, because it is neither a concave nor a convex function for

$\alpha \neq 1$. Although the bulk of the work lies ahead, some progress has been made in this direction. Considering only positive TFRs (smoothed spectrograms in Cohen's class), we defined in [20] a distance measure between two TFRs C_1 and C_2 that is reminiscent of the Jensen divergence

$$J_\alpha(C_1, C_2) \;:=\; H_\alpha\left(\sqrt{C_1\, C_2}\right) \;-\; \frac{H_\alpha(C_1) \;+\; H_\alpha(C_2)}{2}. \tag{55}$$

(Here, $\sqrt{C_1\, C_2}(t,f) := \sqrt{C_1(t,f)\, C_2(t,f)}$.) Currently, we are evaluating the potential of this measure for problems in nonparametric and blind transient detection.

A Signals with Undefined WD-based Rényi Entropy

In this Appendix, we display for any odd integer $\alpha \geq 3$ a signal s such that

$$\iint W_s^\alpha(t,f)\,dt\,df \tag{56}$$

is negative. Hence for such an s, the α-order Rényi entropy $H_\alpha(W_s)$ is not defined. All of our example signals are variations on a theme: peaked, odd functions that create a large negative spike in the Wigner distribution.

First, some background on Hermite functions. The nth-order Hermite function [19, Ch. 1, Sec. 7]

$$h_n(t) \;:=\; (-1)^n\, 2^{1/4}\,(n!)^{-1/2}\,(4\pi)^{-n/2}\, e^{\pi t^2} \left(\frac{d}{dt}\right)^n e^{-2\pi t^2}. \tag{57}$$

has a Wigner distribution that can be written in terms of a Laguerre polynomial. That is [19, p. 66],

$$W_n(t,f) \;:=\; W_{h_n}(t,f) \;=\; 2(-1)^n\, e^{-2\pi r^2}\, L_n(4\pi r^2), \tag{58}$$

with L_n the nth-order Laguerre polynomial

$$L_n(x) \;:=\; \sum_{j=0}^{n} \binom{n}{j} \frac{(-x)^j}{j!} \tag{59}$$

and $r^2 := (t/t_0)^2 + (f/f_0)^2$ with $t_0 := 1$ s and $f_0 := 1$ Hz (we assume this normalization of t and f for the remainder of the paper). As is well-known, the Hermite functions have Wigner distributions that are *(i)* strongly peaked at the origin, with a negative sign when the order n is odd, *(ii)* small but non-negligible away from the origin but inside a circle around the origin of radius somewhat larger than $\left(\left(n+\frac{1}{2}\right)/\pi\right)^{1/2}$, and *(iii)* negligibly small outside that circle. Therefore, the odd-order Hermite functions are natural candidates for yielding negative values in (56).

A.1 Examples for large odd α

Throughout this Appendix, let s be a smooth, rapidly decaying, *odd* signal of unit energy. Then $W_s(t,f)$ is smooth and rapidly decaying as $t^2+f^2 \to \infty$ and

$$|W_s(t,f)| \;<\; 2 \;=\; -\,W_s(0,0), \qquad (t,f) \neq (0,0), \tag{60}$$

as one easily sees from the Cauchy-Schwarz inequality (the fact that $s \in L^2(\mathbb{R})$ causes the inequality to be strict). It thus follows that the asymptotic behavior of (56) as $\alpha \to \infty$, α integer, is determined by the behavior of $W_s(t,f)$ at $(t,f)=(0,0)$. Since

$$\frac{\partial W_s}{\partial t} \;=\; \frac{\partial W_s}{\partial f} \;=\; 0 \;=\; \frac{\partial^2 W_s}{\partial t \partial f} \;=\; \frac{\partial^2 W_s}{\partial f \partial t}, \tag{61}$$

$$\frac{\partial^2 W_s}{\partial t^2} \;=\; 8\|s'\|^2, \qquad \frac{\partial^2 W_s}{\partial f^2} \;=\; 32\pi^2 \|ts(t)\|^2 \tag{62}$$

at $(t,f)=(0,0)$, we have

$$\log\left(-\frac{1}{2}\,W_s(t,f)\right) \;=\; -\,2t^2\|s'\|^2 - 8\pi^2 f^2\|ts(t)\|^2 + o(t^2+f^2) \tag{63}$$

as $t^2+f^2\to 0$. Therefore we have, explicitly,

$$\begin{aligned}\iint W_s^\alpha(t,f)\,dt\,df \;&=\; (-2)^\alpha \iint \exp\left\{\alpha\log\left(-\frac{1}{2}W_s(t,f)\right)\right\} dt\,df \\ &=\; \frac{(-2)^\alpha}{4\alpha\|s'\|\|ts(t)\|}\;(1+o(1))\end{aligned} \tag{64}$$

as $\alpha\to\infty$, α integer. Hence (56) is negative for *any* smooth, rapidly decaying, odd signal s and large odd integer α. □

A.2 Example for $\alpha = 5, 7, \ldots$

Let s be the 1st-order Hermite function

$$s(t) \;=\; h_1(t) \;=\; 2^{5/4}\,\pi^{1/2}\,t\,e^{-\pi t^2}, \qquad t\in\mathbb{R} \tag{65}$$

with Wigner distribution

$$W_s(t,f) \;=\; -\,2\,e^{-2\pi(t^2+f^2)}\left(1-4\pi(t^2+f^2)\right). \tag{66}$$

Using polar coordinates, we have for integer $\alpha\geq 1$

$$\iint W_s^\alpha(t,f)\,dt\,df \;=\; \frac{2^{\alpha-1}}{\alpha}(-1)^\alpha\int_0^\infty e^{-x}\left(1-\frac{2x}{\alpha}\right)^\alpha dx. \tag{67}$$

For odd $\alpha\geq 1$ we have

$$\int_0^\infty e^{-x}\left(1-\frac{2x}{\alpha}\right)^\alpha dx \;=\; \int_0^{\alpha/2} e^{-x}\left(1-\frac{2x}{\alpha}\right)^\alpha dx \;-\; \int_{\alpha/2}^\infty e^{-x}\left(\frac{2x}{\alpha}-1\right)^\alpha dx. \tag{68}$$

The first integral on the right-hand side of (68) increases in $\alpha\geq 1$, since the (non-negative) integrand and integration range increase in $\alpha\geq 1$. The second integral can be evaluated as

$$\int_{\alpha/2}^\infty e^{-x}\left(\frac{2x}{\alpha}-1\right)^\alpha dx \;=\; \left(\frac{4}{e}\right)^{\alpha/2}\frac{\alpha!}{\alpha^\alpha}, \tag{69}$$

and this decreases in $\alpha\geq 1$. Hence the left-hand side of (68) increases in $\alpha\geq 1$, α odd integer. Since

$$\int_0^\infty e^{-x}(1-2x)dx \;=\; -\,1, \qquad \int_0^\infty e^{-x}\left(1-\frac{2x}{3}\right)^3 dx \;=\; -\frac{1}{9} \tag{70}$$

and

$$\int_0^\infty e^{-x}\left(1-\frac{2x}{5}\right)^5 dx \;=\; \frac{127}{625}, \tag{71}$$

we see that (67) is negative for all odd $\alpha\geq 5$. □

A.3 Example for $\alpha = 3$

We will have to spend considerable effort to find an s for which (56) is negative for $\alpha = 3$. Also, a small amount of Gaussian smoothing of W_s is generally enough to make (56) positive for our offending signal. This indicates that the examples for which (56) fails to be positive for $\alpha = 3$ are rather exceptional. Moreover, the results of Appendix C show that the 3rd-order Rényi entropy is well-defined for the sum of two Gaussians, irrespective of their mutual phases. See also the discussion on Hermite functions above.

We shall show that (56) is negative for $\alpha = 3$ and

$$s \;=\; a_3\, h_3 \;+\; a_9\, h_9 \tag{72}$$

with suitably chosen $a_3, a_9 \in \mathbb{R}$ and h_n the nth-order Hermite function.

Using polar coordinates, we obtain (see (58))

$$\iint W_n^3(t,f)\, dt\, df \;=\; 2(-1)^n \int_0^\infty e^{-\frac{3}{2}x}\, L_n^3(x)\, dx \;=:\; 2T_n. \tag{73}$$

It is a quite non-trivial result from Debbi and Gillis [21] that $T_n > 0$ for all $n = 0, 1, \ldots$ Hence we cannot produce negative values in (56) with a single Hermite function.

