Time-scale energy distributions: a general class extending wavelet transforms

Olivier Rioul, Patrick Flandrin

Time-Scale Energy Distributions: A General Class Extending Wavelet Transforms

Olivier Rioul and Patrick Flandrin, *Member, IEEE*

Abstract—**This paper develops the theory of a new general class of signal energy representations depending on time and scale. Time-scale analysis has been introduced recently as a powerful tool through linear representations called (continuous) wavelet transforms (WT's), a concept for which we give an exhaustive bilinear generalization. Although time scale is presented as an alternative method to time frequency, strong links relating the two are emphasized, thus combining both descriptions into a unified perspective. We provide a full characterization of the new class: the result is expressed as an affine smoothing of the Wigner–Ville distribution, on which interesting properties may be further imposed through proper choices of the smoothing function parameters. Not only do specific choices allow recovering known definitions, but they also provide, via separable smoothing, a continuous transition from Wigner–Ville to either spectrograms or scalograms (squared modulus of the WT). This property makes time-scale representations a very flexible tool for nonstationary signal analysis.**

I. Introduction

A NEW method for time-varying signal analysis, called the wavelet transform (WT), has come under investigation during the past few years. It provides a new description of spectral decompositions via the scale concept [33]. The fundamental paper by Goupillaud *et al.* [16] was the first which clearly described linear time-scale decomposition of signals by means of wavelets of constant shape. This concept is attractive in that it is not another formulation of time-frequency ideas (where the Fourier duality is used to introduce the time-varying local frequency parameter), but a new formalism where the basic operator acting on the signal is of time-scaling nature rather than making use of the Fourier variable.

Although the first applications were high resolution seismic analysis and coherent states representations in quantum mechanics, representations of signals' characteristics spread over the time-scale plane seem to be useful in a variety of fields: analysis of speech and sound signals, turbulence flows, to name but a few [33]. By definition, the WT selectively matches, by means of an inner product, transient features characterized by unknown locations and time extents. It is this property that makes it relevant for many nonstationary signal processing tasks, and especially for time-varying spectral analysis. Section I briefly covers the basic properties of the WT in comparison with the short-time Fourier transform (STFT).

The mathematics of the WT has been developed by Grossmann and Morlet, mostly in the language of quantum mechanics [17], which certainly did not look attractive at first for the signal processing community. Various degrees of discretization were formalized by several mathematicians such as Meyer [24], giving rise to new ideas relating discrete WT's to quadrature mirror filter (QMF) filter bank structures [30], [32], relevant for applications such as image coding. In this context, papers by Daubechies [7], [8] and Mallat [23] are among the first that widely caught the attention of the signal processing community, although there had been substantial prior work on this approach.

Even in the early stages of the theory, emphasis was put on the fact that a relevant graphical time-scale representations of signals is the squared modulus of the WT, which is an energetic representation [18], referred to throughout this paper as *scalogram*. Although sole information on the modulus of the transform is not enough to reconstruct the general signal (phase information is also needed), the scalogram provides a graphical picture of the energy of the signal spread over the time-scale plane, in a similar way as a spectrogram spreads the energy over the time-frequency plane. A short comparison is made in Section III.

The purpose of this paper is to provide a full generalization of time-scale energy representation that gives new insights into time-scale signal analysis and better explains the relationship between scalograms and spectrograms. The resulting general framework is mainly theoretical and is expected to serve as a basis for future developments.

Our approach follows the same lines as the derivation of the general class of time-frequency energy representations, referred to as the Cohen's class [5], [6]. We first describe the scalogram by means of members of this class (e.g., Wigner–Ville distributions (WVD's)) and insist on similar properties handled by spectrograms. This is done in Section III. We then give in Section IV a complete characterization of all time-scale energy distributions involving affine smoothing of the WVD. Several equivalent formulations are considered in Section V, more or less

O. Rioul is with the Centre National d'Etudes des Télécommunications, CNET/PAB/RPE, 92131 Issy-les-Moulineaux, France.

P. Flandrin was with LTS-ICPI, Lyons, France, while this work was being done. He is now with the Laboratoire de Physique (URA 1325 CNRS), Ecole Normale Supérieure de Lyon, Lyon Cedex 07, France.

compatible with additional requirements one may impose on the representation.

The flexibility of this approach is finally illustrated in Section VI by recovering several known definitions as special cases, and by deriving a new subclass of (separable) affine smoothed WVD's. This new class offers a great versatility to balance time-frequency resolution and cross-terms reduction from the WVD. A good illustration of the possibilities of our approach is in Section VII: using Gaussian windows, we construct a sole class of time-frequency and time-scale representations of which the three well-known distributions—the WVD, the scalogram, and the spectrogram—are members. It is then possible to pass from the WVD to either the spectrogram or the scalogram by controlled smoothing.

II. The Short-Time Fourier Transform and the Wavelet Transform

A classical linear time-frequency representation, called the short-time Fourier transform (STFT), has been heavily used for nonstationary signal analysis since its introduction by Gabor [15]. The idea is to apply the Fourier transform on a time-varying signal $x(\tau)$, when $x(\tau)$ is seen through a window $h(\tau)$ of limited extent, centered at time location t [1]:

$$F_x(t, f) = \int_{-\infty}^{+\infty} x(\tau) h^*(\tau - t) e^{-2j\pi f\tau}\, d\tau. \qquad (1)$$

This intuitive definition has interesting consequences. For example, provided that the window is of finite energy, there is a one-to-one correspondence of the STFT with the original signal and an exact inversion formula holds.

In recent years, an alternative representation, called the wavelet transform (WT), has been widely addressed in the literature [28], [33]. The fundamental idea here is to replace the frequency shifting operation which occurs in the STFT by a time (or frequency) scaling operation. The resulting definition is

$$T_x(t, a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(\tau) h^* \left(\frac{\tau - t}{a}\right) d\tau \qquad (2)$$

where the function $h(t)$ (called the basic wavelet) is localized in time. This definition depends on a dilation/compression (or scale) parameter a. Note that, in general, negative a's are allowed, just as negative frequencies are allowed in Fourier analysis.

In contrast with the STFT, this transform is no longer adapted to modulations (or frequency shifts) but is now based on scalings by a: this makes the WT a time-scale representation rather than a time-frequency one [12].

As for the STFT, the WT satisfies then an inversion formula, provided that the Fourier transform $H(f)$ of $h(t)$ satisfies the so-called "admissibility condition" [17]

$$\int_{-\infty}^{+\infty} |H(f)|^2 \frac{df}{|f|} = 1. \qquad (3)$$

This means that $h(t)$ is necessarily the impulse response of some bandpass filter. In the time domain, its mean value must be zero, which implies that $h(t)$ oscillates, hence the name wavelet. Several choices exist for $h(t)$; a typical choice is the modulated Gaussian [16], [18].

