Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 673539, 43 pages
doi:10.1155/2009/673539

Review Article

Techniques to Obtain Good Resolution and Concentrated Time-Frequency Distributions: A Review

Imran Shafi,[1,2] Jamil Ahmad,[1,2] Syed Ismail Shah,[1,2] and F. M. Kashif[1,2,3]

[1] *Centre for Advanced Studies in Engineering (CASE), G-5/2 Islamabad, Pakistan*
[2] *Iqra University, H-9 Islamabad, Pakistan*
[3] *Laboratory for Electromagnetic and Electronic Systems (LEES), MIT, Cambridge, MA 02139, USA*

Correspondence should be addressed to Imran Shafi, imran.shafi@gmail.com

Received 12 July 2008; Revised 13 December 2008; Accepted 23 April 2009

Recommended by Ulrich Heute

We present a review of the diversity of concepts and motivations for improving the concentration and resolution of time-frequency distributions (TFDs) along the individual components of the multi-component signals. The central idea has been to obtain a distribution that represents the signal's energy concentration simultaneously in time and frequency without blur and crosscomponents so that closely spaced components can be easily distinguished. The objective is the precise description of spectral content of a signal with respect to time, so that first, necessary mathematical and physical principles may be developed, and second, accurate understanding of a time-varying spectrum may become possible. The fundamentals in this area of research have been found developing steadily, with significant advances in the recent past.

1. Introduction and Historical Perspective

The signals with time-dependant spectral content (STSC) are commonly found in nature or are self-generated for many reasons. The processing of such signals forms the basis of many applications including analysis, synthesis, filtering, characterization or modeling, suppression, cancellation, equalization, modulation, detection, estimation, coding, and synchronization [1]. For a practical application, the STSC can be processed in various ways, other than time-domain, to extract useful information. A classical tool is the Fourier transform (FT) which offers perfect spectral resolution of a signal. However FT possesses intrinsic limitations that depend on the signal to be processed. The instantaneous frequency (IF) [2, 3], generally defined as the first conditional moment in frequency $\langle\omega\rangle_t$, is a useful concept for describing the changing spectral structure of the STSC. A signal processing engineer is mostly confronted with the task of processing frequencies of spectral peaks which require unambiguous and accurate information about the IFs present in the signals. This has made the IF a parameter of practical importance in situations such as seismic, radar, sonar, communications, and biomedical application [2–6].

The introduction of time-frequency (t-f) signal processing has led to represent and characterize the STSC' time-varying contents using TFDs [7, 8]. The TFDs are two-dimensional (2D) functions which provide simultaneously, the temporal and spectral information and thus are used to analyze the STSC. By distributing the signal energy over the t-f plane, the TFDs provide the analyst with information unavailable from the STSC' time or frequency domain representation alone. This includes the number of components present in the signal, the time durations, and frequency bands over which these components are defined, the components' relative amplitudes, phase information, and the IF laws that components follow in the t-f plane. There has been a great surge of activity in the past few years in t-f signal processing domain. The pioneering work is performed by Claasen and Mecklenbrauker [9–11], Janse and Kaizer [12], and Bouachache [13]. They provided the initial impetus, demonstrated useful methods for implementation, and developed ideas uniquely suited to the t-f situation. Also, they innovatively and efficiently made use of the similarities and differences of signal processing fundamentals with quantum mechanics. Claasen and Mecklenbrauker devised many new ideas, procedures and developed a comprehensive

approach for the study of joint distribtutions [9–11]. However Bouachache [13] is believed to be the first researcher, who utilized various distributions for real-world problems. He developed a number of new methods and particularly realized that a distribution may not behave properly in all respects or interpretations, but it could still be used if a particular property such as IF is well described. Flandrin and Escudie [14] and coworkers transcribed directly some of the early quantum mechanical results, particularly the work on the general class of distributions [15, 16] into signal analysis language. The work by Janse and Kaizer [12] developed innovative theoretical and practical techniques for the use of TFDs and introduced new methodologies remarkable in their scope.

Historically the spectrogram [17–23] has been the most widely used tool for the analysis of time-varying spectra and is currently the standard method for the study of nonstationary signals, which is expressed mathematically as the magnitude-square of the short-time Fourier transform (STFT) of the signal, given by

$$S(t,\omega) = \left| \int x(\tau) h(t-\tau) e^{-i\omega\tau} d\tau \right|^2, \tag{1}$$

where $x(t)$ is the signal and $h(t)$ is a window function (throughout the paper that follows, we use both i and j for $\sqrt{-1}$ depending on notational requirements and the limits for $\int$ are from $-\infty$ to ∞, unless otherwise specified). The spectrogram has severe drawbacks, both theoretically, since it provides biased estimators of the signal IF and group delay (GD), and practically, since the Gabor-Heisenberg inequality [24] makes a tradeoff between temporal and spectral resolutions unavoidable. However STFT and its variation, being simple and easy to manipulate, are still the primary methods for analysis of the STSC and most commonly used today.

There are alternative approaches [7, 8, 25] with a motivation to improve upon the important shortcomings of the spectrogram, with an objective to clarify the physical and mathematical ideas needed to understand time-varying spectrum. These techniques generally aim at devising a joint function of time and frequency, a distribution that will be highly concentrated along the IFs present in a signal and cross-terms (CTs) free thus exhibiting good resolution. One form of TFD can be formulated by the multiplicative comparison of a signal with itself, expanded in different directions about each point in time. Such formulations are known as quadratic TFDs (QTFDs) because the representation is quadratic in the signal. This formulation was first described by Wigner in quantum mechanics [26] and introduced in signal analysis by Ville [27] to form what is now known as the Wigner-Ville distribution (WD). The WD is the prototype of distributions that are qualitatively different from the spectrogram, and produces the ideal energy concentration along the IF for linear frequency modulated (FM) signals, given by

$$W(t,\omega) \triangleq \frac{1}{2\pi} \int s^*\left(t - \frac{1}{2}\tau\right) s\left(t + \frac{1}{2}\tau\right) e^{-i\tau\omega} d\tau, \tag{2}$$

where $s(t)$ is the signal, the distribution is said to be bilinear in the signal because the signal enters twice in its calculation. It possesses a high resolution in the t-f plane, and satisfies a large number of desirable theoretical properties [1, 28]. It can be argued that more concentration than in the WD would be undesirable in the sense that it would not preserve the t-f marginals.

It is found that the spectrogram results in a blurred version [1, 3], which can be reduced to some degree by the use of an adaptive window or by the combination of spectrograms. On the other hand, the use of WD in practical applications is limited by the presence of nonnegligible CTs, resulting from interactions between signal components. These CTs may lead to an erroneous visual interpretation of the signal's t-f structure, and are also a hindrance to pattern recognition, since they may overlap with the searched t-f pattern. Moreover, if the IF variations are nonlinear, then the WD cannot produce the ideal concentration. Such impediments, pose difficulties in the STSC' correct analysis, are dealt in various ways and historically many techniques are developed to remove them partially or completely. They were partly addressed by the development of the Choi and Williams distribution [29] in 1989, followed by numerous ideas proposed in literature with an aim to improve the TFDs' concentration and resolution for practical analysis [3, 30–33]. Few other important nonstationary representations among the Cohen's class [1, 15, 34] of bilinear t-f energy distributions include the Margenau and Hill distribution [35], their smoothed versions [9–11, 36, 37] with reduced CTs [29, 38–40] are members of this class. Nearly at the same time, some authors also proposed other time-varying signal analysis tools based on a concept of scale rather than frequency, such as the scalogram [41, 42] (the squared modulus of the wavelet transform), the affine smoothed pseudo-WD (PWD) [43], or the Bertrand distribution [44]. The theoretical properties and the application fields of this large variety of these existing methods are now well determined, and wide-spread [1, 9–11, 28]. Although many other QTFDs have been proposed in literature (an alphabatical list can be found in [45]), no single QTFD can be effectively used in all possible applications. This is because different QTFDs suffer from one or more problems.

Nevertheless, a critical point of these methods is their readability, which means both a good concentration of the signal components and no misleading interference terms. This characteristic is necessary for an easy visual interpretation of their outcomes and a good discrimination between known patterns for nonstationary signal classification tasks. An ideal TFD function roughly requires the following four properties.

(1) *High clarity* which makes it easier to be analyzed. This require high concentration and good resolution along the individual components for the multicomponent signals. Consequently the resultant TFDs are deblurred.

(2) *CTs' elimination* which avoids confusion between noise and real components in a TFD for nonlinear t-f structures and multicomponent signals.

Table 1: Synthesis of main problems related to QTFDs.

Synthesis of major concerns	*Gabor transform*	*WD*	*Gabor-Wigner transform*
Clarity	Worst	Best	Reasonably Good
CTs	Nil	Present for multicomponent signals and nonlinear t-f structures	Almost eliminated
Mathematical properties	Unsatisfactory	Satisfactory	Good
Computational complexity	Quite Low	High	Higher

(3) *Good mathematical properties* which benefit to its application. This requires that TFDs to satisfy total energy constraint, marginal characteristics and positivity issue, and so forth. Positive distributions are everywhere nonnegative, and yield the correct univariate marginal distributions in time and frequency.

(4) *Lower computational complexity* means the time needed to represent a signal on a t-f plane. The signature discontinuity and weak signal mitigation may increase computation complexity in some cases.

A comparison of some popular TFD functions is presented in Table 1. To analyze the signals well, choosing an appropriate TFD function is important. Which TFD function should be used depends on what application it applies on. On the other hand, the short comings make specific TFDs suited only for analyzing STSC with specific types of properties and t-f structures. An obvious question then arise that which distribution is the "best" for a particular situation. Generally there is an attempt to set up a set of desirable conditions and to try to prove that only one distribution fits them. Typically, however, the list is not complete with the obvious requirements, because the author knows that the added desirable properties would not be satisfied by the distribution he/she is advocating. Also these lists very often contain requirements that are questionable and are obviously put in to force an issue. As an illustration, by focusing on the WD and its variants, Jones and Parks [46] have made an interesting comparative study of the resolution properties and have shown that the relative performance of the various distributions depends on the signal. The results show that the pseudo-WD (PWD) is best for the signals with only one frequency component at any one time, the Choi-Williams distribution is most attractive for multicomponent signals in which all components have constant frequency content, and the matched filter STFT is best for signal components with significant frequency modulation. Jones and Parks have concluded that no TFD can be considered as the best approach for all t-f analysis and both concentration and resolution cannot be improved at one time.

In this paper, we will briefly discuss the basic concepts and well-tested algorithms to obtain highly concentrated and good resolution TFDs for an interested reader (although new ideas are coming up rapidly, we cannot discuss all of them due to space limitations). The emphasis will be on the ideas and methods that have been developed steadily so that readily understood by the uninitiated. Unresolved issues are highlighted with stress over the fundamentals to make it interesting for an expert as well. The approaches are presented in a sequence developing the ideas and techniques in a logical sequence rather than historical. The effort is on making sections individually readable.

2. Time-Frequency Analysis

A clear distinction between concentration and resolution is essential to properly evaluate the TFDs' performance. These concepts have generally been considered synonymous or equivalent in literature, and terms are often used interchangeably. Although one intuitively expects higher concentration to imply higher resolution, this is not necessarily the case [46]. In particular, the CTs in the WD do not reduce the auto-component concentration of the WD, which is considered optimal, but they do reduce the resolution. Although high signal concentration is always desired and is often of primary importance, in many applications, signal resolution may be more important, for example, in the analysis of multicomponent dispersive waves and detection and estimation of swell [47–49]. There have generally been two approaches to estimate the time-dependent spectrum of nonstationary processes.

(1) The evolutionay spectrum (ES) approaches [50–53], which model the spectrum as a slowly varying envelope of a complex sinusoid.

(2) The Cohen's bilinear distributions (BDs) [3], including the spectrogram, which provide a general formulation for joint TFDs. Computationally, the ES methods fall within Cohen's class.

There are known limitations and inherent drawbacks associated with these classical approaches. These pheneomena make their interpretation difficult, consequently, estimation of the spectra in the t-f domain displaying good resolution has become a research topic of great interest.

2.1. The Methods Based on Evolutionary Spectrum. The ES was first proposed by Priestley in 1965. The basic idea is to extend the classic Fourier spectral analysis to a more generalized basis: from sine or cosine to a family of orthogonal functions. In his evolutionay spectral theory, Priestely represents nonstationary signals using a general class of oscillatory functions and then defines the spectrum based on this representation [54]. A special case of the ES used the Wold-Cramer representation of nonstationary processes [55–58] to obtain a unique definition of the time-dependent spectral density function. According to the Wold-Cramer decomposition, a discrete time nonstationary process $x[n]$

can be interpreted as the output of a causal, linear, and time-varying (LTV) system with an impulse response $h[n,m]$, at time n to an impulse at time m. It is driven by zero-mean stationary white noise $e[n]$ so that

$$x[n] = \sum_{m=-\infty}^{n} h[n,m]e[m], \qquad H(n,\omega) = \sum_{m=-\infty}^{n} h[n,m]e^{-i\omega(n-m)} \tag{3}$$

is the Zadeh's generalized transfer function (GTF) of the system evaluated on the unit circle.

The Wold-Cramer ES of $x[n]$ [56, 57] shows that the time-varying power spectral density of output is equal to the magnitude squared of time-varying frequency response of the filter. It is defined as

$$S_{\mathrm{ES}}(n,\omega) = \frac{1}{2\pi}|H(n,\omega)|^2. \tag{4}$$

This definition can be viewed as Priestley's ES provided that $H(n,\omega)$ is a slowly varying function of n [57]. This restriction removes possible ambiguities in the definition of the spectrum by selecting the slowest of all the possible time-varying amplitudes for each sinusoid. Without this condition, each signal can have an infinite number of sinusoid/envelope combinations [57]. It has been shown that the ES and the GTF are related to the spectrogram and Cohen's class of BDs [59]. The main objective in deriving and presenting these relations in [59] was to show that the BDs and the spectrogram can be considered estimators of the ES.

A great amount of work is found by Pitton and Loughlin to investigate the positive TFDs and their potential applications [60–65]. Pitton and Loughlin utilized the ES and Thompson's multitaper approach [66, 67] to obtain positive TFDs, but do not discuss the issue of TFDs' concentration and resolution.

Literature indicates that the pioneering work, remarkable in its scope, is performed by Chaparro, Jaroudi, Kayhan, Akan, and Suleesathira. These researchers have not only focused on computing the improved evolutionary spectra of nonstationary signals but also innovatively applied the concepts to application in various practical situations [50–53, 68–91]. Their major work includes, signal-adaptive evolutionary spectral analysis and a parametric approach for data-adaptive evolutionary spectral estimation. An interesting work is performed by Jachan, Matz, and Hlawatsch on the parametric estimations for underspread nonstationary random processes. The necessary description of these methods is presented next.

2.1.1. Signal-Adaptive Evolutionary Spectral Analysis. Although it is well recognized that the spectra of most signals found in practical applications depend on time, estimation of these spectra displaying good t-f resolution is difficult [3]. The problem lies in the adaptation of the analysis methods to the change of frequency in the signal components. Constant-bandwidth mehods, such as the spectrogram and traditional Gabor expansion [92], provide estimates with poor t-f resolution.

The earlier approaches by Akan and Chaparro to obtain high-resolution evolutionary spectral estimates include: averaging estimates obtained using multiple windows [75] and maximizing energy concentration measure [53]. In [53], the authors proposed a modified Gabor expansion that uses multiple windows, dependent on different scales and modulated by linear chirps. Computation of the ES with this expansion provides estimates with good t-f resolution. The difficulties encountered, however, were the choices of scales and in the implementation of the chirping.

The Approach. Akan and Chaparro show that by generalizing and implementing by separating the signal components using evolutionary masking [75], a much improved spectral estimate is obtained by an adaptive algorithm [68]. The adaptation uses estimates of the IF of the signal components. The signal is decomposed into its components by means of masking on an initial spectrum of the signal. However, the masking is implemented manually and there is requirement to perform this action automatically. The estimation of the IF of each of the signal component is accomplished by an averaging procedure. It is shown that using the IF information of the components in the Gabor expansion improves the t-f localization.

Akan defines a finite-extent, discrete-time signal $x(n)$ as a combination of linear chirps with time-varying amplitudes as

$$x(n) = \sum_{p=0}^{P-1}\sum_{k=0}^{K-1} A(n,\omega_k,p)e^{in(\omega_k+(\alpha_p/2)n)}, \tag{5}$$

where $0 \le n \le N-1$, $\omega_k = 2\pi k/K$, and α_p is a parameter for selection of scales and slopes for analysis chirps. The selection of scales and slopes for the analysis chirps can be avoided by considering a more general model for $x(n)$ than the one given in (5) as [68]

$$x(n) = \sum_{p=0}^{P-1}\sum_{k=0}^{K-1} A(n,\omega_k,p)e^{i\phi_p(n,k)}, \tag{6}$$

where each of the signal component, $x_p(n)$, has a phase $\phi_p(n,k)$ to which corresponds an IF $\omega_p(n)$. Mathematically the ES of $x(n)$ comes out to be $S(n,\omega_k) = |\sum_p A(n,\omega_k,p)|^2$. Akan and Chaparro then implements the evolutionary spectral computation using the multi-window warped Gabor expansion [53] for each linear chirp:

$$x(n) = \frac{1}{J}\sum_{j=0}^{J-1}\sum_{m=0}^{M-1}\sum_{k=0}^{K-1} a_p(j,m,k)h_j(n-mL)e^{in\omega_k}, \tag{7}$$

here a_p refers to Gabor coefficients. The synthesis functions may be obtained by scaling a Gaussian window, $g(n)$, as $h_j = 2^{j/2}g(2^j n)$, $j = 0, 1, \ldots, J-1$, where J is the number of scaled windows, and $L < K$ is the time step in the oversampled

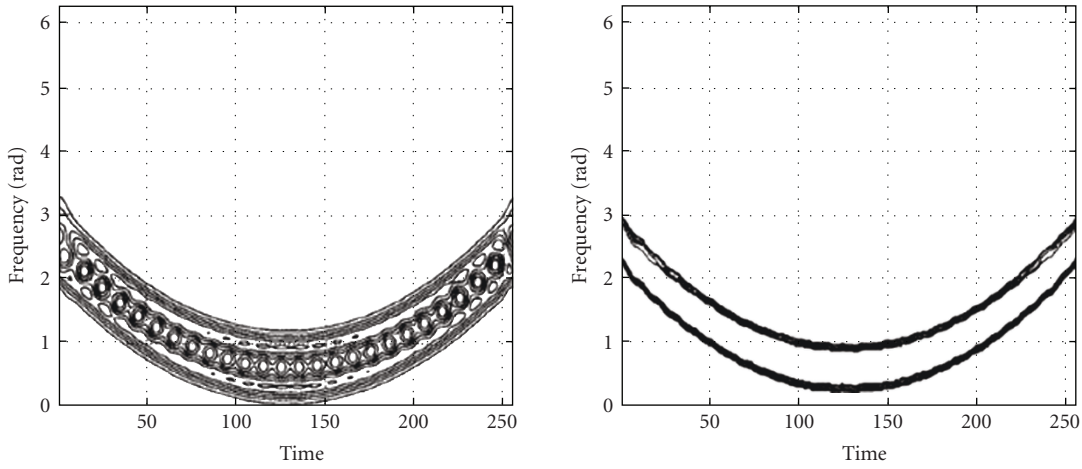

Figure 1: Example 1. A signal consisting of two closely spaced quadratic FM components, (a) initial ES estimate of the signal, (b) the final ES estimate (adopted from Akan [68]).