We now elaborate (56) for the s in (72) and $\alpha = 3$. We have

$$W_s \;=\; |a_3|^2 W_3 + |a_9|^2 W_9 \;+\; a_3\, a_9^* W_{39} \;+\; a_3^* a_9 W_{93}, \tag{74}$$

with W_{nm} the cross Wigner distribution[11] between h_n, h_m

$$W_{nm}(t,f) \;=\; \begin{cases} 2\,(-1)^m \left(\frac{m!}{n!}\right)^{1/2} e^{-2\pi|z|^2}\, (2\sqrt{\pi}z^*)^{n-m}\, L_m^{(n-m)}(4\pi|z|^2)\,, & n \geq m, \\[2ex] 2\,(-1)^n \left(\frac{n!}{m!}\right)^{1/2} e^{-2\pi|z|^2}\, (2\sqrt{\pi}z)^{m-n}\, L_n^{(m-n)}(4\pi|z|^2)\,, & n \leq m. \end{cases} \tag{75}$$

Here $z = t + if$, and $L_k^{(p)}$ are the Laguerre polynomials

$$L_k^{(p)}(x) \;=\; \sum_{j=0}^{k} \frac{(k+p)!}{(k-j)!(p+j)!}\, \frac{(-x)^j}{j!}. \tag{76}$$

Now expanding W_s^3 using (76), introducing polar coordinates $z = re^{i\theta}$ in $\iint W_s^3\, dt\, df$, and retaining only the triple products of (cross) Wigner distributions that are independent of θ (the others cancel upon integration), we obtain

$$\begin{aligned} \iint W_s^3(t,f)\, dt\, df \;=\; \iint \Big(& |a_3|^6 W_3^3 + 3|a_3|^4|a_9|^2 W_3^2 W_9 + 3|a_3|^2|a_9|^4 W_3 W_9^2 \\ & + |a_9|^6 W_9^3 + 6|a_3|^4|a_9|^2 W_3 |W_{39}|^2 \,+\, 6|a_3|^2|a_9|^4 W_9 |W_{39}|^2 \Big)\, dt\, df, \end{aligned} \tag{77}$$

[11] Formula (75) is due to Groenewold (see formula (5.16) in [36]), except that Groenewold has incorrectly a $(-1)^{\max(n,m)}$ instead of $(-1)^{\min(n,m)}$ and calls the $L_k^{(p)}$ Legendre polynomials. Formula (75) can also be found in [19, p. 66, (1.105)], except that there is a complex conjugate missing in the case $n \geq m$ (note that $W_{nm} = W_{mn}^*$; a similar error occurs in [19, p. 64, (1.104)]).

where we also have used that $W_{39}^* = W_{93}$. Inserting the explicit form (75) into the right-hand side of (77), we obtain

$$\iint W_s^3(t,f)\,dt\,df \;=\; -2\left(|a_3|^2 C_{333} + 3|a_3|^4|a_9|^2 C_{339} + 3|a_3|^2|a_9|^4 C_{993} + |a_9|^6 C_{999}\right) \\ -12\,\frac{3!}{9!}\left(|a_3|^4|a_9|^2 C_{333}^{(6)} + |a_3|^2|a_9|^4 C_{339}^{(6)}\right), \tag{78}$$

where we have set

$$C_{kkl}^{(p)} := \int_0^\infty e^{-\frac{3s}{2}} s^p \left(L_k^{(p)}(s)\right)^2 L_l(s)\,ds, \qquad C_{kkl} := C_{kkl}^{(0)}. \tag{79}$$

Thus we have

$$\frac{1}{|a_3|^6}\iint W_s^3(t,f)\,dt\,df \;=\; M_0 + M_1 x + M_2 x^2 + M_3 x^3, \qquad x := \frac{|a_9|^2}{|a_3|^2}, \tag{80}$$

where

$$M_0 := -2\,C_{333}, \qquad M_1 := -6\,C_{339} - 12\,\frac{3!}{9!}\,C_{333}^{(6)}, \\ M_2 := -6\,C_{993} - 12\,\frac{3!}{9!}\,C_{339}^{(6)}, \qquad M_3 := -2\,C_{999}. \tag{81}$$

The computation of the $C_{kkl}^{(p)}$ can be done according to

$$C_{kkl}^{(p)} \;=\; \left(\frac{(p+k)!}{k!}\right)^2 \left(\frac{2}{3}\right)^{p+1} \\ \times \sum_{j_1=0}^{k}\sum_{j_2=0}^{k}\sum_{j_3=0}^{l} \frac{(p+j_1+j_2+j_3)!}{(p+j_1)!\,(p+j_2)!\,j_3!} \binom{k}{j_1}\binom{k}{j_2}\binom{l}{j_3}\left(-\frac{2}{3}\right)^{j_1+j_2+j_3}. \tag{82}$$

For this we have used the explicit forms (59), (76) for the Laguerre polynomials and carried out the integration (see also [37, Sec. 2.a]). It follows that

$$M_0 = 0.107504, \qquad M_1 = 0.198739, \qquad M_2 = -0.364632, \qquad M_3 = 0.067079. \tag{83}$$

The right-hand side of (80) is extremal for

$$x \;=\; -\frac{M_2}{3M_3} \pm \left(\left(\frac{M_2}{3M_3}\right)^2 - \frac{M_1}{3M_3}\right)^{1/2}. \tag{84}$$

Taking the $+$ sign in (84) so that $x = 3.327073$, we obtain

$$M_0 + M_1 x + M_2 x^2 + M_3 x^3 \;=\; -\,0.797104. \tag{85}$$

This completes the construction of the example. □

Figure 2 illustrates the signal (72) and its Wigner distribution for $a_3 = 1$ and $a_9 = 1.824026$.

B Proofs on Component Counting

B.1 Proof of Proposition 1

Let $s \in L^2(\mathbb{R})$ be supported on the interval $[0, \sigma]$, and let $u \in L^2(\mathbb{R})$ be supported on $[\Delta t, \sigma + \Delta t]$. We will show that (22) holds:

$$\Gamma \; := \; \iint X_{s,u}^{\alpha}(t,f)\, dt\, df \; = \; 0 \tag{86}$$

when $\Delta t > \frac{1}{2}(\alpha+1)\sigma$. Here $X_{s,u}$ is the cross-component between s and u (see (19) and (20)). We shall initially assume that s and u are smooth, so that the manipulations below are justified.

We have by binomial expansion that

$$\Gamma \; = \; \sum_{k=0}^{\alpha} \binom{\alpha}{k} \Gamma_k, \tag{87}$$

with

$$\Gamma_k \; := \; \iint W_{s,u}^{k}(t,f)\, W_{u,s}^{\alpha-k}(t,f)\, dt\, df, \qquad k = 0, 1, \ldots, \alpha. \tag{88}$$

We will show that each Γ_k vanishes when $\Delta t > \frac{1}{2}(\alpha+1)\sigma$.

We first consider the case when $k \neq 0, \alpha$. We have by definition

$$\begin{aligned}\Gamma_k \; = \; & \iint \left(\int e^{-j2\pi f\tau} s\left(t+\frac{\tau}{2}\right) u^*\left(t-\frac{\tau}{2}\right) d\tau \right)^k \\ & \times \left(\int e^{-j2\pi f\tau} u\left(t+\frac{\tau}{2}\right) s^*\left(t-\frac{\tau}{2}\right) d\tau \right)^{\alpha-k} dt\, df.\end{aligned} \tag{89}$$

We write this as

$$\begin{aligned}\Gamma_k \; = \; & \iint \underbrace{\int \cdots \int}_{k} \underbrace{\int \cdots \int}_{\alpha-k} e^{-j2\pi f(\tau_1+\ldots+\tau_k+\tau_{k+1}+\ldots+\tau_\alpha)} \\ & \times s\left(t+\frac{\tau_1}{2}\right) u^*\left(t-\frac{\tau_1}{2}\right) \cdots s\left(t+\frac{\tau_k}{2}\right) u^*\left(t-\frac{\tau_k}{2}\right) d\tau_1 \cdots d\tau_k \\ & \times u\left(t+\frac{\tau_{k+1}}{2}\right) s^*\left(t-\frac{\tau_{k+1}}{2}\right) \cdots u\left(t+\frac{\tau_\alpha}{2}\right) s^*\left(t-\frac{\tau_\alpha}{2}\right) d\tau_{k+1} \cdots d\tau_\alpha\, dt\, df.\end{aligned} \tag{90}$$

Integrating over f, a Dirac $\delta(\tau_1 + \cdots + \tau_\alpha)$ term appears. Using this $\delta(\tau_1 + \cdots + \tau_\alpha)$ to cancel the integration over τ_α, we obtain

$$\begin{aligned}\Gamma_k \; = \; & \int \underbrace{\int \cdots \int}_{k} \underbrace{\int \cdots \int}_{\alpha-k} s\left(t+\frac{\tau_1}{2}\right) u^*\left(t-\frac{\tau_1}{2}\right) \cdots s\left(t+\frac{\tau_k}{2}\right) u^*\left(t-\frac{\tau_k}{2}\right) \\ & \times u\left(t+\frac{\tau_{k+1}}{2}\right) s^*\left(t-\frac{\tau_{k+1}}{2}\right) \cdots u\left(t+\frac{\tau_{\alpha-1}}{2}\right) s^*\left(t-\frac{\tau_{\alpha-1}}{2}\right) \\ & \times u\left(t-\frac{\tau_1+\cdots+\tau_{\alpha-1}}{2}\right) s^*\left(t+\frac{\tau_1+\cdots+\tau_{\alpha-1}}{2}\right) d\tau_1 \cdots d\tau_k\, d\tau_{k+1} \cdots d\tau_{\alpha-1}\, dt.\end{aligned} \tag{91}$$

