Time Frequency Signal Analysis:
Past, present and future trends

Boualem Boashash

Signal Processing Research Centre
Queensland University of Technology
2 George street, Brisbane, Qld. 4000, Australia

1 Introduction

This chapter is written to provide both an historical review of past work and an overview of recent advances in time-frequency signal analysis (TFSA). It is aimed at complementing the texts which appeared recently in [1], [2], [3] and [4].

The chapter is organised as follows. Section 1 discusses the need for time-frequency signal analysis, as opposed to either time or frequency analysis. Section 2 traces the early theoretical foundations of TFSA, which were laid prior to 1980. Section 3 covers the many faceted developments which occurred in TFSA in the 1980's and early 1990's. It covers bilinear or *energetic* time-frequency distributions (TFDs). Section 4 deals with a generalisation of bilinear TFDs to multilinear *Polynomial TFDs.* Section 5 provides a coverage of the *Wigner-Ville trispectrum*, which is a particular polynomial TFD, used for analysing Gaussian random amplitude modulated processes. In Section 6, some issues related to multicomponent signals and time-varying polyspectra are addressed. Section 7 is devoted to conclusions.

1.1 An heuristic look at the need for time-frequency signal analysis

The field of time-frequency signal analysis is one of the recent developments in Signal Processing which has come in response to the need to find suitable tools for analysing non-stationary signals. This chapter outlines many of the important concepts underpinning TFSA, and includes an historical perspective of their development. The chapter utilises many concepts and results that were originally reported in [1], [2], [3] and [4] [5], [6], [7].

CONTROL AND DYNAMICS SYSTEMS, VOL. 78

The drawbacks of classical spectral analysis [8], [9], [10], [11], [12], [13] arise largely due to the fact that its principal analysis tool, the Fourier transform, implicitly assumes that the spectral characteristics of the signal are time-invariant, while in reality, signals both natural and man-made, almost always exhibit some degree of non-stationarity. When the important spectral features of the signals are time-varying, the effect of conventional Fourier analysis is to produce an averaged (i.e. smeared or distorted) spectral representation, which leads to a loss in frequency resolution. One way to deal with the spectral smearing is to reduce the effects of the variation in time by taking the spectral estimates over adjacent short time intervals of the signal, centred about particular time instants. Unfortunately, the shortened observation window produces a problem of its own - another smearing caused by the "uncertainty relationship" of time and band-limited signals [14].

Another way to deal with the problem of non-stationarity is to pass the signals through a filter bank composed of adjacent narrow-band bandpass filters, followed by a further analysis of the output of each filter. Again, the same problem described above occurs: the uncertainty principle [14] is encountered this time as a result of the band limitations of the filters. If small bandwidth filters are used, the ability to localise signal features well in time is lost. If large bandwidth filters are used, the fine time domain detail can be obtained, but the frequency resolution becomes poor.

1.2 Problem statement for time-frequency analysis

Classical methods for signal analysis are either based on the analysis of the time signal $s(t)$, or on its Fourier transform defined by

$$S(f) = \int_{-\infty}^{+\infty} s(t) e^{-j2\pi ft} dt \tag{1}$$

The time domain signal reveals information about the presence of a signal, its strengths and temporal evolution. The Fourier transform (FT) indicates which frequencies are present in the signal and their relative magnitudes. For deterministic signals, the representations usually employed for signal analysis are either the instantaneous power (i.e.. the squared modulus of the time signal) or the energy density spectrum (the squared modulus of the Fourier transform of a signal). For random signals, the analysis tools are based on the autocorrelation function (time domain) and its Fourier transform, the power spectrum. These analysis tools have had tremendous success in providing solutions for many problems related to stationary signals. However, they have immediate limitations when applied to non-stationary signals. For example, it is clear that the spectrum gives no indication as to how the frequency content of the signal changes with time, information which is needed when one deals with signals such as frequency modulated

(FM) signals. The chirp signal is an example of such a signal. It is a linear FM signal, used, for example, as a controllable source in seismic processing. It is analogous to a musical note with a steadily rising pitch, and is of the form

$$s(t) = \Pi_T\left(t - \frac{T}{2}\right)\cos\left(f_0 t + \alpha\frac{t^2}{2}\right) \tag{2}$$

where $\Pi_T(t)$ is 1 for $|t| \leq T/2$ and zero elsewhere, f_0 is the centre frequency and α represents the rate of the frequency change.

The fact that the frequency in the signal is steadily rising with time is not revealed by the spectrum; it only reveals a broadband spectrum, (Fig.1, bottom).

It would be desirable to introduce a time variable so as to be able to express the time and frequency-dependence of the signal, as in Fig.1. This figure displays information about the signal in a joint time-frequency domain. The start and stop times are easily identifiable, as is the variation of the spectral behaviour of the signal. This information cannot be retrieved from either the instantaneous power or the spectrum representations. It is lost when the Fourier transform is squared and the phase of the spectrum is thereby discarded. The phase actually contains this information about "the internal organisation" of the signal, as physically displayed in Fig.1. This "internal organisation" includes such details as times at which the signal has energy above or below a particular threshold, and the order of appearance in time of the different frequencies present. The difficulty of interpreting and analysing a phase spectrum makes the concept of a joint time and frequency signal representation attractive. For example, a musician would prefer to interpret a piece of music, which shows the pitch, start time and duration of the notes to be played rather than to be given a magnitude and phase spectrum of that piece of music to decipher [6].

As another illustration of the points raised above, consider the whale signal whose time-frequency (t-f) representation is displayed in Fig.2. By observing this t-f representation, a clear picture of the signal's composition instantly emerges. One can easily distinguish the presence of at least 4 separate components (numbered 1 to 4) that have different start and stop times, and different kinds of energies. One can also notice the presence of harmonics. One could not extract as much information from the time signal (seen at the left hand side in the same figure) or from the spectrum (at the bottom of the same figure).

If such a representation is invertible, the undesirable components of this signal may be filtered out in the time-frequency plane, and the resulting time signal recovered for further use or processing. If only one component of the signal is desired, it can be recognised more easily in such a representation than in either one of the time domain signal or its spectrum. This example illustrates how a time-frequency representation has the potential to be a very powerful tool, due to its ease of interpretation. It is

a time-varying extension of the ordinary spectrum which the engineer is comfortable using as an analysis tool.

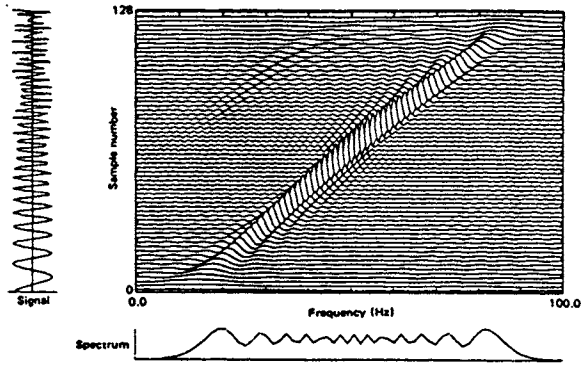


Figure 1. Time-frequency representations of a linear FM signal: the signal appears on the left, and its spectrum on the bottom

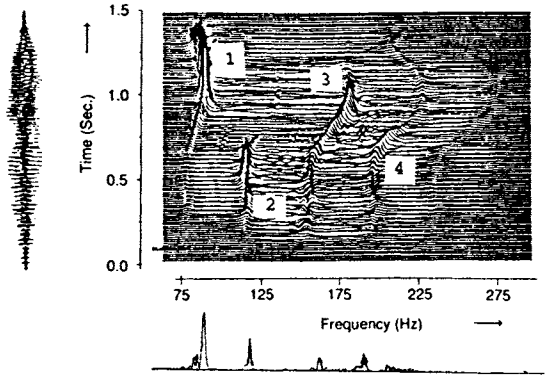


Figure 2. Time-frequency plot of a bowhead whale

2 A review of the early contributions to TFSA

2.1 Gabor's theory of communication

In 1946 Gabor [14] proposed a TFD for the purpose of studying the question of efficient signal transmission. He expressed dissatisfaction with the physical results obtained by using the FT. In particular, the t-f exclusivity of the FT did not fit with his intuitive notions of a time-varying frequency as evident in speech or music. He wanted to be able to represent other signals, not just those limiting cases of a "sudden surge" (delta function) or an infinite duration sinusoidal wave. By looking at the response of a bank of filters which were constrained in time and frequency, Gabor essentially performed a time-frequency analysis. He noted that since there was a resolution limit to the typical resonator, the bank of filters would effectively divide the time-frequency plane into a series of rectangles. He further noted that the dimensions of these rectangles, *tuning width* $\times$ *decay time*, must obey Heisenberg's uncertainty principle which translates in Fourier analysis to:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \tag{3}$$

where Δt and Δf are the equivalent duration and bandwidth of the signal [14]. Gabor believed this relationship to be "at the root of the fundamental principle of communication" [14], since it puts a lower limit on the minimum spread of a signal in time and frequency. The product value of $\Delta t \cdot \Delta f = 1/4\pi$ gives the minimum area unit in this time–frequency information diagram, which is obtained for a complex Gaussian signal.

Gabor's representation divided the time-frequency plane into discrete rectangles of information called logons. Each logon was assigned a complex value, $c_{m,n}$ where m represents the time index and n the frequency index. The $c_{m,n}$ coefficients were weights in the expansion of a signal into a discrete set of shifted and modulated Gaussian windows, which may be expressed as :

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{m,n} \psi(t; m, n) \tag{4}$$

where $\psi(t; m, n)$ are Gaussian functions centred about time, m, and frequency, n [14].

Lerner [15] extended Gabor's work by removing the rectangular constraint on the shape of the elementary cells. Helstrom [16] generalised the expansion by replacing the discrete elementary cell weighting with a continuous function, $\xi(\tau, t, f)$. Wavelet theory was later on developed as a further extension of Gabor's work, but with each partition of the time-frequency plane varying so as to yield a constant Q filtering [17].

2.2 The spectrogram

The spectrogram originated from early speech analysis methods and represents the most intuitive approach to spectrum analysis of non-stationary processes. It represents a natural transition from stationary processing towards time-frequency analysis. In this method, a local power spectrum is calculated from slices of the signal centred around the successive time points of interest, as follows:

$$\rho_{spec}(t,f) = |S(t,f)|^2 = \left| \int_{-\infty}^{\infty} s(\tau) h(t-\tau) e^{-j2\pi f\tau} d\tau \right|^2 \qquad (5)$$

where $h(t-\tau)$ is the time-limiting analysis window, centred at $t = \tau$, and $S(t,f)$ is referred to as the short-time Fourier transform (STFT). The time-frequency character of the spectrogram is given by its display of the signal as a function of the frequency variable, f, and the window centre time. This is a simple and robust method, and has consequently enjoyed continuing popularity. However, it has some inherent problems. The frequency resolution is dependent on the length of the analysis window and thus degrades significantly as the size of the window is reduced, due to the uncertainty relationships. The spectrogram can also be expressed as a windowed transformation of the signal spectrum as follows:

$$\rho_{spec}(t,f) = |S(t,f)|^2 = \left| \int_{-\infty}^{\infty} S(\nu) H(f-\nu) e^{j2\pi\nu t} d\nu \right|^2 \qquad (6)$$

These two representations become identical if $h(t)$ and $H(f)$ are a Fourier transform pair [12]. This indicates that there exists the same compromise for the time resolution; i.e. there is an inherent trade-off between time and frequency resolution. The spectrogram is still one of the most popular tool for TFSA, due to its robustness to noise, linearity property, ease of use and interpretation.

2.3 Page's instantaneous power spectrum

Page [18] was one of the first authors to extend the notion of power spectrum to deal with time-varying signals. He defined the "instantaneous power spectrum" (IPS), $p(t,f)$, which verifies :

$$E_T = \int_{-\infty}^{T} \int_{-\infty}^{\infty} p(t,f) df dt \qquad (7)$$

where E_T represents the total signal energy contained up to time, T, and where $p(t,f)$ represents the distribution of that energy over time and over the frequency. It is a spectral representation of the signal, which varies as

a function of time. In order to obtain an expression for $p(t,f)$, Page first defined a running transform:

$$S_t^-(f) = \int_{-\infty}^{t} s(\tau) e^{-j2\pi f\tau} d\tau \tag{8}$$

which represents the conventional FT of the signal, but calculated only up to time t. This allows the definition of a time-varying FT. He then defined the IPS as the rate of change or gradient in time of $S_t^-(f)$; i.e. the contribution to the overall energy made by each frequency component. This is defined as follows:

$$p(t,f) = \frac{\partial}{\partial t}|S_t^-(f)|^2 \tag{9}$$

It may equivalently be expressed as [18]

$$p(t,f) = 2s(t)\mathcal{R}\{e^{j2\pi ft} S_t^-(f)\} \tag{10}$$

or

$$p(t,f) = 2\int_0^\infty s(t)s(t-\tau)\,\cos 2\pi f\tau\; d\tau \tag{11}$$

where $\mathcal{R}$ denotes the real part.

Since $p(t,f)$ is the gradient of a spectrum, it may contain negative values; it redistributes signal energy as time evolves, compensating for previous values which were either too low or too high. The IPS therefore does not localise the information in time and frequency. Turner [19] has shown that the IPS is not unique, since any complementary function which integrates to zero in frequency can be added to it without changing the distribution. He also proved that the IPS is dependent on the initial time of observation. This indicates that the IPS is not a "true" TFD, i.e.. it does not meet some obvious requirements that a signal analyst expects in order to carry out a practical analysis of the signal. Nevertheless, it represented an important step in the development of ideas which led to our current understanding of TFDs.

Levin [20], following Page's work, defined a forward running (or anti-causal) spectrum $S_t^+(f)$, which is based on future time values, by taking a FT from t to $+\infty$. He also defined a time-frequency representation taking an average of the forward and backward IPS to get :

$$p_L(t,f) = \frac{1}{2}\left[\frac{\partial}{\partial t}|S_t^-(f)|^2 + \frac{\partial}{\partial t}|S_t^+(f)|^2\right] \tag{12}$$

$$= 2s(t)\mathcal{R}\{e^{j2\pi ft} S(f)\} \tag{13}$$

By realising that this combination would lead to an overall time-frequency representation which describes better the signal, Levin defined a distribution that is very similar to Rihaczek's [21] which will be discussed next.

It is worthwhile noting here that we will show in section 3.2 that all the TFDs which have been discussed so far can be written using a general framework provided by a formula borrowed from quantum mechanics.

2.4 Rihaczek's complex energy density

Starting from physical considerations, Rihaczek formed a time-frequency energy density function for a complex deterministic signal, $z(t)$, which, he claimed, was a natural extension of the energy density spectrum, $|Z(f)|^2$, and the instantaneous power, $|z(t)|^2$. His reasoning was as follows: the total energy of a complex signal, $z(t)$, is:

$$E = \frac{1}{2}\int_{-\infty}^{\infty} |z(t)|^2 dt \tag{14}$$

Consider a bandlimited portion of the original signal, around f_0, $z_1(t)$ given as

$$z_1(t) = \mathcal{F}^{-1}\{\Pi_{\Delta B}(f - f_0) \cdot Z(f)\} \tag{15}$$

This portion of the signal, $z_1(t)$, contains the energy

$$E_1 = \frac{1}{2}\int_{-\infty}^{\infty} z(t) z_1^*(t) dt \tag{16}$$

If the bandwidth of $z_1(t)$, ΔB is reduced to δB, then $z_1(t) = Z(f_0)\delta B \cdot e^{j2\pi f_0 t}$. Assuming that $Z(f)$ is constant over the spectral band δB, which is reasonable if $\delta B \to 0$, we then obtain:

$$E_1 = \frac{1}{2}\int_{-\infty}^{\infty} z(t) Z^*(f_0)\delta B e^{-j2\pi f_0 t} dt \tag{17}$$

This quantity in (17) represents the energy in a small spectral band δB, but over all time. To obtain the energy within a small frequency band δB, and a time band ΔT, it suffices to limit the integration in time to ΔT as follows:

$$E_1 = \frac{1}{2}\int_{t_0-\Delta T/2}^{t_0+\Delta T/2} z(t) Z^*(f_0)\delta B e^{-j2\pi f_0 t} dt \tag{18}$$

Taking the limit $\Delta T \to \delta T$ yields

$$E_1 = \frac{1}{2}\, \delta B\, \delta T\, z(t_0)\, Z^*(f_0)\, e^{-j2\pi f_0 t_0} \tag{19}$$

with the resultant time-frequency distribution function being

$$\rho_R(t, f) = z(t) Z^*(f) e^{-j2\pi f t} \tag{20}$$

which is generally referred to as the Rihaczek Distribution (RD). If $z(t)$ is real, one can see that Levin's TFD (which is based on Page's TFD) is simply twice the real part of Rihaczek's TFD.

It is remarkable to see that different approaches to define a TFD that seem to be all natural and straightforward lead to apparently different definitions of a TFD.

2.5 The Wigner-Ville distribution

Ville's work [22] followed Gabor's contribution; he similarly recognised the insufficiency of time analysis and frequency analysis, using the same analogy of a piece of music. He indicated that since a signal has a spectral structure at any given time, there existed the notion of an "instantaneous spectrum" which had the physical attributes of an energy density. Thus the energy within a small portion of the time-frequency plane, $dt \cdot df$ would be

$$E_\delta = W(t,f)\, dt\, df \tag{21}$$

and its integration over f (respectively over t) should yield the instantaneous power $|s(t)|^2$ (respectively the energy spectral density $|S(f)|^2$). Integration over both t and f would yield the energy E.

$$\int_{-\infty}^{\infty} W(t,f)df = |s(t)|^2 \tag{22}$$

$$\int_{-\infty}^{\infty} W(t,f)dt = |S(f)|^2 \tag{23}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(t,f)dtdf = E \tag{24}$$

These desirable properties led Ville to draw an analogy with the probability density function (pdf) of quantum mechanics. i.e. consider that:

1. the distribution $\rho(t,f)$ to be found is the joint pdf in time and frequency,
2. the instantaneous power is one marginal probability of $\rho(t,f)$,
3. the spectrum is the other marginal probability of $\rho(t,f)$.

Then, one could form the characteristic function, $F(u,v)$, of this TFD, and equate the marginal results of $|s(t)|^2$ and $|S(f)|^2$ with the moments generated from the characteristic function (using its moment generating properties):

$$W(t,f) = \underset{t \to u}{\mathcal{F}} \; \underset{f \to v}{\mathcal{F}} \; F(u,v) \tag{25}$$

Using then the framework of quantum mechanical operator theory [23], Ville established that the proper form for the distribution was:

$$W(t,f) = \int_{-\infty}^{+\infty} z(t+\frac{\tau}{2}) \cdot z^*(t-\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau \qquad (26)$$

where $z(t)$ is the analytic complex signal which corresponds to the real signal, $s(t)$ [24]. It is obtained by adding an imaginary part $y(t)$ which is obtained by taking the Hilbert transform of the real signal $s(t)$ [14].