Gabor expansion. Necessary simplification of (7) results in following expression for the evolutionary kernel:

$$A(n,\omega_k,p) = \sum_{l=0}^{N-1} x(l)\xi(n,l)e^{-il\omega_k}, \tag{8}$$

where $\xi(n,l)$ is the time-varying window, defined as $\xi(n,l) = (1/J)\sum_{j=0}^{J-1}\sum_{m=0}^{M-1}\gamma_j^*(l-mL)h_j(n-mL)$ with γ_j being an analysis window biorthogonal to $h_j(n)$ [92]. However (8) can be viewed as short-time chirp FT with a time-varying window.

The adaptive algorithm given by Akan has the following steps.

(1) Computation of an initial ES, $S(n,\omega_k) = |A(n,\omega_k)|^2$ in (7) and (8) by avoiding the selection of scales and slopes for the analysis chirps, that is, taking $\alpha_p = 0$.

(2) Spectral masking of the signal [75], using the initial ES, to obtain signal components, $x_p(n)$. This masking of the signal is accomplished by multiplying its evolutionary kernel $A(n,\omega_k)$ by a masking function defined using the initial ES. Thus to get a component $x_p(n)$, a mask can be defined as

$$M_p(n,k) = \begin{cases} 1, & (n,k) \in R_p, \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where R_p is a region in the initial ES containing a single component. Consequently,

$$x_p(n) = \sum_{k\in R_p} A(n,\omega_k)M_p(n,k)e^{in\omega_k}, \tag{10}$$

this masking is however implemented manually and should be done automatically.

(3) Once each component and its spectral representation, $A(n,\omega_k,p)$ is obtained, the authors proceed with the estimation of the IF of each monocomponent, $\hat{\omega}_p(n)$, and corresponding phase $\hat{\phi}_p(n,k)$. This is performed using numeric integration techniques.

(4) Computation of the final ES, where an estimate of $x_p(n)$ in terms of its signal-adaptive Gabor expansion can be given by

$$\hat{x}_p(n) = \frac{1}{J}\sum_{j=0}^{J-1}\sum_{m=0}^{M-1}\sum_{k=0}^{K-1} a_p(j,m,k)h_j(n-mL)e^{i(n\omega_k+\hat{\phi}_p(n))}, \tag{11}$$

where the Gabor coefficients are calculated according to

$$a_p(j,m,k) = \sum_{n=0}^{N-1}\hat{x}_p(n)\gamma_j^*(n-mL)e^{-\mathrm{i}(n\omega_k+\hat{\phi}_p(n))}. \tag{12}$$

Akan terms the exponential $e^{-i\hat{\phi}_p(n)}$ in (12) as demodulating $x_p(n)$ along its IF, to obtain a signal that is composed of sinusoids and well represented by Gabor bases. After calculating the Gabor coefficients of each component, their spectral representations as in (8) can be obtained. Finally, the estimation of ES of $x(n)$ is possible after compensating for the demodulation as

$$S(n,\omega_k) = \left|\sum_p A\left(n,\omega_k-\hat{\omega}_p(n),p\right)\right|^2. \tag{13}$$

Akan shows by examples that using the IF information of the components in the Gabor expansion, the t-f localization is improved. The results are displayed in Figures 1 and 2 for signals composed of two closely packed quadratic FM components and a smiling face consisting of a quadratic FM component, two sinusoids at different time periods, and a Gaussian function shifted in frequency.

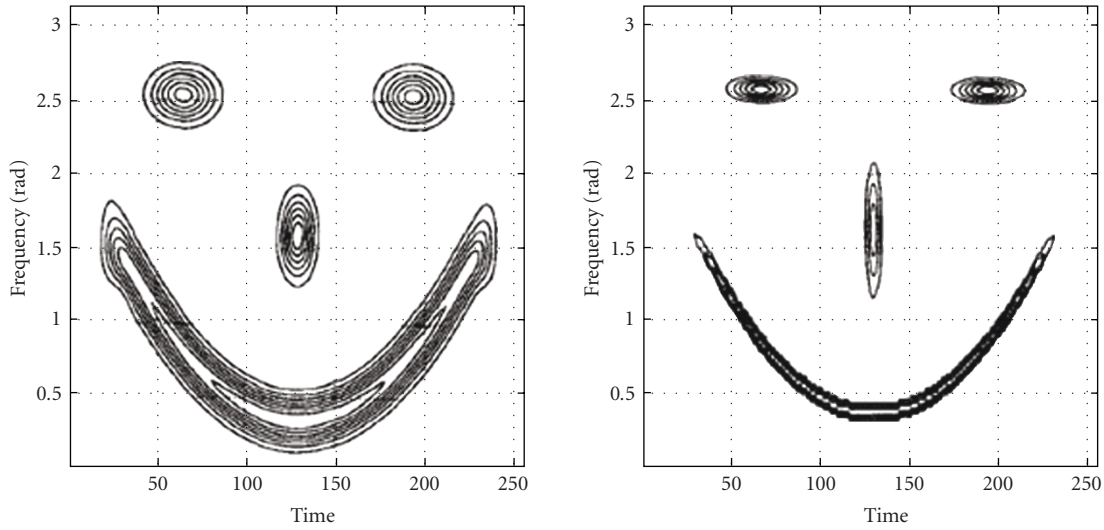

Figure 2: Example 2. A smiling face signal composed of a quadratic FM component, two sinusoids at different time periods, and a Gaussian function shifted in frequency, (a) initial ES estimate of the signal, (b) the final ES estimate (adopted from Akan [68]).

2.1.2. Data-Adaptive Evolutionary Spectral Estimation—A Parametric Approach. The ES theory is though mathematically well grounded, but has suffered from a shortage of estimators. The initial work from Kayhan concentrates on evolutionary periodogram (EP) as an estimator on the line of BDs. His latest work, however, follows a parametric approach in deriving the high-quality estimator for the ES [50, 51]. Parametric approaches to model the nonstationary signal using rational models with time-varying coefficients represented as expansions of orthogonal polynomial have been proposed by various investigators, for example, [93, 94]. However, the validity of their view of a nonstationary spectrum as a concatenation of "frozen-time" spectra has been questioned [57, 95].

In the earlier effort, Kayhan et al. in [50] proposed the evolutionary periodogram (EP) as an estimator of the Wold-Cramer ES. The EP is found to possess many desirable properties and reduces to the conventional periodogram in the stationary case. It is demonstrated by the authors that the EP outperforms the STFT and various BDs in estimating the spectrum of nonstationary signals. The EP estimator can be interpreted as the energy of the output of a time-varying bandpass filter centered around the analysis frequency. To derive the EP, the spectrum at each frequency is found, while minimizing the effect of the signal components at other frequencies under the assumption that these components are uncorrelated or white. Although this assumption is analogous to the one used in deriving the conventional periodogram [96], Kayhan and others realized it to be somewhat unrealistic. The mathematical details and EP's properties are discussed in detail in [50, 97].

Data-Adaptive Evolutionary Spectral Estimator (DASE). In order to improve performance, Kayhan et al. [51] further propose a new estimator that uses information about the signal components at frequencies other than the frequency of interest. The DASE computes the spectrum at each frequency while minimizing the interference from components at other frequencies without making any assumptions regarding these components. This estimator reduces to Capon's maximum likelihood method [98] in the stationary case. The DASE has better t-f resolution than the EP and thus it possesses many desirable properties analogous to those of Capon's method. In particular, it performs more robustly than existing methods when the data is noisy.

The DASE's mathematical derivation alongwith properties can be found in [51], and we present here the examples to demonstrate the performance of the DASE in comparison to other estimators like the EP and BDs. The first example signal is composed of two chirps: one with increasing frequency and one with decreasing frequency. Both components have a quadratic amplitude. Figure 4(c) shows the DASE using the Fourier expansion functions. Figure 4(b) shows the EP spectrum using the same expansion functions. Figure 4(a) shows the BD using exponential kernels. By comparing the three plots, it is clear that the DASE approach produces the best spectral estimate. It outperform the EP by displaying no sidelobes, fewer spurious peaks, and a narrower bandwidth. It also outperforms the BD by producing a nonnegative spectrum with no artifacts and sharper peaks. In the second example, the same signal is imbedded in additive Gaussian white noise. All the parameters from the example above remain unchanged, and the SNR is 24 dB. Figures 4(d)–4(f) show the BD, the EP, and the DASE spectral estimates, respectively. This example serves to demonstrate the effect of noise on each of the methods. Again, the DASE spectrum is found to be the least affected. The EP and the BD spectra display many more spurious peaks than the DASE spectrum.

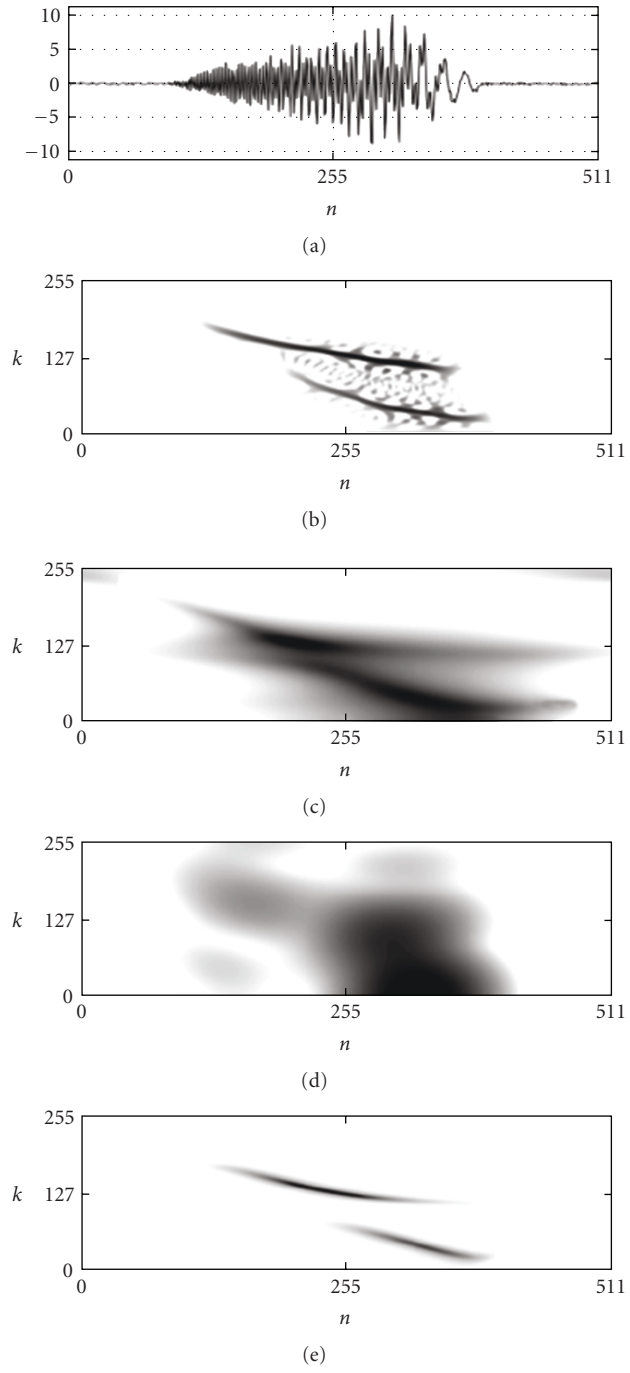

Figure 3: Time-varying parametric spectral analysis of the sum of two bat echolocation signals: (a) time-domain signal; (b) smoothed pseudo-Wigner distribution; (c) TFAR spectral estimate; (d) TFMA spectral estimate; and (e) TFARMA spectral estimate. Logarithmic gray-scale representations are used in (b)–(e) (all adopted from Jachan et al. [76]).

2.1.3. Time-Frequency Models and Parametric Estimators for Random Processes. Nonstationary random processes are more difficult to describe than the stationary processes because their statistics depend on time (or space) [77]. Parsimonious parametric models for nonstationary random processes are useful in many applications such as speech and audio, communications, image processing, computer vision, biomedical engineering, and machine monitoring. A parametric second-order description that is parsimonious in that it captures the time-varying second-order statistics by a small number of parameters is hence of particular interest. Jachan et al. [76] propose the use of frequency shifts in addition to time shifts (delays) for modeling nonstationary process dynamics in a physically intuitive way. The resulting parametric models are shown to be equivalent to specific types of time-varying autoregressive moving-average (TVARMA) models. They are parsimonious for nonstationary processes with small high-lag temporal and spectral correlations (underspread processes), which are frequently encountered in applications. Jachan, Matz, and Hlawatsch also propose efficient order-recursive techniques for model parameter estimation that outperform existing estimators for TVARMA (TVAR,TVMA) models with respect to accuracy and/or complexity

Major Contributions. Jachan et al. [76] consider a special class of TVARMA models that they term t-f ARMA (TFARMA) models. Extending time-invariant ARMA models, which capture temporal dynamics and correlations by representing a process as a weighted sum of time-shifted (delayed) signal components, TFARMA models additionally use frequency shifts to capture a process' nonstationarity and spectral correlations. The lags of the t-f shifts used in the TFARMA model are assumed to be small. This results in nonstationary processes with small high-lag temporal and spectral correlations or, equivalently, with a temporal correlation length that is much smaller than the duration over which the time-varying second-order statistics are approximately constant. Such underspread processes [78, 79] are encountered in many applications. The TFARMA model and its special cases, the TFAR and TFMA models, are shown to be specific types of TVARMA (AR,MA) models. They are attractive because of their parsimony for underspread processes, that is, nonstationary processes with a limited t-f correlation structure.

The underspread assumption results in parsimony which allows an "underspread approximation" that leads to new, computationally efficient parameter estimators for the TFARMA, TFAR, and TFMA model parameters. The authors develop two types of TFAR and TFMA estimators based on linear t-f Yule-Walker equations and on a new t-f cepstrum. Further, it is shown how these estimators can be combined to obtain TFARMA parameter estimators. In particular, TFAR parameter estimation can be accomplished via underspread t-f Yule-Walker equations with Toeplitz/block-Toeplitz structure that can be solved efficiently by means of the Wax-Kailath algorithm [80]. Simulation results demonstrate that the proposed methods perform better than existing TVAR, TVMA, and TVARMA parameter estimators with respect to accuracy and/or complexity. For processes that are not underspread (called "overspread" [78, 79]), the proposed models by Jachan et al. will not be parsimonious and those estimators that involve an underspread approximation exhibit poor performance.

TFARMA models are physically meaningful due to their definition in terms of delays and frequency (Doppler) shifts. This delay-Doppler formulation is also convenient since the nonparametric estimator of the process' second-order statistics that is required for all parametric estimators can be designed and controlled more easily in the delay-Doppler domain. Furthermore, TFARMA models are formulated in a discrete-time, discrete-frequency framework that allows the use of efficient fast FT algorithms. They can be applied in a variety of signal processing tasks, such as time-varying spectral estimation (cf. [81]), time-varying prediction (cf. [82–84]), time-varying system approximation [85], prewhitening of nonstationary processes, and nonstationary feature extraction.

Simulation Results. Jachan et al. check the accuracy of the proposed TFAR, TFMA, and TFARMA parameter estimators by applying them to signals synthetically generated according to the respective model. Here the application of the TFAR, TFMA, and TFARMA models is presented for time-varying spectral analysis of the quasi-natural signal shown in Figure 3(a). The considered signal is the sum of two echolocation chirp signals emitted by a Daubenton's bat (http://www.londonbats.org.uk). A smoothed pseudo-WD (SPWD) [86, 87] of this signal is shown in Figure 3(b).

The analysis based on TFAR, TFMA, and TFARMA is performed on this signal using the parameter estimators. From the estimated TFAR, TFMA, or TFARMA parameters, the corresponding parametric spectral estimates are computed, that is, estimates of the ES (TFMA case) or of its underspread approximation (TFAR and TFARMA cases, resp.). The authors estimate the model orders by means of the AIC [88, 89] and stablize all parameters by means of the technique described in [88], with an appropriate stabilization parameter.

The spectral estimates are depicted in Figures 3(c)–3(e). It is seen that the TFAR spectrum displays the two chirp components fairly well, although there are some spurious peaks (this effect is well known from AR models [90]) and the overall resolution is poorer than that of the nonparametric SPWD in Figure 3(b). The TFMA spectrum, as expected, is unable to resolve the timevarying spectral peaks of the signal. Finally, the TFARMA spectrum exhibits better resolution than the SPWD, and it does not contain any CTs as does the SPWD [87]; on the other hand, the t-f localization of the components deviates slightly from that in the SPWD. As indicated, the important point to note is that these parametric spectra involve only 30 (TFAR and TFARMA) or 42 (TFMA) parameters.