Now suppose that we have a $t \in \mathbb{R}$, $\tau_1, \ldots, \tau_k, \tau_{k+1}, \ldots, \tau_{\alpha-1} \in \mathbb{R}$ such that the integrand in (91) is not zero. Since s is supported on $[0,\sigma]$, and u is supported on $[\Delta t, \sigma + \Delta t]$, we have

$$\begin{array}{rlrl}
t + \frac{\tau_1}{2} & \in [0,\sigma], & t - \Delta t - \frac{\tau_1}{2} & \in [0,\sigma] \\
\vdots & & \vdots & \\
t + \frac{\tau_k}{2} & \in [0,\sigma], & t - \Delta t - \frac{\tau_k}{2} & \in [0,\sigma] \\
t - \frac{\tau_{k+1}}{2} & \in [0,\sigma], & t - \Delta t + \frac{\tau_{k+1}}{2} & \in [0,\sigma] \\
\vdots & & \vdots & \\
t - \frac{\tau_{\alpha-1}}{2} & \in [0,\sigma], & t - \Delta t + \frac{\tau_{\alpha-1}}{2} & \in [0,\sigma] \\
t + \frac{\tau_1 + \cdots + \tau_{\alpha-1}}{2} & \in [0,\sigma], & t - \Delta t - \frac{\tau_1 + \cdots + \tau_{\alpha-1}}{2} & \in [0,\sigma].
\end{array} \tag{92}$$

Adding the first k items in (92) and subtracting the last $\alpha - k$ items, we obtain

$$(2k - \alpha)t \;\in\; [-(\alpha - k)\sigma,\, k\sigma], \tag{93}$$

$$(2k - \alpha)(t - \Delta t) \;\in\; [-(\alpha - k)\sigma,\, k\sigma]. \tag{94}$$

Subtracting (94) from (93) yields

$$(2k - \alpha)\,\Delta t \;\in\; [-l\sigma, l\sigma], \tag{95}$$

with $l = \max(\alpha - k, k) = \frac{1}{2}(\alpha + |2k - \alpha|)$. Since α is odd, we have $|2k - \alpha| \geq 1$. Thus, we finally obtain

$$\Delta t \;\leq\; \frac{1}{2}\left(\frac{\alpha}{|2k - \alpha|} + 1\right)\sigma \;\leq\; \frac{1}{2}(\alpha + 1)\,\sigma. \tag{96}$$

Hence, when $\Delta t > \frac{1}{2}(\alpha + 1)\,\sigma$, the integrand in (91) vanishes identically, so $\Gamma_k = 0$, as required.

Next, consider the case $k = \alpha$. Here the items with index $k+1, \cdots, \alpha - 1$ are absent from (92), but the above argument still yields that $\Delta t \leq \sigma \leq \frac{1}{2}(\alpha + 1)$ whenever the integrand in (91) is non-zero for some $t \in \mathbb{R}$, $\tau_1, \cdots, \tau_{\alpha-1} \in \mathbb{R}$. A similar result holds for the case $k = 1$.

We shall now remove the assumption that s and u are smooth. To this end we give the following Lemma, which will also be used in the proof of Theorem 2 below. We omit the (elementary) proof.

Lemma 7 *Let $s, u, v, w \in L^2(\mathbb{R})$, not necessarily compactly supported. Then for all $t, f \in \mathbb{R}$ we have*

$$|W_{s,u}(t,f) - W_{v,w}(t,f)| \;\leq\; 2\,\|u\|_2\,\|s - v\|_2 \;+\; 2\,\|v\|_2\,\|u - w\|_2. \tag{97}$$

Also, $W_{s,u}$ and $W_{v,w}$ are in $L^2(\mathbb{R}^2)$, and

$$\|W_{s,u} - W_{v,w}\|_2 \;\leq\; \|u\|_2\,\|s - v\|_2 \;+\; \|v\|_2\,\|u - w\|_2. \tag{98}$$

Finally, $W_{s,u} \in C_0(\mathbb{R}^2)$; that is, $W_{s,u}$ is continuous and bounded with $W_{s,u}(t,f) \longrightarrow 0$ as $t^2 + f^2 \to \infty$.

To complete the proof of Proposition 1, we take smooth v, supported on $[0,\sigma]$, and smooth w, supported on $[\Delta t, \sigma+\Delta t]$, such that $\|s-v\|_2$ and $\|u-w\|_2$ are small. Then (97) and (98) show that $W_{s,u}$ is approximated by $W_{v,w}$, both uniformly and in $L^2(\mathbb{R}^2)$ sense. Hence $W_{s,u}$ is approximated by $W_{v,w}$ in $L^\alpha(\mathbf{R}^2)$ sense, since $2 \leq \alpha \leq \infty$. Now the result follows easily from the fact that $\iint X^\alpha_{v,w}(t,f)\,dt\,df$ vanishes for smooth v and w when $\Delta t > \frac{1}{2}(\alpha+1)\,\sigma$. □

B.2 Proof of Theorem 2

For the proof we will need, in addition to Lemma 7, the Riemann-Lebesgue Lemma.

Lemma 8 (Riemann-Lebesgue) *For $g \in L^1(\mathbb{R}^n)$ and $n = 1, 2, \ldots$, we have*

$$\int g(\underline{x})\, e^{-j2\pi \underline{x}\cdot\underline{y}}\, d\underline{x} \longrightarrow 0, \qquad \|\underline{y}\| \longrightarrow \infty. \tag{99}$$

Furthermore, with $u := \mathcal{D}s$ (see (23)), we have the useful formula [5, p. 240],

$$X_{s,u}(t,f) = 2\,W_s\left(t - \frac{\Delta t}{2}, f - \frac{\Delta f}{2}\right) \cos\left(2\pi \left(t\Delta f - f\Delta t + \frac{1}{2}\Delta t \Delta f\right)\right). \tag{100}$$

Wigner distribution case: We assume odd $\alpha \geq 3$ and expand trinomially

$$W^\alpha_{s+u} = W^\alpha_s + W^\alpha_u + X^\alpha_{s,u} + \sum_{\alpha_i} c(\alpha_1,\alpha_2,\alpha_3)\, W^{\alpha_1}_s\, X^{\alpha_2}_{s,u}\, W^{\alpha_3}_u, \tag{101}$$

where in the latter series we have collected the terms in the expansion with $\alpha_1+\alpha_2+\alpha_3 = \alpha$ and at least two of the α_i positive (the $c(\alpha_1,\alpha_2,\alpha_3)$ in the series on the right-hand side are constants).

For the definition of $H_\alpha(W_s)$, we should also say what $\iint W_s(t,f)\,dt\,df$ means in the case that $W_s \in L^2(\mathbb{R}^2)$, $W_s \notin L^1(\mathbb{R}^2)$. Naturally, we define

$$\iint W_s(t,f)\,dt\,df = \|s\|_2^2 \tag{102}$$

for such cases. (When W_s does belong to $L^1(\mathbb{R}^2)$, the identity (102) also holds.) Then it follows that

$$\iint W_{s+u}(t,f)\,dt\,df = \|s+u\|_2^2 \longrightarrow 2\,\|s\|_2^2 = 2\iint W_s(t,f)\,dt\,df \tag{103}$$

when $|\mathcal{D}|^2 = (\Delta t)^2 + (\Delta f)^2 \to \infty$. Here we have used the last item in Lemma 7 and the fact that the inner product $\langle s,u\rangle = \exp^{-j\pi\Delta t\Delta f}\; W_{s,v}\left(\frac{1}{2}\Delta t, \frac{1}{2}\Delta f\right)$, with $v(t) = s(-t)$. Furthermore, we have by shift-invariance that

$$\iint W^\alpha_s(t,f)\,dt\,df = \iint W^\alpha_u(t,f)\,dt\,df, \tag{104}$$

these two numbers being supposed positive. Hence

$$\frac{1}{1-\alpha}\log_2 \frac{\iint W^\alpha_{s+u}(t,f)\,dt\,df}{\left(\iint W_{s+u}(t,f)\,dt\,df\right)^\alpha} \longrightarrow H_\alpha(W_s)+1 \tag{105}$$

when $|\mathcal{D}| \to \infty$, provided we can show that

$$\iint X^{\alpha}_{s,u}(t,f)\,dt\,df \;\longrightarrow\; 0, \tag{106}$$

$$\iint W^{\alpha_1}_s(t,f)\,X^{\alpha_2}_{s,u}(t,f)\,W^{\alpha_3}_u(t,f)\,dt\,df \;\longrightarrow\; 0 \tag{107}$$

as $|\mathcal{D}| \to \infty$ for the relevant set of $(\alpha_1, \alpha_2, \alpha_3)$.