Ville's distribution was derived earlier by Wigner in a quantum mechanical context [25]. For this reason, it is generally referred to as the Wigner-Ville distribution (WVD) and it is the most widely studied of present TFDs. The advantages of the WVD as a signal processing tool are manifold. It is a real joint distribution of the signal in time and frequency. The marginal distributions in time and frequency can be retrieved by integrating the WVD in frequency and time respectively. It achieves optimal energy concentration in the time-frequency plane for linearly frequency modulated signals. It is also time, frequency and scale invariant, and so fits well into the framework of linear filtering theory. The disadvantages of the WVD are chiefly that it is non-positive, that it is "bilinear" and has cross-terms. The non-positivity makes the WVD difficult to interpret as an energy density. The cross-terms cause "ghost" energy to appear mid-way between the true energy components.

A detailed review of the WVD is provided in [2].

3 The second phase of developments in TFSA: 1980's

3.1 Major developments in 1980's

The early research in the 1980's focussed on the WVD as engineers and scientists started to discover that it provided a means to attain good frequency localisation for rapidly time-varying signals. For example, in a seismic context it was shown to be a very effective tool to represent "Vibroseis" chirp signals emitted in seismic processing [26], and hence was used to control the quality of the signal emitted. When the signal emitted was a pure linear FM, the WVD exhibited a sharp peak along the FM law. If the signal was contaminated by harmonic coupling effects and other distortions then this property was lost [27].

The interest in the WVD was fuelled by its good behaviour on chirp signals and by the discovery (and later re-discovery) [25],[22] of its special properties, which made it attractive for the analysis of time-varying signals.

The advance of digital computers also aided its popularity, as the previously prohibitive task of computing a two-dimensional distribution came

within practical reach [28][1].

The WVD of a signal, $z(t)$, is constructed conceptually as the Fourier transform of a "bilinear" kernel[2], as

$$W_z(t,f) = \underset{\tau \rightarrow f}{\mathcal{F}} [K_z(t,\tau)] \tag{27}$$

where $\underset{\tau \rightarrow f}{\mathcal{F}}$ represents a Fourier transformation with respect to the τ variable, and where $K_z(t,\tau)$ is the "bilinear" kernel defined by

$$K_z(t,\tau) \triangleq z(t+\frac{\tau}{2})z^*(t-\frac{\tau}{2}) \tag{28}$$

Most of the early research in the WVD concentrated on the case of deterministic signals, for which the WVD is interpreted as a distribution of the signal in the time-frequency domain. For random signals, it was shown [29] that the expected value of the WVD equals the FT of the time-varying auto-correlation function (when these quantities exist). This gave the WVD an important interpretation as a time-varying Power Spectral Density (PSD) and sparked significant research efforts along this direction.

Filtering and Signal synthesis. It was also realised early that the WVD could be used as a time-varying filter [30]. A simple algorithm was devised which masked (i.e.. filtered) the WVD of the input signal and then performed a least-squares inversion of the WVD to recover the filtered signal [2] [30]. It was also shown that the input-output convolution relationships of filters were preserved when one used the WVD to represent the signals.

Implementation The computational properties of the WVD were further studied and this led to an efficient real-time implementation which exploits the symmetry properties of the Wigner-Ville kernel sequence [31].

Signal Detection, Estimation and Classification. The areas of detection and estimation saw significant theoretical developments based on the WVD [32], [33], [34], motivated by the belief that signal characterisation should be more accurate in a joint t-f domain. A key property helped motivate this interest: the WVD is a unitary (energy preserving) transform.

[1] To the author's best knowledge, the first WVD programme was written by him in APL language in September 1978, for the processing of Vibroseis chirp data [28].

[2] The kernel is actually bilinear only for the cross WVD, introduced in Sec.3.1.

Therefore, many of the classical detection and estimation problem solutions had alternate implementations based on the WVD. The two-dimensional t-f nature of the implementation, however, allowed greater flexibility than did the classical one [35], [36].

The theory and important properties of the WVD which prompted so much of the investigations outlined above were reviewed in detail in [2], and will be briefly summarised in Section 3.3.

A mistake that was made by many of the early researchers was to "sell" uninhibitedly the method as a universal tool, whereas its field of application is really quite specialised. As the WVD became increasingly exposed to the signal processing community, users started to discover the limitations of the method, which are presented below.

Non-linearities One of the main limitations is that the WVD is a method which is non-linear. The WVD performs a "bilinear" transformation of the signal, which is equivalent to a "dechirping" operation. There are serious consequences for multicomponent signals, that is, composite signals such as a sum of FM signals [2]. For such signals, the bilinear nature of the WVD causes it to create cross-terms (or artifacts) which occur in between individual components. This can often render the WVD very difficult to interpret, such as in cases where there are many components or where components are not well separated. In addition, the bilinearity exaggerates the effect of additive noise by creating cross-terms between the signal component and the noise component[3]. At low signal-to-noise ratio (SNR), where the noise dominates, this leads to a very rapid degradation of performance.

Limited duration. Another drawback sometimes attributed to the WVD is that it performs well only for infinite duration signals. Since it is the FT of a bilinear kernel, it is tuned to the presence of infinite duration complex exponentials in the kernel, and hence to linear FM components in the signal. Real life signals are usually time limited, therefore a simple FT of the bilinear kernel does not provide a very effective analysis of the data. There is a need to take a windowed FT [31] of the bilinear kernel or to replace the FT by a high-resolution model-based spectral analysis such as Burg's algorithm [37] [38].

Cohen's bilinear class of smoothed WVDs. A lot of effort went into trying to overcome the drawbacks of the WVD [39], [40], [41], [42], [43], [2]. Researchers realised that although the WVD seemed to be theoretically the best of the TFSA tools available, for practical analysis other TFDs were better because they reduced the cross-terms. They borrowed Cohen's formula from quantum mechanics so as to link all the bilinear TFDs together

[3]This has led to some jokes nick-naming the WVD as a noise generator!

by a 2-D smoothing of the WVD. They then tried to find an optimum smoothing, optimum in the sense of facilitating the job of the analyst in the use of TFSA methods. Since most of the known TFSA methods were obtainable by a 2-D smoothing of the WVD, there was perhaps an ideal TFSA method not yet known which could be discovered by a proper choice of the smoothing window (the spectrogram, Page's running Spectrum,[18], Rihaczek's complex energy distribution [21] could all be achieved via this type of smoothing [26], [11]).

Filtering out cross-terms in the Ambiguity domain. Many researchers turned to 2-D Gaussian smoothing functions to reduce the artifacts [39], [40], because of the Gaussian window property of minimising the bandwidth-time product.

A key development in a more effective effort at trying to reduce artifacts was to correlate this problem with a result from radar theory using the fact that in the ambiguity domain (Doppler-lag), the cross-terms tended to be distant from the origin, while the auto-terms were concentrated around the origin [44], [30]. Understanding this link was very helpful since the WVD was known to be related to the ambiguity function via a 2-D Fourier Transform [2]. The natural way of reducing the cross-terms of the WVD was then simply to filter them out in the ambiguity domain, followed by a 2-D FT inversion.

This led to greater refinements and thought in the design of TFDs. Using this approach Choi-Williams designed a TFD with a variable level smoothing function, so that the artifacts could be reduced more or less depending on the application [43]. Zhao, Atlas and Marks designed smoothing functions in which the artifacts folded back onto the auto-terms [42]. Amin [41] came to a similar result in the context of random signals, with this providing inspiration for the work reported in [45]. There it was shown how one could vary the shape of the cross-terms by appropriate design of the smoothing function.

By making the smoothing function data dependent, Baraniuk and Jones produced a signal dependent TFD which achieved high energy concentration in the t-f plane. This method was further refined by Jones and Boashash [46] who produced a signal dependent TFD with a criterion of local adaptation.

Cross WVD (XWVD). Another approach to reduce or nullify the presence of cross-terms was based on replacing the WVD by the XWVD in order to obtain a distribution which is linear in the signal. The XWVD could be interpreted as an extension of the cross-correlation function for

non-stationary signals. The XWVD is defined as:

$$W_{12}(t,f) = \underset{\tau \rightarrow f}{\mathcal{F}} [K_{12}(t,\tau)] \tag{29}$$

where

$$K_{12}(t,\tau) = z_1(t+\frac{\tau}{2}) z_2^*(t-\frac{\tau}{2}) \tag{30}$$

where $z_1(t)$ is a reference signal and $z_2(t)$ is the signal under analysis. There were then systematic efforts in trying to substitute the use of the XWVD in all areas of application of the WVD. In many cases, this was straightforward, because a reference signal was available. Thus, the XWVD was proposed for optimal detection schemes [32], for sonar and radar applications [47], and for seismic exploration [48]. These schemes were seen to be equivalent to traditional matched filter and ambiguity function based schemes, but their representation in another domain allowed for some flexibility and variation. In other cases, where reference signals were not available, the XWVD could not in general be applied, a fact which prevented the further spread of the XWVD as a replacement for the WVD.

In some applications, however, it is possible to define reference signals from filtered estimates of the original signal, and then use these as if they were the true signal. The filtering procedure uses the IF as a critical feature of the signal. Jones and Parks [49] implicitly used a similar philosophy to estimate their data dependent distributions. They estimated their reference signal as that signal component which maximised the energy concentration in the distribution.

Wideband TFDs. The problems relating to the WVD's poor performance with short duration or wideband signals were addressed in several ways. One way was as mentioned earlier, to use autoregressive modelling techniques. Much more attention, though, was given to designing wideband or *affine* time-frequency representations. The first to be considered was the wavelet transform, which is linear. It was like the Gabor transform in that it obtained its coefficients by projecting the signal onto basis functions corresponding to different positions in time-frequency. The wavelet transform differed from the Gabor transform in that its basis functions all had the same shape. They were simply dilated (or scaled) and time shifted versions of a *mother* wavelet. This feature causes the representation to exhibit a *constant Q* filtering characteristic. That is, at high frequencies the resolution in time is good, while the resolution in frequency is poor. At low frequencies, the converse is the case. Consequently, abrupt or step changes in time may be detected or analysed very well.

Subsequent efforts aimed at incorporating these wideband analysis techniques into bilinear TFDs. One of the early attempts was in [39], and used

the Mellin transform (rather than the Fourier transform) to analyse the bilinear kernel. The Mellin transform is a scale-invariant transform, and as a consequence, is suited to constant Q analysis. A significant contribution was also made by the Bertrands, who used a rigourous application of Group Theory to find the general bilinear class of scale-invariant TFDs [50]. Others showed that this class of TFDs could be considered to be smoothed (in the affine sense) WVDs, and that many properties might be found which were analogous to those of Cohen's class [51].

These techniques were extended for use in wideband sonar detection applications [52], and in speech recognition [53].

Instantaneous Frequency. The development of time-frequency analysis was parallelled by a better understanding of the notion of instantaneous frequency (IF), since in most cases TFSA methods were aiming at a better IF estimation of the FM laws comprising the signal of interest.

For analysts who were used to time-invariant systems and signal theory, the simultaneous use of the terms *instantaneous* and *frequency* appeared paradoxical and contradictory. A comprehensive review of the conceptual and practical aspects of progressing from frequency to instantaneous frequency was given in [3], [4].

The IF is generally defined as the derivative of the phase. For discrete-time signals, the IF is estimated by such estimators as the "central finite difference" (CFD) of the phase [3]. Recently, this was extended by defining a general phase difference estimate of the IF [4]. This allowed us to understand why the WVD performed well for linear FM signals only - it has an in-built CFD IF estimator which is unbiased for linear FM signals only. It is therefore optimally suited for linear FM signals, but only for this class of signals. The remainder of this section will deal with the class of bilinear TFDs which is suited to the analysis of such signals. Section 4 will present a new class of TFDs referred to as polynomial TFDs, which are suited to the analysis of non-linear polynomial FM signals.

3.2 Bilinear class of TFDs

From an engineering point of view, Ville's contribution laid the foundations for a class of TFSA methods which developed in the 1980s. A key step was to realise that if one wanted to implement the WVD, then in practice one can only use a windowed version of the signal and operate within a finite bandwidth. Another key step was to realise that the WVD may be expressed as the FT of the bilinear kernel:

$$W_z(t,f) = \underset{\tau \rightarrow f}{\mathcal{F}} \{z(t+\frac{\tau}{2}) \cdot z^*((t-\frac{\tau}{2})\} \tag{31}$$

If one replaces $z(t)$ by a time windowed and a band-limited version of $z(t)$, then, after taking the FT, one would obtain a new TFD.

$$\rho(t,f) = W_z(t,f) * *\gamma(t,f) \tag{32}$$

where the smoothing function $\gamma(t,f)$ describes the time limitation and frequency limitation of the signal, and $**$ denotes convolution in time and frequency.

If one then decides to vary $\gamma(t,f)$ according to some criteria so as to refine some measurement, one obtains a general time-frequency distribution which could adapt to the signal limitations. These limitations may be inherent to the signal or may be caused by the observation process. If we write in full the double convolution, this then leads to the following formulation:

$$\rho(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{j2\pi\nu(u-t)} g(\nu,\tau) z(u+\frac{\tau}{2})\cdot z^*(u-\frac{\tau}{2}) e^{-j2\pi f\tau}\, d\nu\, du\, d\tau \tag{33}$$

which is generally referred to as Cohen's formula, since it was defined by Cohen in 1966 [54] in a quantum mechanics environment[4]. The function $g(\nu,\tau)$ in (33) is the double FT of the smoothing function $\gamma(t,f)$.

This formula may also be expressed in terms of the signal FT, $Z(f)$, as:

$$\rho_z(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{j2\pi\tau(\eta-f)} g(\nu,\tau) Z(\eta+\frac{\nu}{2})\cdot Z^*(\eta-\frac{\nu}{2}) e^{-j2\pi t\nu}\, d\tau\, d\eta\, d\nu \tag{34}$$

or in the time-lag domain as

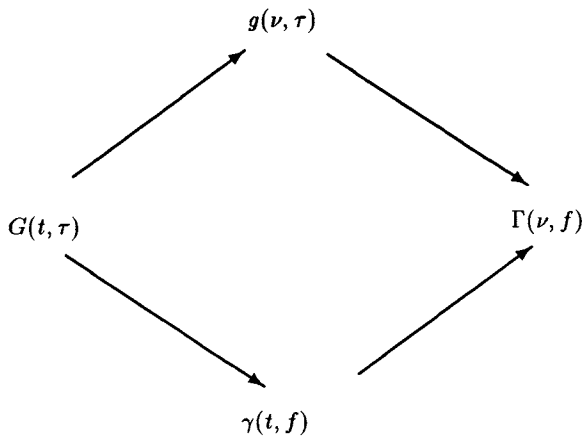
$$\rho_z(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(t-u,\tau) z(u+\frac{\tau}{2})\cdot z^*(u-\frac{\tau}{2}) e^{-j2\pi f\tau}\, du\, d\tau \tag{35}$$

where $G(t,\tau)$ is the inverse Fourier transform of $g(\nu,\tau)$ in the ν variable. Finally, in the frequency-Doppler domain it is given by:

$$\rho_z(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Gamma(f-\eta,\nu) Z(\eta+\frac{\nu}{2})\cdot Z^*(\eta-\frac{\nu}{2}) e^{j2\pi\nu t}\, d\eta\, d\nu \tag{36}$$

The smoothing functions are related by the FT as shown in the following diagram:

[4]It is only recently (about 1984) that Leon Cohen became aware that the formula he devised for quantum mechanics in 1966 [54], was being used by engineers. Since then, he has taken a great and active interest in the use of his formula in signal processing, and has brought the fresh perspective of a physicist to the field of time-frequency signal analysis.



It was shown in the 1980s that nearly all the then known bilinear TFDs were obtainable from Cohen's formula by appropriate choice of the smoothing function, $g(\nu,\tau)$. In this chapter we will refer to this class as the bilinear class of TFDs. Most of the TFDs proposed since then are also members of the bilinear class. Some of those TFDs were discussed earlier. Others have been studied in detail and compared together in [2], [1]. Table 1 lists some of the most common TFDs and their determining functions in the discrete-form, $G(n,m)$. Knowing $G(n,m)$, each TFD can be calculated as:

$$\rho_z(n,k) = \sum_m \sum_p G(p,m)\, z(n+p+m)\, z^*(n+p-m)\, e^{-j4\pi mk/N} \qquad (37)$$

3.3 Properties and limitations of the bilinear class of TFDs

The properties of the bilinear class of TFDs will be listed and compared with the "desirable" properties expected by signal analysts for a TFD to be used as a practical tool. A more complete treatment is provided in [1] and [2].

Realness. To represent the variation in signal energy, the TFD should be real.

Marginal conditions. It is usually desired that integrating the TFD over time (respectively frequency) yields the power spectrum (respectively the instantaneous power). Naturally, the integration over both time and frequency should yield the signal energy.

Time–Frequency Representation	$G(n,m)$
Windowed Discrete WVD	$\delta(n) \quad m \in [\frac{-(M-1)}{2}, \frac{(M-1)}{2}]$ $0 \quad$ otherwise
Pseudo WVD using a rectangular window of odd length P	$\frac{1}{P} \quad n \in [\frac{-(P-1)}{2}, \frac{(P-1)}{2}]$ $0 \quad$ otherwise
Rihaczek–Margenau	$\frac{1}{2}[\delta(n+m)+\delta(n-m)]$
STFT using a Rectangular Window of odd length P.	$\frac{1}{P} \quad \lvert m+n\rvert \leq \frac{(P-1)}{2}$ $0 \quad$ otherwise
Born–Jordan–Cohen	$\frac{1}{\lvert m\rvert+1} \quad \lvert m\rvert \leq \lvert n\rvert$ $0 \quad$ otherwise
Choi–Williams (parameter σ)	$\frac{\sqrt{\sigma/\pi}}{2m} e^{-\sigma n^2/4m^2}$

Table 1: Some TFDs and their determining functions $G(n,m)$. Knowing $G(n,m)$, each TFD can be calculated using eq.(37).

Positivity. It is normally expected that a TFD would be positive. However, we have seen that by construction, Page's TFD was not. It was also shown [12], that:

1) for a TFD to be positive, the smoothing function had to be an ambiguity function, or a linear combination of ambiguity functions, and

2) the property of positivity was incompatible with verifying the marginal condition.

An interpretation of TFDs which is compatible with their non- positivity aspect, is to consider that they provide a measure of the energy flow through the spectral band $[f - \Delta f, f + \Delta f]$ during the time $[t - \Delta t, t + \Delta t]$, [2].

Time shifting and frequency shifting. Time and frequency shifts of amounts, (t_0, f_0), of a signal, $s(t)$, should be reflected in its TFD; i.e. if $\rho_s(t, f)$ is the TFD of $s(t)$, then signal $s(t - t_0)e^{j2\pi f_0 t}$ has a TFD which is $\rho_s(t - t_0, f - f_0)$. All TFDs of the bilinear class satisfy this property.