2.1.4. Miscellaneous Approaches. We find a considerable amount of work by a number of researchers in achieving good resolution ES and applying the results and related theory to many fields, specially where nonstationary signals arise. The purpose of their work has ranged from the simple graphic presentation of the results to sophisticated manipulations of spectra. The authors in [70] propose a new transformation for discrete signals with time-varying spectra. The kernel of this transformation provides the energy density of the signal in t-f with good resolution qualities. With this discrete evolutionary transform a clear representation for the signal as well as its t-f energy density is obtained. The authors suggest the use of either the Gabor or the Malvar discrete signal representations to obtain the kemel of the transformation. The signal adaptive analysis is then possible using modulated or chirped bases, and can be implemented with either masking or image segmentation on the t-f plane. An interesting approach is a piecewise linear approximation of the IF, concentrated along the individual components of signal, using the Hough transform (used in image processing to infer the presence of lines or curves in an image) and the evolutionary spectrum (ES) [71]. The efficiency and practicality of this approach lie in localized processing, linearization of the IF estimate, recursive correction, and minimum problems due to CTs in the TFDs or in the matching of parametric models. This procudere is innovatively used in jammer excision techniques, where unambiguous IF for a jammer composed of chirps can be estimated, using ES and Hough transform. Also Barbarossa in [72] proposed a combination of the WD and the Hough transform for detection and parameter estimation of chirp signals in a problem of detection of lines in an image, which is the WD of the signal under analysis. This method provides a bridge between signal and image processing techniques, is asymptotically efficient, and offers a good rejection capability of the CTs, but it has an increased computational complexity. Barbarossa et al. further proposed an adaptive method for suppressing wideband interferences in spread spectrum communications based on high-resolution TFD of the received signal [73]. The approach is based on the generalized Wigner-Hough Transform as an effective way to estimate the clear picture of the IF of parametric signals embedded in noise. The proposed method provides the advantages like, (1) it is able to reliably estimate the interference parameters at lower SNR, exploiting the signal model, (2) the despreading filter is optimal and takes into account the presence of the excision filter. The disadvantage of the proposed method, besides the higher computational cost, is that it is not robust against mismatching between the observed data and the assumed model.

Chaparro and Alshehri [74], innovatively obtain better spectral esimates and use it for the jammer excision in direct sequence spread spectrum communications when the jammers cannot be parametrically characterized. The authors proceed by representing the nonstationary signals using the t-f and the frequency-frequency evolutionary transformations. One of the methods, based on the frequency-frequency representation of the received signal, uses a deterministic masking approach while the other, based in nonstationary Wiener filtering, reduces interference in a mean-square fashion. Both of these approaches use the fact that the spreading sequence is known at the transmitter and the receiver, and that as such its evolutionary representation can be used to estimate the sent bit. The difference in performance between these two approaches depends on the support rather than on the type of jammer being excised.

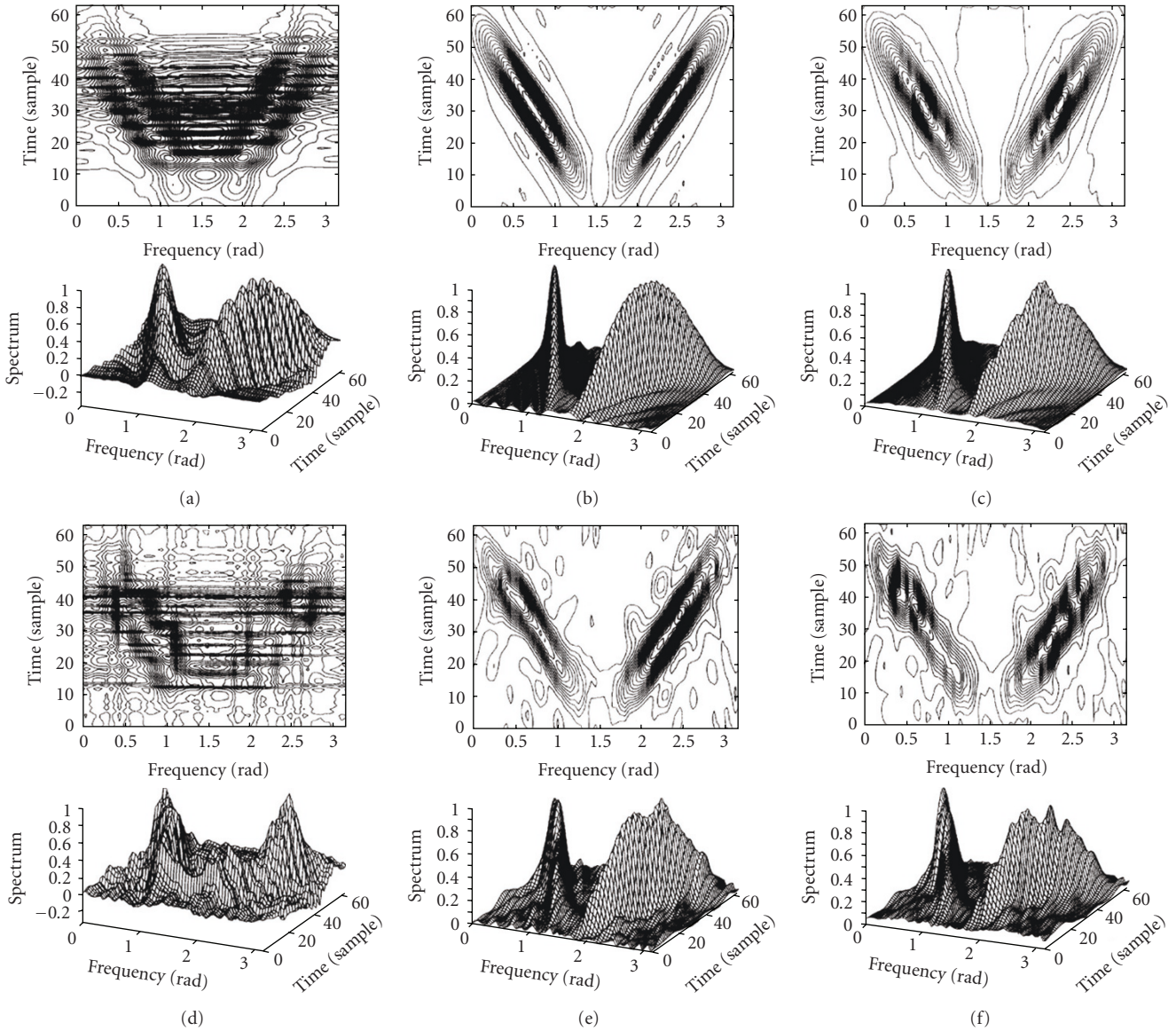

Figure 4: Example signals [51]. Signal 1 composed of two chirps, (a) BD using exponential kernels, (b) EP spectral estimate, and (c) DASE spectral estimate. Signal 2 composed of two chirps with additive Guassian white noise (SNR = 24 dB), (d) BD using exponential kernels, (e) EP spectral estimate, and (f) DASE spectral estimate (all adopted from Kayhan et al. [51]).

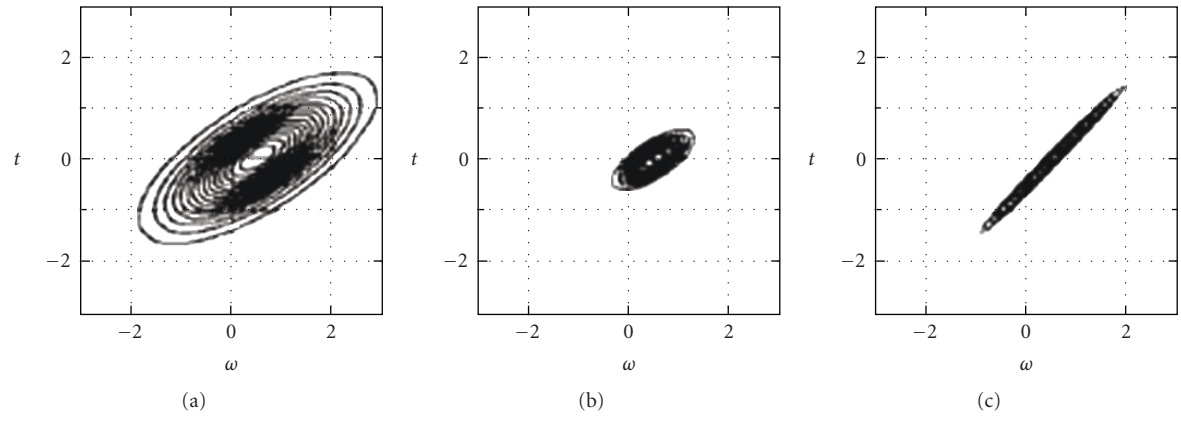

Figure 5: TFD of a Gaussian chirp signal: (a) the WD, (b) the LWD, and (c) the SD distribution with $L = 8$ (adopted from Stanković [108]).

The frequency-frequency masking approach is found to work well when the jammer is narrowly concentrated in parts of the frequency-frequency plane, while the Wiener masking approach works well in situations when the jammer is spread over all frequencies.

Shah et al. [99] developed a method for generating an informative prior when constructing a positive TFD by the method of minimum cross-entropy (MCE). This prior results in a more informative MCE-TFD, as quantified via entropy and mutual information measures. The procedure allows any of the BDs to be used in the prior, and the TFDs obtained by this procedure are close to the ones obtained by the deconvolution procedure at reduced computational cost. Shah along with Chaparro [91, 97] considered the use of the TFDs for the estimation of GTF of an LTV filter with a goal that once it is blurred, it produces the TFD estimate. They used the fact that many of these distributions are written as blurred versions of the GTF and made use of deconvolution technique to obtain the deblurred GTF. The technique is found general and can be based on any TFD with many advantages like (i) it estimates the GTF without the need for orthonormal expansion used in other estimators of the ES, (ii) it does not require the semistationarity assumption used in the existing deconvolution techniques, (iii) it can be used on many TFDs, (iv) the GTF obtained can be used to reconstruct the signal and to model LTV systems, and (v) the resulting ES estimate out performs the ES obtained by using the existing estimation techniques and can be made to satisfy the t-f marginals while maintaining positivity.

The Power Spectral Density of a signal calculated from the second-order statistics can provide valuable information for the characterization of stationary signals. This information is only sufficient for Gaussian and linear processes. Whereas, most real-life signals, such as biomedical, speech, and seismic signals may have non-Gaussian, nonlinear, and nonstationary properties. Addressing this issue, Unsal Artan et al. [100] have combined the higher-order statistics and the t-f approaches and present a method for the calculation of a Time-Dependent Bispectrum based on the positive distributed ES. This idea is particularly useful for the analysis of such signals and to analyze the time-varying properties of nonstationary signals.

2.2. The Methods Based on Cohen's Bilinear Class. In 1966 a method was devised that could generate in a simple manner an infinite number of new ones [3, 15]. The approach characterizes TFDs by an auxiliary function and by the kernel function. We will discuss the significant contributions on high spectral resolution kernels later in this paper. The properties of distribution are reflected by simple constraints on the kernel, and by examining the kernel one readily can ascertain the properties of the distribution. This allows one to pick and choose those kernels that produce distributions with prescribed desirable properties. All TFDs can be obtained from a general expression

$$C(t,\omega) = \frac{1}{4\pi^2}\iiint s^*\left(\mu - \frac{1}{2}\tau\right)s\left(\mu + \frac{1}{2}\tau\right)\Omega(\theta,\tau) \times e^{-i\theta t - i\tau\omega + i\theta\mu}\,d\mu\,d\tau\,d\theta, \tag{14}$$

where $C(t,\omega)$ is the joint distribution of signal $s(t)$, and $\Omega(\theta,\tau)$ is called the kernel. The term kernel was coined by Classen and Mecklenbrauker [9–11]. These two made extensive contributions to general understanding in signal analysis context along with Janssen [101]. Another term, which is brought in (14), is the ambiguity function (AF), for which there are a number of minor differences in terminology. We will use the definition given by Rihaczek, who defines AF as [102]

$$A(t,\tau) = \int s^*(t-\tau)s(t)e^{i\theta t}dt, \tag{15}$$

consequently (14) may be expressed as the FT of product of ambiguity and kernel functions, given as

$$C(t,\omega) = \mathcal{F}\{A(\theta,\tau)\cdot\Omega(\theta,\tau)\}, \tag{16}$$

where $A(\theta,\tau)$ is Woodward AF [103], which has been an important tool in analyzing and constructing signals associated with radar [102]. By constructing signals having a particular AF, desired performance characteristics are achieved. A comprehensive discussion of the AF can be found in [102], and shorter reviews of its properties and applications are found in [104, 105]. Also a number of excellent articles exploring the relationship between AF and the TFDs can be found in [11, 106, 107].

Many divergent attitudes toward the meaning, interpretation and use of Cohen's BDs have arisen over the years, with extensive research for obtaining good resolution and high concentration along the individual components. The divergent viewpoints and interests have led to a better understanding and implementation. The subject is evolving rapidly and most of the issues are open. However it is important to understand the ideas and arguments that have been given, as variations and insights of them have led way to further developments.

2.2.1. The Scaled-Variant Distribution—A TFD Concentrated along the IF. In an important set of papers, Stankovic et al. [33, 108–110] innovatively used the similarities and differences with quantum mechanics and originated many new ideas and procedures to achieve the good resolution and high concentration of joint distributions. Their initial work suggests the use of the polynomial WD [30, 111] to improve the concentration of *monocomponent signals*, taking the IF as polynomial function of time. A similar idea for improving the distribution concentration of the signal whose phase is polynomial up to the fourth order was presented in [25]. In order to improve distribution concentration for a signal with an arbitrary nonlinear IF, the L-Wigner distribution (LWD) was proposed and studied in [25, 112–115]. The polynomial WD, as well as the LWD, are closely related to the time-varying higher-order spectra [111, 114–116]. They were found to satisfy only the generalized forms of marginal and unable to preserve the usual marginal properties [1, 28].

Variant of LWD. Lately Stankovic proposed a variant of LWD obtained by scaling the phase and τ axis by an integer L while keeping the signals' amplitudes unchanged [33, 108].

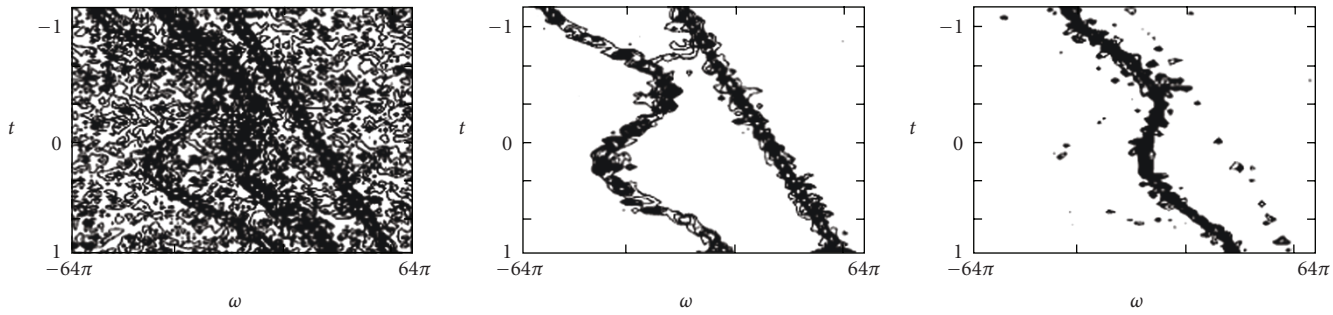

Figure 6: TFD of a signal with fast amplitude variations: (a) the WD, (b) the LWD with $L = 2$, and (c) the SD distribution with $L = 2$ (adopted from Stanković [108]).

He terms this new distribution as the scaled variant of the LWD (SD) of a signal $x(t)$. It is defined, in its pseudo form, as

$$\mathrm{SD}_L(t,\omega) = \int \omega_L(\tau) x^{[L]}\left(t + \frac{\tau}{2L}\right) x^{[L]*}\left(t - \frac{\tau}{2L}\right) e^{-i\omega\tau} d\tau. \tag{17}$$

The word "pseudo" is used to indicate the presence of the window $\omega_L(\tau)$ where $x^{[L]}(t)$ is the modification of $x(t)$ obtained by multiplying the phase function by L while keeping the amplitude unchanged:

$$x^{[L]}(t) = A(t)e^{iL\phi(t)}. \tag{18}$$

The distribution achieves high concentration at the IF—as high as the LWD—while at the same time satisfying time marginal and unbiased energy condition for any L. The frequency marginal is satisfied for asymptotic signals as well.

Simulation Results. The original idea for this distribution stems from the very well-known quantum mechanics forms. There is a partial formal mathematical correspondence between quantum mechanics and signal analysis. Relationship between quantum mechanics and signal analysis may be found in [1] and is beyond the scope of this paper. Historically, work on joint TFDs has often been guided by corresponding developments in quantum mechanics. The similarity comes about because in quantum mechanics the probability distribution for finding the particle at a certain position is the absolute square of the wave function, and the probability for finding the momentum is the absolute square of the FT of the wave function. Thus one can associate the signal with the wave function, time with position, and frequency with momentum. The marginal conditions are formally the same, although the variables are different.

Consequently for a signal $x(t) = A(t)e^{i\phi(t)}$, a function $\Psi(\lambda) = A(\lambda)e^{iL\phi(\lambda)}$ can be formed that corresponds (with $L = 1/\hbar$) to the Wentzel solution of the Schroedinger's equation or to the Feynman's path integral [108]. This form applied to the original quantum mechanics form of the WD $\mathrm{WD}(\lambda, p) = \int \Psi(\lambda + \hbar\tau/2)\Psi^*(\lambda - \hbar\tau/2)e^{-ip\tau}d\tau$ produces the proposed SD exactly. It is shown that significant benefit with respect to the distribution concentration is possible with uncertainty of the order of $1/L^2$ while at the same time keeping other important properties of the TFD invariant by keeping L slightly greater than 1 ($L = 2, 4, \ldots$). It is shown through example of *Gaussian chirpsignal* and a *noisy signal with same order of amplitude and phase variations* (see Figures 5 and 6) that the SD produces the ideal concentration at the IF.