As to (106), we write

$$\cos^{\alpha} x \;=\; \sum_{k=-\alpha}^{\alpha} b_k\,e^{jkx}, \tag{108}$$

where we note that $b_0 = 0$, since α is odd. Then, using (100) we obtain

$$\iint X^{\alpha}_{s,u}(t,f)\,dt\,df \;=\; 2^{\alpha} \sum_{k=-\alpha}^{\alpha} b_k \iint W^{\alpha}_s\left(t - \frac{\Delta t}{2}, f - \frac{\Delta f}{2}\right) e^{j2\pi k(t\Delta f - f\Delta t + \frac{1}{2}\Delta t\Delta f)}\,dt\,df. \tag{109}$$

The substitution $t - \frac{1}{2}\Delta t \to t$, $f - \frac{1}{2}\Delta f \to f$, which leaves the form $t\Delta f - f\Delta t + \frac{1}{2}\Delta t\Delta f$ invariant, then yields integrals of the form

$$\iint W^{\alpha}_s(t,f)\,e^{j2\pi k(t\Delta f - f\Delta t)}\,dt\,df \tag{110}$$

in the right-hand side series in (109). From Lemma 7 and $\alpha \geq 3$, it follows that $W^{\alpha}_s \in L^1(\mathbb{R}^2)$, whence by Lemma 8 we see that the integrals (110) tend to zero when $|\mathcal{D}| \to \infty$ and $k \neq 0$. It follows then that the expression in (109) tends to zero when $|\mathcal{D}| \to \infty$, since $b_0 = 0$.

For the expressions in (107) we argue as follows. By (100) we have

$$\begin{aligned} J \;&:=\; \frac{1}{2^{\alpha_2}} \left| \iint W^{\alpha_1}_s(t,f)\,X^{\alpha_2}_{s,u}(t,f)\,W^{\alpha_3}_u(t,f)\,dt\,df \right| \\ &\leq\; \iint |W^{\alpha_1}_s(t,f)| \left| W^{\alpha_2}_s\left(t - \frac{\Delta t}{2}, f - \frac{\Delta f}{2}\right)\right| \left|W^{\alpha_3}_s\left(t - \Delta t, f - \Delta f\right)\right|\,dt\,df. \end{aligned} \tag{111}$$

Assume that $\alpha_1 > 0$, $\alpha_3 > 0$ (the other cases go in a similar way). By Lemma 7 there is an $A > 0$ such that

$$J \;\leq\; A \iint K(t,f)\,L(\Delta t - t, \Delta f - f)\,dt\,df \;=:\; M(\Delta t, \Delta f), \tag{112}$$

where we have set

$$K(t,f) := |W^{\alpha_1}_s(t,f)|\,, \qquad L(t,f) := |W^{\alpha_3}_s(-t,-f)|\,. \tag{113}$$

By Lemma 7, we have that $K, L \in L^2(\mathbb{R}^2)$. Now let $\varepsilon > 0$ and take smooth, compactly supported $P, Q \in L^2(\mathbb{R}^2)$ such that

$$\|K - P\|_2 < \varepsilon, \qquad \|L - Q\|_2 < \varepsilon. \tag{114}$$

Denote $R := P * Q$, so that R is smooth and compactly supported as well. Then we have for all $\Delta t, \Delta f$ by the Cauchy-Schwarz inequality

$$\begin{aligned} |M(\Delta t, \Delta f) - A\,R(\Delta t, \Delta f)| \;&\leq\; A\,\|K - P\|_2\,\|L\|_2 + A\,\|P\|_2\,\|L - Q\|_2 \\ &\leq\; \varepsilon A\,\left(\|L\|_2 + \|P\|_2\right). \end{aligned} \tag{115}$$

This shows that M can be approximated uniformly and arbitrarily closely by functions $A\,R(\Delta t, \Delta f)$ of compact support. It follows that $M(\Delta t, \Delta f) \to 0$ as $|\mathcal{D}| \to \infty$. Hence J in (111) tends to zero, as $|\mathcal{D}| \to \infty$ and the proof for the Wigner distribution case is complete.

Cohen's class TFR case: Next we consider TFRs of the Cohen type with $\Phi \in L^1(\mathbb{R}^2)$ and $\iint \Phi(t,f)\,dt\,df = 1$. We require Young's Inequality.

Lemma 9 (Young) *Let $1/p + 1/q = 1/r + 1$ with $1 \le p, q, r \le \infty$. When $g \in L^p(\mathbb{R}^n)$ and $h \in L^q(\mathbb{R}^n)$, we have $g * h \in L^r(\mathbb{R}^n)$ and*

$$\|g * h\|_r \;\le\; \|g\|_p\,\|h\|_q. \tag{116}$$

Moreover, when $g \in L^1(\mathbb{R}^n)$ and $h \in C_0(\mathbb{R}^n)$, we have $g * h \in C_0(\mathbb{R}^n)$.

To prove Theorem 2 for the Cohen distributions, we replace all W and X in the expansion (101) by $W*\Phi$ and $X*\Phi$, thereby noting that the latter functions are in $L^2(\mathbb{R}^2)\cap L^\infty(\mathbb{R}^2)\cap C_0(\mathbb{R}^2)$, since the same holds for W and X, and $\Phi \in L^1(\mathbb{R}^2)$ (see Lemmata 7 and 9). As in (103), we have (solving the problem of undefinedness of $\iint (W_s * \Phi)(t,f)\,dt\,df$ by the assumption $\iint \Phi(t,f)\,dt\,df = 1$ in the same way as was done in (102))

$$\iint (W_{s+u} * \Phi)\,(t,f)\,dt\,df \;\longrightarrow\; 2\iint (W_s * \Phi)\,(t,f)\,dt\,df. \tag{117}$$

Also the analogue of (104) holds by shift-invariance. We must therefore show that

$$\iint (X_{s,u} * \Phi)^\alpha\,dt\,df \;\longrightarrow\; 0, \tag{118}$$

$$\iint (W_s * \Phi)^{\alpha_1}\,(X_{s,u} * \Phi)^{\alpha_2}\,(W_u * \Phi)^{\alpha_3}\,dt\,df \;\longrightarrow\; 0 \tag{119}$$

as $|\mathcal{D}| \to \infty$ for the set of relevant $(\alpha_1, \alpha_2, \alpha_3)$.

For (118), we first note that $(X_{s,u} * \Phi)^\alpha \in L^1(\mathbf{R}^2)$, since $\alpha \ge 2$. Next we use (100) to obtain

$$\begin{aligned}
&\frac{1}{2^\alpha}\iint (X_{s,u} * \Phi)^\alpha\,(t,f)\,dt\,df \\
&\qquad = \iint \left\{ \iint W_s\!\left(u - \frac{\Delta t}{2}, v - \frac{\delta f}{2}\right)\,\Phi(t-u, f-v) \right. \\
&\qquad\qquad \left. \times\,\cos\!\left[2\pi\left(u\Delta t - v\Delta f + \frac{1}{2}\Delta t\Delta f\right)\right] du\,dv \right\}^\alpha dt\,df \\
&\qquad = \iint U^\alpha(t,f;\Delta t,\Delta f)\,dt\,df,
\end{aligned} \tag{120}$$

where we have set

$$U(t,f;\Delta t,\Delta f) \;:=\; \iint W_s(u,v)\,\cos\!\left[2\pi\left(u\Delta t - v\Delta f + \frac{1}{2}\Delta t\Delta f\right)\right]\,\Phi(t-u,f-v)\,du\,dv. \tag{121}$$

The identity between the last two lines of (120) is obtained by change of variables according to $u-\frac{1}{2}\Delta t \to u$, $v-\frac{1}{2}\Delta f \to v$ followed by $t-\frac{1}{2}\Delta t \to t$, $f-\frac{1}{2}\Delta t \to f$. Evidently

$$|U(t,f;\Delta t,\Delta f)| \;\le\; (|W_s| * |\Phi|)\,(t,f) \;\in\; L^2(\mathbb{R}^2)\cap L^\infty(\mathbb{R}^2). \tag{122}$$

And also, by Lemma 8, for any $t,f\in\mathbb{R}$

$$U(t,f;\Delta t,\Delta f) \;\longrightarrow\; 0 \tag{123}$$

as $|\mathcal{D}|\to\infty$. Since $\alpha\ge 2$, we conclude from (122), (123) and Lebesgue's theorem on dominated convergence that

$$\iint U^\alpha(t,f;\Delta t,\Delta f) \;\longrightarrow\; 0 \tag{124}$$

as $|D|\to\infty$. This settles (118).