Input-output relationship of a linear filter. Bilinear TFDs are consistent with linear system theory; i.e. if

$$y(t) = s(t) * h(t), \quad \text{that is if} \quad Y(f) = S(f) \cdot H(f)$$

then

$$\rho_y(t, f) = \rho_s(t, f) \underset{t}{*} \rho_h(t, f)$$

Equivalently, if

$$Y(f) = S(f) * H(f), \quad \text{that is if} \quad y(t) = s(t) \cdot h(t)$$

then

$$\rho_y(t, f) = \rho_s(t, f) \underset{f}{*} \rho_h(t, f)$$

Finite support. If a TFD is to represent the instantaneous energy in time and frequency, then it should be zero when the signal is zero; i.e. if $s(t) = 0$ for $t_a < t < t_b$, and $S(f) = 0$ for $f_a < f < f_b$, then, $\rho_s(t, f) = 0$

for $t_a < t < t_b$ or $f_a < f < f_b$. The finite support in time holds provided that the function $g(\nu, \tau)$ introduced in (33) satisfies [11]:

$$\underset{\nu \rightarrow t}{\mathcal{F}^{-1}} \{g(\nu, \tau)\} = 0 \qquad \text{for } |t| > |\tau|/2 \tag{38}$$

and in frequency holds when

$$\underset{\tau \rightarrow f}{\mathcal{F}} \{g(\nu, \tau)\} = 0 \qquad \text{for } |f| > |\nu|/2 \tag{39}$$

Instantaneous frequency and group delay. For FM signals, the non-stationary or time-varying aspect is directly quantified by the variation of the IF, $f_i(t)$, or the group delay (GD), $\tau_g(f)$, expressed as follows:

$$f_i(t) = \frac{1}{2\pi} \frac{d}{dt} arg\{z(t)\} \tag{40}$$

$$\tau_g(f) = -\frac{1}{2\pi} \frac{d}{df} arg\{Z(f)\}. \tag{41}$$

One of the most desirable properties of TFDs is that their first moments in time (respectively frequency) give the GD (respectively the IF), as follows:

$$f_i(t) = \frac{\int_{-\infty}^{\infty} f\rho(t, f)df}{\int_{-\infty}^{\infty} \rho(t, f)df} \tag{42}$$

$$\tau_g(f) = \frac{\int_{-\infty}^{\infty} t\rho(t, f)dt}{\int_{-\infty}^{\infty} \rho(t, f)dt}. \tag{43}$$

These conditions are respected if the partial derivatives of the smoothing function are zero at the origin, $\nu = \tau = 0$. However, they are incompatible with the positivity requirement [2]. The IF and its relationship with TFDs has been the subject of a two-part review paper [3], [4].

Cross-terms, interference terms, artifacts. Since bilinear TFDs are a smoothed version of the FT of the bilinear kernel, $K_z(t, \tau)$ given by (28), i.e. as follows:

$$K_z(t, \tau) = z(t + \frac{\tau}{2}) \cdot z^*(t - \frac{\tau}{2}) \tag{44}$$

there always exist cross-terms which are created in the TFD as a consequence of the interaction of different components of the signal. Consider a signal composed of two complex linear FM signal components

$$z_3(t) = z_1(t) + z_2(t) \tag{45}$$

The bilinear kernel of the signal in the TFD is:

$$K_{z_3}(t,\tau) = K_{z_1}(t,\tau) + K_{z_2}(t,\tau) + K_{z_1 z_2}(t,\tau) + K_{z_2 z_1}(t,\tau) \qquad (46)$$

where the cross-kernels, $K_{z_1 z_2}(t,\tau)$, and $K_{z_2 z_1}(t,\tau)$ are defined respectively by:

$$K_{z_1 z_2}(t,\tau) = z_1(t+\frac{\tau}{2}) \cdot z_2^*(t-\frac{\tau}{2}) \qquad (47)$$

$$K_{z_2 z_1}(t,\tau) = z_2(t+\frac{\tau}{2}) \cdot z_1^*(t-\frac{\tau}{2}) \qquad (48)$$

$$(49)$$

The third and fourth kernel terms comprise the cross-terms, which often manifest in the t-f representation in a very inconvenient and confusing way. Consider the WVD of the signal consisting of two linear FM components, given in Fig.3. It shows three components, when one expects to find only two. The component in the middle exhibits large (positive and negative) amplitude terms in between the linear FM's signal energies where it is expected that there should be no energy at all. These are the cross-terms resulting from the bilinear nature of the TFD, which are often considered to be the fundamental limitations which have prevented more widespread use of time-frequency signal analysis. TFDs have been developed which reduce the magnitude of the cross-terms, but they inevitably compromise some of the other properties of TFDs, such as resolution. The cross-term phenomenon is discussed in more details in [2] and [1].

The analytic signal. We have seen above that the bilinearity causes the appearance of cross-terms for composite signals, i.e. signals with multiple components. A real signal can be decomposed into two complex signals with symmetric spectra. The TFD of a real signal would therefore exhibit cross-terms at the zero frequency position. The analytic signal is normally used in the formation of TFDs to avoid this problem; it is constructed by adding an imaginary part, $y(t)$, to the real signal, $s(t)$, such that $y(t)$ is the Hilbert transform of $s(t)$. This ensures that the resulting complex signal has no energy content for negative frequencies.

Apart from this property, the analytic signal is useful for interpretive reasons and for efficient implementation of TFDs. A full treatment of this question is given in [2], [1], [3].

Multilinearity. Section 4 will show that the "bilinearity" of TFDs makes them optimal only for linear FM signals. For non-linear FM signals, a new class of multilinear TFDs is defined, and presented in the next section.

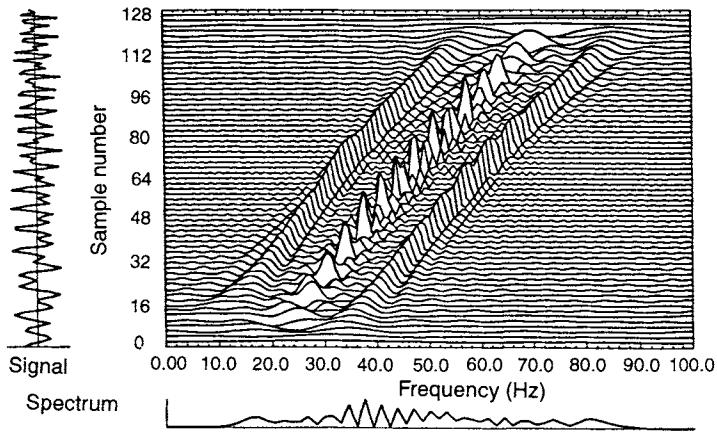


Figure 3.The WVD of a two component signal formed by summing two linear FM signals

Algorithms for bilinear TFDs. In considering the general usefulness of TFDs, it is important to consider the properties of their discrete-time equivalents, and their ease of implementation.

The discrete-time equivalent of the time-lag definition given in (35) leads to the simplest way of implementing TFDs [2, p.444]:

$$\rho_z(n,k) = \underset{m \to k}{\mathcal{F}} \{G(n,m) \underset{n}{*} K_z(n,m)\} \tag{50}$$

The implementation of this discrete-time bilinear TFDs requires three steps:

1. Formation of the bilinear kernel

$$K_z(n,m) = z(n+m) \cdot z^*(n-m)$$

2. Discrete convolution in discrete time n with the smoothing function $G(n,m)$.

3. Discrete FT with respect to discrete delay m.

The implementation of steps 1, 2 and 3 may be simplified by taking advantage of the many symmetries which exist, as explained in [31], [55].

The WVD does not require step 2 because its smoothing function $G(n,m)$ equals $\delta(n)$. Further details of the implementation are contained in [1, chapter 7].

The list of properties of discrete TFDs is given in [2, p.445-451]. In summary the characteristics of $G(n,m)$ determine the properties of a bilinear TFD, such as maximising energy concentration or reducing artifacts. The source code for the implementation of TFDs is listed in [1, chapter 7].

4 Polynomial TFDs.

It was indicated at the end of Section 3.1 that the investigation of the notion of the IF led us to realise that the bilinearity of the WVD makes it suitable for the analysis of linear FM signals only. This observation motivated some work which led to the definition of Polynomial WVDs (PWVDs) which were constructed around an in-built IF estimator which is unbiased for non-linear FM signal [56], [57]. This development led to the definition of a new class of TFDs which behave well for non-linear FM signals, and that are able to solve problems that Cohen's bilinear class of TFDs cannot [58] [59]. Another property of the PWVDs is that they are related to the notion of higher order spectra [57].

The reason why one might be interested in polynomial TFDs and/or time-varying higher order spectra is that 1) many practical signals exhibit some form of non-stationarity *and* some form of non-linearity, and that 2) higher order spectra have in theory the ability to reject Gaussian noise. Since TFDs have been widely used for the analysis of non-stationary signals, and higher order spectra are likewise used for the study of non-linearities and non-Gaussianities, it seems natural to seek the use of time-varying higher order spectra to deal with both phenomena simultaneously. There are many potential applications. For example, in underwater surveillance, ship noise usually manifests itself as a set of harmonically related narrowband signals; if the ship is changing its speed, then the spectral content of these signals is varying over time. Another example is an acoustic wave produced by whales (See the time-frequency representation of a typical whale signal shown in Fig.2). Similar situations occur in vibration analysis, radar surveillance, seismic data processing and other types of engineering applications. For the type of signals described above, the following questions often need to be addressed: what are the best means of analysis, detection and classification; how can one obtain optimal estimates for relevant signal parameters such as the instantaneous frequency of the fundamental component and its harmonics, and the number of non-harmonically related signals ; how can one optimally detect such signals in the presence of noise?

In this section we present first some specific problems and show how one may obtain solutions using the concepts described above. In the re-

mainder of the section, the Polynomial WVDs are introduced as a tool for the analysis of polynomial (non-linear) FM signals of arbitrary order. The next section presents a particular member of this class, referred to as the Wigner-Ville trispectrum. This representation is useful for the analysis of FM signals affected by multiplicative noise.

4.1 Polynomial Wigner-Ville distributions

The problem: In this section we address the problem of representation and analysis of deterministic signals, possibly corrupted by additive noise with high SNR, and which can be modelled as signals having constant amplitude and polynomial phase and frequency modulation.
The model: We consider the signal:

$$z(t) = \Pi_T(t)\ A\ e^{j\,\phi(t)} + w(t) \tag{51}$$

where

$$\phi(t) = \sum_{i=0}^{p} a_i t^i \tag{52}$$

and $w(t)$ is complex noise. Since the polynomial form of the phase, $\phi(t)$, uniformly approximates any continuous function on a closed interval (Weierstrass approximation theorem [60]), the above model can be applied to any continuous phase signal. We are primarily interested in the IF, defined in (40), and rewritten here as:

$$f_i(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{53}$$

The solution: The case where the phase polynomial order, p is equal to 1, belongs to the field of *stationary* spectrum and frequency estimation [61].

The case $p = 2$ corresponds to the class of signals with linear FM and can be handled by the WVD.

The case $p > 2$, corresponds to the case of *non-linear* FM signals, for which the Wigner-Ville transform becomes inappropriate (see Section 4.2) and the ML approach becomes computationally very complicated. However, signals having a non-linear FM law do occur in both nature and engineering applications. For example, the sonar system of some bats often uses *hyperbolic* and *quadratic* FM signals for echo-location [62]. In radar, some of the pulse compression signals are *quadratic* FM signals [63]. In geophysics, in some modes of long-propagation of seismic signals, non-linear signals may occur from earthquakes or underground nuclear tests [64]. In passive acoustics, the estimation of the altitude and speed of a propeller driven aircraft is based on the instantaneous frequency which is a non-linear function of time [65]. Non-linear FM signals also appear in communications,

astronomy, telemetry and other engineering and scientific disciplines. It is therefore important to find appropriate analysis tools for such signals.

The problem of representing signals with polynomial phase was also studied by Peleg and Porat in [66], [67]. They defined the polynomial phase transform (PPT) for estimating the coefficients of the phase, $\phi(n)$, and calculated the Cramer-Rao (CR) lower bounds for the coefficients. A more recent work on the same subject can be found in [68].

A different, yet related approach was taken by this author and his co-workers, extending the conventional WVD to be able to process polynomial FM signals effectively [56], [57]. This approach is developed below.

4.2 The key link: the Wigner-Ville distribution and its inbuilt IF estimator

The WVD has become popular because it has been found to have optimal concentration in the time-frequency plane for linear FM signals [2]. That is, it yields a continuum of delta functions along the signal's instantaneous frequency law as $T \rightarrow \infty$ [2], [3]. For non-linear FM signals this optimal concentration is no longer assured and the WVD based spectral representations become smeared.

We describe in this section the key link between the WVD and the IF that makes it possible to design *Polynomial Wigner-Ville distributions (PWVDs)*[5] which will exhibit a continuum of delta functions along the IF law for polynomial FM signals.

To explain how this is achieved, one needs to look closely at the mechanism by which the WVD attains optimal concentration for a linear FM signal. Consider a unit amplitude analytic signal, $z(t) = e^{j\phi(t)}$. Given that the WVD of this signal is defined by (27) and (28), substitution of $z(t) = e^{j\phi(t)}$, and equation (28) into (27) yields

$$W_z(t,f) = \underset{\tau \rightarrow f}{\mathcal{F}} \left[e^{j(\phi(t+\tau/2)-\phi(t-\tau/2))} \right] \tag{54}$$

Note that the term $\phi(t+\tau/2) - \phi(t-\tau/2)$ in (54) can be re-expressed as

$$\phi(t+\tau/2) - \phi(t-\tau/2) = 2\pi\tau \hat{f}_i(t,\tau) \tag{55}$$

where $\hat{f}_i(t,\tau)$ can be considered to be an instantaneous frequency estimate. This estimate is the difference between two phase values divided by $2\pi\tau$, where τ is the separation in time of the phase values. This estimator is simply a scaled finite difference of phases centrally located about time

[5] In earlier publications, polynomial WVDs were referred to as generalised WVDs.

instant t, and is known as the *central finite difference* estimator [69], [3]. The estimator follows directly from eq.(53):

$$f_i(t) = \frac{1}{2\pi} \lim_{\tau \to 0} \left[\frac{\phi(t+\tau/2) - \phi(t-\tau/2)}{\tau} \right] \tag{56}$$

Eq. (54) can therefore be rewritten as

$$W_z(t,f) = \underset{\tau \to f}{\mathcal{F}} \left[e^{j2\pi\tau \hat{f}_i(t,\tau)} \right] \tag{57}$$

Thus the WVD's bilinear kernel is seen to be a function which is reconstructed from the central finite difference derived IF estimate. It now becomes apparent why the WVD yields good energy concentration for linear FM signals. Namely, the central finite difference estimator is known to be unbiased for such signals [3], and in the absence of noise, $\hat{f}_i(t,\tau) = f_i(t)$. Thus linear FM signals are transformed into sinusoids in the WVD kernel with the frequency of the sinusoid being equal to the instantaneous frequency of the signal, $z(t)$, at that value of time. Fourier transformation of the bilinear kernel then becomes

$$W_z(t,f) = \delta(f - f_i(t)) \tag{58}$$

that is, a row of delta functions along the true IF of the signal. The above equation is valid only for unit amplitude linear FM signals of infinite duration in the absence of noise. For non-linear FM signals a different formulation of the WVD has to be introduced in order to satisfy (58) under the same conditions.

4.3 The design of Polynomial Wigner-Ville distributions

The design of Polynomial WVDs which yield (58) for a non-linear FM signal, is based on replacing the central finite difference estimator, which is inherent in the definition of the WVD, cf.(57), by an estimator which would be unbiased for polynomial FM signals. The general theory of polynomial phase difference coefficients [69], [70] [71] describes the procedure for deriving unbiased IF estimators for arbitrary polynomial phase laws. It is presented below.

4.3.1 Phase difference estimators for polynomial phase laws of arbitrary order

Analysis of the problem. We consider now the discrete-time case where:

$$z(n) = Ae^{j\phi(n)} + w(n), \qquad n = 0, \ldots, N-1 \tag{59}$$

Practical requirements generally necessitate that the IF be determined from discrete time observations, and this means that a discrete approximation to the differentiation operation must be used. This is done by using an FIR differentiating filter. Thus for the discrete time signal whose phase is defined by

$$\phi(n) = \sum_{m=0}^{p} a_m n^m, \tag{60}$$

the IF is computed using the relation:

$$f_i(n) = \frac{1}{2\pi}\phi(n) * d(n) \tag{61}$$

where $d(n)$ is the differentiating filter, which leads to the following estimator:

$$\hat{f}_i(n) = \frac{1}{2\pi}\phi(n) * \hat{d}(n)$$

This section addresses the design of the differentiating filter $d(n)$. For phase laws which are linear or quadratic (i.e. for complex sinusoids or linear FM signals), the differentiating filter needs only to be a two tap filter. It is, in fact, a simple scaled phase difference, known as the *central finite difference*. As the order of the phase polynomial increases, so does the number of taps required in the filter. The filter then becomes a weighted sum of phase differences. The following derivation determines the exact form of these higher order phase difference based IF estimators.

Derivation of the estimates. For the discrete polynomial phase sequence given in (60) the IF is:

$$f_i(n) = \frac{1}{2\pi}\sum_{m=1}^{p} m a_m n^{m-1} \tag{62}$$

For a signal with polynomial phase of order p, a more generalised form of the phase difference estimator is required to perform the desired differentiation. It is defined as [72]:

$$\hat{f}_i^{(q)}(n) = \sum_{l=-q/2}^{q/2} \frac{1}{2\pi} d_l \phi(n+l) \tag{63}$$

where q is the order of the estimator. The d_l coefficients are to be found so that in the absence of noise, $\hat{f}_i^{(q)}(n) = f_i(n)$ for any n, that is:

$$\sum_{l=-q/2}^{q/2} d_l \phi(n+l) = \sum_{i=1}^{p} i a_i n^{i-1} \tag{64}$$

q	d_{-3}	d_{-2}	d_{-1}	d_0	d_1	d_2	d_3
2			-1/2	0	1/2		
4		1/12	-2/3	0	2/3	-1/12	
6	-1/60	3/20	-3/4	0	3/4	-3/20	1/60

Table 2: The values of differentiating filter coefficients for $q = 2, 4, 6$.

$$\sum_{l=-q/2}^{q/2} d_l \sum_{i=0}^{p} a_i \cdot (n+l)^i \quad = \quad \sum_{i=1}^{p} i\, a_i n^{i-1} \tag{65}$$

Because a polynomial's order is invariant to the choice of its origin, without any loss of generality we can set $n = 0$. Then (65) becomes

$$\sum_{l=-q/2}^{q/2} d_l \sum_{i=0}^{p} a_i l^i = a_1 \tag{66}$$

Then by equating coefficients for each of the a_i, we obtain $p+1$ equations. In matrix form this is given by:

$$Q\underline{d} = \underline{\xi} \tag{67}$$

where

$$Q = \begin{bmatrix} 1 & 1 & \dots & 1 \\ (-q/2) & (-q/2+1) & \dots & (q/2) \\ \dots & \dots & \dots & \dots \\ (-q/2)^p & (-q/2+1)^p & \dots & (q/2)^p \end{bmatrix} \tag{68}$$

$$\underline{d} = [d_{-q/2} \ \dots \ d_0 \ \dots \ d_{q/2}]^T \tag{69}$$

$$\underline{\xi} = [0\ 1\ 0\ \dots\ 0]^T \tag{70}$$

The matrix equation, (67), has a unique solution for $q = p$. The coefficients of the differentiating filter are given in Table 2, and are derived by solving the matrix equation (67) for $p = q = 2, 4, 6$. It is obviously most convenient if q is an even number. This will ensure that for estimating the IF at a given value of n, one does not need to evaluate the phase at non-integer sample points. That is, one does not need to interpolate the signal to obtain the required phase values. In practice, the use of estimators with odd valued q brings unnecessary implementational problems without any benefit in return. Therefore only even valued q estimators are used.