Realization of the SD. A method for the direct realization of the SD, based on the straightforward application of a distribution definition, is presented in [110]. In the case of multicomponent signals, it may be equal to the sum of the SDs of each component separately. For the SD in (17), signal $x(t)$ is modified into $x^L(t)$, oversampled L times, while the number of samples that are used for calculation is kept unchanged. This method is not computationally much more demanding than the realization of any ordinary ($L = 1$) distribution. In the case of multicomponent signals, this method produces signal power concentrated at the resulting IF, according to the theorem presented in [108]. Theory is illustrated on the numerical examples of multicomponent real signal, real noisy multicomponent signal, and a multicomponent signal whose components intersect (see Figures 7 and 8). The proposed distributions may achieve arbitrary high concentration at the IF, satisfying the marginal properties. Till the publication of [110], this was possible only in a very special case of the linear frequency modulated signal using the WD.

2.2.2. Reassigned TFDs. Some TFDs were proposed to adapt to the signal t-f changes. In particular, an adaptive TFD can be obtained by estimating some pertinent parameters of a signal-dependant function at different time intervals [45]. Such TFDs provide highly localized representations without suffering QTFDs' CTs. The tradeoff is that these TFDs may not satisfy some desirable properties such as energy preservation. Examples of adaptive TFDs include the high resolution TFD [117], the signal-adaptive optimal-kernel TFDs [118, 119], the optimal radially Gaussian TFD [120], and Cohen's nonnegative distribution [34]. Reassigned TFDs also adapt to the signal by employing other QTFDs of the signal such as the spectrogram, the WD, or the scalogram

[121–127]. The former types of adaptive TFDs are discussed under the name Optimal-kernel TFDs in Section 2.2.3.

The method of reassignment improves considerably the t-f concentration and sharpens a TFD by mapping the data to t-f coordinates that are nearer to the true region of support of the analyzed signal. The method has been independently introduced by several researchers under various names [121–127], including method of reassignment, remapping, t-f reassignment, and modified moving-window method. In the case of the spectrogram or the STFT, the method of reassignment sharpens blurry t-f data by relocating the data according to local estimates of the IF and GD. This mapping to reassigned t-f coordinates is very precise for signals that are separable in time and frequency with respect to the analysis window.

The Reassignment Method. Pioneering work on the method of reassignment was first published by Kodera et al. under the name of *modified moving window method* [124]. Their technique enhances the resolution in time and frequency of the classical moving window method (equivalent to the spectrogram) by assigning to each data point a new t-f coordinate that better reflects the distribution of energy in the analyzed signal. This clever modification of the spectrogram unfortunately remained unused because of implementation difficulties and because its efficiency was not proved theoretically. Later on, Auger and Flandrin [121] showed that this method, which they called *the reassignment method*, can be applied advantageously to all the bilinear t-f and time-scale representations, and can be easily computed for the most common ones. Independently of Kodera et al., Nelson arrived at a similar method for improving the t-f precision of short-time spectral data from partial derivatives of the short-time phase spectrum [125]. It is easily shown that Nelson's cross-spectral surfaces compute an approximation of the derivatives that is equivalent to the finite differences method.

In the classical moving window method [128], a time-domain signal, $x(t)$, is decomposed into a set of coefficients, $\in (t, \omega)$, based on a set of elementary signals, $h_\omega(t)$, defined as

$$h_\omega(t) = h(t)e^{j\omega t}, \tag{19}$$

where $h(t)$ is a (real-valued) low-pass kernel function, like the window function in the STFT. The coefficients in this decomposition are defined as

$$\begin{aligned} \in (t,\omega) &= \int x(\tau)h(t-\tau)e^{-j\omega(\tau-t)}d\tau \\ &= e^{j\omega t}\int x(\tau)h(t-\tau)e^{-j\omega\tau}d\tau \\ &= e^{j\omega t}X(t,\omega) \\ &= X_t(t,\omega) = M_t(\omega)e^{j\varphi_\tau(\omega)}, \end{aligned} \tag{20}$$

where $M_t(\omega)$ is the magnitude, and $\varphi_\tau(\omega)$ is the phase, of $X_t(\omega)$, the FT of the signal $x(t)$ shifted in time by t and windowed by $h(t)$.

However $x(t)$ can be reconstructed from the moving window coefficients by

$$\begin{aligned} x(t) &= \iint X_\tau(\omega)h_\omega^*(\tau-t)d\omega d\tau \\ &= \iint X_\tau(\omega)h(\tau-t)e^{-j\omega(\tau-t)}d\omega d\tau \\ &= \iint M_\tau(\omega)e^{j\varphi_\tau(\omega)}h(\tau-t)e^{-j\omega(\tau-t)}d\omega d\tau \\ &= \iint M_\tau(\omega)h(\tau-t)e^{j[\varphi_\tau(\omega)-\omega\tau+\omega t]}d\omega d\tau. \end{aligned} \tag{21}$$

For signals having magnitude spectra, $M(t, \omega)$, whose time variation is slow relative to the phase variation, the maximum contribution to the reconstruction integral comes from the vicinity of the point t, ω satisfying the phase stationarity condition

$$\begin{aligned} \frac{\partial}{\partial\omega}\left[\varphi_\tau(\omega)-\omega\tau+\omega t\right] &= 0, \\ \frac{\partial}{\partial\tau}\left[\varphi_\tau(\omega)-\omega\tau+\omega t\right] &= 0, \end{aligned} \tag{22}$$

or equivalently, around the point $\hat{t}$, $\hat{\omega}$ defined by

$$\begin{aligned} \hat{t}(\tau,\omega) &= \tau - \frac{\partial\varphi_\tau(\omega)}{\partial\omega} = -\frac{\partial\varphi_\tau(\tau,\omega)}{\partial\omega}, \\ \hat{\omega} &= \frac{\partial\varphi_\tau(\omega)}{\partial\tau} = \omega + \frac{\partial\varphi_\tau(\tau,\omega)}{\partial\tau}. \end{aligned} \tag{23}$$

The t-f coordinates thus computed are equal to the local GD, $\hat{t}_g(t, \omega)$, and local IF, $\hat{\omega}_i(t, \omega)$, and are computed from the phase of the STFT, which is normally ignored when constructing the spectrogram. These quantities are local in the sense that they are represent a windowed and filtered signal that is localized in time and frequency, and are not global properties of the signal under analysis.

The modified moving window method, or method of reassignment, changes (reassigns) the point of attribution of $\in (t, \omega)$ to this point of maximum contribution $t_g(t, \omega)$, $\omega_i(t, \omega)$, rather than to the point t, ω at which it is computed. This point is sometimes called the *center of gravity* of the distribution, by way of analogy to a mass distribution. This analogy is a useful reminder that the attribution of spectral energy to the center of gravity of its distribution only makes sense when there is energy to attribute, so the method of reassignment has no meaning at points where the spectrogram is zero valued.

Efficient Computation of Reassigned Times and Frequencies. The reassignment operations proposed by Kodera et al. cannot be applied directly to the discrete STFT data, because partial derivatives cannot be computed directly on data that is discrete in time and frequency, and it has been suggested that this difficulty has been the primary barrier to wider use of the method of reassignment.

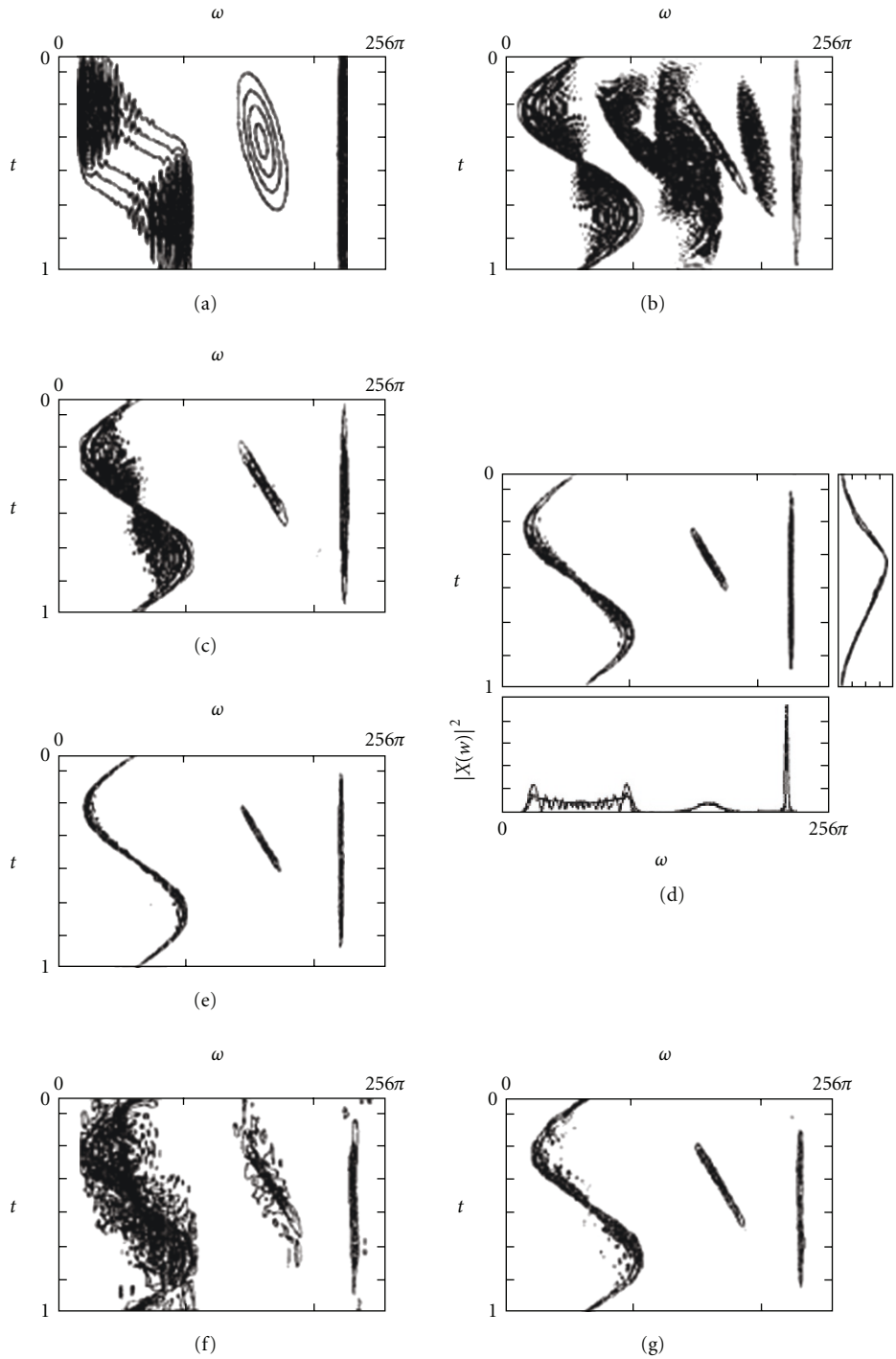

Figure 7: TFD of a multicomponent signal: (a) the spectrogram, (b) the WD, (c) the S-method, (d) the S-distribution with $L = 2$, including marginal properties, (e) the S-distribution with $L = 2$, (f) the S-method of noisy signal, (g) the S-distribution with $L = 2$ of noisy signal (adopted from Stanković [110]).

Auger and Flandrin [121] showed that the method of reassignment, proposed in the context of the spectrogram by Kodera et al., could be extended to any member of Cohen's class of TFDs by generalizing the reassignment operations. Auger and Flandrin's starting point of the efficient reassignment method is

$$\mathrm{TFD}(x;t,\omega) = \iint \phi_{\mathrm{TF}}(u,\Omega)\mathrm{WD}(x;t-u,\omega-\Omega)du\frac{d\Omega}{2\pi}, \tag{24}$$

which shows the 2D low-pass filtering of the WD, leading to a TFD of the Cohen's class [3, 15]. However, this smoothing

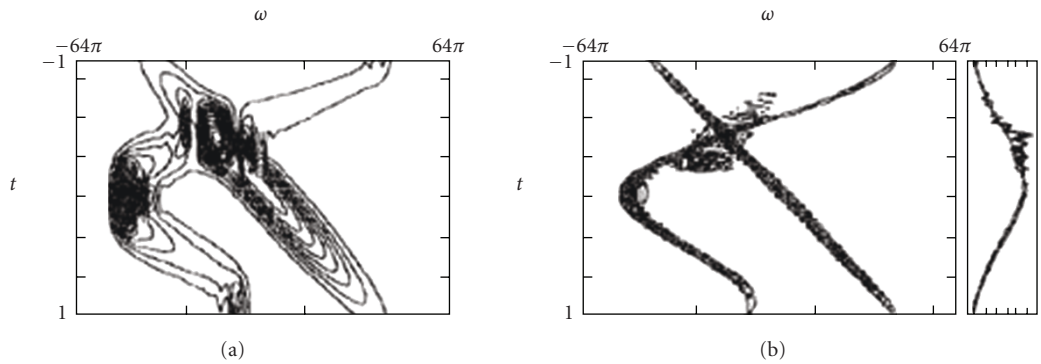

FIGURE 8: TFD of a multicomponent signal whose components intersects: (a) spectrogram, (b) S-distribution with $L = 2$ (adopted from Stanković [110]).

also produces a less accurate t-f localization of the signal components. Its shape and spread must therefore be properly determined so as to produce a suitable tradeoff between good interference attenuation and good t-f concentration [1, 28, 29]. Interesting examples of smoothings are the PWD [9–11], the SPWD [36], and all the reduced interference distributions [29, 39, 40]. As a complement to this smoothing, other processings can be used to improve the readability of a signal representation. A kind of signal representation processing, to which the reassignment method belongs, is to perform an increase of the signal components concentration.

The above expression shows that the value of a TFD at any point (t, ω) of the t-f plane is the sum of all the terms $\phi_{\mathrm{TF}}(u, \Omega)\mathrm{WD}(x; t-u, \omega-\Omega)$, which can be considered as the contributions of the weighted WD values at the neighboring points $(t-u, \omega-\Omega)$. $\mathrm{TFD}(x; t, \omega)$ is then the average of the signal energy located in a domain centered on (t, ω) and delimited by the essential support of $\phi_{\mathrm{TF}}(u, \Omega)$. This averaging leads to the attenuation of the oscillating CTs, but also a signal components broadening. The TFD can hence be nonzero on a point (t, ω) where the WD indicates no energy, if there are some nonzero WD values around. Therefore, one way to avoid this is to change the attribution point of this average, and to assign it to the center of gravity of these energy contributions, whose coordinates are

$$\hat{t}(x; t, \omega) = t - \frac{\iint u \cdot \phi_{\mathrm{TF}}(u, \Omega)\mathrm{WD}(x; t-u, \omega-\Omega)du(d\Omega/2\pi)}{\iint \phi_{\mathrm{TF}}(u, \Omega)\mathrm{WD}(x; t-u, \omega-\Omega)du(d\Omega/2\pi)},$$
$$\hat{\omega}(x; t, \omega) = \omega - \frac{\iint \Omega \cdot \phi_{\mathrm{TF}}(u, \Omega)\mathrm{WD}(x; t-u, \omega-\Omega)du(d\Omega/2\pi)}{\iint \phi_{\mathrm{TF}}(u, \Omega)\mathrm{WD}(x; t-u, \omega-\Omega)du(d\Omega/2\pi)}, \tag{25}$$

rather than to the point (t, ω) where it is computed. This reassignment leads to the construction of a modified version of this TFD, whose value at any point (t', ω') is therefore the sum of all the representation values moved to this point:

$$\mathrm{MTFD}(x; t', \omega') = \iint \mathrm{TFD}(x; t, \omega)\delta\left(t' - \hat{t}(x; t, \omega)\right) \times \delta(\omega' - \hat{\omega}(x; t, \omega))dt\frac{d\omega}{2\pi}, \tag{26}$$

where $\delta(t)$ denotes the Dirac impulse. It should be noticed that the aim of the reassignment method is to improve the sharpness of the localization of the signal components by reallocating its energy distribution in the t-f plane. Thus, when the representation value is zero at one point, it is useless to reassign it. Equations (25), the reassignment operators, have therefore neither sense nor use in this case. It should be also noticed that if the smoothing kernel $\phi_{\mathrm{TF}}(u, \Omega)$ is real valued, the reassignment operators (25) are also real valued, since the WD is always real valued.