As to (119), we can literally repeat the argument used for the Wigner distribution case (see just after Lemma 9). This proves Theorem 2 for the case of Cohen TFRs with $\Phi\in L^1(\mathbb{R}^2)$, $\iint \Phi(t,f)\,dt\,df = 1$. □

Finally, we note that the arguments to prove Theorem 2 remain valid when $s+\mathcal{D}s = s+u$ is replaced by $s_1+\mathcal{D}s_2$, where $s_1,s_2\in L^2(\mathbb{R})$ are unrelated. In particular, for odd $\alpha\ge 3$, we have that

$$\lim_{|\mathcal{D}|\to\infty}\iint W^\alpha_{s_1+\mathcal{D}s_2}(t,f)\,dt\,df \;=\; \iint W^\alpha_{s_1}(t,f)\,dt\,df + \iint W^\alpha_{s_2}(t,f)\,dt\,df. \tag{125}$$

C Third-order Rényi Entropy for the Sum of Two Gaussians

In this appendix we consider the 3rd-order Rényi entropy of the sum of two Gaussian pulses in (32)–(33). The parameters γ and/or ξ in (33) will be suppressed in the case that $\gamma=1$ and/or $\xi=0$; that is

$$g_\gamma(\nu) := g_\gamma(\nu,0), \qquad g(\nu,\xi) := g_1(\nu,\xi), \qquad g(\nu) := g_1(\nu,0). \tag{126}$$

C.1 Simplification via symplectic transformation

We first note that it is sufficient to consider the case

$$\gamma=1,\ a>0,\ b=0,\ w=1,\ c=-a,\ d=0,\ z\in\mathbf{C}\setminus\{0\}. \tag{127}$$

To see this, write s as

$$s \;=\; z'\,Z_{\sqrt{\gamma}}\,R_{b'}\,T_{a'}\,\Gamma(-\theta)\,g(|a''+jb''|) \;+\; z''\,g(-|a''+jb''|) \tag{128}$$

where

$$z' = w\,e^{j\frac{\pi}{2}(b-d)(a+c)}, \qquad b' = \frac{b+d}{2\sqrt{\gamma}}, \qquad a' = \frac{(a+c)\sqrt{\gamma}}{2} \tag{129}$$

$$z'' = \frac{z}{w}e^{-j\pi(b-d)(a+c)}, \qquad a''+jb'' = \frac{(a-c)\sqrt{\gamma}}{2} + j\,\frac{b-d}{2\sqrt{\gamma}} = |a''+jb''|\,e^{j\theta}. \tag{130}$$

The operators Z_λ, R_y, T_x and $\Gamma(\varphi)$ in (128) are given by

$$(Z_\lambda s)(t) := \lambda^{1/2}\, s(\lambda t), \qquad (R_y s)(t) := e^{j2\pi yt}\, s(t), \qquad (T_x s)(t) := s(t-x) \tag{131}$$

and

$$(\Gamma(\varphi)s)(t) \;:=\; \pm\left(\frac{1}{j\sin\varphi}\right)^{\frac{1}{2}} \int \exp\!\left(\frac{-\pi}{j\sin\varphi}\left(u^2\cos\varphi - 2ut + t^2\cos\varphi\right)\right) s(u)\, du \tag{132}$$

when $\sin\varphi \neq 0$ and

$$\Gamma(\varphi)s \;:=\; \pm\, s \tag{133}$$

when $\sin\varphi = 0$. In (132)–(133), the operator $\Gamma(\varphi)$ is given apart from a sign that is irrelevant in the present context (just like the number z' in (128)). The operators in (131)–(133) are members of the *metaplectic group,* [19, Ch. 4], and their action on signals is reflected by certain symplectic linear transforms of the time-frequency plane. We have, explicitly,

$$\begin{aligned} W_{Z_\lambda s}(t,f) &= W_s(\lambda t, \lambda^{-1}t), && (134)\\ W_{R_y s}(t,f) &= W_s(t, f-y), && (135)\\ W_{T_x s}(t,f) &= W_s(t-x, f), && (136)\\ W_{\Gamma(\varphi)s}(t,f) &= W_s(t\cos\varphi - f\sin\varphi, t\sin\varphi + f\cos\varphi), && (137) \end{aligned}$$

showing that integrals of functions of W_s over the entire time-frequency plane are invariant under application to s of any of these operators. (See also [38, Secs. 27.3, 27.4.2, 27.12.2].) Hence s and the signal $g\left(|a''+jb''|\right) + z''\, g\left(-|a''+jb''|\right)$ (see (128)) yield the same value for the right-hand side of (35). That is,

$$Q_\gamma(a,b,w;c,d,z) \;=\; Q_1\left(|a''+jb''|, 0, 1; -|a''+jb''|, 0, z''\right), \tag{138}$$

with z'' and $a''+jb''$ given in (130).

C.2 Proof of Proposition 4

In this case, the signal in (35) is given by

$$s(t) \;=\; g(a)(t) \;+\; r\, e^{j\varphi} g(-a)(t) \;=\; 2^{\frac{1}{4}} e^{-\pi(t-a)^2} + 2^{\frac{1}{4}} r\, e^{-\pi(t+a)^2 + j\varphi}. \tag{139}$$

We first note that

$$\begin{aligned} W_{g(a)}(t,f) &= 2\exp\!\left(-2\pi(t-a)^2 - 2\pi f^2\right), && (140)\\ W_{g(a),g(-a)}(t,f) &= 2\exp\!\left(-j4\pi af - 2\pi t^2 - 2\pi f^2\right). && (141) \end{aligned}$$

Therefore,

$$\begin{aligned} W_s(t,f) &= 2e^{-2\pi f^2}\left(e^{-2\pi(t-a)^2} + r^2 e^{-2\pi(t+a)^2} + 2re^{-2\pi t^2}\cos(4\pi af + \varphi)\right) && (142)\\ &=: 2e^{-2\pi f^2}(A_1 + A_2 + 2A_3) && (143) \end{aligned}$$

with the obvious identifications for A_1, A_2, and A_3. We then obtain

$$\iint W_s(t,f)\,dt\,df \;=\; \|s\|^2 \;=\; 1+r^2+2rx^3y, \tag{144}$$

the definitions of x and y from Proposition 4. Expanding,

$$(A_1+A_2+2A_3)^3 \;=\; (A_1+A_2)^3+6(A_1+A_2)^2A_3+12(A_1+A_2)A_3^2+8A_3^3, \tag{145}$$

we note that A_1+A_2 depends on t only. Calculating

$$\begin{aligned}
\iint e^{-6\pi f^2}(A_1+A_2)^3\,dt\,df &= \frac{1}{6}(1+r^6)+\frac{1}{2}(r^2+r^4)x^8, &(146)\\
\iint e^{-6\pi f^2}(A_1+A_2)^2A_3\,dt\,df &= \frac{1}{6}(rx^3+r^5x^3+2r^3x^7)\cos\varphi, &(147)\\
\iint e^{-6\pi f^2}(A_1+A_2)A_3^2\,dt\,df &= \frac{1}{12}(r^2+r^4)(1+x^4\cos 2\varphi)x^2, &(148)\\
\iint e^{-6\pi f^2}A_3^3\,dt\,df &= \frac{1}{8}r^3\left(\frac{1}{3}x^9\cos 3\varphi+x\cos\varphi\right) &(149)
\end{aligned}$$

and using

$$y=\cos\varphi,\qquad 2y^2-1=\cos 2\varphi,\qquad 4y^3-3y=\cos 3\varphi, \tag{150}$$

leads us to

$$\begin{aligned}
\iint W_s^3(t,f)\,dt\,df \;=\; 8\Bigg[&\frac{1}{6}(1+r^6)+\frac{1}{2}(r^2+r^4)x^8+(r^2+r^4)(x^2-x^6) \\
&+\;(rx^3+r^5x^3+2r^3x^7-r^3x^9+r^3x)y \\
&+\;2(r^2+r^4)x^6y^2+\frac{4}{3}r^3x^9y^3\Bigg].
\end{aligned} \tag{151}$$

Therefore, from (144) and (151), we have

$$\begin{aligned}
\frac{\iint W_s^3(t,f)\,dt\,df}{\left(\iint W_s(t,f)\,dt\,df\right)^3} &= \frac{4}{3}+\frac{\iint W_s^3(t,f)\,dt\,df-\frac{4}{3}\left(\iint W_s(t,f)\,dt\,df\right)^3}{\left(\iint W_s(t,f)\,dt\,df\right)^3} \\
&= \frac{4}{3}-\left(1-2x^2+2x^6-x^8\right)\left[4(r^2+r^4)-8xyr^3\right]\left(1+r^2+2rx^3y\right)^{-3} \\
&= \frac{4}{3}-\left(1-x^2\right)^3\left(1+x^2\right)\left(\frac{r^{-1}+r}{2}-xy\right)\left(\frac{r^{-1}+r}{2}+x^3y\right)^{-3},
\end{aligned} \tag{152}$$

and (38) follows from the definition of v in (37) and from (34). □

C.3 Properties of $Q(x,y,v)$

The form (38) is very convenient for finding the minimum and maximum of Q_1 and for studying the behavior of Q_1 as a function of a,r, and φ.

We have, for instance, that $Q(x,y,v)$ decreases for $y \in [-1,1]$ for fixed $x \in (0,1]$, $v \geq 1$. Hence the minimum of Q equals the minimum of

$$Q(x,-1,v) \;=\; \frac{4}{3} - (1-x^2)^3\,(1+x^2)\frac{v+x}{(v-x^3)^3} \tag{153}$$

over $x \in (0,1]$, $v \geq 1$. The right-hand side of (153) is increasing in $v \geq 1$ when $x \in (0,1]$ is fixed. Hence the minimum of Q equals the minimum of

$$Q(x,-1,1) \;=\; \frac{4}{3} - \left(\frac{1-x^2}{1-x^3}\right)^3 (1+x^2)(1+x) \tag{154}$$

over $x \in (0,1]$. Since

$$\frac{\partial Q}{\partial x}(x,-1,1) \;=\; -\left(\frac{1-x}{1-x^3}\right)^4 (1-x^2)^3, \tag{155}$$

we see that $Q(x,-1,1)$ decreases for $x \in (0,1]$, so that its minimum value occurs at $x = 1$. Therefore,

$$Q(x,y,v) \;\geq\; \frac{4}{27} \;=\; Q(1,-1,1) \tag{156}$$

for $x \in (0,1]$, $y \in [-1,1]$, $v \geq 1$. Thus the maximum value of $H_3(s) = -\frac{1}{2}\log_2(\frac{4}{27}) \approx 1.377$ (see Figure 3).