For the case where $p > q$, the matrix equation (67) can be approximately solved assuming that Q has a full rank. This represents an overdetermined problem, and the least-squares solution is given by:

$$\underline{d} = (Q^T Q)^{-1} Q^T \underline{\xi} \tag{71}$$

Choice of p and q. In analysing practical signals, the first task is to estimate the true order of the signal's polynomial phase p. This may involve a priori information, some form of training scheme, or even an educated guess. Once p has been estimated, the order q of the estimator $\hat{f}_i^{(q)}(n)$, has to be chosen. For an exact solution of (67), the rule is to chose q to be the least even number which is greater than p. In some situations, however, it may be preferable to use a lower value of q (and hance only to approximate the solution of eq.(67)) because the differentiating filter will be less susceptible to noise. This is due to the fact that as the polynomial order increases, there is increased likelihood that the noise will be modelled as well as the signal.

The next section uses these generalised (or polynomial) phase difference IF estimators, to replace the central finite difference based IF estimator which is built in to the WVD. The result of this replacement is a class of *polynomial WVDs* which ideally concentrate energy for polynomial phase signals.

4.3.2 Non-integer powers form for Polynomial WVDs (form I)

The q-th order unbiased IF estimator for polynomial phase signals can be expressed by [73]:

$$\hat{f}_i^{(q)}(t) = \frac{1}{2\pi\tau} \sum_{l=-q/2}^{q/2} d_l \, \phi(t + l\tau/q) \tag{72}$$

where $q \geq p$. Now it is straightforward to define Polynomial Wigner-Ville distributions with fractional powers as a generalisation of eq.(57):

$$W_z^{(q)}(t,f) = \underset{\tau \to f}{\mathcal{F}} \left\{ \exp\{j2\pi\tau \hat{f}_i^{(q)}(t,\tau)\} \right\} = \underset{\tau \to f}{\mathcal{F}} \{K_z^{(q)}(t,\tau)\} \tag{73}$$

where $\hat{f}_i^{(q)}(t,\tau)$ is the estimator given by eq.(72), centrally located about time instant, t. For a unit amplitude signal, i.e. $A = 1$ in (51), it follows from (73) and (72) that:

$$\begin{aligned} K_z^{(q)}(t,\tau) &= \exp\left\{ j \sum_{l=-q/2}^{q/2} d_l \, \phi(t + l\tau/q) \right\} = \prod_{l=-q/2}^{q/2} \exp\{j d_l \phi(t + l\tau/q)\} \\ &= \prod_{l=-q/2}^{q/2} [z(t + l\tau/q)]^{d_l} \end{aligned} \tag{74}$$

We refer to this form of Polynomial WVDs as the "fractional powers form" since the coefficients d_l are in general, rational numbers[6].

Example 1: Suppose the order of polynomial in (52) is $p = 2$ (linear FM signal). Then for $q = 2$ and $A = 1$, we get from (74):

$$K_z^{(2)}(t,\tau) = z(t+\tau/2)z^*(t-\tau/2) \tag{75}$$

Thus, the PWVD of order $q = 2$ is identical to the conventional WVD.

Example 2: Suppose $p = 3$ (quadratic FM) or $p = 4$ (cubic FM). Then if we set $q = 4$ (such that $q \geq p$) and for $A = 1$ we obtain from (74):

$$K_z^{(4)}(t,\tau) = \left[z^*(t+\frac{\tau}{2})\right]^{1/12} \left[z(t-\frac{\tau}{2})\right]^{1/12} \left[z(t+\frac{\tau}{4})\right]^{2/3} \left[z^*(t-\frac{\tau}{4})\right]^{2/3} \tag{76}$$

It is easy to verify that for cubic FM signals when $T \to \infty$,

$$W_z^{(4)}(t,f) = \delta(f - \frac{1}{2\pi}(a_1 + 2a_2 t + 3a_3 t^2 + 4a_4 t^3)) = \delta(f - f_i(t)) \tag{77}$$

Although derived for a polynomial phase signal model, the PWVD with a fixed value of q, (PWVD_q) can be used as a non-parametric analysis tool, in much the same way as the short-time Fourier transform or the conventional WVD is used.

Discrete implementation For computer implementation, it is necessary that the discrete form for the signal kernel and resulting polynomial WVD be used. The discrete form for the kernel and distribution are therefore presented here. The discrete form for the multilinear kernel is

$$K_z^{(q)}(n,m) = \prod_{l=-q/2}^{q/2} [z(n+lm)]^{d_l} \tag{78}$$

where $n = tf_s$, $m = \tau f_s$ and f_s is the sampling frequency assumed to be equal to 1 for simplification. The resulting time-frequency distribution is

[6]Note also that $K_z^{(q)}(t,\tau)$ is a *multi-linear kernel* if the coefficients d_l are integers. While the WVD's (bilinear) kernel transforms linear FM signals into sinusoids, the PWVD (multi-linear) kernel can be designed to transform higher order polynomial FM signals into sinusoids. These sinusoids manifest as delta functions about the IF when Fourier transformed. Thus the WVD may be interpreted as a method based on just the first order approximation in a polynomial expansion of phase differences.

given by:

$$W_z^{(q)}(n,k) = \underset{m \to k}{\mathcal{F}} \left\{ K_z^{(q)}(n,m) \right\} = \underset{m \to k}{\mathcal{F}} \left\{ \prod_{l=-q/2}^{q/2} [z(n+lm)]^{d_l} \right\} \tag{79}$$

where k is the discrete frequency variable.

Implementation difficulties. The implementation of the Polynomial WVD with signal, $z(n)$, raised to fractional powers, requires the use of phase unwrapping procedure (calculation of the phase sequence from the discrete-time signal $z(n)$). However, phase unwrapping performs well only for high SNRs and mono-component signals.

Since the implementation of the "non-integer powers" form of the PWVD is problematic and since its expected value cannot be interpreted as conventional time-varying polyspectra, we present an alternative form of the PWVD, where the signal, $z(n)$, is raised to integer powers.

4.3.3 Integer powers form for polynomial WVDs (form II)

The alternative way of implementing "unbiased" IF estimators for arbitrary polynomial phase laws requires that we weight the phases at unequally spaced samples and then take their sum. This allows the weights, (b_l), to be prespecified to integer values. The IF estimator of this type is defined as [56]:

$$\hat{f}_i^{(q)}(t,\tau) = \frac{1}{2\pi\tau} \sum_{l=-q/2}^{q/2} b_l \, \phi(t + c_l\tau) \tag{80}$$

Here c_l are coefficients which control the separation of the different phase values used to construct the IF estimator. Coefficients b_l and c_l may be varied to yield unbiased IF estimates for signals with an arbitrary polynomial FM law. The procedure for determining the b_l and c_l coefficients for the case $q = 4$ is illustrated in the Example 3, given below. While the b_l may theoretically take any values, they are practically constrained to be integers, since the use of integer b_l enables the expected values of the PWVD to be interpreted as time-varying higher order spectra. This important fact will make the form II of the PWVD preferable, and further on in the text, form II will be assumed unless otherwise stated.

The Polynomial Wigner-Ville distributions which result from incorporating the estimator in (80), are defined analogously to (73), again assuming

constant amplitude A. The multilinear kernel of the PWVD is given by

$$K_z^{(q)}(t,\tau) = \prod_{l=-q/2}^{q/2} [z(t+c_l\tau)]^{b_l} \tag{81}$$

The above expression for the kernel may be rewritten in a symmetric type form according to:

$$K_z^{(q)}(t,\tau) = \prod_{l=0}^{q/2} [z(t+c_l\tau)]^{b_l}\ [z^*(t+c_{-l}\tau)]^{-b_{-l}} \tag{82}$$

The discrete time version of the PWVD is given by the Discrete FT of:

$$K_z^{(q)}(n,m) = \prod_{l=0}^{q/2} [z(n+c_l m)]^{b_l}\ [z^*(n+c_{-l}m)]^{-b_{-l}} \tag{83}$$

where $n = tf_s$, $m = \tau f_s$ and f_s is the sampling frequency.

We have already mentioned earlier that the conventional WVD is a special case of the Polynomial WVD and may be recovered by setting $q = 2$, $b_{-1} = -1$, $b_0 = 0$, $b_1 = 1$, $c_{-1} = -1/2$, $c_0 = 0$,$c_1 = 1/2$.

Example 3: Design of the PWVD form II, for quadratic and cubic FM signals

Since $p = 3$ for quadratic FM, or $p = 4$ for cubic FM, we set $q = 4$ to account for both cases. The set of coefficients b_l and c_l must be found to completely specify the new kernel. In deciding on integer values to be assigned to the b_l it is also desired that the sum of all the $|b_l|$ be as small as possible. This criteria is used because the greater the sum of the $|b_l|$ the greater will be the deviation of the kernel from linearity, since the b_l coefficients which multiply the phases translate into powers of $z(t + c_l\tau)$. The extent of the multilinearity in the kernel should be limited as much as possible to prevent excessively poor performance of the PWVD in noise.

To be able to transform second order phase law signals into sinusoids (via the conventional WVD's kernel), it is known that the b_l must take on the values, $b_{-1} = -1$, $b_0 = 0$ and $b_1 = +1$. To transform third and fourth order phase law signals into sinusoids, it is necessary to incorporate two extra b_l terms (i.e. the phase differentiating filter must have two extra taps [69]). An attempt to adopt ± 1 for these two extra b_l terms values would fail since the procedure for determining a_l coefficients, eq. (89) would yield an inconsistent set of equations. As a consequence, the IF estimator would be biased. The simplest values that these terms can assume are $+2$ and -2,

and therefore the simplest possible kernel satisfying the criteria specified above is characterised by:

$$b_2 = -b_{-2} = 1, \quad b_1 = -b_{-1} = 2, \quad b_0 = 0 \tag{84}$$

The c_l coefficients must then be found such that the PWVD kernel transforms unit amplitude cubic, quadratic or linear frequency modulated signals into sine waves. The design procedure necessitates setting up a system of equations which relate the polynomial IF of the signal to the IF estimates obtained from the polynomial phase differences, and solving for the c_l. It is described below.

In setting up the design equations it is assumed that the signal phase in discrete-time form is a p-th order polynomial, given by:

$$\phi(n) = \sum_{i=0}^{p} a_i \, n^i \tag{85}$$

where the a_i are the polynomial coefficients. The corresponding IF is then given by [69]:

$$f_i(n) = \frac{1}{2\pi} \sum_{i=1}^{p} i \, a_i \, n^{i-1} \tag{86}$$

A q-th order phase difference estimator ($q \geq p$) is applied to the signal and it is required that, at any discrete-time index, n, the output of this estimator gives the true IF. The required system of equations to ensure this is:

$$\frac{1}{2\pi m} \sum_{l=-q/2}^{q/2} b_l \, \phi(n + c_l m) = f_i(n) \tag{87}$$

that is:

$$\frac{1}{2\pi m} \sum_{l=-q/2}^{q/2} b_l \sum_{i=0}^{p} a_i \, (n + c_l m)^i = \frac{1}{2\pi} \sum_{i=1}^{p} i \, a_i \, n^{i-1} \tag{88}$$

Note that because of the invariance of a polynomial's order to its origin, n may be set equal to zero without loss of generality. Setting n equal to zero in (88), then, yields

$$\frac{1}{m} \sum_{i=0}^{p} a_i \, m^i \sum_{l=-q/2}^{q/2} b_l \, c_l^i = a_1 \tag{89}$$

All of the a_i coefficients on the left and right hand side of (89) may be equated to yield a set of $p + 1$ equations. Doing this for the values of b_l

specified in (84) and for $p = q = 4$ yields:

$$a_0[1-1+2-2] = 0 \times a_0 \quad (90)$$

$$a_1[c_2 - c_{-2} + 2c_1 - 2c_{-1}] = 1 \times a_1 \quad (91)$$

$$a_2[c_2^2 - c_{-2}^2 + 2c_1^2 - 2c_{-1}^2] = 0 \times a_2 \quad (92)$$

$$a_3[c_2^3 - c_{-2}^3 + 2c_1^3 - 2c_{-1}^3] = 0 \times a_3 \quad (93)$$

$$a_4[c_2^4 - c_{-2}^4 + 2c_1^4 - 2c_{-1}^4] = 0 \times a_4 \quad (94)$$

It is obvious that (90) is always true, and if $c_1 = -c_{-1}$ and $c_2 = -c_{-2}$, eqns. (92) and (94) are satisfied too. This condition amounts to verifying the symmetry property of the FIR filter. Solving for c_1, c_{-1}, c_2 and c_{-2} then becomes straightforward from (91) and (93) subject to the condition that $c_1 = -c_{-1}$ and $c_2 = -c_{-2}$. The solution is:

$$c_1 = -c_{-1} = \frac{1}{2(2-2^{1/3})} \approx 0.675 \quad (95)$$

$$c_2 = -c_{-2} = -2^{1/3} c_1 \approx -0.85 \quad (96)$$

The resulting discrete-time kernel is then given by:

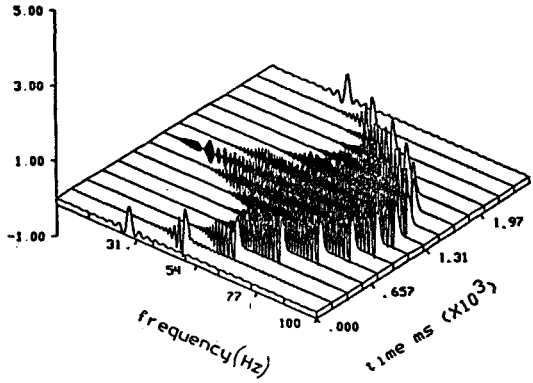
$$K_z^{(4)}(n,m) = [z(n+0.675m)\, z^*(n-0.675m)]^2\, z^*(n+0.85m)\, z(n-0.85m) \quad (97)$$

Note. It was recently shown that for $p = q = 4$, the solution given by (95) and (96) is just one of an infinite number of possible solutions. The full details appeared in [74].

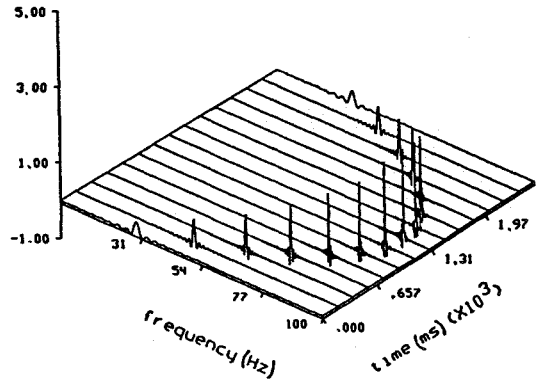
Fig.4(a) and 4(b) illustrate the conventional WVD and the $\text{PWVD}_{q=4}$ (form I or form II) of the same quadratic FM signal (noiseless case) respectively. The superior behaviour of the latter is indicated by the sharpness of the peaks in Fig.4(b). From the peaks of the PWVD the quadratic IF law can be recovered easily. The conventional WVD, on the other hand, shows many oscillations that tend to degrade its performance.

Implementation. Several important points need to be made concerning the practical implementation of the kernel in (97). Firstly, to form the discrete kernel one must have signal values at non-integer time positions. The signal must therefore be sampled or interpolated reasonably densely. The interpolation can be performed by use of an FFT based interpolation filter. Secondly, it is crucial to use analytic signals, so that the multiple artifacts between positive and negative frequencies are suppressed [24]. Thirdly, the PWVD is best implemented by actually calculating a frequency scaled version of the kernel in (97) and then accounting for the scaling in the Fourier transform operation on the kernel. That is, the PWVD is best implemented as

$$W_z^{(4)}(n,k) = \underset{m \to \frac{1}{0.85}k}{DFT} \left\{ [z(n+0.794m)z^*(n-0.794m)]^2\, z^*(n+m)z(n-m) \right\} \quad (98)$$



(a)



(b)

Figure 4.Time-frequency representations of a quadratic FM signal: (a) WVD; (b) $\mathrm{PWVD}_{q=4}$ (form I or form II)

This formulation, because it causes some of the terms within the kernel to occur at integer lags, reduces errors which arise from inaccuracies in the interpolation process.

4.4 Some properties of a class of PWVDs

The Polynomial WVD preserves and/or generalises most of the properties that characterise the WVD. In the derivation of the following properties we have assumed that the PWVD kernel is given by

$$K_z^{(q)}(t,\tau) = \prod_{l=1}^{q/2} [z(t+c_l\tau)]^{b_l}\ [z^*(t+c_{-l}\tau)]^{-b_{-l}} \tag{99}$$

and that the following conditions apply:

$$b_i = -b_{-i} \qquad i = 1,\ldots,q/2 \tag{100}$$

$$c_i = -c_{-i} \qquad i = 1,\ldots,q/2 \tag{101}$$

$$\sum_{i=1}^{q/2} b_i c_i = 1/2 \tag{102}$$

These limitations define the class of PWVDs we are considering. For consistency of notations used in higher-order spectra, it is important to introduce a parameter, which is alternative to the order, q. The parameter used here is defined as:

$$k = 2 \cdot \sum_{i=1}^{q/2} b_i \tag{103}$$

and corresponds to the order of multi-linearity of the PWVD kernel, or in the case of random signals, the order of polyspectra. Note that this represents a slight change of notation.

The following properties, are valid $\forall k \in \mathrm{N}$, and $\forall (t,f) \in \mathrm{R}^2$ (see Appendix C for proofs):

P-1. The PWVD is real for any signal $x(t)$:

$$\left[W_{\{x(t)\}}^{(k)}(t,f)\right]^* = W_{\{x(t)\}}^{(k)}(t,f) \tag{104}$$

P-2. The PWVD is an even function of frequency if $x(t)$ is real:

$$W_{\{x^*(t)\}}^{(k)}(t,-f) = W_{\{x(t)\}}^{(k)}(t,f) \tag{105}$$

P-3. A shift in time by t_0 and in frequency by f_0 (i.e. modulation by $e^{j2\pi f_0 t}$) of a signal $x(t)$ results in the same shift in time and frequency of the PWVD ($\forall t_0, f_0 \in R$):

$$W^{(k)}_{\{x(t-t_0)e^{j2\pi f_0(t-t_0)}\}}(t,f) = W^{(k)}_{\{x(t)\}}(t-t_0, f-f_0) \tag{106}$$

P-4. If $y(t) = w(t)z(t)$ then:

$$W^{(k)}_{\{y(t)\}}(t,f) = W^{(k)}_{\{w(t)\}}(t,f) *_f W^{(k)}_{\{z(t)\}}(t,f) \tag{107}$$

where $*_f$ denotes the convolution in frequency.