As can be seen in definitions (1) and (2), both the STFT and the WT analyze signals by means of inner products with analyzing waveforms depending on two parameters. In fact, the main difference between the STFT and the WT is related to the structure of these analyzing waveforms:

1) The STFT uses an analyzing waveform of constant envelope with an increasing number of oscillations as higher frequencies are analyzed. This amounts, in the frequency domain, to using modulated versions of a low-pass filter to explore the spectral content of the analyzed signal ("uniform filterbank"). This is clearly evidenced when rewriting (1), using Parseval equality, as

$$F_x(t, f) = e^{-2j\pi tf} \int_{-\infty}^{+\infty} [X(\nu) e^{2j\pi\nu t}] H^*(\nu - f)\, d\nu.$$

2) In the WT case, the waveforms, referred to as wavelets, are of the form $1/\sqrt{|a|}\; h((\tau - t)/a)$. They are generated from the basic wavelet $h(\tau)$ by time-shift (t) and dilation (a) operations: their envelopes are narrowed as higher frequencies are analyzed, whereas their number of oscillations, hence their shape, remains constant. The WT therefore uses dilated or compressed versions of a bandpass filter, whose relative bandwidths are constant ("constant-Q-filterbank"). This structure is evidenced when rewriting (2) as

$$T_x(t, a) = \int_{-\infty}^{+\infty} [X(\nu) e^{2j\pi\nu t}]\, \sqrt{|a|}\; H^*(a\nu)\, d\nu.$$

The STFT and the WT share another common property: they both permit displaying the energy of the signal in the analysis plane (t, f) or (t, a). A precise comparison is done in the next section. We shall see that the energetic point of view moreover provides a framework in which the two time-frequency and time-scale representations are comparable.

III. Spectrograms and Scalograms as Smoothed Wigner–Ville

Owing to their definition, STFT's and WT's are complex-valued functions and they convey both modulus and phase informations. Both are necessary to reconstruct the signal, and the phase information is sometimes desirable for analysis purposes [18]. However, a description based only on the squared modulus, providing an energy density distribution, is often preferred.

In this context, the spectrogram, defined as $|F_x(t, f)|^2$, has been widely used for many signal processing tasks [26]. Since it can be shown that

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |F_x(t, f)|^2\, dt\, df = E_x = \int_{-\infty}^{+\infty} |x(t)|^2\, dt$$

the spectrogram distributes the energy E_x of the signal all over the time-frequency plane: it is the energy distribution associated to the STFT.

A similar quantity, $|T_x(t, a)|^2$, can be defined in the case of the WT: we propose to refer to it as a scalogram. One has [17], [18]

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |T_x(t, a)|^2 \frac{dt\, da}{a^2} = E_x$$

i.e., the scalogram distributes the energy E_x of the signal all over the time-scale plane: it is the energy distribution associated to the WT.

In the STFT case, it was soon recognized that the analysis was faced with a classical time and frequency resolution tradeoff [26]. In fact, the time and frequency resolution capabilities of either the STFT or the WT at a given point in the analysis plane can be measured by means of the time extent Δt and frequency extent Δf of the corresponding analyzing waveform at this point. The shorter the time duration Δt, the higher the time resolution, and, by the uncertainty principle [9], the poorer the frequency resolution.

In a spectrogram, once an analyzing window has been chosen, the resolution capabilities of the spectrogram remain fixed all over the time-frequency plane. Fig. 1[1] shows that the situation is different for scalograms: owing to the constant-Q structure described in Section II, resolution capabilities depend on the analyzed frequency. This follows from the fact that, for a bandpass filter $H(f)$ of central frequency f_0 and bandwidth Δf, changing the scale parameter corresponds to explore the frequency axis with a relative bandwidth $\Delta f/f_0$ kept constant.

We have seen how time and frequency resolution tradeoffs underlie the structures of the spectrogram and scalogram. In fact, such structures can be used to compare the two distributions locally in frequency or scale.

If a pure sine wave is analyzed, scale and frequency are

[1]The best time-frequency joint resolution which can be achieved by spectrograms and scalograms is limited by Heisenberg's uncertainty principle. It can be represented by a (time-frequency) resolution cell whose equivalent area is constant but whose shape may vary, depending on the analysis. The figure is intended to provide a symbolic illustration of this fact by representing resolution cells as black rectangular domains, at different locations in the time-frequency plane and for different analysis parameters.

In the case of spectrograms (Figs. 1(a) to (c)), the joint resolution is fixed by the choice of the window length and it is the same at any point in the time-frequency plane. If the window is neither short nor long (Fig. 1(a)), a medium resolution is achieved in both time and frequency. However, if a long window is used (Fig. 1(b)) the spectrogram exhibits a low (respectively, high) resolution in time (respectively, frequency). Conversely, if a short window is used (Fig. 1(c)), the spectrogram exhibits a high (respectively, low) resolution in time (respectively, frequency).

In the case of scalograms (Fig. 1(d)), the joint resolution is frequency dependent, according to the constant-Q structure of the wavelet transform. Time (respectively, frequency) resolution is increased at high (respectively, low) frequencies, with a corresponding decrease of the associated frequency (respectively, time) resolution.

Comparing Figs. 1(a) to (c) on one hand, and Fig. 1(d) on the other hand, it turns out that, at high (respectively, medium or low) frequencies, a scalogram behaves approximately as a spectrogram with a short (respectively, medium or large) window.

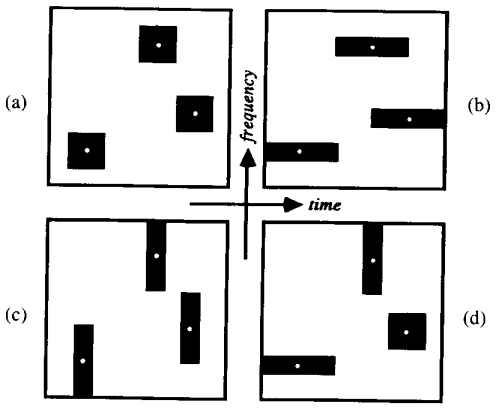


Fig. 1. Compared time-frequency resolution of spectograms and scalograms. (a) Spectrogram with window neither short nor long; (b) Spectrogram with long window; (c) Spectrogram with short window. (d) Scalograms.

are inversely proportional

$$a \equiv f_0/f. \tag{4}$$

It follows that for a general signal, a scalogram approximatively behaves, around a given analyzed frequency f associated to a via (4), as a spectrogram whose frequency resolution, hence Δf, is fitted to f in such a way that $\Delta f/f$ is kept constant:

$$|T_x(t, a = f_0/f)|^2 \cong |F_x(t, f)|^2.$$

The aim of this paper is to find a deeper relationship between spectrograms and scalograms. More precisely, we wish to provide a simple interpretation for both spectrograms and scalograms within a wider general class. To do so, we base our developments on a common property that both share: a bilinear dependence on the analyzed signal.

In fact, an inclusive class for bilinear time-frequency energy distributions has already been defined by Cohen [5]. This general Cohen's class includes all bilinear energy distributions that are covariant under time and frequency shifts, i.e., such that a time (or frequency) delay in the signal accordingly shifts the representation in the time-frequency plane [22]. A general formulation can be written as

$$C_x(t, f; \Pi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_x(\tau, \nu) \Pi(\tau - t, \nu - f)\, d\tau\, d\nu \tag{5}$$

where $\Pi(t, \nu)$ is some arbitrary time-frequency characterization function and where

$$W_x(t, f) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-2j\pi f\tau}\, d\tau$$

is the well-known Wigner–Ville distribution (WVD) [4]. If $\Pi(t, f)$ is a low-pass function in the time-frequency plane, the general class (5) may be considered as composed of smoothed versions of the WVD. More generally,

all members of the Cohen's class result from the WVD by a correlation operation (we have chosen to write (5) as a two-dimensional correlation rather than the usual convolution notation for reasons to be made clear later).