Simulation Results. In order to evaluate the benefits of the reassignment method in practical applications, a comparison of the experimental results provided by some TFDs and their modified versions is shown in this section, adopted from Auger and Flandrin [121]. Auger and Flandrin analyze a 256-point computer-generated signal made up of one sine wave component, one chirp component, one chirped Gaussian packet, and one signal with constant amplitude and an instantaneous frequency describing half a sine period. Figure 9(a) shows the SPWD, adding a time-direction smoothing to PWD. There are very few CTs, but the signal components concentration is still weaker. Its modified version (shown in Figure 9(b)) is nearly ideal: all CTs are removed by the smoothings, and the signal components are strongly localized by the reassignment method. If the time and frequency smoothing windows are equal, the representation becomes then the spectrogram (Figure 9(c)), whose modified version (Figure 9(d)) perfectly localizes the chirp component. Finally, the next figures show time-scale representations. The affine PWD performs a scale-invariant frequency direction smoothing of the WD. Its modified version yields much more concentrated signal components, but still retains some CTs. An additional scale-invariant time direction smoothing removes nearly all CTs, yielding an affine SPWD (Figure 9(e)) with less concentrated signal components, and a nearly ideal modified affine SPWD (Figure 9(f)). Figure 9(g) now shows a scalogram whose window length was chosen to provide the same frequency direction smoothing, but (consequently) an approximately two times longer time direction smoothing than the previous affine SPWD. All the WD CTs have been removed, but

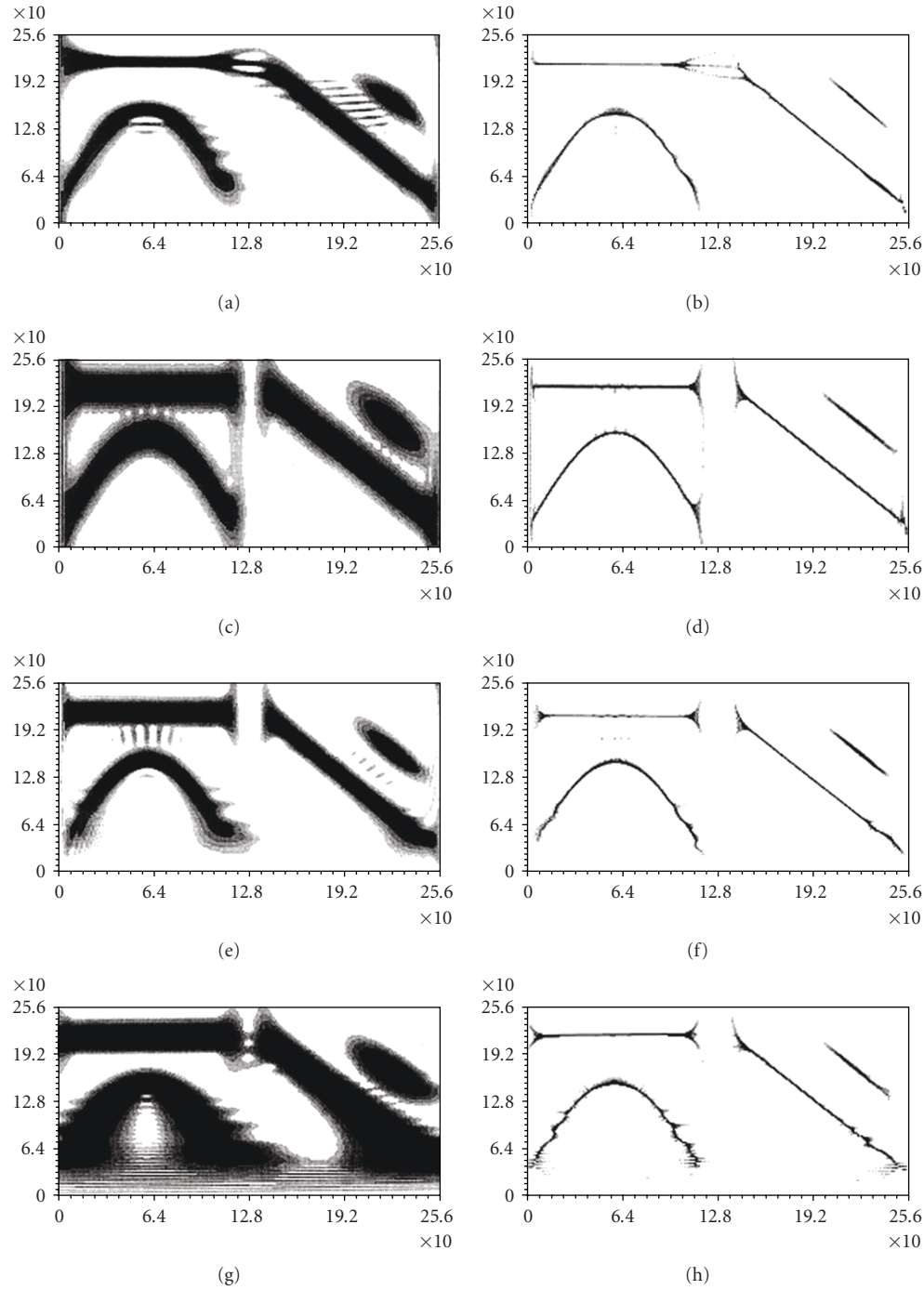

Figure 9: Numerical examples for reassignment method for a 4-component signal made up of a sine wave component, a chirp component, a chirped Gaussian packet, and one signal with constant amplitude and an IF describing half a sine period, (a) SPWD. h : 79-point Gaussian window, (b) modified version of the SPWD, (c) spectrogram. h : 31-point Gaussian window, (d) modified version of the spectrogram, (e) affine SPWD. h: Gaussian window with $F_0 \cdot Th = 4.24$, (f) modified version of affine SPWD, (f) scalogram. h: Gaussian window with $F_0 \cdot Th = 3.0$, (h) modified version of the scalogram (all adopted from Auger and Flandrin [121]).

the time resolution is really inadequate, especially at low frequencies. Its modified version is much easier to interpret, but the localization of the component with sinusoidal frequency modulationcy seems weaker than on the affine SPWD.

2.2.3. Optimal-Kernel TFDs. The result in (14) to (16) indicate that a quadratic TFD is obtained by first smoothing the symmetric AF (using the kernel function) and then by taking a 2D FT of the result. This result is equivalent to a 2D filtering in the ambiguity domain. The properties of

distribution are reflected by simple constraints on the kernel and have been used advantageously to develop practical methods for analysis and filtering, as was done by Eichmann and Dong [129]. Excellent reviews relating the properties of the kernel to the properties of the distribution have been given by Janse and Kaizer [12], Janssen [101], Claasen and Meclenbrauker [11], and Boashash [7]. By examining the kernel one readily can ascertain the properties of the distribution. This allows one to pick and choose those kernels that produce distributions with prescribed desirable properties. Thus, by a proper choice of kernel function, one can reduce or remove the CTs in the analysis of multicomponent signal. This unified approach is simple with an advantage that all distributions can be studied together in a consistent way. Since for any given signal some TFDs are "better than others," kernel design has become an important research area. Generally the optimum kernel TFDs can be achieved by three different approaches to optimizing the kernel with an aim to improve the resolution of resulting TFDs, which are

(1) high resolution TFDs based on high spectral resolution kernels;

(2) high resolution TFDs based on signal-independant kernels;

(3) high resolution TFDs based on signal-dependant kernels;

High Resolution TFDs-High Spectral Resolution Kernels. TFDs along with their temporal and spectral resolutions are uniquely defined by the employed t-f kernels. Potential kernels seek to map, at every time sample, the time-varying signals in the data into approximately fixed frequency sinusoids in the local autocorrelation function (LAF). Applying the FT to the LAF, therefore, provides a peaky spectrum where the location of the peaks is indicative to the signals' instantaneous power concentrations. The sinusoidal components in the LAF, however, generally appear with some type of amplitude modulations (AMs), which are highly dependent on the kernel composition [130]. Such modulation presents a limitation on spectral resolution in the t-f plane, as it is likely to spread both the auto and CTs to localizations over a wide range of frequencies.

A. *Improving TFDs' Spectral Resolution*. Because of the kernel modulation effects on the various terms, closely spaced frequencies may not be resolved. Further, since TFDs are Fourier based, then in addition to the AM imposed by the kernels, the spectral resolution is limited by and highly dependent on the extent of LAF, that is, the lag window employed [130]. However, increasing the length of the LAF will not always yield improved resolution. Events occurring over short periods of time do not require large kernels, which may only lead to increased CT contributions from distant events and obscure the local autoterms. Limited availability of data samples may also provide another reason for using small extent kernels. In these cases, improving spectral resolution of a TFD can be achieved by parameterizing its local autocorrelation function via autoregressive (AR) modeling techniques [131–135]. Such parameterization seeks to fit a least-squares random model to the second-order statistics of the LAF at different time instants. The AR modeling techniques, however, view the LAF as a stationary process along the lag dimension. Since t-f distribution kernels translate deterministic signals into others of deterministic nature, it will be more appropriate to fit a deterministic, rather than a stochastic, model to the LAF. Further, all modeling techniques applied in the TFD context mostly have only dealt with PWD or the SPWD kernels.

Amin and Williams [130] have maintained that in addition to PWD and SPWD of separable time and lag windows, there exists a large class of t-f kernels for which the LAF is amenable to high spectral resolution techniques. The members of this class satisfy the desirable t-f properties for power localization in nonstationary environment, yet they produce local autocorrelation functions that are amenable to exponential deterministic modeling during periods of stationarity. The proposed high spectral resolution kernels are, however, required to meet two basic conditions [130]:

(1) *the frequency marginal*,

(2) *an exponential behavior in the ambiguity domain for constant values of few parameters*.

In dealing with sinusoidal data, the first property guarantees that the autoterm sinusoids in the LAF are undamped. The second property enforces an exponential damping on all CTs. As a result, the sinusoidal components in the data translate into damped/undamped sinusoids in the local autocorrelation function. High-resolution techniques such as reduced rank approximation of the backward linear prediction data matrix can then be applied for frequency estimation. The authors use Prony's method and its least squares reduced-order approximation based on the singular value decomposition (SVD) [136, 137] in the t-f context. This method is shown to be applicable to high spectral resolution TFD problems, specifically when the underlying LAF is made up of a sum of exponentially damped/undamped sinusoids or chirp-like signals. The authors derive a class of TFD kernels in which the autoterms and the CTs of the sinusoidal components in the data are, respectively, mapped into undamped and damped sinusoids. By using the backward linear prediction frequency estimation approach [136], these two sets of components produce a linear predictor error filter whose zeros lie on and outside the unit circle, respectively. With the extraneous zeros of the polynomial lying inside the unit circle, fitting a deterministic model to the LAF of the proposed class of t-f kernels not only yields accurate estimates of the frequencies of the sinusoids but also provides a mechanism to distinguish between the true and false distribution terms.

B. *Simulation Results*. The simulations in Figures 10 and 11 high spectral resolution kernels illustrate the effectiveness of the high-resolution TFDs achieved by the high spectral resolution kernels. A test signal is constructed that consisted

of two complex exponentials as

$$x(n) = e^{i2\pi 12.8(n-30)/128} + e^{i2\pi 51.2n/128}, \quad n = 0, 1, 2, \ldots, 127. \tag{27}$$

These two signals' components have normalized frequencies of 0.1 and 0.4 Hz, respectively. First, the authors compute the binomial TFD using the alias free formulation [39] for comparison. The LAF, which extended to 128 points, is computed, and the binomial kernel is applied to it. Applying an FFT across the lags produced the result shown in Figure 10(a). The two components are well resolved, and the CT interference is low. Figure 10(b) shows the high-resolution TFD result using the binomial kernel. Only even lag terms are used in the LAF. The results are similar to the binomial TFD, but the resolution is higher. In addition, the CTs are small and generally fall between the autoterms and are not spread, as is the case for the binomial TFD. Figure 11(a) shows the results obtained using the raw LAF values, which is equivalent to the PWD. The autoterms are well resolved, but the CTs are as large as in the conventional PWD and fall between the CTs. A 20-*point* analysis window is used to find the Hankel structure for the odd positive lags obtained from the same LAF used to form the binomial TFD. The authors limit the number of terms included from the SVD computation by excluding all terms with magnitudes less than 15% of the largest singular value. The effectiveness of the approach with a nonstationary chirp is shown in Figure 11(b). The authors analyze a complex exponential with a starting frequency of 0.05 Hz and a positive chirp rate of 0.6×10^{-4} Hz/sample using the alias-free binomial LAF. Here, 0.1 Hz spans 100 frequency samples. We can see that the method provides a very nice estimate of the t-f course of the signal.

A High-Resolution QTFD—Signal-Independent Kernel. A signal independent kernel for the design of a high-resolution and CTs free quadratic TFD is proposed in [138]. The filtering of the CTs in the ambiguity domain that reduces (or removes) the CTs in the t-f domain results in a lower t-f resolution. That is, there is tradeoff between CTs suppression and t-f resolution in the design of a given quadratic TFD. Barkat and Boashash propose a kernel that allows retaining as many autoterms energy as possible while filtering out as much CTs energy as possible. The kernel is defined in the time lag domain keeping in view the implementation of the resultant TFD.

Beginning from a time function $(1/\cosh^2(t))$ whose spectrum presents the narrowest mainlobe compared with many other considered time function for the same signal duration. Barkat and Boashash extend it to a 2D quantity $(|\tau|/\cosh^2(t))$ and then taking it to a power α; they obtain two desirable characteristics. First, its FT (kernel function), which is centered around the origin, presents sharp cutoff edges. Second, the volume beneath it can be controlled by varying the value of α. Consequently, the proposed time-lag kernel is given by

$$G(t, \tau) = \left[\frac{|\tau|}{\cosh^2(t)}\right]^{\alpha}, \tag{28}$$

further it can be written as 2D ambiguity domain filter, given by

$$g(\upsilon, \tau) = \int G(t, \tau) e^{-i2\pi\upsilon t} dt, \tag{29}$$

where υ and τ are the two usual variables in the ambiguity domain. Using (29) in the general formula of the QTFDs, the authors come up with the following discrete-time version of the proposed TFD on simplification:

$$\rho_z(n, f) = \sum_{m=-M}^{M} \sum_{p=-M}^{M} G(p, m) z(n + p + m) \times z^*(n + p - m) e^{-i2\pi m f}, \tag{30}$$

where $z(t)$ is the analytic multicomponent signal under consideration, and the discrete-time expressions $G(n, m)$ and $z(n)$ are obtained by sampling $G(t, \tau)$ and $z(t)$ at a frequency $f_s = 1/T$ such that $t = n \cdot T$ and $\tau = m \cdot T$. The resulting TFD in (30) is alias-free and periodic in f with a period equal to unity.

Simulation Results. The distribution in (29) is claimed to solve problems that the WD or the spectrogram cannot. In particular, the proposed distribution is shown to resolve two close signals in the t-f domain that the two other distributions cannot. Further synthetic and real data collected from real-world applications are used to validate the proposed distribution (see Figures 12–14).

Adaptive TFDs—Signal-Dependant Kernel. Adaptive TFDs are highly localized t-f representations without suffering from CTs, and they can generally be obtained by estimating some pertinent parameters of time-varying signal-dependant function. A great amount of work is performed by Baraniuk and Jones, who have developed several different approaches optimizing the signal-dependant kernel t-f analysis [118–120, 139], including the following:

(1) *1/0 optimal kernel TFD* [118] formulation in which the optimal kernel turns out to have a special binary structure: it takes on only the values 1 and 0;

(2) *optimal radially Gaussian kernel TFD* [120] tempering the "*1/0 kernel*" optimization formulation with an additional smoothness constraint that forces the optimal kernel to be Gaussian along radial profiles;

(3) *signal adaptive optimal kernal TFDs* [119].

Baraniuk and Jones have made use of the fact that symmetric AF is the characteristic function of the WD. The mathematical and possible physical analogy between the two enhances the interpretation of the properties of the AF. As an illustration, consider the example of the AF of the bat chirp in Figure 15(a). The FT maps the WD autocomponents

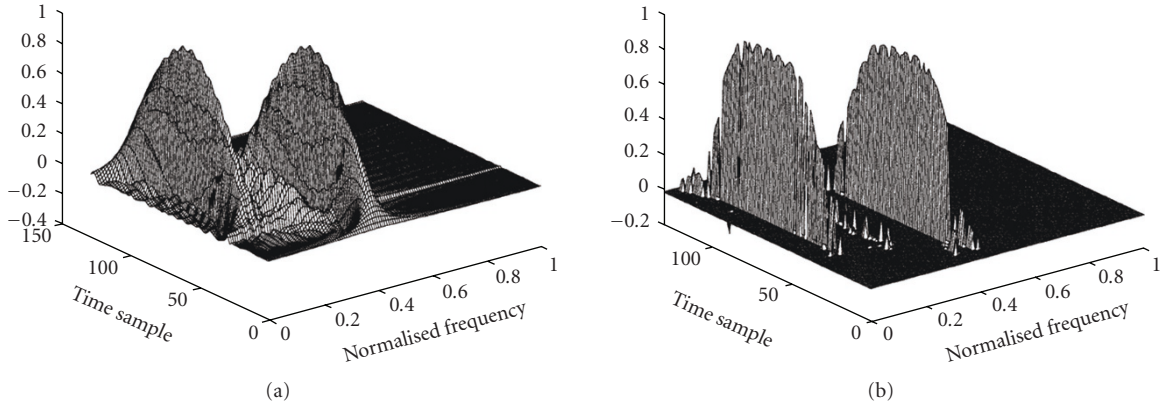

Figure 10: TFD results for $x(n)$: (a) bionomial TFD (b) high-resolution bionomial TFD (adopted from Amin and Williams [130]).

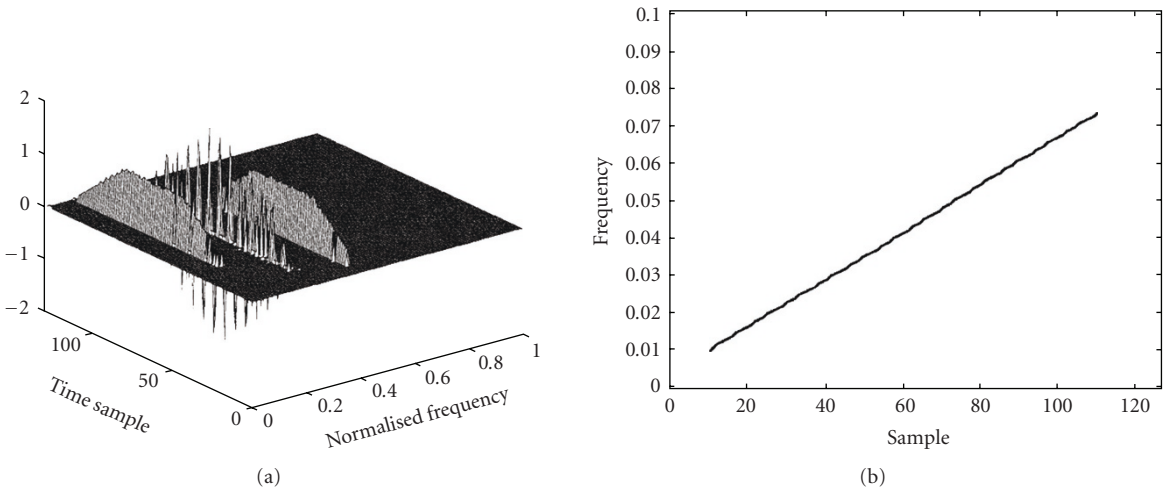

Figure 11: (a) High-resolution TFD for $x(n)$ using the raw LAF values (PWD equivalent). (b) High-resolution result for a complex exponential signal over 128 time samples (adopted from Amin and Williams [130]).

to a region centered on the origin of the AF plane, whereas it maps the oscillatory WD cross-components away from the origin. In the AF image 15(a), the AF autocomponents corresponding to the three harmonics of the bat chirp lie superimposed at the center of the AF image, while the AF cross-components lie to either side. The components slant in the AF because the bat signal is chirping. The fact that the auto- and cross-components are spatially separated in the AF domain facilitates optimization of the kernel function, which is used as a masking function to the AF to suppress the CTs. The two later concepts based on optimal radially Gaussian and signal adaptive optimal kernels are discussed next to illustrate the work of Baraniuk and Jones.