P-5. Projection of the $W^{(k)}_{\{x(t)\}}(t,f)$ to the time axis (time marginal):

$$\int_{-\infty}^{\infty} W^{(k)}_{\{x(t)\}}(t,f)df = |x(t)|^k \tag{108}$$

P-6. The local moment of the PWVD in frequency gives the instantaneous frequency of the signal $x(t)$:

$$\frac{\int_{-\infty}^{\infty} fW^{(k)}_{\{x(t)\}}(t,f)df}{\int_{-\infty}^{\infty} W^{(k)}_{\{x(t)\}}(t,f)df} = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{109}$$

P-7. Time-frequency scaling: for $y(t) = \sqrt[k]{|a|} \cdot x(at)$

$$W^{(k)}_{\{y(t)\}}(t,f) = W^{(k)}_{\{x(t)\}}(at, \frac{f}{a}) \tag{110}$$

P-8. Finite time support: $W^{(k)}_{\{x(t)\}}(t,f) = 0$ for t outside $[t_1, t_2]$ if $x(t) = 0$ outside $[t_1, t_2]$.

4.5 Instantaneous frequency estimation at high SNR

Consider signal $z(n)$ as given by (59) and (60), where additive noise $w(n)$ is complex white Gaussian.

Since the WVD is very effective for estimating the IF of signals with linear ($p = 2$) FM laws [3], a natural question which arises is whether the PWVD can be used for accurate estimation of the IF of non-linear polynomial FM signals in noise, as expected by construction. The peaks of the PWVD can in fact be used for computationally efficient IF estimation. This is shown in Appendix A for polynomial phase laws up to order $p = 4$. For higher order polynomial phase laws the SNR operational thresholds can become quite high and methods based on unwrapping the phase are often simpler [69].

Fig.5 summarises results for a quadratic ($p = 3$) FM signal in complex additive white Gaussian noise ($N = 64$ points), as specified in eqn.(59). The curves showing the reciprocal value of the mean square IF estimate error for $\text{PWVD}_{q=4}$ (solid line) and the WVD (dashed line) were obtained by Monte Carlo simulations and plotted against the Cramer-Rao (CR) lower variance bound for the IF [7]. One can observe that the $\text{PWVD}_{q=4}$ peak based IF estimates meet the CR bound at high SNRs and thus it shows that PWVD peak based IF estimates provide a very accurate means for IF estimation. On the other hand, the WVD peak based IF estimate is biased and that is why its mean square error (MSE) is always greater than the CR lower bound.

For polynomial phase laws of order higher than $p = 4$, the SNR threshold for PWVD based IF estimation becomes comparatively high. As mentioned earlier, in these circumstances, alternative computationally simpler methods based on unwrapping the phase (or a smoothed version of it) tend to be just as effective [3], [69].

The question remains as to how much we loose by choosing the order q of the PWVD higher than necessary. In Fig.6 we summarise the results for a linear FM signal ($p = 2$) in additive Gaussian noise ($N = 64$ points). The dashed curve shows the performance of the conventional WVD (or $\text{PWVD}_{q=2}$), while the solid line shows the inverse of the MSE curve for $\text{PWVD}_{q=4}$. Both curves were obtained by Monte Carlo simulations[8] and plotted against the CR bound (which is in this case about 8 dB higher than for $p = 3$, Fig.5). One can observe from Fig.6 that if the value of q is chosen higher than required ($q = 4$), the variance of the $\text{PWVD}_{q=4}$ based estimate never reaches the CR bound (it is about 8 dB below) and its SNR threshold appears at a higher SNR. This observation is not surprising since going to higher-order non-linearities always causes higher variances of estimates.

4.6 Higher order TFDs

In the same way as the WVD can be interpreted as the core of the class of bilinear time-frequency distributions, the PWVD can be used to define a class of multilinear or higher-order TFDs [56]. Alternative forms of higher-order TFDs, as extensions of Cohen's class, were proposed by several authors [76], [77], [78], [79]. Note that the general class of higher order TFDs can be defined in the multitime-multifrequency space, as a method for time-varying higher order spectral analysis. However, in our approach, we choose to project the full multidimensional space onto the t-f subspace, in order to obtain specific properties (such as t-f representation of polynomial FM signals). The implication of the projection operation is currently

[7] Expressions for CR bounds can be found in [69] and [67].

[8] The performance of the WVD peak IF estimator is also confirmed analytically in [75]

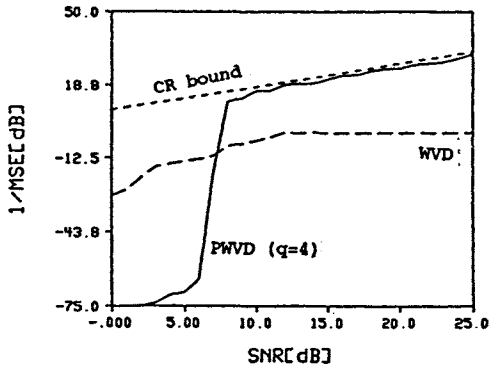


Figure 5. Statistical performance of the $\mathrm{PWVD}_{q=4}$ (solid line) and the WVD (dashed line) IF estimators vs CR bound. The signal is a *quadratic* FM in additive white Gaussian noise (64 points).

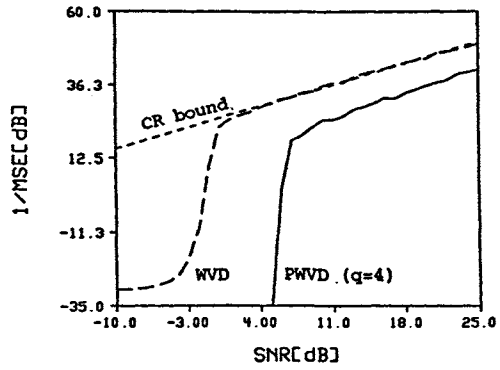


Figure 6. Statistical performance of the $\mathrm{PWVD}_{q=4}$ (solid line) and the WVD (dashed line) IF estimators vs CR bound. The signal is a *linear* FM in additive white Gaussian noise (64 points).

under investigation and further results are expected to appear in [74]. Section 6 will briefly discuss the case of multicomponent (or composite) signals. Before that, the next section will present one particular PWVD.

5 The Wigner-Ville trispectrum

This section presents a particular member of the class of polynomial TFDs which can solve a problem for which bilinear TFDs are ineffective: namely the analysis and representation of FM signals affected by multiplicative noise.

5.1 Definition

Many signals in nature and in engineering applications can be modelled as amplitude modulated FM signals. For example, in radar and sonar applications, in addition to the non-stationarity, the signal is subjected to an amplitude modulation which results in Doppler spreading. In communications, the change of the reflective characteristics of channel during the signal interval, causes amplitude modulation referred to as time-selective fading. Recently Dwyer [80] showed that a reduced form (or slice) of the trispectrum clearly reveals the presence of Gaussian amplitude modulated (GAM) tones. This was shown to be the case even if the noise was white. The conventional power spectrum is unable to perform this discrimination, because the white noise smears the spectrum. The Wigner-Ville distribution (WVD) would have the same limitation being a second-order quantity. A fourth order TFD, however, is able to detect not only GAM tones, but also GAM linear FM signals. Ideally one would like to detect GAM signals of arbitrarily high order polynomial phase signals. This, however, is beyond the scope of this chapter.

This extension of Dwyer's fourth order to a higher order TFD which could reveal GAM linear FM signals has since been called the *Wigner-Ville trispectrum* (WVT) [58], [7].

The WVT thus defined is a member of the class of Wigner-Ville polyspectra based on the PWVDs. The values of the parameters q, b_i and c_i can be derived by requiring that the WVT is an "optimal" t-f representation for linear FM signals and at the same time a fourth-order spectrum (as it name suggests). Furher discussion of these two requirements follows.

(i) $k = 4$

The fourth-order spectrum or the trispectrum, was shown [80] to be very effective for dealing with amplitude modulated sinusoids. The lowest value of q that can be chosen in (99) is $q = 2$. Then we have: $|b_1| + |b_{-1}| = k = 4$. In order to obtain a real WVT, condition (100) should be satisfied and thus we get: $b_1 = -b_{-1} = 2$.

(ii) Optimality for linear FM signals.

The WVT of a noiseless deterministic linear FM signal, $y(t)$, with a unit amplitude and infinite duration should give a row of delta impulses along the signal's instantaneous frequency:

$$W_y^{(4)}(t,f) = \delta(f - f_i(t)) \tag{111}$$

Suppose a signal, $y(t)$, is given by:

$$y(t) = e^{j2\pi(f_0+\alpha t)t} \tag{112}$$

Then:

$$K_y^{(4)}(t,\tau) = y^2(t+c_1\tau)[y^*(t+c_{-1}\tau)]^2 \tag{113}$$

that is:

$$\arg\{K_y^{(4)}(t,\tau)\} = 2\pi[f_0\tau(2c_1-2c_{-1})+2\alpha t\tau(2c_1-2c_{-1})+\alpha\tau^2(2c_1^2-2c_{-1}^2)] \tag{114}$$

In order to satisfy (111) notice that the following must hold:

$$\arg\{K_y^{(4)}(t,\tau)\} = 2\pi\tau(f_0+2\alpha t) \tag{115}$$

From (114) and (115) we obtain a set of two equations:

$$2c_1 - 2c_{-1} = 1 \tag{116}$$

$$2c_1^2 - 2c_{-1}^2 = 0 \tag{117}$$

Solving for c_1 and c_{-1} we get $c_1 = -c_{-1} = 1/4$. Thus the remaining two conditions for the properties of the PWVDs to be valid, namely (101) and (102), are thus satisfied.

Definition. The Wigner-Ville trispectrum of a random signal, $z(t)$, is defined as:

$$\mathcal{W}_z^{(4)}(t,f) = \mathcal{E}\left\{\int_{-\infty}^{\infty} z^2(t+\frac{\tau}{4})\,[z^*(t-\frac{\tau}{4})]^2 e^{-j2\pi f\tau}d\tau\right\} \tag{118}$$

where $\mathcal{E}$ is the expected value.

Note. The WVT is actually a reduced form of the full Wigner-Ville trispectrum that was defined in [81] and [79] as follows:

$$\begin{aligned}\mathcal{W}_z^{(4)}(t,f_1,f_2,f_3) &= \mathcal{E}\int_{\tau_1}\int_{\tau_2}\int_{\tau_3} z^*(t-\alpha_3)z(t-\alpha_3+\tau_1) \\ &\quad z(t-\alpha_3+\tau_2)z^*(t-\alpha_3+\tau_3)\prod_{i=1}^{3} e^{-j2\pi f_i\tau_i}d\tau_i\end{aligned} \tag{119}$$

where $\alpha_3 = (\tau_1 + \tau_2 + \tau_3)/4$. Equation (118) is obtained by selecting: $\tau_1 = \tau_2 = \tau/2$; $\tau_3 = 0$, and $f_1 = f_2 = f_3 = f$ in (119). For simplification, we use the term WVT to refer to the reduced form. The WVT satisfies all the properties listed in Sec.4.4. Its relationship with the signal group delay is given in Appendix B.

Cumulant based 4th order spectra. There are a number of ways of forming a cumulant based WVT. One definition has been provided in [76]. A second way to define the cumulant based WVT, assuming a zero-mean random signal, $z(t)$, was given in [74], where the time-varying fourth order cumulant function is given by:

$$\begin{aligned} C_z^{(4)}(t,\tau) &= \mathcal{E}\{z^2(t+\frac{\tau}{4})\,[z^*(t-\frac{\tau}{4})]^2\} - \\ &\quad 2\left[\mathcal{E}\{z(t+\frac{\tau}{4})z^*(t-\frac{\tau}{4})\}\right]^2 - \mathcal{E}\{z^2(t)\}\cdot\mathcal{E}\{z^*(t)^2\} \end{aligned}$$

The corresponding cumulant WVT then is defined as:

$$cW_z^{(4)}(t,f) = \underset{\tau \rightarrow f}{\mathcal{F}} \left\{C_z^{(4)}(t,\tau)\right\} \tag{120}$$

This definition has the advantage that it is a natural extension of Dwyer's fourth order spectrum, and hence can detect GAM linear FM signals.

5.2 Analysis of FM signals affected by Gaussian multiplicative noise

Let us assume a signal model:

$$z(t) = a(t)e^{j\,\phi(t)} \tag{121}$$

where $a(t)$ is a real zero-mean Gaussian multiplicative noise process with covariance function, $R_a(\tau) = v\lambda e^{-2\lambda|\tau|}$ and $\phi(t)$ is the phase of the signal, given by: $\phi(t) = 2\pi(\theta + f_0 t + \alpha t^2)$. The covariance function was chosen in order to describe a white noise process as $\lambda \rightarrow \infty$. The initial phase θ is a random variable uniformly distributed on $(-\pi, \pi]$. In radar the real Gaussian modulating process, $a(t)$, represents a model for a time-fluctuating target where the pulse length of transmitted signal is longer than the reciprocal of the reflection process [82].

The problem that we investigate is that of recovering the IF of the signal, $z(t)$. For this class of signals we show that for an asymptotically white process, $a(t)$, describing the envelope of $z(t)$, we have:

(A.) The expected value of the WVD is:

$$\mathcal{W}_z^{(2)}(t,f) = v \tag{122}$$

(B.) The Wigner-Ville trispectrum is:

$$\mathcal{W}_z^{(4)}(t,f) = v^2\lambda^2\delta(f - f_i(t)) + 2v^2\lambda \tag{123}$$

Proof: The power spectral density of $a(t)$ is $S_a(f) = v\,\lambda^2/(\lambda^2 + \pi^2 f^2)$. For $\lambda \to \infty$ (asymptotically white $a(t)$), $R_a(\tau) = v\delta(\tau)$ and $S_a(f) = v$.
(A.)

$$\begin{aligned}
\mathcal{K}_z^{(2)}(t,\tau) &= \mathcal{E}\{z(t+\tau/2)z^*(t-\tau/2)\} && (124)\\
&= R_a(\tau)e^{j\,2\pi(f_0+2\alpha t)\tau} && (125)\\
&= v\delta(\tau) \qquad (\lambda \to \infty) && (126)
\end{aligned}$$

The FT of $\mathcal{K}_z^{(2)}(t,\tau)$ with respect to τ gives (122).
(B.)

$$\begin{aligned}
\mathcal{K}_z^{(4)}(t,\tau) &= \mathcal{E}\{z^2(t+\tau/4)[z^*(t-\tau/4)]^2\} && (127)\\
&= [R_a^2(0) + 2R_a^2(\tau/2)]\; e^{j\,2\pi(f_0+2\alpha t)\tau} && (128)\\
&= [v^2\lambda^2 + 2v^2\lambda^2 e^{-2\lambda|\tau|}]\; e^{j\,2\pi(f_0+2\alpha t)\tau} && (129)\\
&= [v^2\lambda^2 + 2v^2\lambda\delta(\tau)]\; e^{j\,2\pi(f_0+2\alpha t)\tau} \quad (\lambda \to \infty) && (130)
\end{aligned}$$

Since the IF of the signal, $z(t)$, is $f_0 + 2\alpha t$, the FT of the above expression leads to (123).

In summary, the WVD of a signal, $z(t)$, affected by white multiplicative noise cannot describe the instantaneous frequency law of the signal in the time-frequency plane, while the WVT can. In order to confirm statements (A) and (B), the following experiment was performed:

Experiment. A complex signal, $e^{j\,2\pi(\theta+f_0 t+\alpha t^2)}$, is simulated with parameters, $f_0 = 50$ Hz, $\alpha = 47.25$ Hz/s, and where θ is a random variable uniformly distributed over $(-\pi, \pi]$. The sampling frequency and the number of samples are 400 Hz and 256 respectively. The signal is modulated by white Gaussian noise, and the real part of the resulting AM signal, $z(t)$, is shown in Fig.7. A gray-scale plot of the WVD (single realization) of the signal, $z(t)$, is shown in Fig.8(a). Notice that no clear feature can be extracted from this time-frequency representation. The WVT (single realization) of the same signal is presented in Fig.8(b). The linear time-varying frequency component appears clearly. Figs.9(a) and (b) show the WVD and the WVT (respectively) averaged over 10 realizations.

Additional results relevant to this section can be found in [74], where the cumulant WVT was studied and compared to the (moment) WVT defined in (118).

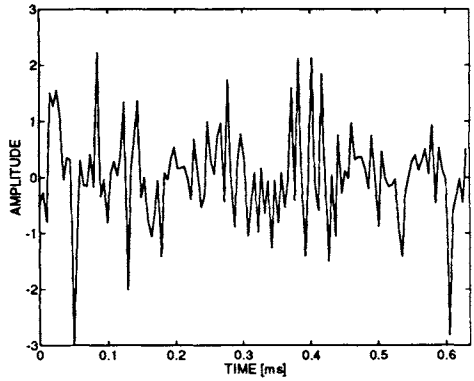


Figure 7. A linear FM signal modulated by white Gaussian noise

5.3 Instantaneous frequency estimation in the presence of multiplicative and additive Gaussian noise

This section discusses the problem of estimating the instantaneous frequency law of a discrete-time complex signal embedded in noise, as follows:

$$z(n) = a(n)e^{j\phi(n)} + w(n), \qquad n = 0, \ldots, N-1 \tag{131}$$

where $w(n)$ is complex white Gaussian noise with zero-mean and variance σ_w^2; $\phi(n)$ is the signal phase with a quadratic polynomial law:

$$\phi(n) = 2\pi(\theta + f_0 n + \alpha n^2) \tag{132}$$

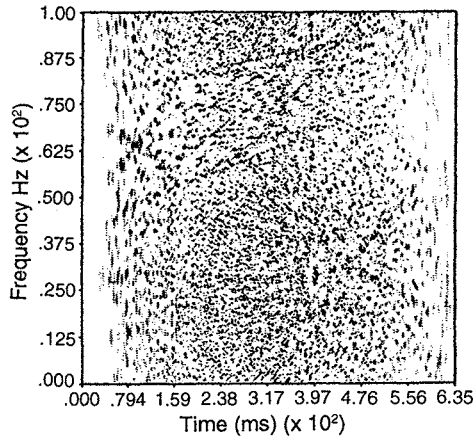
and $a(n)$ is real white Gaussian noise with mean, μ_a, and variance, σ_a^2, independent of $w(n)$. A further assumption is that only a single set of observations, $z(n)$, is available. The instantaneous frequency is estimated from the peak of the discrete WVT. Three separate cases are considered:

1. $a(n) = \mu_a = A = const$, that is $\sigma_a^2 = 0$;
2. $\mu_a = 0$; $\sigma_a^2 \neq 0$.
3. $\mu_a \neq 0$; $\sigma_a^2 \neq 0$.