The maximum of Q clearly equals $4/3$ and occurs at $x = 1$, $(-1,1) \neq (y,v) \in [-1,1] \times [1,\infty)$. This bound and (156) combine to give (39). Thus the minimum value of $H_3(s) = -\frac{1}{2}\log_2(\frac{4}{3}) \approx -0.208$ (see Figure 3).

The behavior of Q at $(1,-1,1)$ is somewhat irregular. We have for $\theta \geq 0$ that

$$\lim_{x\uparrow 1,(v+y)(1-x)^{-1}\to\theta} Q(x,y,v) \;=\; \frac{4}{3} - \frac{32}{(3+\theta)^3}. \tag{157}$$

Using (38) it is obvious that Q increases for $x \in (0,1]$ when $v \geq 1$ and $y \in (0,1]$ are fixed. However, from (38) and

$$\frac{d}{dx}\left(\frac{v-xy}{(v+x^3y)^3}\right) \;=\; -\,y\,\frac{v(1+9x^2)-8x^3y}{(v+x^3y)^3}, \tag{158}$$

it follows that Q decreases in x near $x = 0$ and increases in x near $x = 1$ when $y \in (-1,0)$ and $v \geq 1$ are fixed (see Figure 3).

Finally, Q decreases or increases in v for $x \in (0,1]$ and $y \in [-1,1]$ according to whether $v <$ or $> \frac{1}{2}xy(x^2+3)$. Furthermore, we note that

$$Q(0,y,v) \;=\; \frac{4}{3} - \frac{1}{v^2}, \qquad -1 \leq y \leq 1, \quad v \geq 1, \tag{159}$$

is the limiting value of Q_1 in (38) as $a \to \infty$.

For the cases that $y = \pm 1$, $v = 1$, we have the special results

$$Q(x,1,1) \;=\; \frac{1}{3} + \frac{1}{4w^2-3} + \left(\frac{1}{4w^2-3}\right)^2 - \left(\frac{1}{4w^2-3}\right)^3, \tag{160}$$

$$Q(x,-1,1) \;=\; \frac{1}{3} - \frac{1}{4w^2-1} + \left(\frac{1}{4w^2-1}\right)^2 + \left(\frac{1}{4w^2-1}\right)^3, \tag{161}$$

where (see Proposition 4)

$$w \;=\; \frac{1}{2}\left(x^{1/2}+x^{-1/2}\right) \;=\; \cosh\frac{1}{3}\pi a^2, \tag{162}$$

so that

$$\frac{1}{4w^2-3} \;=\; \frac{\cosh\frac{1}{3}\pi a^2}{\cosh \pi a^2}, \qquad \frac{1}{4w^2-1} \;=\; \frac{\sinh\frac{1}{3}\pi a^2}{\sinh \pi a^2}. \tag{163}$$

This shows, for instance, that when x increases from 0 to 1, $Q(x,1,1)$ increases from 1/3 to 4/3 and $Q(x,-1,1)$ decreases from 1/3 to 4/27.

C.4 Effects of Gaussian smoothing

We next present formulas for the quantities

$$\frac{\iint (W_s * \Phi_\rho)^n(t,f)\,dt\,df}{(\iint (W_s * \Phi_\rho)(t,f)\,dt\,df)^n}, \qquad n=2,3 \tag{164}$$

which are required in Sections 3.3 and 4.1. Here we take the signal s to be the sum of two Gaussians as in (32) with $w=z=\gamma=1$ and Φ_ρ to be the 2-d Gaussian (41). Using symplectic transformations as in Section C.1 and the radial symmetry of both the Wigner distribution of $g(0,0)$ and Φ_ρ, it can be shown that the quantities in (164) remain the same when s is replaced by the signal

$$2^{\frac{1}{4}}\, e^{-\pi(t-|a''+jb''|)^2} \;+\; 2^{\frac{1}{4}}\, e^{-\pi(t+|a''+jb''|)^2+j\varphi}, \tag{165}$$

with $a''+jb''=\frac{1}{2}(a-c+j(b-d))$ and $\varphi=\psi-\pi(a+c)(b-d)$. For the resulting signal (139) we compute

$$(W_s * \Phi_\rho)(t,f) \;=\; \eta\, e^{-\pi\eta a f^2}\left(e^{-\pi\eta(t-a)^2}+e^{-\pi\eta(t+a)^2}+2Re^{-\pi\eta t^2}\cos(2\pi a f+\varphi)\right), \tag{166}$$

with

$$\eta=\frac{1}{1+\rho^2}, \qquad R=e^{-\pi\eta\rho^2a^2}. \tag{167}$$

Next, with x and y as in (37) and

$$x_1 \;=\; e^{-\frac{1}{3}\pi\eta a^2}, \tag{168}$$

the same methods as those employed in Section C.2 yield

$$\iint (W_s * \Phi_\rho)(t,f)\,dt\,df \;=\; 2(1+x^3\,y), \tag{169}$$

$$\iint (W_s * \Phi_\rho)^2(t,f)\,dt\,df \;=\; \eta\left(1+x_1^6+4Ryx_1^3+R^2(1+(2y^2-1)x_1^6)\right), \tag{170}$$

$$\iint (W_s * \Phi_\rho)^3(t,f)\,dt\,df \;= \tag{171}$$

$$2\eta^2\left[\frac{1}{3}+x_1^8+R^2(x_1^2-x_1^6)+(2Rx_1^3+2Rx_1^7+R^3x_1-R^3x_1^9)y+4R^2x_1^6y^2+\frac{4}{3}R^3x_1^9y^3\right]. \tag{172}$$

A few additional simplifications lead to the forms

$$\frac{\iint (W_s * \Phi_\rho)^2(t,f)\,dt\,df}{\left(\iint (W_s * \Phi_\rho)(t,f)\,dt\,df\right)^2} = \frac{1}{2}\eta\left[1 - \frac{1}{2}(1-R^2)(1-x_1^6)(1+x^3y)^{-2}\right] \tag{173}$$

and

$$\frac{\iint (W_s * \Phi_\rho)^3(t,f)\,dt\,df}{\left(\iint (W_s * \Phi_\rho)(t,f)\,dt\,df\right)^3} = \frac{\eta^2}{4}\left[\frac{4}{3} - \frac{1-x_1^4}{(1+x^3y)^3}\left(x_1^4 - 2R^2x_1^2 + 1 - (x_1^4 - 2R^{-2}x_1^2+1)R^3x_1y\right)\right]. \tag{174}$$

To obtain (42), we substitute $\eta = 1$, $R = x^{3/2}$, and $x_1 = x^{1/2}$.

Finally, we turn to the expression (53) in Section 4.1. The definitions of x in (37) and of x_1 in (168) show that we can ignore x and x_1 in (173), (174) when a is sufficiently large (in Figure 6 we have $a > 2$, and this is sufficiently large). Hence, replacing the right-hand sides of (173) and (174) by $\frac{1}{12}\eta^2$ and $\frac{1}{4}\eta(1+R)$, respectively, and using (167), we obtain, to a good approximation, (53).

D Entropy Lower Bounds

D.1 Proof of Theorem 5

In addition to Young's Inequality (Lemma 9 in Appendix B above), we will need a relatively recent result of Lieb [23].[12] Recall that C_s^{Gspect} denotes the spectrogram of the signal s computed using Gaussian window g.

Lemma 10 (Lieb) *Given $s \in L^2(\mathbb{R})$ and $p \geq 2$, then*

$$\|W_s\|_p^p \;\leq\; (2^{p-1}/p)\,\|s\|_2^{2p} \tag{175}$$

with equality if and only if s is a Gaussian. In addition,

$$\|C_s^{\mathrm{Gspect}}\|_p^p \;\leq\; (2/p)\,\|s\|_2^p\,\|g\|_2^p \tag{176}$$

with equality if and only if s is a Gaussian of the same form as g (see [23] for more details).

For unit-energy s and a kernel such that $\iint \Phi(t,f)\,dt\,df = 1$, we have $\iint C_s(t,f)\,dt\,df = 1$. Using first Lemma 9 and then (175) from Lemma 10, we obtain

$$\iint C_s^\alpha(t,f)\,dt\,df \;\leq\; \|C_s\|_\alpha^\alpha \;\leq\; \|W_s\|_\alpha^\alpha\,\|\Phi\|_1^\alpha \;\leq\; \frac{2^{\alpha-1}}{\alpha}\,\|\Phi\|_1^\alpha. \tag{177}$$

Thus,

$$H_\alpha(C_s) \;\geq\; \frac{1}{1-\alpha}\log_2\frac{2^{\alpha-1}}{\alpha}\,\|\Phi\|_1^\alpha, \tag{178}$$

and (44) follows.