In fact, the Wigner–Ville distribution is not the only time-frequency representation which may serve as the basic distribution from which all others are deduced by correlation as in (5). Other choices include the Rihaczek's distribution [6]

$$R_x(t, f) = x(t)X^*(f)e^{-2j\pi ft}$$

and, more generally, any generalized Wigner distribution [21] parameterized by a real number α

$$W_x^{(\alpha)}(t, f) = \int_{-\infty}^{+\infty} x(t - (\alpha - \tfrac{1}{2})\tau) \, x^*(t - (\alpha + \tfrac{1}{2})\tau)e^{-2j\pi f\tau} \, d\tau. \quad (6)$$

Both the WVD and the Rihaczek's distribution are special cases associated with the particular values $\alpha = 0$ and $\alpha = 1/2$, respectively. We have chosen the WVD in (5) because it has several interesting properties that the others do not share, such as being real-valued, perfectly localized for chirp signals [13] and maximally concentrated in the time-frequency plane [21]. Our choice is justified because whenever Π is low-pass function, (5) is a smoothing operation: hence it is appropriate to choose the basic distribution on which smoothing is applied to as maximally concentrated.

Since the Cohen's class includes all time-frequency energy distributions, it should of course include the spectrogram as a special case. It turns out [4] that the corresponding characterization function Π is the WVD of the analyzing window itself! This yields the remarkable well-known formula

$$|F_x(t, f)|^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(\tau, \nu) \cdot W_h(\tau - t, \nu - f) \, d\tau \, d\nu. \quad (7)$$

which can be extended by replacing WVD's by generalized WVD's (6), or to any member of the Cohen's class for which the two-dimensional Fourier transform $\Phi(\nu, \tau)$ of the characterization function $\Pi(t, f)$ is of modulus one [21] (generalized WVD's (6) correspond to the choice $\Phi_\alpha(\nu, \tau) = e^{2j\pi\alpha\nu\tau}$):

$$\text{if } |\Phi(\nu, \tau)| = 1, \quad \text{then}$$

$$|F_x(t, f)|^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C_x(\tau, \nu; \Pi)C_h^*(\tau - t, \nu - f; \Pi) \, d\tau \, d\nu. \quad (8)$$

This equation shows that structure constraints of spectrograms are such that they can be related to members of the Cohen's class. We now show that something similar happens for scalograms, as a prerequisite for deriving the time-scale counterpart of the Cohen's class in the next section. The new results, counterparts of (7), (8) for scalograms $|T_x(t, a)|^2$, are provided by the following proposition.

Proposition: The scalogram (squared modulus of the wavelet transform) results from the "affine" correlation of the signal WVD by the wavelet WVD:

$$|T_x(t, a)|^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(\tau, \nu)W_h\left(\frac{\tau - t}{a}, a\nu\right) d\tau \, d\nu. \quad (9)$$

More generally, WVD's in (9) can be replaced by any time-frequency energy distribution for which the two-dimensional Fourier transform $\Phi(\nu, \tau)$ of the characterization function $\Pi(t, f)$ is of modulus one and depends on its variables only through their product:

If $\Phi(\nu, \tau)$ is of the form $\varphi(\nu\tau)$ and $|\varphi(\nu\tau)| = 1$, then

$$|T_x(t, a)|^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C_x(\tau, \nu; \Pi)C_h^* \cdot \left(\frac{\tau - t}{a}, a\nu; \Pi\right) d\tau \, d\nu.$$

This shows that scalograms can be characterized from WVD's, or more generally, from members of the Cohen's class, in a way similar to what was done for the spectrogram. This result was first stated in [29] for WVD's, and then generalized by Posch [27] and Flandrin and Rioul [10]. A simple proof is given in Appendix A.

Note that several distributions $C_x(\tau, \nu; \Pi)$ satisfy the two conditions of the proposition, such as the generalized WVD's (6), and other distributions whose $\Phi(\nu, \tau)$ are complex exponentials depending on $\nu|\tau|$ (e.g., Page's distribution [6]) or $|\nu|\tau$.

IV. Time-Scale Energy Distributions

The formalism of the Cohen's class (5) has been widely used for some time [6], [12], [3]. It is indeed attractive because of the following:

1) it allows the recovery of *all* bilinear time-frequency energy distributions as special cases, through proper choices of characterization functions Π;
2) special properties of a given time-frequency distribution can be handled directly via corresponding constraints of their associated characterization functions; and
3) new, specific definitions of time-frequency distributions $C_x(t, \nu; \Pi)$ can be derived simply by first imposing constraints on the distribution, then translating them into requirements to be fulfilled by the characterization function Π.

We now want to define a time-scale counterpart of the Cohen's class, for which the above three points apply. Consider the proposition of Section III in the restrictive case where WVD's are used: mimicking what happens for

time-frequency distributions (spectrograms (7) can be generalized to the Cohen's class (5) by extending W_h to any Π), a simple generalization would be

$$\Omega_x(t; a; \Pi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_x(\tau, \nu) \Pi \left(\frac{\tau - t}{a}, a\nu \right) d\tau \, d\nu. \tag{10}$$

The basic difference with the definition of the Cohen's class (5), of course, is the affine correlation which is used in (10). We have to justify that (10) indeed provides the general formulation of time-scale energy distributions. To do that, consider the affine transformation:

$$x_{t,a}(\tau) = \frac{1}{\sqrt{|a|}} x \left(\frac{\tau - t}{a} \right) \tag{11}$$

where the factor $1/\sqrt{|a|}$ is introduced for normalization purposes. The following theorem is the main result of this paper:

Theorem: Let $\Omega_x(t, a)$ be any bilinear distribution such that

$$\Omega_{x_{\theta,a'}}(t, a) = \Omega_x \left(\frac{t - \theta}{a'}; \frac{a}{a'} \right). \tag{12}$$

Such distributions are said to be covariant to affine transformations and are called time-scale distributions. Then, $\Omega_x(t, a)$ is necessarily of the form (10), *where $\Pi(t, f)$ is an arbitrary time-frequency function.*

The proof is given in Appendix B. This is a powerful result; it shows that any time-scale energy distribution one can imagine always falls in the class (10) and therefore can always be associated to some characterization function Π! This formulation is not only a trivial generalization of (9); it truly characterizes all time-scale distributions.

Also, the formulation (10) clearly reveals the "affine smoothing" concept underlying time-scale distributions. This will be useful for combining time scale and time frequency into a unified perspective. It can be shown that (10) describes a correlation in the affine group [17], [29], and that this correlation cannot be written in the form of a convolution. This justifies our choice to describe both general classes (5) and (10) using correlations. Again, the WVD may be replaced in (10) by any generalized WVD (6), and we have chosen, as in the Cohen's class, to restrict the description to the WVD.

We should insist on the fact that the key point here is not to provide a specific new time-scale representation with special interesting properties; instead, what is important in this approach is the use of the general formalism which results from the theorem. This new class of time-scale distributions (10) is as general as the Cohen's class of time-frequency distributions (5), and include several interesting special examples, which are derived in the sequel.

V. Imposing Properties on Time-Scale Energy Distributions

Similar to Cohen's class (5), the new class of time-scale energy distributions (10) has a very important property: a number of specific requirements imposed on a time-scale distribution can be translated to a specific constraint on the characterization function Π. This will be very useful for deriving special examples of time-scale distributions. We therefore want to list some corresponding pairs of desired properties/constraints on the characterization functions.

However, again similar to Cohen's class, the constraints obtained are sometimes not easily written in terms of the characterization function Π, but rather in terms of Fourier transforms of Π. We therefore first define a few alternative forms to (10).

Weighting Function: If we want to characterize the class (10) with the help of the two-dimensional Fourier transform $\Phi(\nu, \tau)$ of the characterization function $\Pi(t, f)$, the equivalent formulation reads

$$\Omega_x(t, a; \Pi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi \left(a\nu, \frac{\tau}{a} \right) A_x(\nu, \tau) e^{-2j\pi\nu t} \, d\nu \, d\tau$$

where $\Phi(\nu, \tau)$ is referred to as the *weighting function*, and where

$$A_x(\nu, \tau) = \int_{-\infty}^{+\infty} x \left(t + \frac{\tau}{2} \right) x^* \left(t - \frac{\tau}{2} \right) e^{2j\pi\nu t} \, dt$$

is the (narrow-band) *ambiguity function*, defined as the inverse 2D Fourier transform of the WVD.