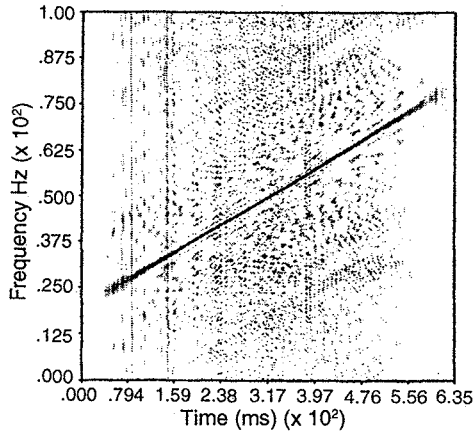
Case 3 describes a general case. In the first case, multiplicative noise is absent. In the second case the multiplicative noise plays a dominant role.

5.3.1 Performance of the estimator for case 1

Expressions for the variance of the estimate of the instantaneous frequency (IF), for signals given by (131) and (132) and for the case $a(n) = A$, are

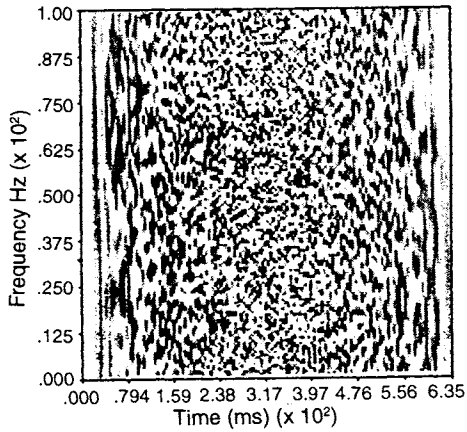


(a)

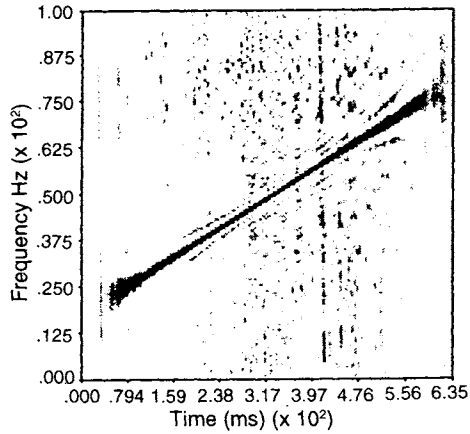


(b)

Figure 8. The WVD (a) and the WVT (b) of a linear FM modulated by white Gaussian noise (one realization)



(a)



(b)

Figure 9. The WVD (a) and the WVT (b) of a linear FM modulated by white Gaussian noise (ten realizations)

derived in [75] using the peak of the discrete WVD. These results are also confirmed in [83]. Following a similar approach, we derived expressions for the algorithm based on the peak of the WVT [59]. The variance of the WVT peak estimator of the IF (for a linear FM signal in additive white Gaussian noise) is shown to be [59]:

$$\sigma^2_{\hat{f}_i} = \frac{6\sigma_w^2}{A^2 N(N^2-1)(2\pi)^2} \tag{133}$$

This expression is identical to the CR lower bound of the variance of an estimate of the frequency for a stationary sinusoid in white Gaussian noise as given in [84]. The same result is obtained for a discrete WVD peak estimate [75]. As SNR decreases, a threshold effect occurs due to the non-linear nature of the estimation algorithm. As reported in [84], this threshold appears when $SNR_{DFT} \approx 15dB$, for a frequency estimate of a stationary sinusoid in noise. The threshold SNR for the discrete WVT peak estimate is shown to be [59]:

$$SNR = (27 - 10\log N)dB \tag{134}$$

This equation can be used to determine the minimum input SNR (for a given sample length, N) which is required for an accurate IF estimation.

Computer simulations were performed in order to verify the results given by eqs. (133) and (134). The results are shown in Fig.10 for $N = 128$ points. The x-axis shows the input SNR, defined as $10\log A^2/\sigma_w^2$. The y-axis represents the reciprocal of the mean-square error in dB. The curves for the WVD and WVT based estimates are obtained by averaging over 100 trials. Notice that the variance of the WVT based estimate meets the CR bound as predicted by (133) and that the threshold SNR for the WVT appears at $6dB$, as predicted by (134). This threshold is about $6dB$ higher than the threshold SNR for the WVD peak estimate, the higher threshold being due to the increased non-linearity of the kernel.

5.3.2 Performance of the estimator for case 2

Suppose that $a(n)$ is a real white Gaussian process with zero-mean and variance, σ_a^2, such that $a(n) \neq 0$, $(n = 0, \ldots, N-1)$. It is shown in [59] that the expression for the variance of the IF estimate is:

$$\sigma^2_{\hat{f}_i} = \frac{18\sigma_w^2}{(2\pi)^2 \sigma_a^2 N(N^2-1)} \tag{135}$$

Computer simulations have confirmed expression (135). The results are shown in Fig.11 for $N = 128$ points. The axes are the same as in Fig.10, except that the input SNR in dB is assumed to be $10\log\sigma_a^2/\sigma_w^2$.The curves for the WVT and WVD were obtained over 100 trials. We observe that the variance of the IF estimate given by (135) is three times ($4.7dB$) greater

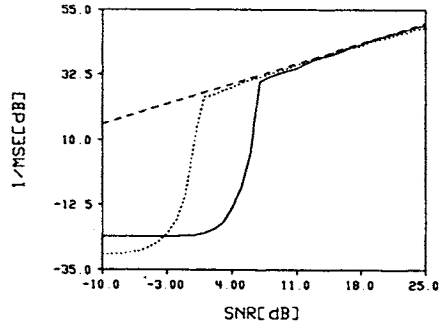


Figure 10. Statistical performance of the WVD (dotted line) and the WVT (solid line) vs CR bound (dashed line) for a constant amplitude linear FM signal in additive white Gaussian noise

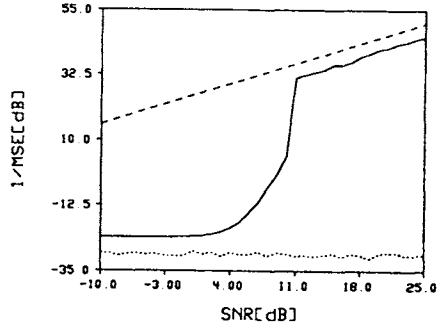


Figure 11. Statistical performance of the WVD (dotted line) and the WVT (solid line) vs CR bound (dashed line) for a linear FM signal modulated by real white zero-mean Gaussian noise and affected by additive white Gaussian noise

than the one expressed by (133). The SNR threshold for the WVT is at $10dB$.

5.3.3 Performance of the estimator for case 3

In the general case, where $a(n) \sim \mathcal{N}(\mu_a, \sigma_a)$, the expression for the variance of the IF estimate is given in [5] as:

$$\sigma^2_{\hat{f}_i} = \frac{6\,(3\sigma_a^4 + 6\mu_a^2\sigma_a^2 + \mu_a^4)\cdot\sigma_w^2}{(2\pi)^2(\mu_a^2+\sigma_a^2)^3\,N\,(N^2-1)} \tag{136}$$

This expression was confirmed by simulations, with the results being shown in Fig.12(a). There the reciprocal of the MSE is plotted as a function of the input SNR defined as: $10\log(\mu_a^2+\sigma_a^2)/\sigma_w^2$ and the quantity R defined as: $R = \sigma_a/(\sigma_a+\mu_a)$. Note that $R=0$ and $R=1$ correspond to the case 1 and 2 respectively. Fig.12(b) shows the behaviour of the WVD peak IF estimator for this case. One can observe that for $R > 0.25$ (i.e. $\mu_a < 3\sigma_a$) the WVT outperforms the WVD based IF estimator.

In summary, random amplitude modulation (here modelled as a Gaussian process) of an FM signal, behaves as multiplicative *noise* for second-order statistics, while in the special case of the fourth-order statistics it contributes to the signal power. In practical situations, the choice of the method (second- or fourth-order) depends on the input SNR and the ratio between the mean (μ_a) and the standard deviation (σ_a) of Gaussian AM.

6 Multicomponent signals and Polynomial TFDs

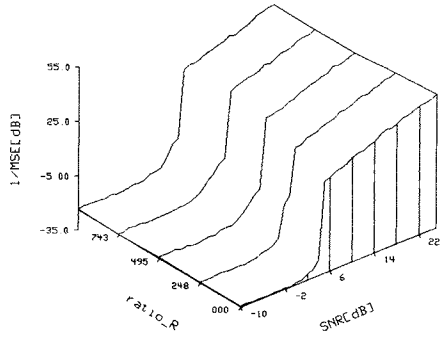
Until now we have considered only single component FM signals; that is, signals limited to only one time-varying feature in the frequency domain. In both natural and man-made signals, it is much more common to encounter *multicomponent* signals. It is therefore important to see if and how Polynomial TFDs can be used for multicomponent signal analysis. This is the problem which is investigated in this section[9].

6.1 Analysis of the cross-terms

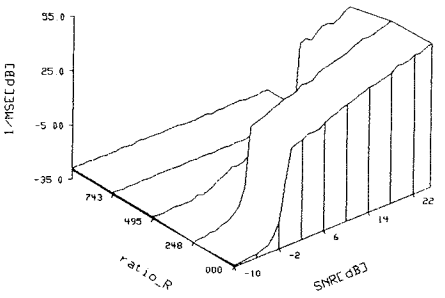
Let us consider a composite FM signal which may be modelled as follows:

$$z_M(t) = \sum_{i=1}^{M} a_i(t)e^{j[\Theta_i + 2\pi\int_0^t f_i(u)du]} = \sum_{i=1}^{M} y_i(t) \tag{137}$$

[9]The results presented in this section were obtained while finishing this manuscript



(a)



(b)

Figure 12. Statistical performance of the peak based IF estimator for a Gaussian multiplicative process with mean μ_a and variance σ_a^2, where $R = \sigma_a/(\sigma_a + \mu_a)$: (a) WVT; (b) WVD

where each $y_i(t)$ is an FM signal with random amplitude $a_i(t)$; Θ_i are random variables, such that $\Theta_i \sim \mathcal{U}[-\pi, \pi)$; and $a_i(t)$ and Θ_i are all mutually independent for any i and t. The (moment) WVT given by eq.(118), of $z_M(t)$ can be expressed as:

$$W^{(4)}_{z_M}(t,f) = \sum_{i=1}^{M} W^{(4)}_{y_i}(t,f) + \sum_{i=1}^{M} \sum_{j=1, j\neq i}^{M} W^{(4)}_{y_i,y_i,y_j,y_j}(t,f) + \\ +4\sum_{i=1}^{M} \sum_{j=i+1}^{M} W^{(4)}_{y_i y_j, y_i, y_j}(t,f) \tag{138}$$

while the cumulant WVT defined in (120) is given by:

$$cW^{(4)}_{z_M}(t,f) = \sum_{i=1}^{M} cW^{(4)}_{y_i}(t,f) \tag{139}$$

since all components $y_i(t)$ are zero-mean and mutually independent. Hence, only the moment WVT is affected by the cross-terms which are represented by the second and the third summand in (138). The cross-terms are artificially created by the method and they have no physical correspondent in the signal. Hence, the cross-terms are generally treated as undesirable in time-frequency analysis as discussed in section 3.3 [2]. The cross-terms of the (moment) WVT (138) can be divided into two groups:

- those given by:

$$\sum_{i=1}^{M} \sum_{j=1, j\neq i}^{M} W^{(4)}_{y_i,y_i,y_j,y_j}(t,f)$$

  which have *oscillatory* amplitude in the time-frequency plane.

These cross-terms can be suppressed by time-frequency smoothing of the WVT using methods equivalent to that of Choi-Williams or Zhao-Atlas-Marks [85]. These cross-terms correspond to well-studied cross-terms generated by quadratic t-f methods.

- those with *constant* amplitude in the t-f plane.

This class of cross-terms can be expressed in the form:

$$4\sum_{i=1}^{M} \sum_{j=i+1}^{M} W^{(4)}_{y_i y_j, y_i, y_j}(t,f) \tag{140}$$

and as (140) suggests, they have 4 time greater amplitude than the auto-terms, and frequency contents along:

$$f_{ij}(t) = \frac{f_i(t) + f_j(t)}{2} \qquad (i = 1, \ldots, M;\ \ j = i+1, \ldots, M)$$

Note that if $a_i(t)$ $(i = 1, \ldots, M)$ are white processes, then each cross-term:

$$W^{(4)}_{y_i y_j, y_i, y_j}(t, f) = \text{const}$$

Then these cross-terms will be spread in the entire t-f plane, rather than concentrated along $f_{ij}(t)$.

The most serious problem in the application of the moment WVT to composite FM signals is that of distinguishing these constant or "non-oscillating" cross-terms from the auto-terms.

In the next subsection, we consider methods for elimination of the "non-oscillating" cross-terms based on alternative forms of reducing the tri-frequency space to the frequency subspace.

6.2 Non-oscillating cross-terms and slices of the moment WVT

Consider the WVT, defined in the time–tri-frequency space (t, f_1, f_2, f_3) [81], of a deterministic signal $z_M(t)$. One definition is given in (119), and repeated here:

$$\begin{aligned} W^{(4)}_{z_M}(t, f_1, f_2, f_3) &= \int_{\tau_1}\int_{\tau_2}\int_{\tau_3} z^*_M(t - \alpha_3) z_M(t + \tau_1 - \alpha_3) z_M(t + \tau_2 - \alpha_3) \\ &\quad z^*_M(t + \tau_3 - \alpha_3) \prod_{i=1}^{3} e^{-j2\pi f_i \tau_i} d\tau_i \end{aligned} \tag{141}$$

where $\alpha_3 = (\tau_1 + \tau_2 + \tau_3)/4$. The WVT can be equivalently expressed in terms of $Z_M(f)$, FT of $z_M(t)$ as [81]:

$$\begin{aligned} W^{(4)}_{z_M}(t, f_1, f_2, f_3) &= \int_{\nu} Z^*_M(f_1 + f_2 + f_3 - \frac{\nu}{4})\, Z_M(f_1 + \frac{\nu}{4})\, Z_M(f_2 + \frac{\nu}{4}) \\ &\quad Z^*_M(-f_3 - \frac{\nu}{4})\, e^{j2\pi\nu t} d\nu \end{aligned} \tag{142}$$

We postulate that:

Postulate 1 *If the signal $z_M(t)$ has no overlap between its components in the* time *domain, then all slices of the WVT expressed by (141) such that:*

$$\tau_1 + \tau_2 - \tau_3 = \tau \tag{143}$$

are free from "non-oscillating cross-terms".

In addition, one can show that if

$$\tau_1 - \tau_2 \pm \tau_3 = 0 \tag{144}$$

then for any deterministic signal $z_1(t) = e^{j2\pi(f_0 t + \alpha t^2/2)}$, the WVT sliced as above yields: $W_{z_1}^{(4)} = \delta[f - (f_0 + \alpha t)]$. Obviously, the WVT defined by (118) can be derived from (141) satisfying both (143) and (144) by selecting $\tau_1 = \tau_2 = \tau/2$ and $\tau_3 = 0$. We refer to this form of the WVT (given below):

$$W_{z_M}^{(4)}(t,f) = \int_\tau [z_M(t + \frac{\tau}{4})]^2 \, [z_M(t - \frac{\tau}{4})]^2 e^{-j2\pi f\tau} d\tau \qquad (145)$$

as the *lag-reduced form.*

Postulate 2 *If the signal $z_M(t)$ has no overlap between its components in the* frequency *domain, then a (single) slice of the WVT expressed by (142) along $f_1 = f_2 = -f_3 = \Omega$ and given by:*

$$W_{z_M}^{(4)}(t,\Omega) = \int_\nu [Z_M(\Omega + \frac{\nu}{4})]^2 \, [Z_M^*(\Omega - \frac{\nu}{4})]^2 e^{j2\pi\nu t} d\nu \qquad (146)$$

is free from "non-oscillating cross-terms".

We refer to this form of the WVT as to the *frequency-reduced form.*

Example 1. Consider the particular deterministic composite signal given by

$$z_2(t) = e^{j2\pi F_1 t} + e^{j2\pi F_2 t} \qquad (147)$$

[$M = 2$, $a_i = 1$, $\Theta_i = 0$, $f_i(t) = F_i$ in (137)]. This is an example of a signal with non-overlapping frequency content. It is straightforward to show that:

$$\int_t W_{z_2}^{(4)}(t,f)dt = \delta(f - F_1) + \delta(f - F_2) + 4\delta[f - (F_1 + F_2)/2]$$

where the third summand above is the non-oscillating cross-term. Smoothing the WVT $W_{z_M}^{(4)}(t,f)$ (in the t-f plane) cannot eliminate this cross-term. On the contrary, the frequency reduced form of the WVT yields:

$$\int_t W_{z_M}^{(4)}(t,\Omega)dt = \delta(\Omega - F_1) + \delta(\Omega - F_2)$$

which is the correct result. Smoothing $W_{z_M}^{(4)}(t,\Omega)$ is necessary to suppress the oscillating cross-terms. Fig.13 illustrates a similar example of two *linear FM* signals with non-overlapping components in frequency. Smoothing of the WVT is performed using an adaptive kernel based on the Radon transform [86]. As postulate 1 claims, only the t-f representation in Fig.13(b) allows an accurate description and analysis of the signal.

Example 2. Consider a composite signal given by

$$z_2'(t) = \delta(t - T_1) + \delta(t - T_2)$$

which is dual to (147). This is an example of a signal with non-overlapping content in the time domain. One can show that:

$$\int_f W_{z_2'}^{(4)}(t,f)df = \delta(t-T_1) + \delta(t-T_2)$$

The frequency reduced form of the WVT yields:

$$\int_\Omega W_{z_2'}^{(4)}(t,\Omega)d\Omega = \delta(t-T_1) + \delta(t-T_2) + 4\delta[t-(T_1+T_2)/2]$$

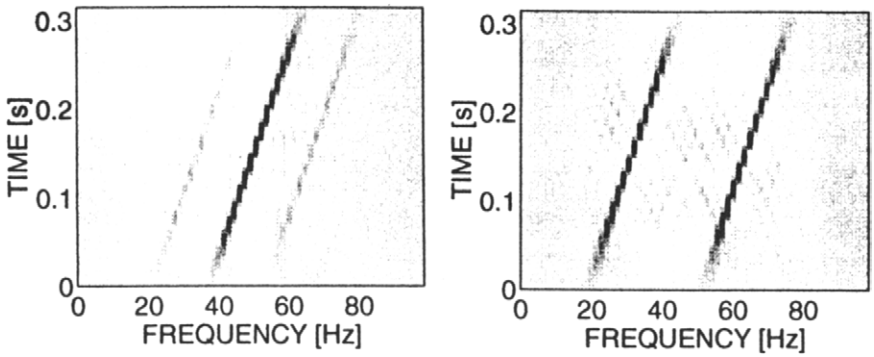
Fig.14 illustrates a similar example of two *linear FM* signals with non-overlapping components in the time domain. Smoothing of the WVT is performed by the same method as in Fig.13. As postulate 2 claims, only the t-f representation in Fig.14(a) allows an accurate description and analysis of the signal.

For general composite FM signals with possible time *and* frequency overlap it is necessary to initially perform an automatic segmentation of data[2] so that the problem is either reduced to the monocomponent case or to one of the two cases covered by the postulates stated above. The general case will appear elsewhere. Additional material related to multicomponent signals and PWVDs can be found in [5].