[12] Lieb sharpens Lemma 9 further in [23].

The bound (45) for the Wigner distribution follows from the same argument but omitting the kernel Φ. While Gaussian signals saturate the bound (45) for the Wigner distribution, the more general bound (44) may be unattainable for other Cohen's class TFRs.

The bound (46) for the Gaussian-windowed spectrogram follows from the same argument as the Wigner distribution but using (176) from Lemma 10. □

D.2 Proof of Theorem 6

Since the classical Young's theorem (Lemma 9) does not apply to the affine smoothing of (48), we begin by stating an analogue matched to the affine convolution

$$(g\#h)(c,d) \;:=\; \iint g(a,b)\, h\Big(b\,(c-a)\,,\, \frac{d}{b}\Big)\, da\, db \tag{179}$$

defined on the affine group. The following was obtained by specializing the general results of [39, pp. 293–8] to the scalar affine group having group operation "$\circ$" defined by $(a,b)\circ(c,d) := (a+c/b, bd)$, $b,d>0$, and left Haar measure $da\,db$. All integrals and norms in the following can be interpreted to run over the upper half-plane $\mathbf{U} := \mathbb{R}\times\mathbb{R}_+$ to account for $b>0$ in (a,b).

Lemma 11 *Let $1/p+1/q = 1/r+1$ with $1\le p,q,r\le\infty$. When $g\in L^p(\mathbf{U})$ and $h\in L^q(\mathbf{U})$, we have $g\#h \in L^r(\mathbf{U})$ and*

$$\|g\#h\|_r \;\le\; \|g\|_p\, \|h\|_q. \tag{180}$$

While the affine smoothing (48) is not a group convolution proper, the condition for existence and integrability of an affine class TFR follows immediately from this Lemma. Substituting $\lambda(tf, 1/f) := \Pi(t,f)$ into (48) immediately yields the form (179) and the conclusion that $\Omega_s \in L^p(\mathbf{U})$, $1\le p\le\infty$, provided $W_s\in L^p(\mathbf{U})$ and $\lambda\in L^1(\mathbf{U})$. A change of variable converts the constraint on λ into a constraint on the original kernel Π

$$\lambda(t,f)\in L^1(\mathbf{U}) \quad\Leftrightarrow\quad \frac{1}{f}\,\Pi(t,f)\in L^1(\mathbf{U}). \tag{181}$$

Now, using first Lemma 11 and then Lemma 10, we have for unit-energy s

$$\iint \Omega_s^\alpha(t,f)\, dt\, df \;\le\; \|\Omega_s\|_\alpha^\alpha \;\le\; \|W_s\|_\alpha^\alpha \left\|\frac{1}{f}\,\Pi(t,f)\right\|_1^\alpha \;\le\; \frac{2^{\alpha-1}}{\alpha}\left\|\frac{1}{f}\,\Pi(t,f)\right\|_1^\alpha. \tag{182}$$

Taking logarithms yields the result. □

Acknowledgements

The authors wish to thank Richard Askey, Michèle Basseville, Paulo Gonçalvès, Alfred Hero III, Douglas Jones, Richard Orr, and William Williams for stimulating discussions at various stages of this work. Thanks also to the Isaac Newton Institute of Cambridge University, where the final version of the paper was completed.

References

[1] W. J. Williams, M. L. Brown, and A. O. Hero, "Uncertainty, information, and time-frequency distributions," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1566, pp. 144–156, 1991.

[2] R. Orr, "Dimensionality of signal sets," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1565, pp. 435–446, 1991.

[3] L. Cohen, "What is a multicomponent signal?," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '92*, vol. V, pp. 113–116, 1992.

[4] L. Cohen, *Time-Frequency Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1995.

[5] P. Flandrin, *Temps-Fréquence.* Paris: Hermès, second ed., 1998.

[6] P. Flandrin, *Time-Frequency and Time-Scale Analysis.* Academic Press, 1999.

[7] T. Cover and J. Thomas, *Elements of Information Theory.* Wiley, 1991.

[8] R. G. Baraniuk, P. Flandrin, and O. Michel, "Information and complexity on the time-frequency plane," in *14ème Colloque GRETSI*, (Juan Les Pins, France), pp. 359–362, 1993.

[9] P. Flandrin, R. G. Baraniuk, and O. Michel, "Time-Frequency complexity and information," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '94*, vol. III, pp. 329–332, 1994.

[10] C. E. Shannon, "A mathematical theory of communication, Part I," *Bell Sys. Tech J.*, vol. 27, pp. 379–423, July 1948.

[11] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Stat. and Prob.*, vol. 1, pp. 547–561, 1961.

[12] W. J. Williams, "Reduced interference distributions: Biological applications and interpretations," *Proc. IEEE*, vol. 84, pp. 1264–1280, Sept. 1996.

[13] T.-H. Sang and W. J. Williams, "Rényi information and signal-dependent optimal kernel design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '95*, vol. 2, pp. 997–1000, 1995.

[14] O. Rioul and P. Flandrin, "Time-scale energy distributions: A general class extending wavelet transforms," *IEEE Trans. Signal Processing*, vol. 40, pp. 1746–1757, July 1992.

[15] A. Papandreou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "The hyperbolic class of quadratic time-frequency representations. Part I: Constant-Q warping, the hyperbolic paradigm, properties, and members," *IEEE Trans. Signal Processing*, vol. 41, pp. 3425–3444, Dec. 1993.

[16] F. Hlawatsch, A. Papandreou-Suppappola, and G. F. Boudreaux-Bartels, "The power classes — Quadratic time-frequency representations with scale covariance and dispersive time-shift covariance," *IEEE Trans. Signal Processing*, vol. 47, pp. 3067–3083, Nov. 1999.

[17] R. G. Baraniuk and D. L. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. Signal Processing*, vol. 43, pp. 2269–2282, Oct. 1995.

[18] F. Hlawatsch and P. Flandrin, "The interference structure of the Wigner distribution and related time-frequency signal representations," in *The Wigner Distribution — Theory and Applications in Signal Processing* (W. Mecklenbräuker and F. Hlawatsch, eds.), Amsterdam: Elsevier, 1997.

[19] G. B. Folland, *Harmonic Analysis in Phase Space.* Princeton, NJ: Princeton University Press, 1989.

[20] O. Michel, R. G. Baraniuk, and P. Flandrin, "Time-frequency based distance and divergence measures," *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, pp. 64–67, Oct. 1994.

[21] O. Debbi and J. Gillis, "An inequality relating to triple integrals of Laguerre functions," *Israel J. Math.*, vol. 18, pp. 45–52, 1974.

[22] A. J. E. M. Janssen, "Positivity properties of phase-plane distribution functions," *J. Math. Phys.*, vol. 25, pp. 2240–2252, July 1984.

[23] E. Lieb, "Integral bounds for radar ambiguity functions and the Wigner distribution," *J. Math. Phys.*, vol. 31, pp. 594–599, Mar. 1990.

[24] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pp. 429–457, 1946.

[25] W. Beckner, "Inequalities in Fourier analysis," *Annals Math.*, vol. 102, pp. 159–182, 1975.

[26] J. Bertrand and P. Bertrand, "A class of affine Wigner functions with extended covariance properties," *J. Math. Phys.*, vol. 33, pp. 2515–2527, July 1992.

[27] D. L. Jones and T. W. Parks, "A resolution comparison of several time-frequency representations," *IEEE Trans. Signal Processing*, vol. 40, pp. 413–420, Feb. 1992.

[28] W. J. Williams and T. H. Sang, "Adaptive RID kernels which minimize time-frequency uncertainty," *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, pp. 96–99, Oct. 1994.

[29] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: Optimal kernel design," *IEEE Trans. Signal Processing*, vol. 41, pp. 1589–1602, Apr. 1993.

[30] R. G. Baraniuk and D. L. Jones, "A radially Gaussian, signal-dependent time-frequency representation," *Signal Processing*, vol. 32, pp. 263–284, June 1993.

[31] D. L. Jones and R. G. Baraniuk, "An adaptive optimal-kernel time-frequency representation," *IEEE Trans. Signal Processing*, vol. 43, pp. 2361–2371, Oct. 1995.

[32] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2127–2135, Dec. 1990.

[33] D. L. Jones and R. G. Baraniuk, "A simple scheme for adapting time-frequency representations," *IEEE Trans. Signal Processing*, vol. 42, pp. 3530–3535, Dec. 1994.

[34] I. Daubechies, *Ten Lectures on Wavelets.* New York: SIAM, 1992.

[35] R. Coifman and V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. IT–38, pp. 713–718, Mar. 1992.

[36] H. J. Groenewold, "On the principles of elementary quantum mechanics," *Physica*, vol. 21, pp. 405–460, 1946.