A. *The Optimal Radially Gaussian TFD*. The signal-dependent TFD proposed in [120] is based on kernels with Gaussian radial cross section:

$$\Phi(\theta,\tau) = \exp\left(-\frac{\theta^2+\tau^2}{2\sigma^2(\psi)}\right), \tag{31}$$

where $\Phi(\theta,\tau)$ is the kernel function, and the $\sigma(\psi)$ is the *spread function* that controls the spread of the Gaussian at radial angle ψ. The angle $\psi \equiv \arctan(\tau/\theta)$ is measured between the radial line through the point (θ,τ) and the θ axis. Radially Gaussian kernels can be expressed in polar coordinates, using $\xi = \sqrt{\theta^2+\tau^2}$ as radius variable:

$$\Phi(\xi,\psi) = \exp\left(-\frac{\xi^2}{2\sigma^2(\psi)}\right). \tag{32}$$

A high-quality TFD results when the kernel is well matched to the components of a given signal. This technique is in contrast to the approach considered by Barkat and Boashash [138] described in the last section. This is because the radially Gaussian kernel is adapted to a signal by solving the following optimization problem:

$$\max_{\Phi} \int_0^{2\pi}\int_0^{\infty} |A(\xi,\psi)\Phi(\xi,\psi)|^2 \xi\, d\xi\, d\psi \tag{33}$$

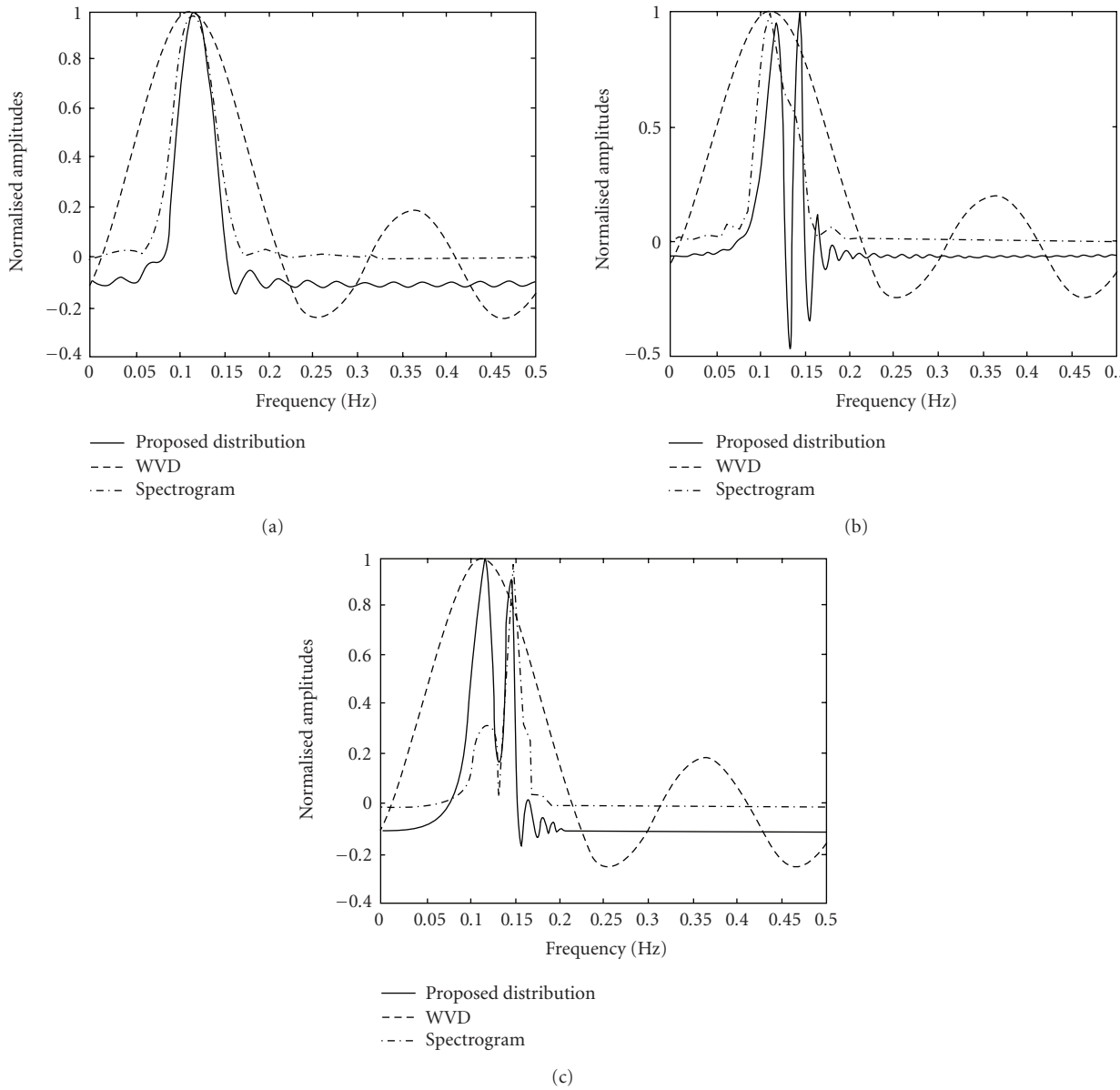

FIGURE 12: Slices taken at the same time instant $n = 3$ (signal start time instant) of the WD, the spectrogram, and the proposed TFD for a multicomponent signal composed of two parallel linear FM components using (a) a small size window length, (b) a medium size window length, and (c) a large size window length (adopted from Barkat and Boashash [138]).

subject to (32) and

$$\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{\infty} |\Phi(\xi,\psi)|^2 \xi\, d\xi\, d\psi = \frac{1}{4\pi^2}\int_0^{2\pi} \sigma^2(\psi)\, d\psi \le \alpha, \quad \alpha \ge 0, \tag{34}$$

where $A(\xi,\psi)$ represent the AF of the signal in polar coordinates. The solution to the above optimization problem is denoted by Φ_{opt}. The constraints and performance index are motivated by a desire to suppress cross-components and to pass autocomponents with as little distortion as possible. The performance measure in (33) determines the shape of the pass-band of the optimal radially Gaussian kernel. By this, it is desired that as much autocomponent energy as possible can be passed into the TFD for a kernel of fixed volume thus autocomponent distortion can be minimized. In most cases, authors have preferred this TFD to the *1/0 optimal-kernel TFD* [118]. The *optimal Radially Gaussian kernel* of the bat chirp is well matched to the AF autocomponents as shown in Figure 15(b). As a result a high-resolution TFD is obtained shown in Figure 15(c).

B. *Signal-Adaptive Optimal-Kernel TFD*. In another approach by Jones and Baraniuk, it is argued that TFDs with fixed windows or kernels figure prominently in many applications but perform well only for limited classes of signals [119]. Representations with signal-dependent kernels can overcome this limitation. However, while they often

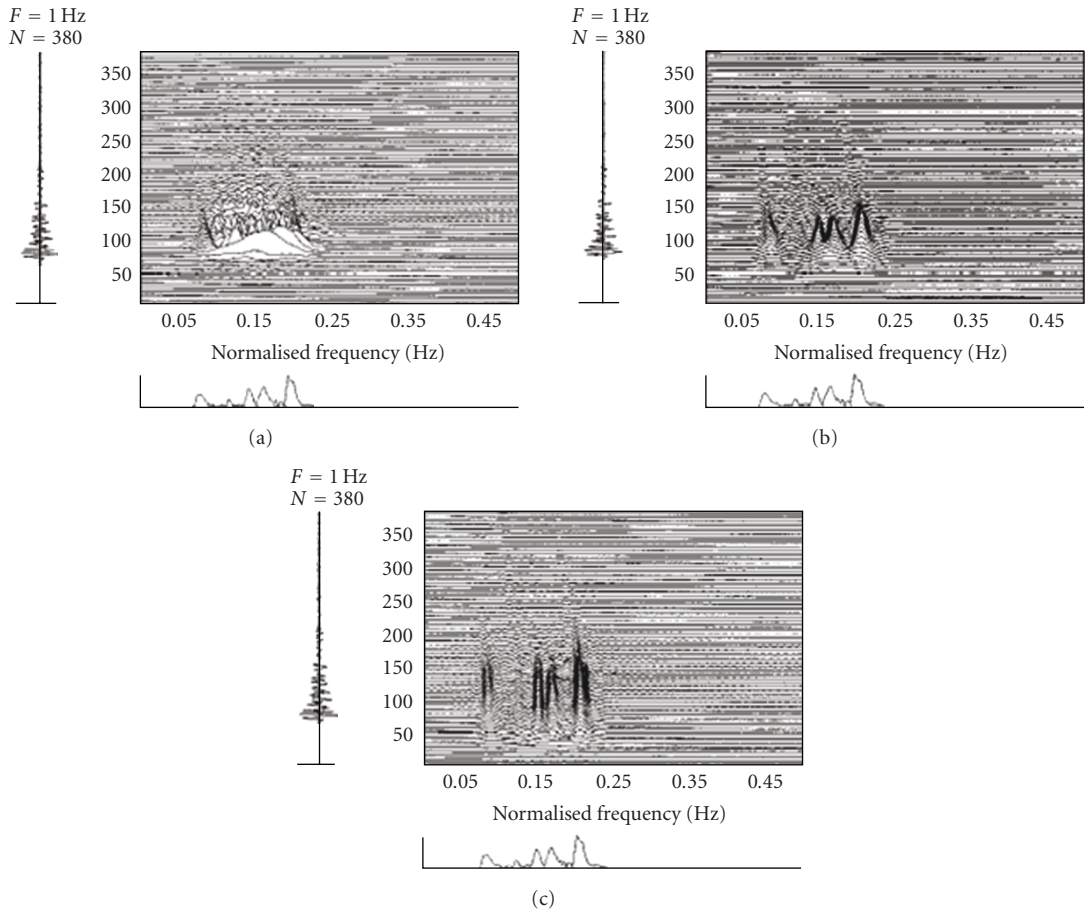

Figure 13: (a) WVD, (b) the spectrogram, and (c) the proposed TFD for a real-life automotive signal (adopted from Barkat and Boashash [138]).

perform well, most existing schemes are block-oriented techniques unsuitable for online implementation or for tracking signal components with characteristics that change with time. By adapting the radially gaussian kernel over time to maximize performance, the resulting *adaptive optimal-kernel (AOK) TFD* [119] is found suitable for online operation with long signals whose t-f characteristics change over time. The method employs a short-time AF (STAF) both for kernel optimization and as an intermediate step in computing constant-time slices of the representation.

Jones and Baraniuk adopt a general approach by deriving time-dependant spectra through generalizing the relationship between the power spectrum and the autocorrelation function. The concept of a local autocorrelation function was developed by Fano [140] and Schroeder and Atal [141], and the relationship of their work to time-varying spectra was considered by Ackroyd [142, 143]. A local autocorrelation method was used by Lampard [144] for deriving the Page distribution, and subsequently other investigators have pointed out the relation to other distributions. The basic idea is to write the joint TFD, as

$$P(t,\omega) = \frac{1}{2\pi}\int R_t(\tau)e^{-i\omega\tau}d\tau, \tag{35}$$

where $R_t(\tau)$ is a time-dependant or local autocorrelation function. Many expressions for $R_t(\tau)$ have been proposed. Jones and Baraniuk chose the instantaneous correlation of signal $s(t)$ as

$$R(t,\tau) \equiv s^*\left(u-\frac{\tau}{2}\right)s\left(u+\frac{\tau}{2}\right), \tag{36}$$

which also gives the WD and is argued by Mark [145] for symmetry. The authors give the time localized STAF by

$$\begin{aligned} A(t;\theta,\tau) &\equiv \int R(u,\tau)\omega^*\left(u-t-\frac{\tau}{2}\right)\omega\left(u-t+\frac{\tau}{2}\right)e^{i\theta u}du \\ &= \int s^*\left(u-\frac{\tau}{2}\right)\omega^*\left(u-t-\frac{\tau}{2}\right)s\left(u+\frac{\tau}{2}\right) \\ &\quad\cdot\,\omega\left(u-t+\frac{\tau}{2}\right)e^{i\theta u}du, \end{aligned} \tag{37}$$

where $\omega(u)$ is a symmetrical window function equal to zero for $|u| > T$. The variables τ and θ are the usual ambiguity plane parameters; the variable t indicates the center position of the signal window. Only the portion of the signal in the interval $[t-T, t+T]$ with $|\tau| < 2T$ is incorporated into $A(t;\theta,\tau)$ [119].

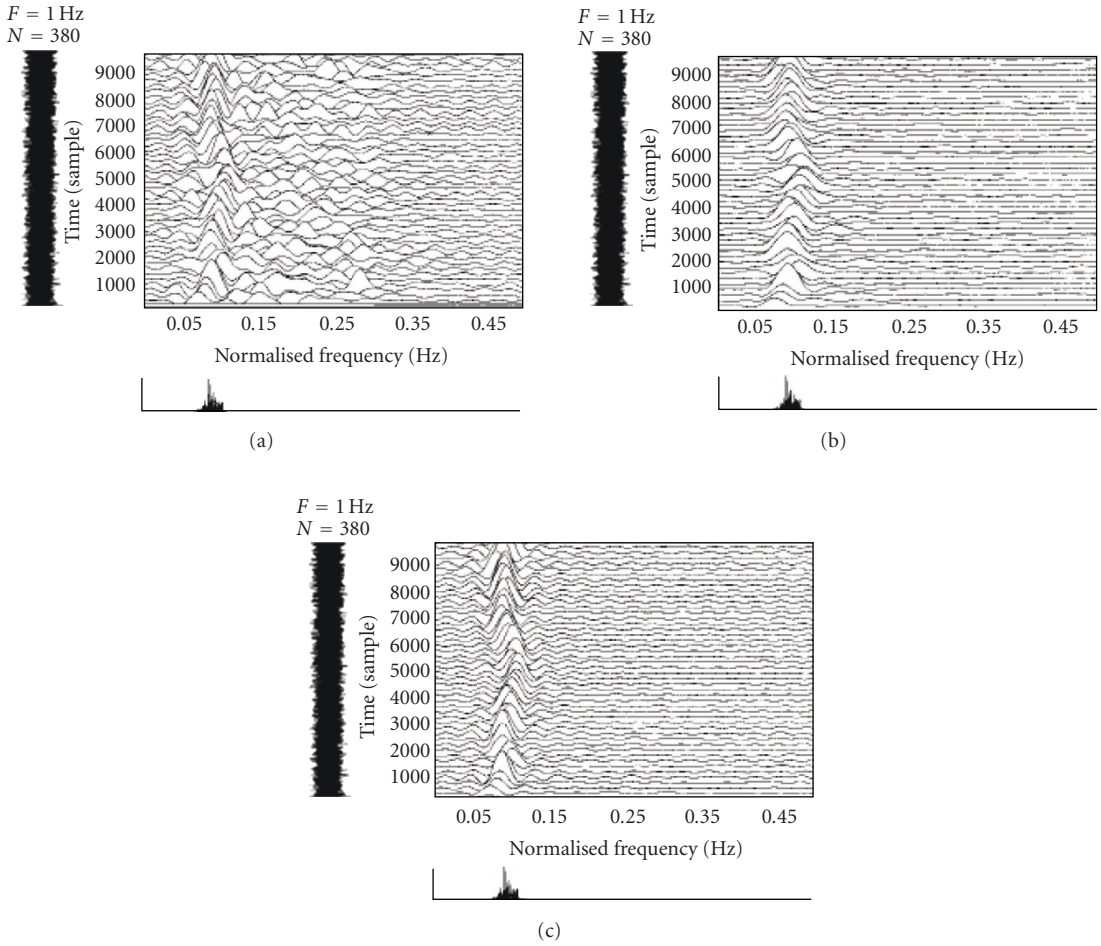

Figure 14: (a) WVD, (b) the spectrogram, and (c) the proposed distribution of a real-life acoustic signal (adopted from Barkat and Boashash [138]).

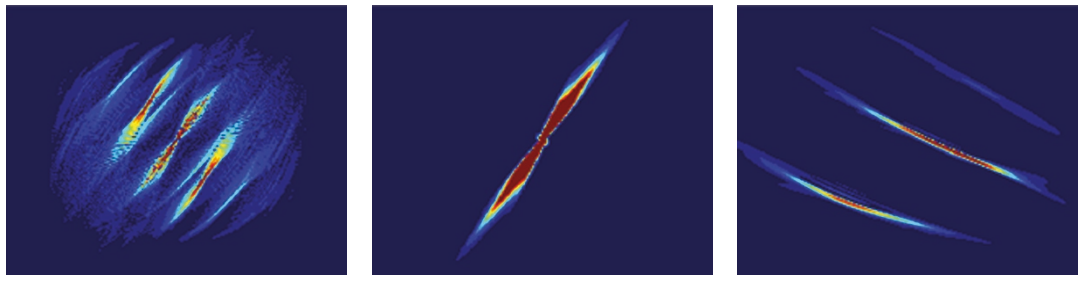

Figure 15: (a) Ambiguity function of bat chirp, (b) optimal radially Gaussian kernel, and (c) the optimal radially Gaussian TFD (all adopted from Baraniuk [120]).

Conceptually, the algorithm presented in [119] computes the STAF centered at time in both rectangular and polar coordinates and solves the optimization problem in (33) and (34) to obtain the optimal kernel. Once the optimal kernel has been determined, a single, current-time slice of the AOK TFD is computed as one slice (at time t only) of the 2D FT of the STAF-kernel product:

$$P_{\mathrm{AOK}}(t,\omega) = \frac{1}{4\pi^2}\iint A(t;\theta,\tau)\Phi_{\mathrm{opt}}(t;\theta,\tau)e^{-i\theta t - i\tau\omega}\,d\theta\,d\tau. \tag{38}$$

Jones and Baraniuk also suggest that certain enhancements, such as cone-kernel constraints and approximate retention of marginals, are easily incorporated with little

additional computation. While somewhat more expensive than fixed-kernel representations, this new technique often provides much better performance, as shown in Figures 16-17; however, it is limited to processing relatively short signals in an offline fashion.