7 Conclusions

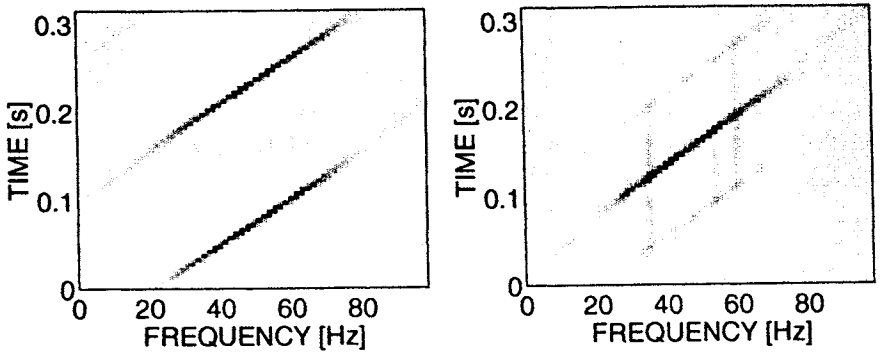
This chapter has presented a review of the important issues of time-frequency analysis, and an overview of recent advances based on multilinear representations.

It was shown in this chapter that the bilinear class of TFDs is suited only to the analysis of linear FM signals, i.e. for signals with a first order degree of non-stationarity. These bilinear TFDs, however, are not appropriate for the analysis of non-linear FM signals. For these signals, *Polynomial TFDs* have been proposed which are suitable for the analysis of such signals. In this chapter we have considered in particular, a sub-class of Polynomial TFDs, namely the Wigner-Ville trispectrum, which revealed to be a very efficient tool for the analysis of FM signals affected by multiplicative noise. The issue of multicomponent signal analysis using time-varying polyspectra has been briefly addressed.



(a) (b)

Figure 13. Smoothed moment WVT in (a) the lag-reduced form; (b) the frequency-reduced form of a signal with two linear FMs with frequency non-overlapping content



(a) (b)

Figure 14.Smoothed moment WVT in (a) the lag-reduced form; (b) the frequency-reduced form of a signal with two linear FMs with non-overlapping content in the time domain

Appendices

A Noise performance of the instantaneous frequency estimator for cubic FM signals

The following derivation was initiated by P. O'Shea [7], with some later modifications being done by B. Ristic and B. Boashash. Consider a constant amplitude second or third order polynomial FM signal, $z_s[n]$ embedded in white Gaussian noise. N samples of the observation are available, and the observed signal is given, after an amplitude normalisation, by:

$$z_r[n] = z_s[n] + z_w[n] = e^{j\phi[n]} + z_w[n] \tag{148}$$

where $\phi[n]$ is the time–varying phase function and $z_w[n]$ is complex white Gaussian noise of zero mean and variance $2\sigma^2$. Then the PWVD kernel defined in (98) is

$$\begin{aligned}
K_{z_r}[n,m] &= z_r^2[n+0.794m]\,(z_r^*[n-0.794m])^2\; z_r^*[n+m] z_r[n-m] \\
&= (z_s[n+0.794m] + z_w[n+0.794m])^2 \cdot \\
&\quad (z_s^*[n-0.794m] + z_w^*[n-0.794m])^2 \cdot \\
&\quad (z_s^*[n+m] + z_w^*[n+m]) \cdot (z_s[n-m] + z_w[n-m]) \\
&= z_s^2[n+0.794m]\,(z_s^*[n-0.794m])^2\; z_s^*[n+m]\, z_s[n-m] \\
&\quad + z_s^2[n+0.794m]\,(z_s^*[n-0.794m])^2\; z_s^*[n+m]\, z_w[n-m] \\
&\quad + z_s^2[n+0.794m]\,(z_s^*[n-0.794m])^2\; z_s^*[n-m]\, z_w[n+m] \\
&\quad + 2 z_s[n+0.794m]\, z_w[n+0.794m]\,(z_s^*[n-0.794m])^2 \\
&\quad \cdot z_s[n-m]\, z_s^*[n+m] \\
&\quad + 2 z_s[n-0.794m]\, z_w[n-0.794m]\,(z_s^*[n+0.794m])^2 \\
&\quad \cdot z_s[n-m]\, z_s^*[n+m] \\
&\quad + \ldots\ldots \\
&\quad + z_w^2[n+0.794m]\,(z_w^*[n-0.794m])^2\; z_w^*[n+m]\, z_w[n-m]
\end{aligned} \tag{149}$$

The kernel expansion in (149) shows three types of terms. The first (on line 1) is due to the signal, the second (on line 2–6) is due to the cross–terms between signal and noise, and the third (on line 7) is due to the noise. The term due to the signal is simply the expression for the PWVD kernel of a noiseless complex exponential; this has been seen in Example 3. Since it has amplitude, A^6, it will have power A^{12}. The power of the term due to the noise (only) is given by

$$NP_{noise} = \mathcal{E}\left\{|z_w[n+0.794m]|^4 |z_w[n-0.794m]|^4\; |z_w[n+m]|^2\; |z_w[n-m]|^2\right\} \tag{150}$$

Since the noise is zero-mean Gaussian, it can be shown that the above expression reduces to:

$$NP_{noise} = \begin{cases} 256\sigma^{12} & \text{if } m \neq 0 \\ 46080\sigma^{12} & \text{if } m = 0 \end{cases} \tag{151}$$

that is, the power of noise is not stationary (with respect to the lag index m). Note that this noise is white for $m \neq 0$.

The power of the second type of terms in (149), due to the cross-components between signal and noise, can be expressed as:

$$NP_{cross-terms} = \begin{cases} 12A^{10}\sigma^2 + 68A^8\sigma^4 + 224A^6\sigma^6 + \ldots & \text{if } m \neq 0 \\ 12A^{10}\sigma^2 + 120A^8\sigma^4 + 920A^6\sigma^6 + \ldots & \text{if } m = 0 \end{cases} \tag{152}$$

At high SNRs ($A^2 >> \sigma^2$) the power of cross-terms reduces to

$$NP_{cross-terms} \approx 12A^{10}\sigma^2 \tag{153}$$

The total noise power in the kernel

$$NP_{kernel} = NP_{noise} + NP_{cross-terms} \tag{154}$$

at high SNRs reduces to:

$$NP_{kernel} \approx 12A^{10}\sigma^2 \tag{155}$$

Since the PWVD is the Fourier transform of the kernel, it is in fact the Fourier transform of a complex exponential of amplitude A^6, in white noise of variance given by (155). To determine the variance of the PWVD peak based IF estimate, one may follow the approach Rao and Taylor [10] [75] used for the conventional WVD estimate, that is, one can use the formula for the variance of the DFT peak based frequency estimate. This variance for white noise at high SNR, is given by

$$var_{DFT}(\hat{f}) = \frac{6}{(2\pi)^2(SNR)(N^2-1)} \tag{156}$$

where the "SNR" term in the above equation is the SNR in the DFT. Now since the PWVD kernel is conjugate symmetric, at most only $N/2$ samples can be independent. Thus the SNR in the PWVD spectrum at high SNR is

$$SNR_{PWVD} = \frac{A^{12}(N/2)}{12A^{10}\sigma^2} \tag{157}$$

[10] A small correction has to be made in eq.(4) of [75]. Namely, the noise of $z(n+k)z^*(n-k)$ term is expressed by σ_z^4 which is correct only for $k \neq 0$.

Substituting this expression for SNR in (156), and introducing a variance reduction factor of 0.85^2 to account for the overall 0.85 frequency axis scaling, the following result is obtained

$$var_{PWVD}(\hat{f}) = \frac{2 \cdot 6 \cdot 12A^{10}\sigma^2(0.85)^2}{(2\pi)^2 A^{12} N(N^2-1)} = \frac{104.04\sigma^2}{(2\pi)^2 A^2 N(N^2-1)} \quad (158)$$

Now by comparison, the CR bound for a complex sinusoid in complex white Gaussian noise of variance, $2\sigma^2$, is given by:

$$var_{CR}(\hat{f}) = \frac{12\sigma^2}{(2\pi)^2 A^2(N^2-1)} \quad (159)$$

It can be seen that the $PWVD_4$ based variance in (158) corresponds to 8.67 times the CR lower variance bound for estimating a sinusoid in white Gaussian noise. That is, it is approximately 9 dB higher than the stationary CR bound. Additionally, there will need to be some adjustment in the variance in (158) due to a small degree of interdependence between the $N/2$ samples of the half-kernel. Simulation has shown, however, that this adjustment is negligible.

Thus the variance for the PWVD peak based IF estimate for a fourth order polynomial phase law is seen to be 9 dB higher than the CR bound for estimating the frequency of a *stationary tone*. The problem of determining the CR bound on IF estimates for *polynomial phase signals* has been addressed in [69], [67]. For fourth order polynomial phase signals the CR bound may be shown to be 9 dB higher than for the stationary bounds, i.e exactly the same as the variance of the PWVD peak based IF estimate at high SNR. Thus the PWVD peak based IF estimator yields estimates which meet the CR bound.

Fig.5 shows the actual variance of the PWVD peak based IF estimator plotted against the CR bound for 64 points. The correspondence is seen to be quite close at high SNR.

The approximate SNR "threshold" for the PWVD based IF estimator may be determined by noting that for stationary frequency estimation, deterioration occurs when the SNR in the DFT is 15dB [84]. Thus the approximate threshold for the PWVD (at high SNR) is given by

$$\frac{A^{12}(N/2)}{12A^{10}\sigma^2} \approx 15dB \quad (160)$$

or

$$\frac{A^2(N)}{2\sigma^2} \approx 26dB \quad (161)$$

The threshold of 8 dB seen in Fig.5 is exactly as predicted by eq.(161).

B Group delay of the WVT

The local moment of the WVT with respect to time is by definition:

$$< t >_f = \frac{\int_{-\infty}^{\infty} t W_x^{(4)}(t,f)dt}{\int_{-\infty}^{\infty} W_x^{(4)}(t,f)dt} \tag{162}$$

Consider a deterministic signal, $x(t) = a(t)e^{j\phi(t)}$. Its Fourier transform can be represented by $X(f) = A(f)e^{j\varphi(f)}$. The WVT is defined as:

$$W_x^{(4)}(t,f) = \int_{-\infty}^{\infty} [x(t+\frac{\tau}{4})]^2 \, [x^*(t-\frac{\tau}{4})]^2 e^{-j2\pi f\tau} d\tau \tag{163}$$

If we substitute $y(t) = x^2(t)$ it is easy to show that

$$W_x^{(4)}(t,f) = 2\int_{-\infty}^{\infty} Y(2f+\frac{\theta}{2}) \, Y^*(2f-\frac{\theta}{2}) e^{j2\pi\theta t} d\theta \tag{164}$$

After several lines of mathematical manipulations based on the same derivation as for the WVD [9], one can show that:

$$< t >_f = -\frac{1}{2\pi} \Im m \left\{ \frac{d}{df} \ln[X(2f) * X(2f)] \right\} \tag{165}$$

Now we can observe that the local moment in time of the WVT is equal to the *group delay* of the signal if and only if:

$$\arg\{X(2f) * X(2f)\} = \arg\{X(f)\} \tag{166}$$

The proof is given in [5]. Almost all practical signals do not satisfy condition (166).

C Properties of PWVDs

The PWVDs satisfy the following properties which were originally derived by B. Ristic, see [87]:

P-1. The PWVD is real for any signal $x(t)$:

$$\left[W_{\{x(t)\}}^{(k)}(t,f)\right]^* = W_{\{x(t)\}}^{(k)}(t,f) \tag{167}$$

Proof:

$$\begin{aligned} \left[W_{\{x(t)\}}^{(k)}(t,f)\right]^* &= \left[\int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [x(t+c_l\tau)]^{b_l} \, [x^*(t+c_{-l}\tau)]^{-b_{-l}} \, e^{-j2\pi f\tau} d\tau\right]^* \\ &= \int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [x^*(t+c_l\tau)]^{b_l} \, [x(t+c_{-l}\tau)]^{-b_{-l}} \, e^{+j2\pi f\tau} d\tau \end{aligned}$$

Substitution of τ by $-u$ yields:

$$\left[W^{(k)}_{\{x(t)\}}(t,f)\right]^* = -\int_{\infty}^{-\infty} \prod_{l=1}^{q/2} [x^*(t-c_l u)]^{b_l} [x(t-c_{-l}u)]^{-b_{-l}} e^{-j2\pi f u} du$$

Since coefficients b_i and c_i obey (100) and (101) we have:

$$\begin{aligned}\left[W^{(k)}_{\{x(t)\}}(t,f)\right]^* &= \int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [x^*(t+c_{-l}u)]^{-b_{-l}} [x(t+c_l u)]^{b_l} e^{-j2\pi f u} du \\ &= W^{(k)}_{\{x(t)\}}(t,f) \quad \blacksquare \end{aligned} \tag{168}$$

P-2. The PWVD is an even function of frequency if $x(t)$ is real:

$$W^{(k)}_{\{x^*(t)\}}(t,-f) = W^{(k)}_{\{x(t)\}}(t,f) \tag{169}$$

Proof:

$$W^{(k)}_{\{x(t)\}}(t,-f) = \int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [x(t+c_l\tau)]^{b_l} [x^*(t+c_{-l}\tau)]^{-b_{-l}} e^{-j2\pi(-f)\tau} d\tau$$

Substitution of τ by $-u$ yields:

$$\begin{aligned} W^{(k)}_{\{x(t)\}}(t,-f) &= -\int_{\infty}^{-\infty} \prod_{l=1}^{q/2} [x(t-c_l u)]^{b_l} [x^*(t-c_{-l}u)]^{-b_{-l}} e^{-j2\pi f u} du \\ &= W^{(k)}_{\{x(t)\}}(t,f) \end{aligned} \tag{170}$$

since coefficients b_i and c_i satisfy (100) and (101).■

P-3. A shift in time by t_0 and in frequency by f_0 (i.e. modulation by $e^{j2\pi f_0 t}$) of signal $x(t)$ results in the same shift in time and frequency of the PWVD ($\forall t_0, f_0 \in R$):

$$W^{(k)}_{\{x(t-t_0)e^{j2\pi f_0(t-t_0)}\}}(t,f) = W^{(k)}_{\{x(t)\}}(t-t_0, f-f_0) \tag{171}$$

Proof:

$$\begin{aligned} &W^{(k)}_{\{x(t-t_0)e^{j2\pi f_0(t-t_0)}\}}(t,f) = \\ &\int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [x(t-t_0+c_l\tau)]^{b_l} [x^*(t-t_0+c_{-l}\tau)]^{-b_{-l}} \\ &\cdot \exp\{j2\pi f_0[t\cdot\textstyle\sum_{l=1}^{q/2}(b_l+b_{-l}) + \tau\cdot\sum_{l=1}^{q/2}(c_l b_l + c_{-l}b_{-l})]\}\, e^{-j2\pi f\tau} d\tau \end{aligned}$$

If

$$\sum_{l=1}^{q/2}(b_l + b_{-l}) = 0 \tag{172}$$

and

$$\sum_{l=1}^{q/2}(c_l b_l + c_{-l}b_{-l}) = 1 \tag{173}$$

then:

$$W^{(k)}_{\{x(t-t_0)e^{j2\pi f_0(t-t_0)}\}}(t,f) = W^{(k)}_{\{x(t)\}}(t-t_0, f-f_0) \tag{174}$$

Note that (172) and (173) are satisfied if conditions (100), (101) and (102) are valid.■

P-4. If $y(t) = w(t)z(t)$ then:

$$W^{(k)}_{\{y(t)\}}(t,f) = W^{(k)}_{\{w(t)\}}(t,f) *_f W^{(k)}_{\{z(t)\}}(t,f) \tag{175}$$

where $*_f$ denotes the convolution in frequency.

Proof:

$$\begin{aligned} K^{(k)}_{\{y(t)\}}(t,\tau) &= \prod_{l=1}^{q/2}[y(t+c_l\tau)]^{b_l}[y^*(t+c_{-l}\tau)]^{-b_{-l}} \\ &= K^{(k)}_{\{w(t)\}}(t,\tau)\cdot K^{(k)}_{\{z(t)\}}(t,\tau) \end{aligned} \tag{176}$$

Now, the FT of the PWVD kernel:

$$W^{(k)}_{\{y(t)\}}(t,f) = \underset{\tau \to f}{\mathcal{F}}\left[K^{(k)}_{\{w(t)\}}(t,\tau)\cdot K^{(k)}_{\{z(t)\}}(t,\tau)\right] \tag{177}$$

$$= W^{(k)}_{\{w(t)\}}(t,f) *_f W^{(k)}_{\{z(t)\}}(t,f) \quad ■ \tag{178}$$

P-5. Projection of the $W^{(k)}_{\{x(t)\}}(t,f)$ onto the time axis (time marginal):

$$\int_{-\infty}^{\infty} W^{(k)}_{\{x(t)\}}(t,f)df = |x(t)|^k \tag{179}$$

Proof:

$$\underset{f \to \tau}{\mathcal{F}^{-1}}\left[W^{(k)}_{\{x(t)\}}(t,f)\right] = \int_{-\infty}^{\infty} W^{(k)}_{\{x(t)\}}(t,f)e^{j2\pi f\tau}df \tag{180}$$

$$= K^{(k)}_{\{x(t)\}}(t,\tau) \tag{181}$$

For $\tau = 0$ it becomes:

$$\int_{-\infty}^{\infty} W^{(k)}_{\{x(t)\}}(t,f)df = K^{(k)}_{\{x(t)\}}(t,0) \tag{182}$$

$$= \prod_{l=1}^{q/2}[x(t)]^{b_l}\cdot[x^*(t)]^{-b_{-l}} \tag{183}$$

$$= |x(t)|^k \tag{184}$$

since coefficients b_l satisfy (100) and k is defined by (103).■

P-6. The local moment of the PWVD with respect to frequency gives the instantaneous frequency of the signal $x(t)$:

$$\frac{\int_{-\infty}^{\infty} f W^{(k)}_{\{x(t)\}}(t,f)df}{\int_{-\infty}^{\infty} W^{(k)}_{\{x(t)\}}(t,f)df} = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{185}$$

Proof: The local moment of the PWVD in the frequency is:

$$< f >_t = \frac{\int_{-\infty}^{\infty} f W^{(k)}_{\{x(t)\}}(t,f)df}{\int_{-\infty}^{\infty} W^{(k)}_{\{x(t)\}}(t,f)df} \tag{186}$$

$$= \frac{\frac{1}{2\pi j}\frac{\partial K^{(k)}_{\{x(t)\}}(t,\tau)}{\partial \tau}|_{\tau=0}}{K^{(k)}_{\{x(t)\}}(t,\tau)\,|_{\tau=0}} \tag{187}$$

Since:

$$\frac{d\{\prod_{l=1}^{q/2}\Psi_l(\tau)\}}{d\tau} = \sum_{j=1}^{q/2}\left(\prod_{i=1,\, i\neq j}^{q/2}\Psi_i(\tau)\right)\cdot\frac{d\Psi_j(\tau)}{d\tau}$$

and assuming that coefficients b_i and c_i satisfy (100) and (101), it follows that:

$$\frac{\partial K^{(k)}_{\{x(t)\}}(t,\tau)}{\partial \tau}|_{\tau=0} = [x'(t)x^*(t) - x(t)\,(x^*(t))']\cdot|x(t)|^{k-2}\cdot\sum_{j=1}^{q/2}c_j b_j \tag{188}$$

Thus we have:

$$< f >_t = \frac{1}{2}\frac{1}{2\pi j}\frac{x'(t)x^*(t) - x(t)\,(x^*(t))'}{x(t)\,x^*(t)} \tag{189}$$

The eq.(189) is identical to the corresponding one obtained for the conventional WVD [9]. Thus it is straightforward to show that:

$$< f >_t = \frac{1}{2\pi}\mathcal{I}m\left\{\frac{d}{dt}[\ln\, z(t)]\right\} \tag{190}$$

For a complex signal, $z(t) = A(t)e^{j\phi(t)}$, the average frequency of the PWVD at time instant, t, is:

$$< f >_t = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \quad ■ \tag{191}$$

P-7. Time-frequency scaling: for $y(t) = \sqrt[k]{|a|}\cdot x(at)$

$$W^{(k)}_{\{y(t)\}}(t,f) = W^{(k)}_{\{x(t)\}}(at, \frac{f}{a}) \tag{192}$$

Proof:

$$W^{(k)}_{\{y(t)\}}(t,f) = \int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [y(t+c_l\tau)]^{b_l} [y^*(t+c_{-l}\tau)]^{-b_{-l}} e^{-j2\pi f\tau} d\tau \tag{193}$$

$$= a \int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [x(at+ac_l\tau)]^{b_l} [x^*(at+ac_{-l}\tau)]^{-b_{-l}} e^{-j2\pi f\tau} d\tau$$

Substitution of $a\tau$ by u yields:

$$W^{(k)}_{\{y(t)\}}(t,f) = \int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [x(at+c_l u)]^{b_l} [x^*(at+c_{-l}u)]^{-b_{-l}} e^{-j2\pi \frac{f}{a}u} du$$

$$= W^{(k)}_{\{x(t)\}}(at, \frac{f}{a}) \quad \blacksquare \tag{194}$$

P-8. Finite time support: $W^{(k)}_{\{x(t)\}}(t,f) = 0$ for t outside $[t_1, t_2]$ if $x(t) = 0$ outside $[t_1, t_2]$.