[37] J. Gillis and H. Shimshoni, "Triple product integrals of Laguerre functions," *Math. Comp.*, vol. 16, pp. 50–62, 1962.

[38] N. G. de Bruijn, "A theory of generalized functions with applications to Wigner distribution and Weyl correspondence," *Nieuw Archief voor Wiskunde (3)*, vol. 21, pp. 205–280, 1973.

[39] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis*, vol. I. New York: Academic Press, 1963.

Figure Captions

Figure 1: Solid curve and axis on left: The 3rd-order Rényi entropy $H_3(W_s)$ of the Wigner distribution of the sum $s(t) := g(t-\Delta t) + g(t+\Delta t)$ of two Gaussian components plotted versus the displacement parameter Δt (see (32) in Section 3.2 for the exact signal definition). (The asymptotic $H_3(W_s)$ levels are $-\frac{1}{2}\log_2\frac{4}{3} \approx -0.208$ and $\frac{1}{2}\log_2 3 \approx 0.792$ bits.) Dotted curve and axis on right: Time-bandwidth product of s.

Figure 2: Example of a signal for which the 3rd-order Rényi entropy $H_3(W_s)$ of the Wigner distribution is not defined. (a) The signal s consists of a special combination of two odd-order Hermite functions (see Appendix A.3). Its Wigner distribution (b) as an image and (c) in 3-d (note the large negative peak).

Figure 3: The 3rd-order Rényi entropy $H_3(W_s)$ of the Wigner distribution of the two-component Gaussian signal (32) plotted versus displacement parameter a and phase φ (in rads). We use (38) with $v = 1$. Comparison with Figure 1 (which coincides with the $\varphi = 0$ slice of the surface) illustrates the sensitivity of $H_3(W_s)$ to relative phase. (The asymptotic levels here are the same as those in Figure 1; the (overestimation) peak value is $\frac{1}{2}\log_2 27 - 1 \approx 1.377$ bits.)

Figure 4: Signals and positive parts of the Wigner distributions from the experiment illustrated in Figure 3 corresponding to a single fixed separation a and two different phases φ. (a) Signal (left) and Wigner distribution (right) for $\varphi = \frac{\pi}{2}$; $H_3(W_s) = 2.741$ bits. (b) Signal and Wigner distribution for $\varphi = \frac{3\pi}{2}$; $H_3(W_s) = 3.857$ bits.

Figure 5: The 3rd-order Rényi entropy $H_3(C_s)$ of the matched-window spectrogram plotted versus displacement parameter Δt and phase φ (in rads) for the same signal utilized in Figure 3 (see (42)). The reduced sensitivity of the spectrogram to relative phase results in swifter saturation with reduced overshoot. Furthermore, for small φ and Δt, $H_3(C_s)$ actually responds sooner than $H_3(W_s)$ to small increases in Δt. Note also the 1-bit bias in the asymptotic levels of $H_3(C_s)$ versus those of $H_3(W_s)$. (The asymptotic levels here are $\frac{1}{2}\log_2 3 \approx 0.792$ and $\frac{1}{2}\log_2 3 + 1 \approx 1.792$.)

Figure 6: The effect of time-frequency smoothing on (52) as a function of the parameter β. For a signal consisting of two well-separated Gaussian pulses ($a > 2$ in (37)), we form the TFR $C_s = W_s * \Phi_\rho$, with Φ_ρ from (41). The parameter ρ controls the degree of time-frequency smoothing: $\rho = 0$ generates the Wigner distribution; $\rho = 1$ generates the matched-window spectrogram. We plot $H_3(C_s) - \beta H_2(C_s)$ versus the smoothing parameter ρ for the eleven values $\beta = 0, 0.1, \ldots, 1$. (The upper curve corresponds to $\beta = 0$, while the lower curve corresponds to $\beta = 1$.) The minimum point of each curve, marked by a circle, corresponds to an "information optimal TFR." Data obtained by numerical simulation; results from the analytical approximation (53) correspond closely.

Biographies

Richard G. Baraniuk received the BSc degree in 1987 from the University of Manitoba (Canada), the MSc degree in 1988 from the University of Wisconsin-Madison, and the PhD degree in 1992 from the University of Illinois at Urbana-Champaign, all in Electrical Engineering. In 1986, he was a research engineer with Omron Tateisi Electronics in Kyoto, Japan. After spending 1992–1993 with the Signal Processing Laboratory of Ecole Normale Supérieure, in Lyon, France, he joined Rice University, where he is currently a Professor of Electrical and Computer Engineering and DJ at KTRU. He spent Autumn 1998 at the Isaac Newton Institute of Cambridge University, England as the Rosenbaum Fellow.

Dr. Baraniuk's research interests lie in the area of signal and image processing and include wavelets, probabilistic models, networks, and time-frequency analysis. He received a NATO post-doctoral fellowship from NSERC in 1992, the National Young Investigator award from the National Science Foundation in 1994, a Young Investigator Award from the Office of Naval Research in 1995, the Rosenbaum Fellowship from the Newton Institute in 1998, the C. Holmes MacDonald National Outstanding Teaching Award from Eta Kappa Nu in 1999, the Charles Duncan Junior Faculty Achievement Award from Rice in 2000, and the ECE Young Achievement Award from the University of Illinois in 2000. He is a member of the "Signal Processing Theory and Methods" Technical Committee of the IEEE Signal Processing Society and serves on the editorial board of the journal *Applied and Computational Harmonic Analysis*.

Patrick Flandrin was born in Bron, France, in 1955. He received the Ingénieur degree from the Institut de Chimie et Physique Industrielles de Lyon in 1978 and the Docteur-Ingénieur Doctorat d'Etat ès Sciences Physiques degrees from the Institut National Polytechnique de Grenoble in 1982 and 1987, respectively. He joined the Centre National de la Recherche Scientifique (CNRS) in 1982, where he is currently Directeur de Recherches. Until 1990, he was with ICPI Lyon, where he was Head of the Signal Processing Department from 1987 to 1990. He is now with the Physics Department at Ecole Normale Supérieure de Lyon, in charge of the group "Signals and Systems in Physics." He is also adjunct-director of the CNRS cooperative structure "GdR Information, Signaux, Images et Vision."

Dr. Flandrin's research interests include nonstationary signal processing (with emphasis on time-frequency and time-scale methods), the study of self-similar stochastic processes, and the analysis of chaotic signals and systems. He has published over 100 journal or conference papers in these areas and contributed several chapters to collective books. He is the author of the book *Temps-Fréquence* (Paris: Hermès, 1993 and 1998), translated into English as *Time-Frequency/Time-Scale Analysis* (San Diego: Academic Press, 1999). He has been a guest co-editor of the Special Issue "Wavelets and Signal Processing" of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the Technical Program Chairman of the 1994 IEEE-SP *International Symposium on Time-Frequency and Time-Scale Analysis*. Past Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, he is currently on the Editorial Board of APPLIED AND COMPUTATIONAL HARMONIC ANALYSIS and of SIGNAL PROCESSING. He is also a member of the "Signal Processing Theory and

Methods" Technical Committee of the IEEE Signal Processing Society. Dr. Flandrin is a Senior Member of the IEEE and was awarded the Philip Morris Scientific Prize in Mathematics in 1991.

Augustus J. E. M. Janssen received the Eng. and Ph.D. degrees in mathematics from the Eindhoven University of Technology, Eindhoven, The Netherlands in October 1976 and June 1979, respectively. From 1979 to 1981, he was a Bateman Research Instructor at the Mathematics Department of California Institute of Technology, Pasadena, USA. In 1981, he joined the Philips Research Laboratories, Eindhoven, where his principal responsibility is to provide high-level mathematical service and consultancy in mathematical analysis. His research interest is in Fourier analysis, with emphasis on time-frequency analysis, in particular Gabor analysis. His current research interests include the Fourier analysis of non-linear devices such as quantizers. Dr. Janssen has published some 95 papers in the fields of signal analysis, mathematical analysis, Wigner distribution and Gabor analysis, information theory, and electron microscopy. Furthermore, he has published 35 internal Philips reports, and he holds 5 US patents. He received the prize for the best contribution to the Mathematical Entertainments column of the Mathematical Intelligencer in 1987 and the EURASIP 1988 Award for the Best Paper of the Year in Signal Processing.

Olvier J. J. Michel (S'84, M'85) was born in Mont Saint Martin, France, in 1963. He completed his studies at Ecole Normale Supérieure de Cachan in the department of Applied Physics, where he received the "Agrégation de Physique" in 1986. He received the Ph.D. degree from University Paris-XI Orsay in 1991 in signal processing. In 1991, he joined the Physics Department at Ecole Normale Supérieure de Lyon, France as an Assistant Professor. He spent 1993–94 with the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, on a NATO fellowship. Since 2000, he has been a Professor in the Laboratoire d'Astrophysique at the University of Nice-Sophia Antipolis, France.

Dr. Michel's research interests include non-stationnary spectral analysis, array processing, non-linear time series analysis, information theory, and dynamical systems studies, in close relationship with physical experiments in the field of chaos and hydrodynamical turbulence.