Bifrequency Kernels: Assume the domain in which the arbitrary function characterizing the distribution is expressed should correspond to two frequency parameters in both dimensions, then the equivalent formulation to (10) is

$$\Omega_x(t, a; \Pi) = \frac{1}{|a|} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \psi(\nu, f) X \left(\frac{1}{a} \left(f - \frac{\nu}{2} \right) \right) \cdot X^* \left(\frac{1}{a} \left(f + \frac{\nu}{2} \right) \right) e^{-2j\pi(t/a)\nu} \, d\nu \, df \tag{13}$$

where the bifrequency characterization function ψ is defined as

$$\psi(\nu, f) = \int_{-\infty}^{+\infty} \Pi(t, f) e^{-2j\pi\nu t} \, dt. \tag{14}$$

We now list a few useful pairs of desired properties/corresponding constraints on the characterization functions Π, Φ, or ψ.

Energy: The term "energy distribution" is justified by the following equality:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi) \frac{dt \, da}{a^2} = \left[\int_{-\infty}^{+\infty} \psi(0, f) \frac{df}{|f|} \right] E_x \tag{15}$$

with ψ as in (14). This means that energy is properly spread over the time-scale plane if the quantity into brackets in the right-hand side of (15) is unity.

Marginal in Frequency: The spectral energy density of x is recovered from the marginal in frequency, i.e.,

$$\int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi)\, dt = \left|X\left(\frac{f_0}{a}\right)\right|^2 \tag{16}$$

as long as $\Phi(0, \tau) = e^{-2j\pi f_0 \tau}$.

Marginal in Time: Similarly, the instantaneous power of x is obtained as time marginal, i.e.,

$$\int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi)\, \frac{da}{a^2} = |x(t)|^2$$

as long as

$$\int_{-\infty}^{+\infty} \Phi\left(a\nu, \frac{\tau}{a}\right) \frac{da}{a^2} = \delta(\tau)$$

for any ν.

Moyal-Type Formula: Finally, a Moyal-type formula relating inner products of signals and distributions may be obtained as

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi) \Omega_y^*(t, a; \Pi)\, \frac{dt\, da}{a^2} = \left|\int_{-\infty}^{+\infty} x(t) y^*(t)\, dt\right|^2 \tag{17}$$

if

$$\int_{-\infty}^{+\infty} \Phi\left(a\nu, \frac{\tau}{a}\right) \Phi^*\left(a\nu, \frac{\tau'}{a}\right) \frac{da}{a^2} = \delta(\tau - \tau'), \qquad \text{for any } \nu. \tag{18}$$

These correspondences between properties of the representation and constraints on characterization functions are used, among others, in the next section which derives a few examples of time-scale representations.

VI. Some Examples of the Time-Scale Energy Distributions

We have just seen that, in a way similar to Cohen's class, specific structures of characterization functions (Φ, Π, or ψ) permit obtaining specific properties of the associated time-scale distributions. As a result, they allow the derivation of specific definitions as special cases of time-scale energy distributions. The design procedure is very simple: 1) list the different desired properties of the distribution, 2) look up the corresponding constraints that the characterization function Π must meet, and 3) deduce the particular form of the distribution. Note that it may happen that several constraints, when taken together, admit no solution. In this case some properties have to be given up in step 1. We now review several examples of design.

A. Product Kernels Allowing the Identification "Time-Scale = Time-Frequency"

In this section we want to determine the conditions such that a time-scale distribution can equivalently be seen as a time-frequency distribution.

Just as the WT uses bandpass filters, the smoothing function Π is preferably chosen to be bandpass as a function of frequency. Define Π_0 such that

$$\Pi(t, f) = \Pi_0(t, f - f_0) \Leftrightarrow \Phi(\nu, \tau) = \Phi_0(\nu, \tau) e^{-2j\pi f_0 \tau}$$

where f_0 is some nonzero frequency. Using this notation, an interesting identification between time-scale and time-frequency distributions may be found, provided that the weighting function $\Phi_0(\nu, \tau)$ depends only on the product $\nu\tau$:

$$\text{If } \Phi_0(\nu, \tau) = \varphi(\nu\tau), \qquad \text{then}$$

$$\Omega_x(t, a; \Pi) = C_x\left(t, \frac{f_0}{a}; \Pi_0\right). \tag{19}$$

As a special case, it can easily be checked that

$$\Pi(t, f) = \delta(t)\, \delta(f - f_0) \qquad \text{implies}$$

$$\Omega_x(t, a; \Pi) = W_x\left(t, \frac{f_0}{a}\right) \tag{20}$$

which means that the usual WVD can be recovered as the "no smoothing" limit case, with the identification "frequency = inverse of scale."

The condition under which the identification (19) holds is met by numerous distributions. In addition to the class of generalized Wigner distributions (6), we can mention the Choï–Williams' distribution [6], which has recently received special attention and which is associated to the choice $\Phi_0(\nu, \tau) = e^{-(\nu\tau)^2/\sigma}$, where σ is a positive parameter.

B. Scalograms

A simple example of time-scale energy distributions is the scalogram which, according to (9), can be seen as the affine smoothing of the WVD of the analyzed signal by the WVD of the analyzing wavelet [29]

$$|T_x(t, a)|^2 = \Omega_x(t, a; W_h).$$

Because the smoothing function associated to the scalogram is itself a WVD, it cannot be perfectly concentrated in both time and frequency. As a result, several properties, such as marginals, are lost. We get for instance

$$\int_{-\infty}^{+\infty} |T_x(t, a)|^2\, dt = \int_{-\infty}^{+\infty} \left|X\left(\frac{f}{a}\right)\right|^2 |H_0(f - f_0)|^2\, df$$

where we have noted $H(f) = H_0(f - f_0)$ to make explicit the dependence of the bandpass filter transfer function on its central frequency f_0. This has to be compared to (16): clearly, a good estimation of the spectral energy density would be obtained with a highly frequency-selec-

tive analyzing wavelet. This corresponds necessarily to a large amount of time spreading and, hence, to some loss in time resolution.

Also note that the admissibility condition for scalograms (3) is easily recovered from the energy condition (15).

C. Bertrands' Class and Localized Bifrequency Kernels

Several researchers, some of them working outside the area of signal processing, have considered time-frequency energy distributions aimed at "broad-band" signals [2], [19], [31]. For example, Bertrands' approach [2] assumes, as in the theorem of Section IV, the basic covariance requirement with respect to affine transformations (12) applied to bilinear forms of the signal, yet the parameterization they retain is different. Rewritten in time-scale terms, this latter reads

$$P_x(t, a; K_B) = \frac{1}{|a|}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} K_B(\nu, f) X\left(\frac{1}{a}\nu\right) \cdot X^*\left(\frac{1}{a}f\right) \exp\left[-2j\pi(t/a) \cdot (f-\nu)\right] d\nu\, df \tag{21}$$

where a bifrequency kernel $K_B(\nu, f)$ appears. Consequently, the result (21) is identical to the frequency-domain formulation (13) of the general class (10), up to the change of variables

$$\psi(\nu, f) = K_B\left(f - \frac{\nu}{2}, f + \frac{\nu}{2}\right).$$

We believe that our approach, although quite similar to Bertrands', allows a more simple interpretation of (21) as a natural affine counterpart of the Cohen's class (see (10)). Moreover, it does not necessitate any *a priori* distinction between the "narrow-band" or "broad-band" nature of the analyzed signal.