Proof: Suppose $t < t_1$. Since coefficients c_l satisfy (101) we have,

$$W^{(k)}_{\{x(t)\}}(t,f) = \int_{-\infty}^{\infty} \prod_{l=1}^{q/2} [x(t+c_l\tau)]^{b_l} [x^*(t-c_l\tau)]^{-b_{-l}} e^{-j2\pi f\tau} d\tau \tag{195}$$

$$= \int_{-\infty}^{0} \prod_{l=1}^{q/2} [x(t+c_l\tau)]^{b_l} [x^*(t-c_l\tau)]^{-b_{-l}} e^{-j2\pi f\tau} d\tau$$

$$+ \int_{0}^{\infty} \prod_{l=1}^{q/2} [x(t+c_l\tau)]^{b_l} [x^*(t-c_l\tau)]^{-b_{-l}} e^{-j2\pi f\tau} d\tau$$

$$= I_1 + I_2 \tag{196}$$

Integral $I_1 = 0$ since $x(t+c_l\tau) = 0$; integral $I_2 = 0$ since $x^*(t-c_l\tau) = 0$. Therefore $W^{(k)}_{\{x(t)\}}(t,f) = 0$. Similarly, for $t > t_2$, it can be shown that $W^{(k)}_{\{x(t)\}}(t,f) = 0$. $\blacksquare$

References

[1] B. Boashash. *Methods and Applications of Time-Frequency Signal Analysis.* Longman Cheshire, ISBN No. 0-13-007444-6, Melbourne, Australia, 1991.

[2] B. Boashash. Time-frequency signal analysis. In S. Haykin, editor, *Advances in Spectral Estimation and Array Processing*, volume 1 of 2, chapter 9, pages 418–517. Prentice Hall, Englewood Cliffs, New Jersey, 1991.

[3] B. Boashash. Interpreting and estimating the instantaneous frequency of a signal – Part I: Fundamentals. *Proceedings of the IEEE*, pages 519–538, April 1992.

[4] B. Boashash. Interpreting and estimating the instantaneous frequency of a signal – Part II: Algorithms. *Proceedings of the IEEE*, pages 539–569, April 1992.

[5] B. Boashash and B. Ristich. Polynomial Wigner-Ville distributions and time-varying polyspectra. In B. Boashash, E. J. Powers, and A. M. Zoubir, editors, *Higher Order Statistical Signal Processing*. Longman Cheshire, Melbourne, Australia, 1993.

[6] G. Jones. *Time-frequency analysis and the analysis of multicomponent signals*. PhD Thesis, Queensland University of Technology, Australia, 1992.

[7] P.J. O'Shea. *Detection and estimation methods for non-stationary signals*. PhD Thesis, University of Queensland, 1991.

[8] B. Boashash, B. Escudie, and J. M. Komatitsch. Sur la possibilite d'utiliser la representation conjointe en temps et frequence dans l'analyse des signaux modules en frequence emis en vibrosismiques. In *7th Symposium on Signal Processing and its Applications*, pages 121–126, Nice, France, 1979. GRETSI. in French.

[9] T.A.C.M. Classen and W.F.G. Mecklenbrauker. The Wigner distribution - Part I. *Phillips Journal of Research*, 35:217–250, 1980.

[10] T.A.C.M. Classen and W.F.G. Mecklenbrauker. The Wigner distribution - Part II. *Phillips Journal of Research*, 35:276–300, 1980.

[11] T.A.C.M. Classen and W.F.G. Mecklenbrauker. The Wigner distribution - Part III. *Phillips Journal of Research*, 35:372–389, 1980.

[12] B. Boashash, P. Flandrin, B. Escudie, and J. Grea. Positivity of time-frequency distributions. *Compte Rendus Acad. des Sciences de Paris*, Series A(t288):307–309, January 1979.

[13] B. Boashash. Wigner analysis of time-varying signals - Its application in seismic prospecting. In *Proceedings of EUSIPCO*, pages 703–706, Nuernberg, West Germany, September 1983.

[14] D. Gabor. Theory of communication. *Journal of the IEE*, 93:429–457, 1946.

[15] R. Lerner. Representation of signals. In E. Baghdady, editor, *Lectures on Communications System Theory*, pages 203–242. McGraw-Hill, 1990.

[16] C. Helstrom. An expansion of a signal into gaussian elementary signals. *IEEE Trans. Information Theory*, 13:344–345, 1966.

[17] I. Daubeshies. The wavelet transform: A method for time-frequency localisation. In S. Haykin, editor, *Advances in Spectral Estimation and Array Processing*, volume 1 of 2. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1990.

[18] C.H. Page. Instantaneous power spectra. *Journal of Applied Physics*, 23(1):103–106, 1953.

[19] C. Turner. On the concept of an instantaneous spectrum and its relation to the autocorrelation function. *Journal of Applied Puysics*, 25:1347–1351, 1954.

[20] M. Levin. Instantaneous spectra and ambiguity functions. *IEEE Transactions on Information Theory*, 13:95–97, 1967.

[21] A.W. Rihaczek. Signal energy distribution in time and frequency. *IEEE Transactions on Information Theory*, 14(3):369–374, 1968.

[22] J. Ville. Theorie et application de la notion de signal analytique. *Cables et Transmissions*, 2A(1):61–74, 1948.

[23] L. Cohen. Time-frequency distributions - A review. *Proceedings of the IEEE*, 77(7):941–981, July, 1989.

[24] B. Boashash. Note on the use of the Wigner distribution. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(9):1518–1521, September 1988.

[25] E.P. Wigner. On the quantum correction for thermodynamic equilibrium. *Physics Review*, 40:748–759, 1932.

[26] B. Boashash. *Representation Temps-Frequence.* Dipl. de Docteur-Ingenieur these, University of Grenoble, France, 1982.

[27] B. Boashash. Note D'information sur la representation des signaux dans le domaine temps-frequence. Technical Report 135 81, Elf-Aquitaine Research Publication, 1981.

[28] B. Boashash. Representation conjointe en temps et en frequence des signaux d'energie finie. Technical Report 373 78, Elf-Aquitaine Research Publication, 1978.

[29] W. Martin. Time-frequency analysis of random signals. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1325–1328, Paris, France, April 1982.

[30] G.F. Boudreaux-Bartels. *Time-frequency signal processing algorithms: Analysis and synthesis using Wigner distributions.* PhD Thesis, Rice University, Houston, Texas, 1983.

[31] B. Boashash and P. J. Black. An efficient real-time implementation of the Wigner-Ville distribution. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-35(11):1611–1618, November 1987.

[32] V.J. Kumar and C. Carroll. Performance of Wigner distribution function based detection methods. *Optical Engineering*, 23:732–737, 1984.

[33] S. Kay and G.F. Boudreaux-Bartels. On the optimality of the Wigner distribution for detection. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1263–1265, Tampa, Florida, USA, 1985.

[34] B. Boashash and F. Rodriguez. Recognition of time-varying signals in the time-frequency domain by means of the Wigner distribution. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 22.5.1–22.5.4, San Diego, USA, April 1984.

[35] B. Boashash and P. J. O'Shea. A methodology for detection and classification of some underwater acoustic signals using time-frequency analysis techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(11):1829–1841, November 1990.

[36] B. Boashash and P. J. O'Shea. Signal detection and classification by time-frequency distributions. In B. Boashash, editor, *Methods and Applications of Time-Frequency Signal Analysis*, chapter 12. Longman Cheshire,, Melbourne, Australia, 1991.

[37] B. Boashash and H. J. Whitehouse. High resolution Wigner-Ville analysis. In *11th GRETSI Symposium on Signal Processing and its Applications*, pages 205–208, Nice, France, June 1987.

[38] H. J. Whitehouse, B. Boashash, and J. M. Speiser. High resolution processing techniques for temporal and spatial signals. In *High Resolution Techniques in Underwater Acoustics*. Springer-Verlag, 1990. Lecture Notes in Control and Information Science.

[39] N. Marinovic. *The Wigner distribution and the ambiguity function: generalisations, enhancement, compression and some applications.* PhD Thesis, City University of New York, 1986.

[40] A.J. Janssen. *Application of the Wigner distribution to harmonic analysis of generalised stochastic processes.* PhD Thesis, Amsterdam, 1990.

[41] M. Amin. Time-frequency spectrum analysis and estimation for non-stationary random processes. In B. Boashash, editor, *Methods and Applications of Time-Frequency Signal Analysis*, chapter 9. Longman Cheshire,, Melbourne, Australia, 1992.

[42] Y. Zhao, L.E. Atlas, and R.J. Marks II. The use of cone-shaped kernels for generalised time-frequency representation of non-stationary singals. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 38(7), June 1990.

[43] I. Choi and W. Williams. Improved time-frequency representation of multi-component signals using exponential kernels. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(4):862–871, April 1990.

[44] P. Flandrin. Some features of time-frequency representations of multicomponent signals. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 41B.1.4–41B.4.4, San Diego, USA, 1984.

[45] P. J. Kootsookos, B. C. Lovell, and B. Boashash. A unified approach to the STFT, TFD's and instantaneous frequency. *IEEE Transactions on Acoustics, Speech and Signal Processing*, August 1991.

[46] G. Jones and B. Boashash. Instantaneous quantities and uncertainty concepts for signal dependent time frequency distributions. In Franklin T. Luk, editor, *Advanced Signal Processing Algorithms, Architectures and Implementations*, San Diego, USA, July 1991. Proceedings of SPIE.

[47] H.H. Szu. Two-dimensional optical processing of one-dimensional acoustic data. *Optical Engineering*, 21(5):804–813, September/October 1982.

[48] P. J. Boles and B. Boashash. Application of the cross Wigner-Ville distribution to seismic surveying. In B. Boashash, editor, *Methods and Applications of Time-Frequency Signal Analysis*, chapter 20. Longman Cheshire, Melbourne, Australia, 1992.

[49] D. L. Jones and T. W. Parks. A high resolution data-adaptive time-frequency representaion. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(12):2127–2135, December 1990.

[50] J. Bertrand and P. Bertrand. Time-frequency representations of broad-band signals. In *Proc. of Intern. Conf. on Acoust. Speech and Signal Processing*, pages 2196–2199, New York, USA, 1988.

[51] O. Rioul and P. Flandrin. Time-scale energy distributions: a general class extending wavelet transforms. *IEEE Transactions on Acoustics, Speech and Signal Processing*, pages 1746–1757, July 1992.

[52] R. Altes. Detection, estimation and classification with spectrograms. *Journal of the Acoustical Society of America*, 67:1232–1246, 1980.

[53] T.E. Posch. Kernels, wavelets and time-frequency distributions. *IEEE Transactions on Information Theory*. submitted.

[54] L. Cohen. Generalised phase space distributions. *Journal of Mathematical Physics*, 7:181–186, 1967.

[55] B. Boashash and A. P. Reilly. Algorithms for time-frequency signal analysis. In B. Boashash, editor, *Methods and Applications of Time-Frequency Signal Analysis*, chapter 7. Longman Cheshire,, Melbourne, Australia, 1991.

[56] B. Boashash and P. O'Shea. Time-varying higher order spectra. In Franklin T. Luk, editor, *Advanced Signal Processing Algorithms, Architectures and Implementations*, San Diego, USA, July 1991. Proceedings of SPIE.

[57] B. Boashash and P. J. O'Shea. Polynomial Wigner-Ville distributions and their relationship with time-varying higher spectra. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1993.

[58] B. Boashash and B. Ristich. Time-varying higher order spectra and the reduced Wigner trispectrum. In Franklin T. Luk, editor, *Advanced Signal Processing Algorithms, Architectures and Implementations*, volume 1770, pages 268–280, San Diego, USA, July 1992. Proceedings of SPIE.

[59] B. Boashash and B. Ristich. Analysis of FM signals affected by gaussian AM using the reduced Wigner-Ville trispectrum. In *Proc. of the Intern. Conf. Acoustic, Speech and Signal Processing*, Mineapolis, Minnesota, April, 1993.

[60] J.F. Randolf. *Basic Real and Abstract Analysis*. Academic Press, New York, 1968.

[61] S. M. Kay. *Modern Spectral Estimation: Theory and Application*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1987.

[62] R. Altes. Sonar for generalized target description and its similarity to animal echolocation systems. *J. Acoust. Soc. Am.*, 59(1):97–105, Jan. 1976.

[63] A.W. Rihaczek. *Principles of High-Resolution Radar*. Pensinsula Publishing, Los Altos, 1985.

[64] A. Dziewonski, S. Bloch, and M. Landisman. A technique for the anlysis of the transient signals. *Bull. Seismolog. Soc. Am.*, pages 427–449, Feb. 1969.

[65] B. Ferguson. A ground based narrow-band passive acoustic technique for estimating the altitude and speed of a propeler driven aircraft. *J. Acoust. Soc. Am.*, 92(3), September 1992.

[66] S. Peleg and B. Porat. Estimation and classification of polynomial phase signals. *IEEE Trans. Information Theory*, 37:422–429, March 1991.

[67] S. Peleg and Porat. The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase. *IEEE Transactions on Signal Processing*, 39(3):749–752, March 1991.

[68] Z. Faraj and F. Castanie. Polynomial phase signal estimation. In *Signal Processing IV: Theories and Applications*, pages 795–798, Proc. EUSIPCO-92, August 1992.

[69] B. Boashash, P. J. O'Shea, and M. J. Arnold. Algorithms for instantaneous frequency estimation: A comparative study. In Franklin T. Luk, editor, *Advanced Signal Processing Algorithms, Architectures and Implementations*, pages 24–46, San Diego, USA, August 1990. Proceedings of SPIE 1348.

[70] M. Arnold. *Estimation of the instantaneous parameters of a signal.* Thesis, University of Queensland, Australia, 1991.

[71] T. Soderstrom. On the design of digital differentiating filters. Technical Report UPTEC 8017, University of Technology, Uppsala University, March 1980.

[72] M. J. Arnold and B. Boashash. The generalised theory of phase difference estimators. *IEEE Tran. Signal Processing*, 1993. Submitted.

[73] B. Boashash and B. Ristic. Time varying higher order spectra. In *Proceedings of 25th Asilomar Conference*, Pacific Grove, California, Nov. 1991.

[74] B. Boashash and B. Ristich. Application of cumulant tvhos to the analysis of composite fm signals in multiplicative and additive noise. In Franklin T. Luk, editor, *Advanced Signal Processing Algorithms, Architectures and Implementations*, San Diego, USA, July 1993. Proceedings of SPIE.

[75] P. Rao and F. J. Taylor. Estimation of instantaneous frequency using the Wigner distribution. *Electronics Letters*, 26:246–248, 1990.

[76] A. Dandawate and G. B. Giannakis. Consistent kth order time-frequency representations for (almost) cyclostationary processes. In *Proc. Ann. Conference on Information Sciences and Systems*, pages 976–984, Johns Hopkins University, March, 1991.

[77] J. R. Fonollosa and C. L. Nikias. General class of time-frequency higher-order spectra: Definitions, properties, computation and applications to transient signals. In *Proc. Int. Signal Processing Workshop on Higher-Order Statistics*, pages 132–135, Chamrousse, France, July 1991.

[78] A. Swami. Third-order Wigner distributions: definitions and properties. In *Proc. Int. Conf Acoustic, Speech and Signal Processing (ICASSP)*, pages 3081–4, Toronto, Canada, May, 1991.

[79] P. O. Amblard and J. L. Lacoume. Construction of fourth-order Cohen's class: A deductive approach. In *Proc. Int. Symp. Time-Frequency Time-Scale Analysis*, Victoria, Canada, October 1992.

[80] R. F. Dwyer. Fourth-order spectra of Gaussian amplitude-modulated sinusoids. *J. Acoust. Soc. Am.*, pages 919–926, August 1991.

[81] J.R. Fonollosa and C.L.Nikias. Analysis of transient signals using higher-order time- frequency distributions. In *Proc. Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, pages V–197 - V–200, San Francisco, March, 1992.

[82] H. Van Trees. *Detection, Estimation and Modulation Theory: Part III.* John Wiley, New York, 1971.

[83] B. Boashash, P. J. O'Shea, and B. Ristic. A statistical/computational comparison of some algorithms for instantaneous frequency estimation. In *ICASSP*, Toronto, May, 1991.

[84] D.C. Rife and R.R. Boorstyn. Single tone parameter estimation from discrete-time observations. *IEEE Transactions on Information Theory*, 20(5):591–598, 1974.

[85] F. Hlawatsch and G. F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9(2):21–67, April 1991.

[86] B. Ristic and B. Boashash. Kernel design for time-frequency analysis using Radon transform. *IEEE Transactions on Signal Processing*, 41(5):1996–2008, May 1993.

[87] B. Ristic. *Adaptive and higher-order time-frequency analysis methods for nonstationary signals.* PhD Thesis, Queensland University of Technology, Australia, (to appear).