One of the recent application based on AOK-based technique is source detection and classification in over-the-horizon radar (OTHR) systems [147]. The application makes use of a data-dependent kernel in the ambiguity domain to capture the target signal components, which are then resolved using coherent spectrum estimation. By using the different time-Doppler signatures, important target maneuvering information, which is difficult to extract using other linear and bilinear t-f representation methods, can be easily revealed.

2.2.4. Dispersive Class TFDs. These TFDs are also termed *warping-based TFDs* which provide a very good concentration for STSC having nonlinear t-f characterstics, such as dolphin and whale whistles, radar and sonar waveforms, and shock waves in fault structures. To improve the processing of such signals, QTFDs that satisfy the dispersive GD shift covariance property are designed by Papandreou, Hlawatsch and Boudreaux-Bartels in [148–152]. The dispersive GD shift covariance property for a QTFD $Q_x(t,f)$ is defined as [45]

$$\begin{aligned} Q_{\Re_c X}(t,f) &= Q_x(t-c\tau(f),f) \text{ with} \\ (\Re_c X)(f) &= e^{j2\pi c\xi(f/f_r)}X(f), \end{aligned} \tag{39}$$

where $\Re_c, X(f), f, f_r, \tau(f)$, and $\xi(f/f_r)$ are the dispersive GD shift operator, FT of the signal $x(t)$, frequency variable, positive normalization frequency, the change in GD, and the phase spectrum or characteristic basis function, respectively.

This property, defined in (39), is important for analyzing signals propagating through systems with dispersive characteristics or equivalently, with nonlinear GD functions. If the signal spectrum $X(f)$ is passed through an all pass dispersive system with output $Y(f) = e^{-j2\pi\xi(f/f_r)}X(f)$, then the change in GD, $\tau(f) = (d/df)\xi(f/f_r)$, is proportional to the derivative of the one-to-one phase function $\xi(f/f_r)$. Because GD is a measure of the time delay introduced in each sinusoidal component of the signal at frequency f, the ideal QTFD should preserve this change in GD or frequency-dependent time shift $\tau(f)$ as indicated in (39). Different dispersive GD shifts can be obtained by fixing $\xi(f/f_r)$ in (39). In particular, Papandreou, Hlawatsch, and Boudreaux-Bartels obtain the following.

(1) Linear dispersive GD shift [148] $\tau(f) = (2/f_r)|f/f_r|$.

(2) Hyperbolic dispersive GD shift [153] $\tau(f) = 1/f$.

(3) The kth power dispersive GD shift [154] $\tau(f) = (k/f_r)|f/f_r|^{k-1}$.

(4) Exponential dispersive GD shift [150] $\tau(f) = (k/f_r)e^{kf/f_r}$.

Dispersive GD shift covariant QTFD classes are unitarily equivalent to known QTFD classes because they can be obtained by warping existing time shift convariant classes such as Cohen's class or the affine class. The dispersive class QTFDs $Q^{(D)}$ may be obtained by warping Cohen's class QTFDs $Q^{(C)}$ or affine class QTFDs $Q^{(A)}$ using [148–152, 155]

$$\begin{aligned} Q_X^{(D)}(t,f) &= Q_{W_\xi X}\left(\frac{t}{f_r\tau(f)}, f_r\xi\left(\frac{f}{f_r}\right)\right) \text{ with} \\ (W_\xi X)(f) &= \left|f_r\tau\left(f_r\xi^{-1}\left(\frac{f}{f_r}\right)\right)\right|^{-1/2} X\left(f_r\xi^{-1}\left(\frac{f}{f_r}\right)\right), \end{aligned} \tag{40}$$

where W_ξ is the warping operator. A list of properties that different dispersive QTFDs satisfy is provided in [156].

Matched Time-Frequency Processing. Papandreou and Boudreaux-Bartels prove that, for successful t-f analysis, it is advantageous to match the specific time shift of a QTFD in (39) with changes in the GD of the signal. In some applications, signals with known GD, $\tau(f) = (d/df)\xi(f/f_r)$, need to be processed. As a result, a matched QTFD can be designed as in (40) with a characteristic function. When the signal GD is not known *a priori*, some preprocessing is necessary before designing a well-matched QTFD. A rough GD estimate can be obtained by fitting a curve through the spectrogram of the signal or by using one of the many porposed algorithms to estimate GD or IF characteristics [2, 31, 157–159]. Because the phase function of the signal needs to be one-to-one for designing its matched QTFD by appropriately warping the WD or its smoothed versions, approximations of the GD function can also be used.

Warping-Based TFDs—Theoretical Examples and Advantages. Different dispersive QTFDs can be obtained simply by choosing the function $\xi(f/f_r)$ or its derivative $\tau(f)$ in (40). Once the function is fixed, the resulting QTFDs can be useful in analyzing signals whose GD characteristics are the same or approximately the same as $\tau(f)$. Some examples of these classes include the linear chirp class (warped affine class) with linear GD, the hyperbolic class (warped Cohen's class) with hyperbolic GD, the kth power class (warped affine class) with kth order power GD, and the exponential class (warped affine class) with exponential GD.

Papandreau, Boudreaux-Bartels, and coworkers [160–163] demonstrate the effectiveness of dispersive class QTFDs and the importance of matching STSC with QTFDs using various simulations including constant and linear, constant and hyperbolic, constant and exponential, and constant and power t-f structures and power t-f structures with real data. QTFDs in all these cases show better resolution and CT suppression. For example, it is demonstrated that the dispersive WD is highly localized for the time modulation signal $G(f) = \sqrt{|\tau(f)|}e^{-j2\pi c\xi(f/f_r)}$. Specifically, it is found that dispersive WD is a dirac delta function at GD $\tau(f)$ of the signal, that is, $\mathrm{WD}_G^{(D)}(t,f) = |\tau(f)|\delta(t-c\tau(f))$. This means that the dispersive WD is ideally matched to time modulation signals when the $\tau(f)$ in the dispersive WD formulation matches the GD of the signal. It is important to note that

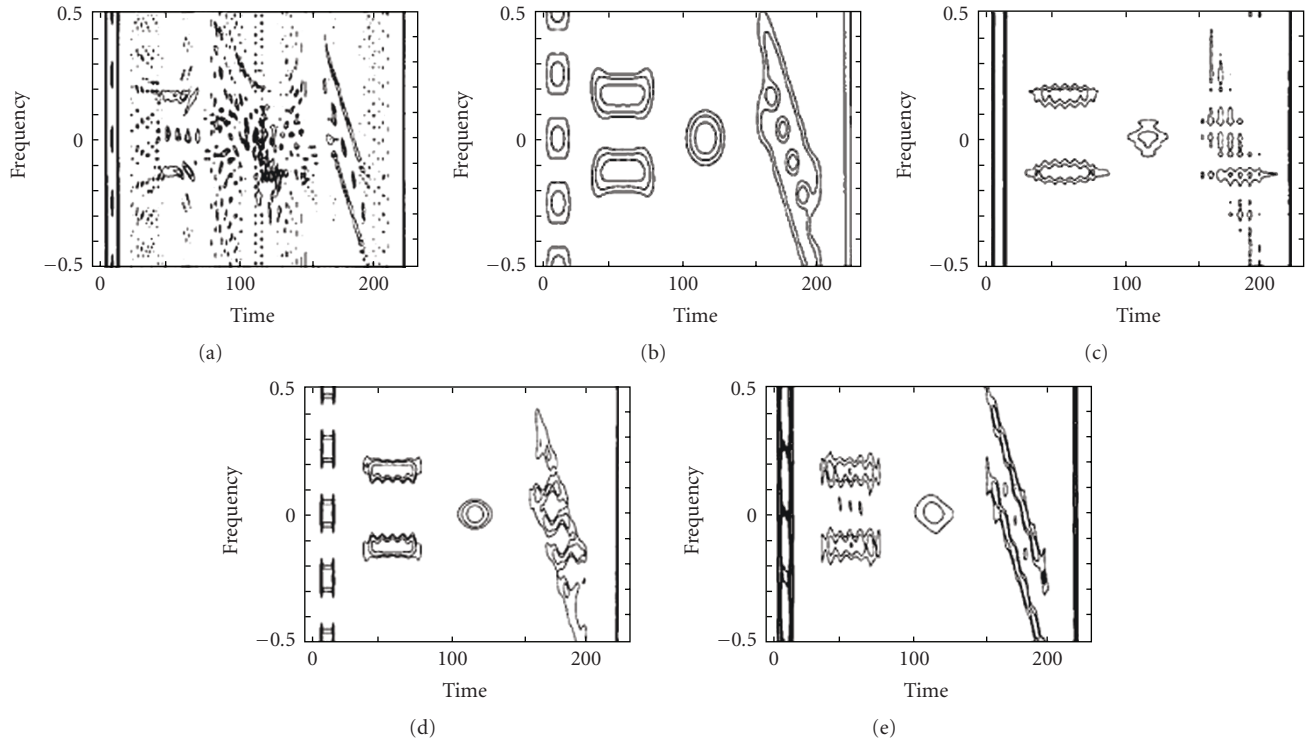

Figure 16: Comparison of several state-of-the-art TFDs of the synthetic test signal of three impulses, two simultaneous sinusoidal pulses, a Gaussian pulse, and two parallel linear chirps: (a) WD; (b) spectrogram computed with a 27-*point* Gaussian window; (c) Choi-Williams distribution with smoothing parameter $\alpha = 1$ [29]; (d) cone-kernel representation with τ-extent parameter $= 33$ [146]; (e) radially Gaussian optimal-kernel TFD computed using volume $\alpha = 2$ [118–120, 139] (adopted from Jones and Baraniuk [119]).

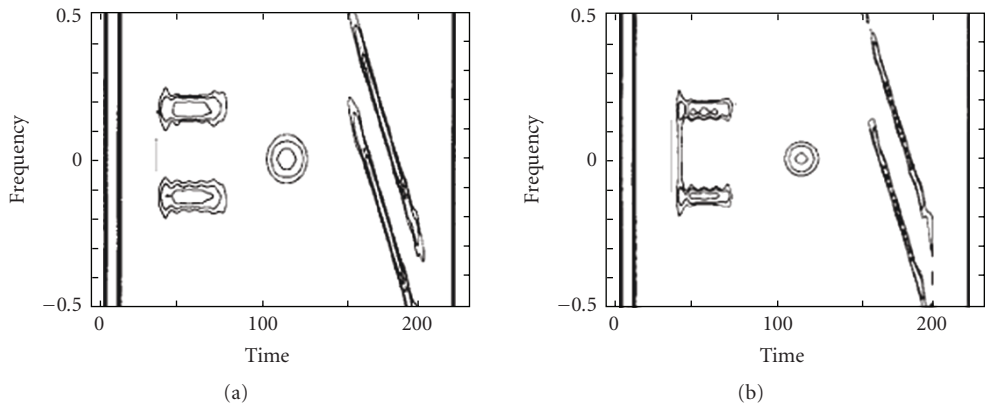

Figure 17: New adaptive TFDs of the synthetic test signal of three impulses, two simultaneous sinusoidal pulses, a Gaussian pulse, and two parallel linear chirps: (a) adaptive optimal-kernel (AOK) TFD computed using volume $\alpha = 1.4$, window width $2T = 32$, and one kernel optimization iteration per input time sample, (the AOK TFD is quite insensitive to these values); (b) AOK TFD with additional time-support (cone) constraint [146] (adopted from Jones and Baraniuk [119]).

a dual dispersive class can be similarly obtained to match the dispersive FM signals by preserving dispersive IF shift [164].

Another example is the affine class that is actually the power class with $k = 1$. When $k = 2$, the corresponding power class is the linear chirp class that is well matched to signals with linear t-f characteristics. Two QTFDs from the linear chirp class are the linearly warped WD and the chirpogram. These are obtained when the WD and the spectrogram, respectively, are warped as in (40) with quadratic characteristics function. The linearly warped WD is found to provide high localized representations when analyzing linear time mudulation signals. On the other hand, the chirpogram has a definite t-f resolution advantage over the spectrogram when analyzing multicomponent signals with linear characteristics. This is because the smoothing operation of the chirpogram is performed along lines of

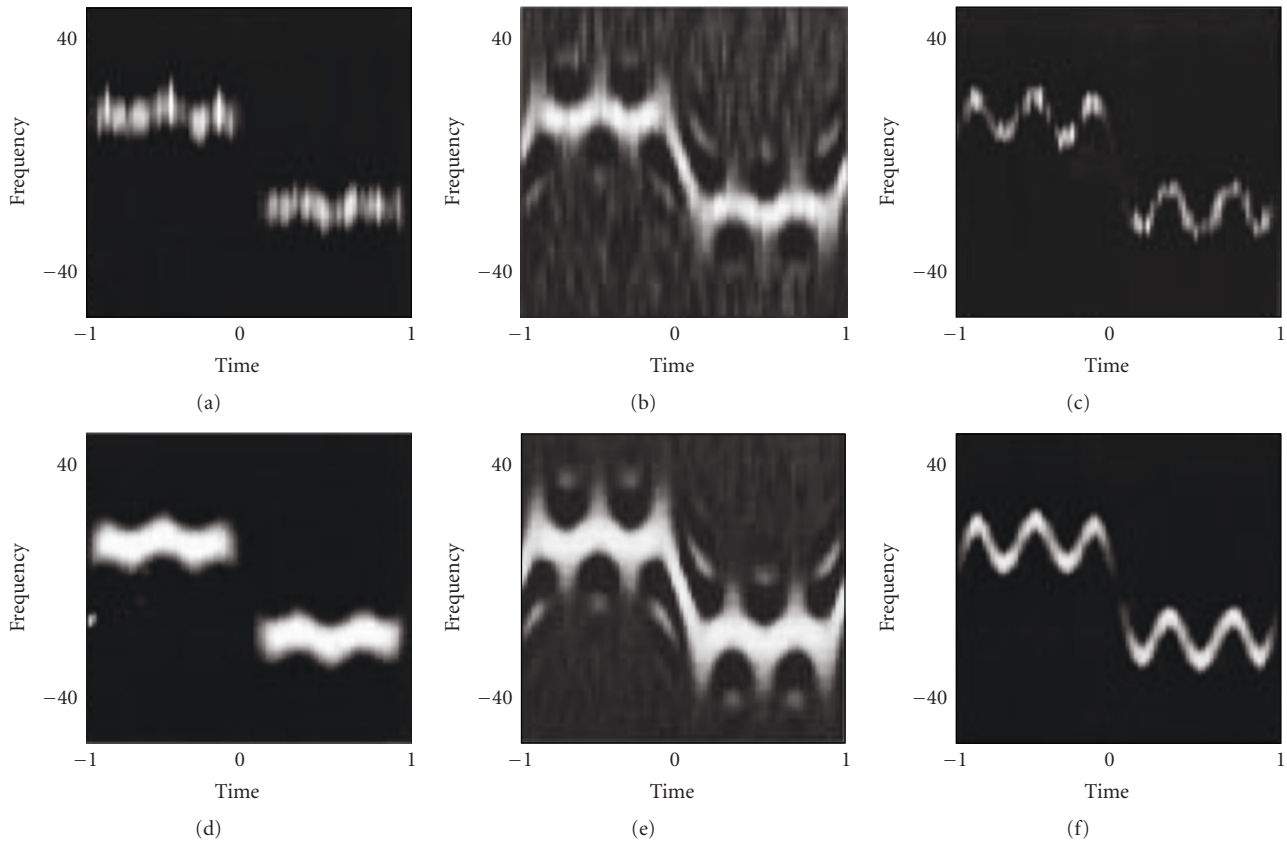

Figure 18: TFD of a signal with highly nonlinear IF. Spectrogram (a), (d), WD (b), (e), complex-lag distribution (c), (f). (a), (b), and (c) signals with a relatively high noise SNR = 10 dB. (d), (e), and (f) signals with a small noise SNR = 30 dB (all adopted from Stankovic [165]).

any slope in the t-f plane whereas the smoothing of the spectrogram is only along horizontal or vertical lines [166].

Through various examples, Hlawatsch et al. prove the power QTFDs as ideal for signals that propagate through linear systems with specific power GD characteristics such as when a wave propagates through a dispersive medium [160]. Other signals that are matched to kth power QTFDs include the dispersive propagation of shock wave in a steel beam ($k = 0.5$) [167, 168], transionospheric signals mesured by satellites ($k = -1$) [169], acoustical waves reflected from a spherical shell immersed in water [170], some cetacean mammal whistles [171], and diffusion equation-based waveforms ($k = 0.5$) [172] (e.g., waves from uniform distinbuted radio communication transmission lines [173]).

Limitations. Warping-based or Dispersive QTFDs could be computationally intensive when implemented directly using numerical integration as in the case of warping WD to obtain power WD. Hlawatsch et al. suggest an alternative implementation scheme based on the warping formulation in (40) that allows the use of existing efficient algorithms for computing Cohen's class or Affine class QTFDs as done by them in [160] for power QTFDs. However the increased computational complexity of the dispersive QTFDs is the trade-off for the improved performance in analyzing signals with matched dispesive GD characteristics.

2.2.5. TFDs with Complex Arguments. One of the most important concept to improve concentration in case of nonlinear structures is the complex argument distributions (CADs) introduced by Stankovic [165] and generalized later by Cornu et al. [174]. The purpose is to develop a distribution highly concentrated along the GD or the IF for the mono- and multicomponent signals. The CADs use complex-frequency argument (in the Laplace domain) and a corresponding complex-lag argument in the time domain [165]. These two new forms are able to produce almost completely concentrated representations along the GD or the IF. Since the signal is available along the real time axis only, the tools for calculation of a complex-valued argument form of the signal are also considered and proposed by Stankovic in [165]. These tools make use of the relation between the FT and the Laplace transform and the analytic extension of the signal [175]. Stankovic [165] also presents a procedure for application of CADs on the t-f analysis of multicomponent signals, which may produce CTs reduced forms of the CADs. The procedure is shown to be efficient in reducing the noise influence on the t-f representation of monocomponent noisy signals. Stankovic further analyzes

the CADs as an IF estimator in the noisy signal cases and derives the estimator's variance and bias.