As an example, we now derive a design procedure which can be applied to the definitions occurring in [2], [19], [31]. Define a subclass of (13) (or (21)) consisting in characterization functions which are perfectly localized on some curve $f = F(\nu)$ in their bifrequency representation:

$$\psi(\nu, f) = G(\nu)\,\delta(f - F(\nu)) \Leftrightarrow \Phi(\nu, \tau) = G(\nu)e^{-2j\pi F(\nu)\tau} \tag{22}$$

where $G(\nu)$ is an arbitrary function. The associated time-scale distributions, with localized bifrequency kernels then read

$$\Omega_x(t, a; \Pi) = \frac{1}{|a|}\int_{-\infty}^{+\infty} G(\nu) X\left(\frac{1}{a}\left[F(\nu) - \frac{\nu}{2}\right]\right) \cdot X^*\left(\frac{1}{a}\left[F(\nu) + \frac{\nu}{2}\right]\right) e^{-2j\pi(t/a)\nu}\, d\nu. \tag{23}$$

Specifying

$$G(\nu) = \frac{(\nu/2)}{\sinh(\nu/2)}; \qquad F(\nu) = (\nu/2)\coth(\nu/2) \tag{24}$$

allows recovering a particular distribution which was often used by the Bertrands [2]

$$B_x(t, a) = \frac{1}{|a|}\int_{-\infty}^{+\infty} \frac{(\nu/2)}{\sinh(\nu/2)} X\left(\frac{1}{a}\frac{(\nu/2)e^{-(\nu/2)}}{\sinh(\nu/2)}\right) \cdot X^*\left(\frac{1}{a}\frac{(\nu/2)e^{+(\nu/2)}}{\sinh(\nu/2)}\right) e^{-2j\pi(t/a)\nu}\, d\nu.$$

Other choices [14] yield distributions used in [19], [31]. As an exercise, we have recovered the above specific definition associated to (22)–(24) using the design procedure mentioned previously: this is detailed in Appendix C which obtains (24) starting from a localized bifrequency kernel and imposing several *a priori* requirements (time-localization and a Moyal-type formula).

D. Separable Kernels and Affine Smoothed Wigner–Ville

It is known [11] that the tradeoff underlying the time and frequency behaviors of the spectrogram can be overcome if we replace the associated WVD smoothing by a smoothing function which is separable in time and frequency

$$\Pi_0(t, f) = g(t)H_0(f).$$

The resulting distribution (called the smoothed pseudo-WVD) reads

$$C_x(t, f; \Pi_0) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(\tau, \nu) g(\tau - t) H_0(\nu - f)\, d\tau \cdot d\nu \tag{25}$$

$$= \int_{-\infty}^{+\infty} h_0(\tau)\left[\int_{-\infty}^{+\infty} g(\theta - t) x\left(\theta + \frac{\tau}{2}\right) \cdot x^*\left(\theta - \frac{\tau}{2}\right) d\theta\right] e^{-2j\pi f\tau}\, d\tau. \tag{26}$$

This offers a great versatility for balancing, e.g., time-frequency resolution and cross-terms reduction [11], although this is necessarily obtained at the expense of the loss of other properties such as marginals.

We propose a similar approach for time-scale distributions and define the affine smoothed WVD by taking

$$\Pi(t, f) = \Pi_0(t, f - f_0) = g(t)H_0(f - f_0).$$

The resulting definition is

$$\Omega_x(t, a; \Pi) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(\tau, \nu) g\left(\frac{\tau - t}{a}\right) \cdot H_0(a\nu - \nu_0)\, d\tau\, d\nu. \tag{27}$$

For practical computations, an equivalent form, which parallels that of (26) may be preferred:

$$\Omega_x(t, a; \Pi) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{|a|}} h_0\left(\frac{\tau}{a}\right) \left[\int_{-\infty}^{+\infty} \frac{1}{\sqrt{|a|}} g\left(\frac{\theta - t}{a}\right) \cdot x\left(\theta + \frac{\tau}{2}\right) x^*\left(\theta - \frac{\tau}{2}\right) d\theta \right] \cdot \exp\left[-2j\pi (f_0/a)\tau \right] d\tau.$$

The behavior of affine smoothed WVD's is illustrated for particular examples in the next section.

VII. From Wigner–Ville to Spectrograms or Scalograms

We have seen in Section III that on one hand, the spectrogram is a time-frequency distribution obtained from the WVD by smoothing and that on the other hand, the scalogram is a time-scale distribution also obtained from the WVD by affine smoothing. The WVD is, therefore, at the intersection of both classes of time-frequency and time-scale distributions. Note that we have already determined this intersection in Section VI-A, and indeed we have seen that the WVD can be seen as the no-smoothing case of both classes, allowing the simple identification "scale = inverse of frequency" (20).

The aim of this section is to fill the gap between the "unsmoothed" WVD and the spectrogram or the scalogram through a continuous transition governed by the characterization functions Π acting on the WVD. To do that, it is convenient to parameterize Π is such a way that we control separately the smoothing in time and in frequency. As a result, the smoothing functions are separable, and their equivalent area vary from zero ("unsmoothed" WVD) to the a limit fixed by the Heisenberg's uncertainty principle [9] (spectrogram or scalogram). This choice corresponds to using smoothed-pseudo WVD's (25) or affine smoothed WVD's (27).

The following proposition shows that this can be achieved exactly only using Gaussian smoothing.

Proposition: A continuous passage from Wigner–Ville to the spectrogram or scalogram using separable smoothing functions requires the latter to be Gaussian:

$$\Pi(t, f) = \frac{\sqrt{\alpha\beta}}{\pi} e^{-\alpha t^2} e^{-\beta(f - f_0)^2}$$

where α, β and f_0 are positive parameters.

The proof is given in Appendix D.

The important parameter here is the equivalent area BT $= 2\pi/\sqrt{\alpha\beta}$, which runs from 0 (WVD) to 1 (spectrogram/scalogram) and truly controls both transitions. The role of α and β for fixed $\alpha\beta$ does not influence the joint resolution of the analysis but rather modifies the individual time and frequency resolutions. Fig. 2 illustrates such transitions by providing several analyses of a synthetic signal for different values of BT.

In practical implementations of spectrograms, scalograms, or smoothed WVD's, the smoothing functions have no reason to be strictly Gaussian. However, the proposition gives a good idea of what is happening for smoothing functions commonly used in signal analysis: instead of looking at the two extreme representations (spectrogram and scalogram) separately, a deeper insight can be gained by considering a whole continuum between these two extremes, with the WVD as a necessary intermediate step. Moreover, as mentioned in Section VI-D, the transition permits to trading off joint resolutions and interferences reduction (see Fig. 2).

Such a general framework, along with its interpretation, allows us to think that by suitably controlling only the BT parameter, we can adapt the analysis tool to various situations occurring in different applications of signal analysis.

VIII. Conclusion

The material presented in this paper is based on similar properties of two notions that are both defined by covariance requirements: 1) *local frequency*, covariant under modulations (or Fourier-frequency shifts); 2) *time scaling* (or frequency scaling by Fourier duality), covariant under dilations or contractions.

This similarity is evidenced by describing general time-frequency and time-scale energy distributions in a unified way as a result of some 2D correlation acting on the WVD. The WVD thus appears to play a central role in both analyses since it is (among a few other members of the Cohen's class, see Section III) covariant under frequency shifts as well as under scale dilations.

We have thereby derived a large class of time-scale and time-frequency representations, on which many possible (and sometimes, exclusive) properties may be imposed. We have studied several specific requirements (such as energy normalization, time marginal, etc.) and associated parameterizations of the representation. There is obviously a great versatility for the choice of representations, which may be appropriate for various applications. This suggests the analysis tool has to be carefully designed in order to express particular needs: starting from the most general formulation, one can, for instance, build a subset of time-scale energy representations, suitable for a given application, by imposing specific requirements. Controlling a few parameters on this set of representations should help in many ways, e.g., for determining which representation best reveals a given time-scale signature.

This is well illustrated by the last result presented in this paper: a continuous transition from the WVD to either spectrograms or scalograms is possible, and permits balancing time-frequency resolution and cross-terms reduction in the time-scale representation, in a similar (but different) way as for the corresponding time-frequency representation. In light of this, we recommend that various properties of time-frequency and time-scale methods

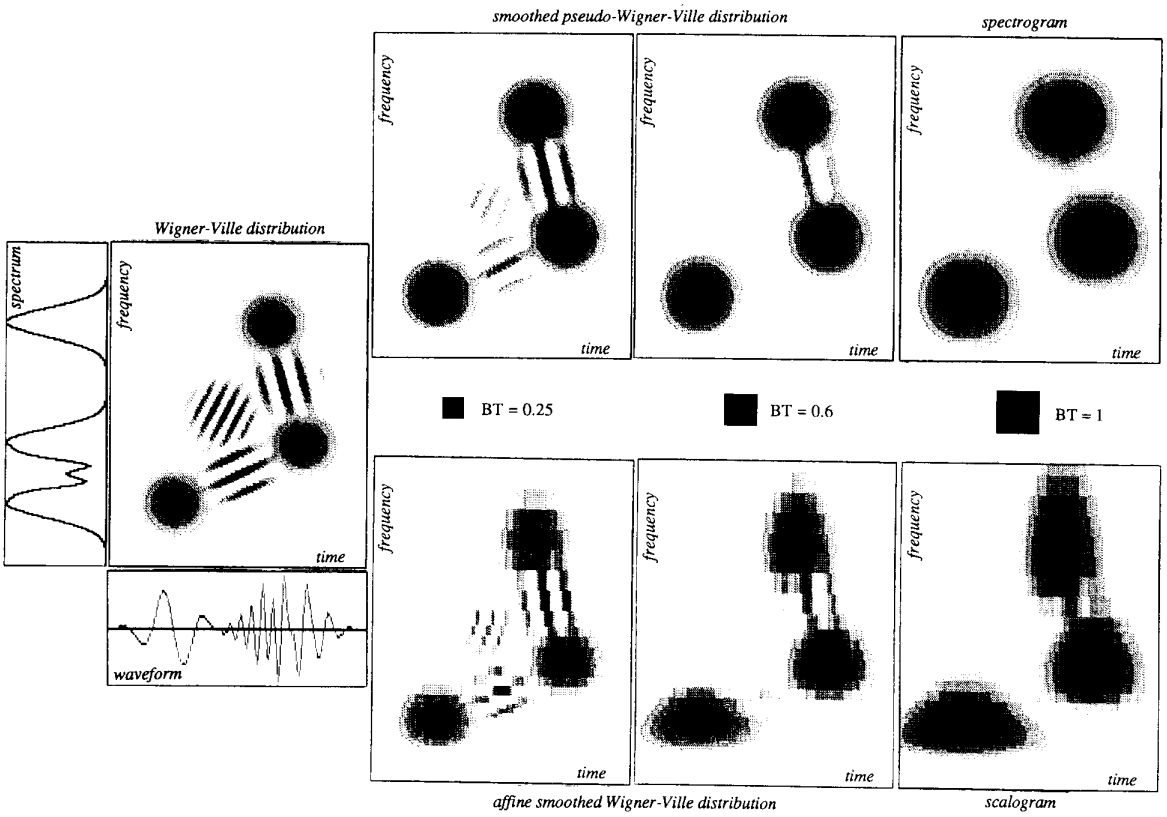


Fig. 2. From Wigner-Ville to spectrograms or scalograms. A continuous transition from Wigner-Ville to spectrograms (respectively, scalograms) is possible via smoothed pseudo-Wigner-Ville (respectively, affine smoothed Wigner-Ville) distributions, when they are based on separable Gaussian functions. The area of the associated resolution cells can be measured by a dimensionless parameter BT, which varies from zero ("unsmoothed" Wigner-Ville) to 1 (spectrogram or scalogram). The figure presents an example of such transitions in the case of a situation resembling the one used symbolically in Fig. 1.

A synthetic signal consisting of three Gaussian wave packets (with different central times and different central frequencies) has been generated. The time-domain waveform and its spectrum (modulus of the Fourier transform) are plotted along the time and frequency axes of the corresponding Wigner-Ville distribution. Different analyses from the WVD to the spectrogram or the scalogram (both associated to BT = 1) via smoothed pseudo-Wigner-Ville and affine smoothed Wigner-Ville (both with BT = 0.25 and BT = 0.6) are presented and the corresponding resolution cells are represented as black rectangular domains in the center of the diagram.

be compared keeping in mind that both result from a smoothing operation acting on the WVD. The difference is related to the nature of the smoothing operation used: time-frequency or affine (time-scale) smoothing.

Appendix A
Scalograms as Affine-Smoothed Wigner-Ville

It is convenient to introduce 2D Fourier transformations in (5). Changing variables accordingly yields a dual characterization

$$C_x(t, f; \Pi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(\nu, \tau) A_x(\nu, \tau) \cdot \exp\left[-2j\pi(\nu t + \tau f)\right] d\nu\, d\tau \quad \text{(A1)}$$

where Φ is the 2D Fourier transform of the characterization function Π. One has

$$C_h\left(\frac{\tau - \theta}{a}, af; \Pi\right) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(\nu, \tau) A_h(\nu, \tau) \cdot \exp\left[-2j\pi(\nu(t - \theta)/a + \tau a f)\right] d\nu\, d\tau$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi\left(a\nu, \frac{\tau}{a}\right) \left[A_h\left(a\nu, \frac{\tau}{a}\right) e^{2j\pi\nu\theta}\right] \cdot \exp\left[-2j\pi(\nu t + \tau f)\right] d\nu\, d\tau.$$

It can be easily checked that the quantity into brackets corresponds to the ambiguity function of $h(t)$ after action of the affine transformation (11). Therefore we obtain

$$C_h\left(\frac{t - \theta}{a}, af; \Pi\right) = C_{h_{\theta,a}}(t, f; \Pi)$$

provided that $\Phi(a\nu, \tau/a) = \Phi(\nu, \tau)$ for any a [27], [10].

Any weighting function satisfying this is the function of the product of its variables. On the other hand, Moyal's formula [4] guarantees that, for any two finite energy signals,

$$\left|\int_{-\infty}^{+\infty} x(t)y^*(t)\,dt\right|^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(t,f)W_y(t,f)\,dt\,df = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} A_x(\nu,\tau)A_y^*(\nu,\tau)\,d\nu\,d\tau. \tag{A2}$$

Using (A1) and applying Parseval's relation, we easily obtain

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C_x(t,f;\Pi)C_y^*(t,f;\Pi)\,dt\,df = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |\Phi(\nu,\tau)|^2 A_x(\nu,\tau)A_y^*(\nu,\tau)\,d\nu\,d\tau$$

and therefore a generalized Moyal's formula holds whenever the weighting function has modulus one [21]. This is true especially if $y(t)$ is chosen as $h_{\tau,a}(t)$ (see (11)). In such a case, the left-hand side of (A2) identifies to the scalogram, which completes the proof.

Appendix B
General Formulation of Time-Scale Energy Distributions

Assume the affine covariance requirement

$$\Omega_{h_{\theta,a'}}(t,a) = \Omega_x\left(\frac{t-\theta}{a'};\frac{a}{a'}\right) \tag{B1}$$

(where $h_{\theta,a'}$ is defined as in (11)) is imposed to a bilinear distribution Ω_x:

$$\Omega_x(t,a) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} K(u,u';t,a)x(u)x^*(u')\,du\,du'.$$

This yields the condition

$$a'K(a'u+\theta, a'u'+\theta; t, a) = K\left(u,u';\frac{t-\theta}{a'},\frac{a}{a'}\right)$$

for any a', any θ, and any a, or, equivalently,

$$K(u,u';t,a) = \frac{1}{a}K\left(\frac{u-\theta}{a'},\frac{u'-\theta}{a'},\frac{t-\theta}{a'},\frac{a}{a'}\right).$$

Fix $\theta = t$ and $a' = a$. We obtain

$$K(u,u';t,a) = \frac{1}{a}K\left(\frac{u-t}{a},\frac{u'-t}{a};0,1\right)$$

and the distribution takes the form

$$\begin{aligned}\Omega_x(t,a) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{a}K\left(\frac{u-t}{a},\frac{u'-t}{a};0,1\right)\cdot x(u)x^*(u')\,du\,du'.\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{a}K\left(\frac{\theta-t}{a}+\frac{\tau}{2a},\frac{\theta-t}{a}-\frac{\tau}{2a};0,1\right)x\left(\theta+\frac{\tau}{2}\right)x^*\left(\theta-\frac{\tau}{2}\right)d\theta\,d\tau\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(\tau,\nu)\Pi\left(\frac{\tau-t}{a},a\nu\right)d\tau\,d\nu\end{aligned}$$

with the smoothing function Π defined as

$$\Pi(t,f) = \int_{-\infty}^{+\infty} K\left(t+\frac{\tau}{2},t-\frac{\tau}{2};0,1\right)e^{2j\pi f\tau}\,d\tau.$$

Appendix C
Derivation of Bertrands' Distribution

Assume our time-scale distributions are characterized by a kernel function whose bifrequency representation is localized according to (22). Further specifications of G and F can be obtained by imposing specific requirements to the corresponding distribution.

1) Time Localization: Imposing time localization (in the Bertrands' sense [2])

$$X(f) = \frac{1}{\sqrt{|f|}}e^{-2j\pi ft_0} \Rightarrow \Omega_x(t,a;\Pi_\delta) = |a|\,\delta(t-t_0)$$

yields

$$\Omega_x(t,a;\Pi_\delta) = |a|\int_{-\infty}^{+\infty}\frac{G(a\nu)}{\sqrt{F^2(a\nu)-(a\nu/2)^2}}e^{2j\pi\nu(t-t_0)}\,d\nu$$

and, hence, the condition

$$G^2(\nu) = F^2(\nu) - (\nu/2)^2. \tag{C1}$$

2) Moyal-Type Formula: If we impose the condition (18) for the Moyal-type formula (17) to hold, we obtain

$$\begin{aligned}\int_{-\infty}^{+\infty} f\left(a\nu,\frac{\tau}{a}\right)f^*\left(a\nu,\frac{\tau'}{a}\right)\frac{da}{a^2} &= \int_{-\infty}^{+\infty} G^2(a\nu)\exp\left[2j\pi\frac{F(a\nu)}{a}(\tau-\tau')\right]\frac{da}{a^2}\\ &= \int_{-\infty}^{+\infty}\frac{G^2(a\nu)}{a^2\frac{d}{da}\left(\frac{F(a\nu)}{a}\right)}\cdot\exp\left[2j\pi\frac{F(a\nu)}{a}(\tau-\tau')\right]d\left(\frac{F(a\nu)}{a}\right).\end{aligned}$$

Therefore, (18) is satisfied if

$$G^2(\nu) = F(\nu) - \nu\frac{dF}{d\nu}(\nu). \tag{C2}$$

The simultaneous requirements of time localization (C1) and Moyal-type formula (C2) lead to the differential equation

$$F(\nu) - \nu \frac{dF}{d\nu}(\nu) = F^2(\nu) - (\nu/2)^2$$

which is written, using the auxiliary functions $U(\nu) = F(\nu) - (\nu/2)$, $V(\nu) = F(\nu) + (\nu/2)$, and $W(\nu) = U(\nu)/V(\nu)$, as $U(\nu)\,(dV/d\nu)(\nu) - V(\nu)\,(dU/d\nu)(\nu) = U(\nu)\,V(\nu)$, i.e., as $(dW/d\nu)(\nu) = W(\nu)$. The solution is

$$W(\nu) = \frac{F(\nu) - (\nu/2)}{F(\nu) + (\nu/2)} = ce^{\nu}.$$

From this it follows that $W(0) = 1$. This implies $c = 1$ which, in turn, implies $F(\nu) = (\nu/2)\coth(\nu/2)$. Substituting this into (C2) yields

$$G^2(\nu) = \left((\nu/2)\frac{e^{\nu}+1}{e^{\nu}-1} - (\nu/2)\right) \cdot \left((\nu/2)\frac{e^{\nu}+1}{e^{\nu}-1} + (\nu/2)\right) = \frac{\nu^2 e^{\nu}}{e^{\nu}-1}$$

hence $G(\nu) = (\nu/2)/(\sinh(\nu/2))$, which completes the proof.

Appendix D
From Spectrograms to Scalograms via Wigner–Ville, Using Gaussian Smoothing

For a sake of convenience, introduce the notation

$$H(f) = H_0(f - f_0)$$

which allows us to consider (25) and (27) in a common framework, with either H_0 for the smoothed pseudo-WVD or H for the affine smoothed WVD. If both spectrograms and scalograms are supposed to be attainable through separable kernels, then their associated smoothing function, which is a WVD, must necessarily be itself a separable function of time and frequency. However, if we impose to a WVD to be separable, e.g.,

$$W_x(t, f) = g(t)H_0(f) \tag{D1}$$

we readily obtain

$$|X(f)|^2 = \int_{-\infty}^{+\infty} W_x(t, f)\,dt = G(0)H_0(f)$$

and

$$|x(t)|^2 = \int_{-\infty}^{+\infty} W_x(t, f)\,df = g(t)h_0(0).$$

Therefore, a separable WVD is necessarily of the form

$$W_x(t, f) = \frac{|x(t)|^2 |X(f)|^2}{G(0)h_0(0)}. \tag{D2}$$

From (D1), it follows that

$$G(0)h_0(0) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(t, f)\,dt\,df = E_x \geq 0$$

and, hence (D2) is a nonnegative quantity:

$$W_x(t, f) \geq 0.$$

This means that, if separable WVD's exist, they are necessarily everywhere nonnegative [25]. The nonnegativity condition being imposed, we know from Hudson's theorem [20] that the only signals which are admissible are exponentials of quadratic forms in t (with possibly complex-valued coefficients) such that

$$x(t) = e^{-(\alpha t^2 + \beta t + \gamma)}, \qquad \mathrm{Re}\,\{\alpha\} > 0.$$

However, since separability is imposed too, no coupling between time and frequency is allowed, which restricts the class of solutions to

$$x(t) = (2\alpha/\pi)^{1/4} e^{-\alpha t^2} e^{-2j\pi\zeta t} e^{i\psi} \Rightarrow W_x(t, f) = 2e^{-2\alpha t^2} \exp\left[-(2\pi^2/\alpha)(f - \zeta)^2\right]$$

where α and ζ are real-valued, ψ is a pure phase factor, and the normalization has been chosen for ensuring energy conservation. Therefore, a suitable choice of separable smoothing functions which allows a continuous passage from Wigner–Ville to spectrograms or spectrograms is of the form of a (normalized) product of Gaussians, i.e.,

$$\Pi(t, f) = \frac{\sqrt{\alpha\beta}}{\pi} e^{-\alpha t^2} e^{-\beta(f - f_0)^2}.$$

This completes the proof.

Acknowledgment

The authors are indebted to T. Doligez and B. Vidalie who developed the software used for producing the pictures in Fig. 2. Useful discussions, at an early stage of this work, with I. Daubechies and T. Paul are also greatfully acknowledged. Finally, the authors would like to thank the anonymous reviewers for helpful remarks and comments.

References

[1] J. B. Allen and L. R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558-1564, 1977.

[2] J. Bertrand and P. Bertrand, "Time-frequency representations of broad-band signals," in *Proc. 1988 IEEE Int. Conf. Acoust., Speech, Signal Processing* (New York, NY), Apr. 11-14, 1988, pp. 2196-2199; also in J. M. Combes, A. Grossman, and P. Tchamitchian, Eds., *Wavelets, Time-Frequency Methods, and Phase Space*. Berlin: Springer, 1989, pp. 164-171, 1989.

[3] H. I. Choï and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 6, pp. 862-871, 1989.

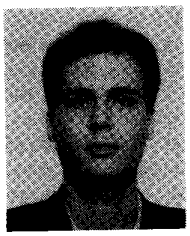
[4] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution—a tool for time-frequency signal analysis. Part I: Continuous-time signals," *Philips J. Res.*, vol. 35, no. 3, pp. 217-250, 1980.

[5] L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7, no. 5, pp. 781-786, 1966.

[6] L. Cohen, "Time-frequency distribution—a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941-981, 1989.

[7] I. Daubechies, "Orthonormal bases of wavelets with finite support—

connection with discrete filters," in *Wavelets, Time-Frequency Methods, and Phase Space*, J. M. Combes, A. Grossman, and P. Tchamitchian, Eds. Berlin: Springer, 1989, pp. 38-66.
[8] I. Daubechies, "The wavelet transform, time-frequency localization, and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 961-1005, Sept. 1990.
[9] N. G. de Bruijn, "Uncertainty principles in Fourier analysis," in *Inequalities II*, O. Shisha, Ed. New York: Academic, 1967.
[10] P. Flandrin and O. Rioul, "Affine smoothing of the Wigner-Ville distribution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing ICASSP-90* (Albuquerque, NM), 1990, pp. 2455-2458.
[11] P. Flandrin, "Some features of time-frequency representations of multicomponent signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing ICASSP-84* (San Diego, CA), 1984, pp. 41B.4.1-41B.4.4.
[12] P. Flandrin, "Time-frequency and time-scale," in *Proc. IEEE 4th ASSP Workshop Spectrum Estimation Modeling* (Minneapolis, MN), 1988, pp. 77-80.
[13] P. Flandrin and B. Escudié, "Sur la Localisation des Représentations Conjointes dans le Plan Temps-Fréquence," *C. R. Acad. Sci.*, série I, tome 295, pp. 475-478, 1982 (in French).
[14] P. Flandrin, "Sur une Classe Générale d'Extensions Affines de la Distribution de Wigner-Ville," Int. Rep. ICPI-TS 9001, Jan. 1990 (in French).
[15] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.*, vol. 93, pp. 429-457, 1946.
[16] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related tranforms in seismic signal analysis," *Geoexploration*, vol. 23, pp. 85-102, 1984.
[17] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. Anal.*, vol. 15, no. 4, pp. 723-736, 1984.
[18] A. Grossmann, R. Kronland-Martinet, and J. Morlet, "Reading and understanding continuous wavelet transforms," in *Wavelets, Time-Frequency Methods, and Phase Space*, J. M. Combes, A. Grossman, and P. Tchamitchian, Eds. Berlin: Springer, 1989, pp. 2-20.
[19] A. Grossmann, personnal communication.
[20] R. L. Hudson, "When is the Wigner quasi-probability density non-negative?" *Rep. Math. Phys.*, vol. 6, no. 2, pp. 249-252, 1974.
[21] A. J. E. M. Janssen, "On the locus and spread of pseudodensity functions in the time-frequency plane," *Philips J. Res.*, vol. 37, no. 3, pp. 79-110, 1982.
[22] J. G. Krüger and A. Poffyn, "Quantum mechanics in phase space, I. Unicity of the Wigner distribution function," *Physica*, vol. 85A, pp. 84-100, 1976.
[23] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, no. 7, pp. 674-693, 1989.
[24] Y. Meyer, "Orthonormal wavelets," in J. M. Combes, A. Grossman, and P. Tchamitchian, Eds., *Wavelets, Time-Frequency Methods, and Phase Space*. Berlin: Springer, 1989, pp. 21-37.
[25] C. Nahum, "Etude de la Transformation de Wigner," Mémoire ENST-SYC, 1984 (in French).
[26] S. N. Nawab and T. F. Quatieri, "Short-time Fourier transform," in *Advanced Topics in Signal Processing*, J. S. Lim and A. V. Oppenheim, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[27] T. Posch, "Wavelet transform and time-frequency distributions," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1152, pp. 477-482, 1988.
[28] O. Rioul and M. Vetterli, "Wavelet transforms in signal processing," *IEEE SP Mag.*, 1991.
[29] O. Rioul, "Wigner-Ville representations of signals adapted to shifts and dilations," Tech. Memo. 112277-880422-03-TM, AT&T Bell Labs, 1988.
[30] O. Rioul, "A unifying multiresolution theory for the discrete wavelet transform, regular filter banks, and pyramid transforms," Int. Rep. CRPE-188, CNET/RPE/ETP; also *IEEE Trans. Signal Processing*, submitted for publication.
[31] A. Unterberger, "The calculus of pseudodifferential operators of Fuchs type," *Commun. Part. Diff. Eq.*, vol. 9, no. 12, pp. 1179-1236, 1984.
[32] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Processing*, submitted for publication.
[33] J. M. Combes, A. Grossman, and P. Tchamitchian, Eds., *Wavelets, Time-Frequency Methods, and Phase Space*. Berlin: Springer, 1989.



Olivier Rioul was born in Strasbourg, France, on July 4, 1964. He received the diploma degrees in electrical engineering from the Ecole Polytechnique, Palaiseau, France, and from Telecom University, Paris, in 1987 and 1989, respectively. He is currently working toward the Ph.D. degree in signal processing at Télécom University. His areas of special interest are wavelet theory, image coding, and fast signal processing algorithms.

Patrick Flandrin (M'85) was born in Bron, France, on June 2, 1955. He received the Ingénieur degree from Institut de Chimie et Physique Industrielles de Lyon in 1978, the Docteur-Ingénieur degree and the Doctorat d'Etat és Sciences Physique from 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 Chargé de Recherches. Until 1990, he was with the Laboratoire de Traitement du Signal at ICPI Lyon, and was head of this department from 1987 to 1989. He is now with the Laboratoire de Physique at Ecole Normale Supérieure de Lyon, in charge of the signal processing group. He is also leader of a working group within the CNRS cooperative structure "GdR Traitement du Signal et Images." His research interests include mainly 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.

Dr. Flandrin was awarded the Philip Morris Scientific Prize in Mathematics in 1991. He is a member of EURASIP and SEE.