Complex Argument Distribution—Definitions. A. *Complex-frequency distribution*. It is defined as [165]

$$\mathrm{CAD}_F(t.\omega) = \frac{1}{2\pi}\int X\left(j\omega + j\frac{\pi}{4}\right)X^*\left(j\omega - j\frac{\pi}{4}\right) \times X^j\left(j\omega + \frac{\theta}{4}\right)X^{-j}\left(j\omega - \frac{\theta}{4}\right)e^{j\theta t}d\theta, \tag{41}$$

where $x(t)$, $X(j\omega) = A(\omega)e^{j\varphi(\omega)}$ and $X(j\omega \pm j(\pi/4))$ are the signal, its FT pair, and its Laplace transform at $s = \pm\theta/4 + j\omega$. Stankovic numerically proves that the CAD of a signal $X(j\omega)$ is concentrated along the GD $t_g(\omega) = -\varphi'(\omega)$ with the lowest spreading term being of the fifth order. This means that a completely concentrated distribution for the phase of up to the fourth-order polynomial function of time can be obtained.

B. *Complex-Lag Distribution*. The IF is more commonly used signal parameter than the GD. Stankovic [165] introduces the TFD producing improved concentration along the IF by replacing frequency with time since any definition in frequency domain can be reintroduced in its dual form. In order to define a representation with complex-valued argument, mathematical quantity introduced is the "complex time" or complex-lag argument which is related to the time axis in the same way as the complex frequency is related to the frequency axis. The complex-lag distribution is defined by

$$\mathrm{CAD}_T(t.\omega) = \int x\left(t + \frac{\tau}{4}\right)x^*\left(t - \frac{\tau}{4}\right) \times x^{-j}\left(t + j\frac{\tau}{4}\right)x^j\left(t - j\frac{\tau}{4}\right)e^{-j\omega\tau}d\tau. \tag{42}$$

Simulation Examples. It has been shown through simulations that the inner artifacts in the representations of signals with fast varying frequency or GD could be significantly reduced. First example is a noisy monocomponent signal considered by Stankovic, which is expressed as [165]

$$x(t) = \exp\left[j\left(6\cos(\pi t) + \frac{2}{3}\cos(3\pi t) + \frac{2}{3}\cos(5\pi t)\right)\right] + n(t) \tag{43}$$

within the interval $t \in [-1, 1]$, with $\Delta t = 2/N$, and $N = 128$. The spectrogram, WD, and the complex-lag distribution are shown in Figure 18. Based on the distributions from Figure 18, the IF is estimated for various values of the noise standard deviation: $\sigma_n = 0.031$, $\sigma_n = 0.31$, $\sigma_n = 0.707$, and $\sigma_n = 1$. The estimated IF is shown in Figure 19. It can be seen that although the lag window is quite narrow, the bias in the WD and the spectrogram is significant and dominates in the estimation error. Variance in the complex-lag distribution is slightly higher, whereas the bias is significantly lower, thus improving the overall estimation. When the noise is increased, the number of instants where the IF estimator completely misses the IF increases [176] in the WD and the spectrogram. Therefore, although the variance of the estimation should not be large, the misses degrade the performance of these distributions. However, the performance of the complex-lag distribution is far ahead of all considered cases.

Second example considered by Stankovic is multicomponent signal given by

$$x(t) = \exp\left(j7.5\pi(0.5t^4 - 0.8t^2) - j8.5\pi t\right) + \exp j3\cos(\pi t) + \frac{\cos(3\pi t)}{2} + \frac{\cos(5\pi t)}{2} + 8.5\pi t + n(t). \tag{44}$$

According to the procedure for a multicomponent signal realization illustrated by Stankovic, the representations are obtained shown in Figure 20. The case of real-valued signal

$$x(t) = \cos\left[9\cos(\pi t) + \frac{2}{3}\cos(3\pi t) + \frac{5}{7}\cos(5\pi t)\right] + n(t) \tag{45}$$

whose components intersect, is shown in Figure 20, as well. CTs between positive and negative frequencies are removed in the same way as other CTs.

Generalized Representations of Phase Derivative for Regular Signals. Recently a class of generalized complex-lag distributions (GCADs) is proposed by Cornu et al. in [174]. These distributions based on generalized complex-lag moment (GCM) provide the representation of arbitrary instantaneous phase derivative (IPD). They can estimate any order of IPDs and produce high concentration. These GCADs can provide an accurate IF estimation, even in the case of significant IF variation within only a few signal samples. Moreover the IF rate estimation introduced by some of the existing methods (e.g., [177]) can also be obtained with a slight modification of a particular form of the GCAD. In addition, better IFR concentration can be achieved by using higher order t-f rate distributions belonging to the proposed class of distributions. These distributions are parameterized by two integers K and N. One important property is that they provide high concentration along the Kth derivative of the phase. A special case of this class of distributions, for $K = 1$, provides distributions for t-f analysis. Among them, the WD is one of the special cases. The most interesting distributions for $K = 2$ (time-"frequency rate" analysis) and for $K = 3$ are analyzed too by Cornu et al.

A. *Concept*. Cornu et al. make use of the ideal representation of an arbitrary IPD, given as [174]

$$\mathrm{IPD}_K(t, \Omega) = \delta\left(\Omega - \Phi^{(K)}(t)\right), \tag{46}$$

where Ω is the axis which corresponds to the Kth derivative of the phase function $\Phi(t)$ for the signals of the form $s(t) = A\exp(j\Phi(t))$. The authors term this distribution in general the ideal time "phase derivatives" distribution, that provides this representation. The IF representation and the IF rate

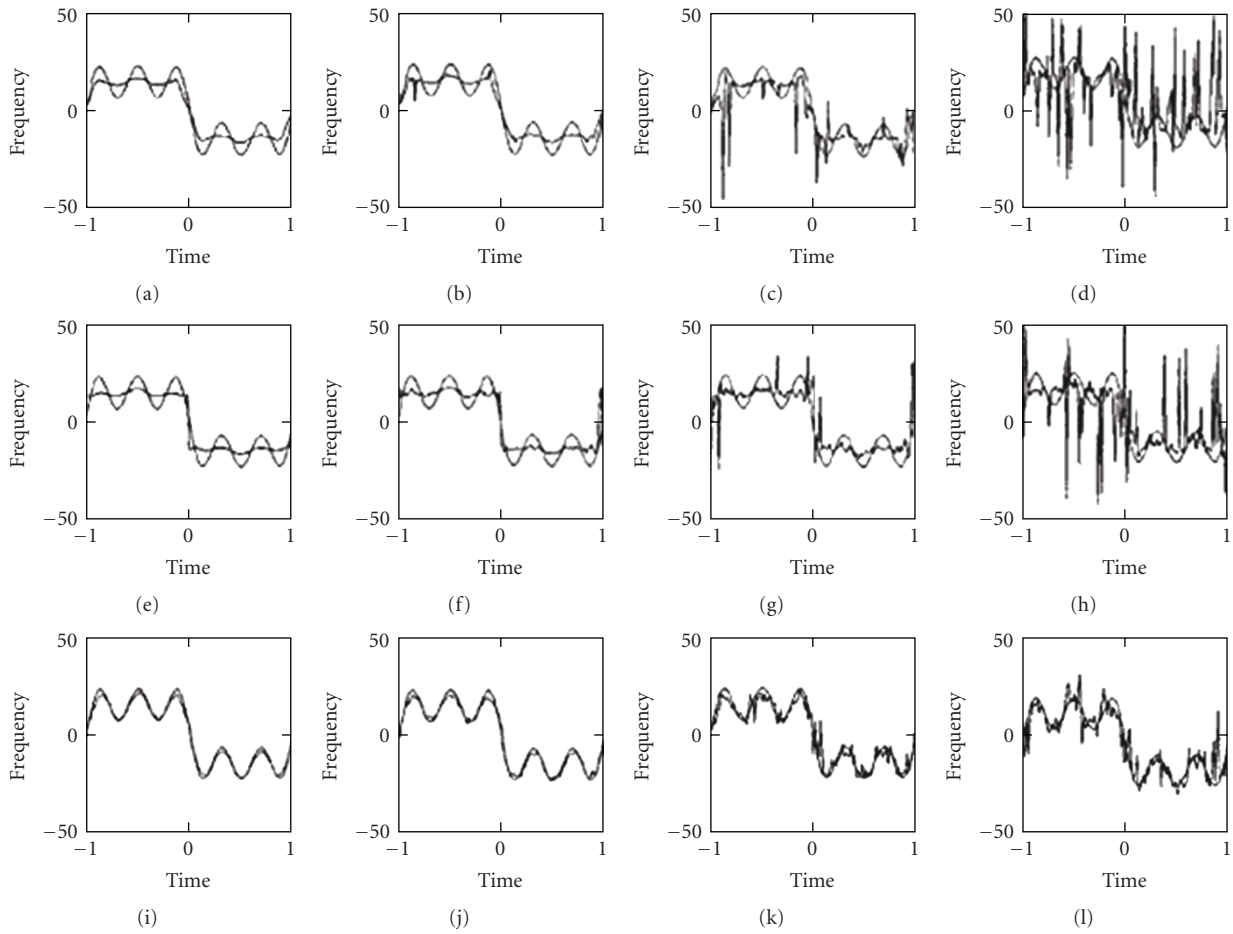

Figure 19: (a)–(d) IF estimation based on WD, (e)–(h) spectrogram, (i)–(l) complex lag-distribution. First column: SNR = 30 dB. Second column: SNR = 10 dB. Third column: SNR = 3 dB. Fourth column: SNR = 0 dB (all adopted from Stankovic [165]).

can subsequently be obtained by replacing $K = 1$ and 2, respectively, in (46).

By making use of complex variable analysis [175], Cornu et al. define complex moment in order to otbain distribution concentrated along Kth phase derivative, as

$$\mathrm{GCM}_N^K[s](t,\tau) = \prod_{k=0}^{N-1} s^{\omega_{N,k}^{N-K}}(t+\omega_{N,k}\tau), \tag{47}$$

where $\omega_{N,k} = e^{j2\pi k/N}$, and (47) is introduced to produce phase in discrete form as

$$\mathrm{Angle}\Big[\mathrm{GCM}_N^K[s](t,\tau)\Big] = \sum_{k=0}^{N-1} \Phi(t+\omega_{N,k}\tau)\omega_{N,k}^{N-K} \tag{48}$$

by expanding Φ into Taylor's series, and subsequent simplication leads to the following expression for the discrete phase:

$$\mathrm{Angle}\Big[\mathrm{GCM}_N^K[s](t,\tau)\Big] = \Phi^{(K)}(t)\frac{N\tau^K}{K!} + \mathbb{R}(t,\tau), \tag{49}$$

where the remainder $\mathbb{R}(t,\tau)$, the spread factor of the GCAD, is written as

$$\mathbb{R}(t,\tau) = N\sum_{k=1}^{\infty}\Phi^{(Nk+K)}(t)\frac{\tau^{Nk+K}}{(Nk+K)!}. \tag{50}$$

To obtain a distribution concentrated along along Kth phase derivative, Cornu et al. linearizes the first term in (49) with respect to τ. As a result, the GCM becomes

$$\mathrm{GCM}_N^K[s](t,\tau) = \prod_{k=0}^{N-1} s^{\omega_{N,k}^{N-K}}\left(t+\omega_{N,k}\sqrt[K]{\frac{K!}{N}}\tau\right). \tag{51}$$

By taking the FT of the GCM produces the GCAD, whose mathematical expression is given as [174]:

$$\begin{aligned}\mathrm{GCAD}_N^K[s](t,\omega) &= \mathrm{F}_\tau\Big[\mathrm{GCM}_N^K[s](t,\tau)\Big]\\ &= \mathrm{F}_\tau\Big[e^{j\Phi^{(K)}(t)\tau}e^{j\mathbb{R}(t,\tau)}\Big]\\ &= \delta\Big(\omega-\Phi^{(K)}(t)\Big)\omega^*\mathrm{F}_\tau\Big[e^{j\mathbb{R}(t,\tau)}\Big].\end{aligned} \tag{52}$$

It can be concluded that a high value of N reduces interferences in the resultant GCAD, since the spread factor $\mathbb{R}(t,\tau)$ is reduced.

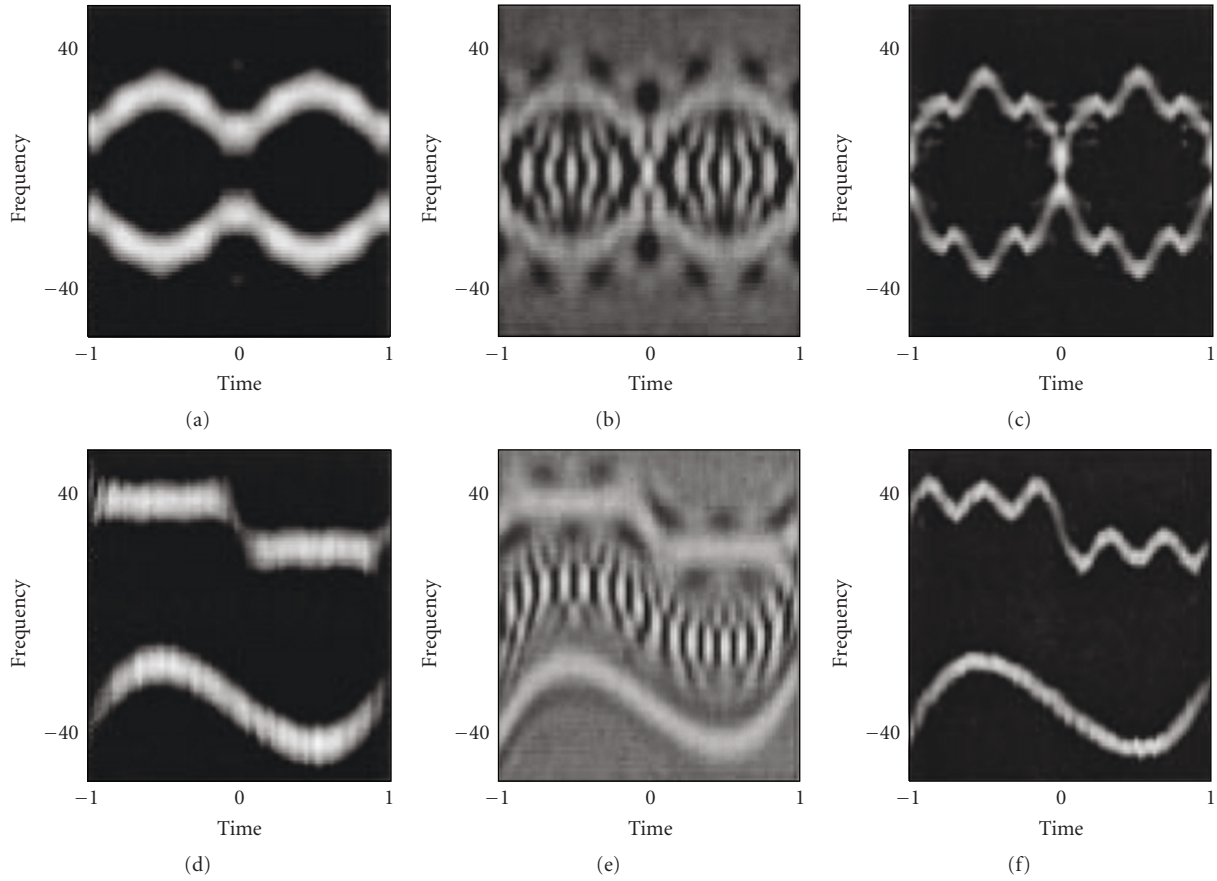

Figure 20: TFD of a multicomponent signal with highly nonlinear IF. Spectrogram (a), (d), WD (b) , (e), complex-lag distribution (c) , (f). (a) , (b), (c): noisy signal with SNR = 30 dB. (d), (e) , (f): noisy signal with SNR = 20 dB (all adopted from Stankovic [165]).

B. *IF Representation Using Test Signals*. Cornu, Stankovic, Ioana, and Ljubisa Stankovic test the proposed GCADs on several signals and compare the results with conventional representations. Two test cases are shown here to determine the effectiveness of proposed approach. The first one is a periodically FM signal with rapid frequency variations added with noise. The second one is still a periodically FM signal with faster frequency variations without noise. The WD and the GCAD of first signal are depicted in Figures 21(a)-21(b). The signal is contaminated with a white Gaussian noise. The SNR is about 10 dB. The WD cannot follow the frequency variation of the signal since it is highly nonlinear. The sixth-order GCAD is naturally more robust to noise and exhibits a better signal representation. It is almost interference free and has no artifacts in this case. The second signal, has very rapid frequency variations (TFDs shown in Figures 21(c)-21(d)). The sixth-order GCAD is still well fitted to the theoretical IF, and the interferences level is negligible compared with the WD.

2.2.6. Neural Network-Based TFDs. A neural network- (NN-) based method to compute TFDs is proposed by Shafi et al. [178, 179]. The output TFDs (henceforth neural network TFDs (NTFDs)) are free of any blurring effect, without any prior knowledge of the components in the signal. Figure 22 is the general block form representation of the method. The authors treat TFDs as 2D images and recognize vectors of different nature based on various edges. These vectors are separated and clustered according to the elbow criterion. The multiple Bayesian regularized neural networks (BRNNs) are trained for each group of vectors in a cluster and by keeping track of the network error or performance, accessible via the training record, the best network is selected in terms of training performance. These selected networks are termed the localized neural networks (LNNs), specialized for one type of vectors each.

The Method. The NN-based model can be described as the combination of three processes which include the preprocessing of input data, training/testing of the BRNNM, and the postprocessing of output data. The spectrogram and highly concentrated WD of known signals are used to train the BRNNs. The presence of CTs in the target TFDs make them unsuitable to be presented as target to NN [180], and are therefore removed before the TFD is fed to NNs. This is achieved by multiplying the WD with Gabor transform of the signal obtained with reasonably sized hamming window [181].

The input and target TFDs are converted into vectors. The mean of pixel values so obtained are computed with a

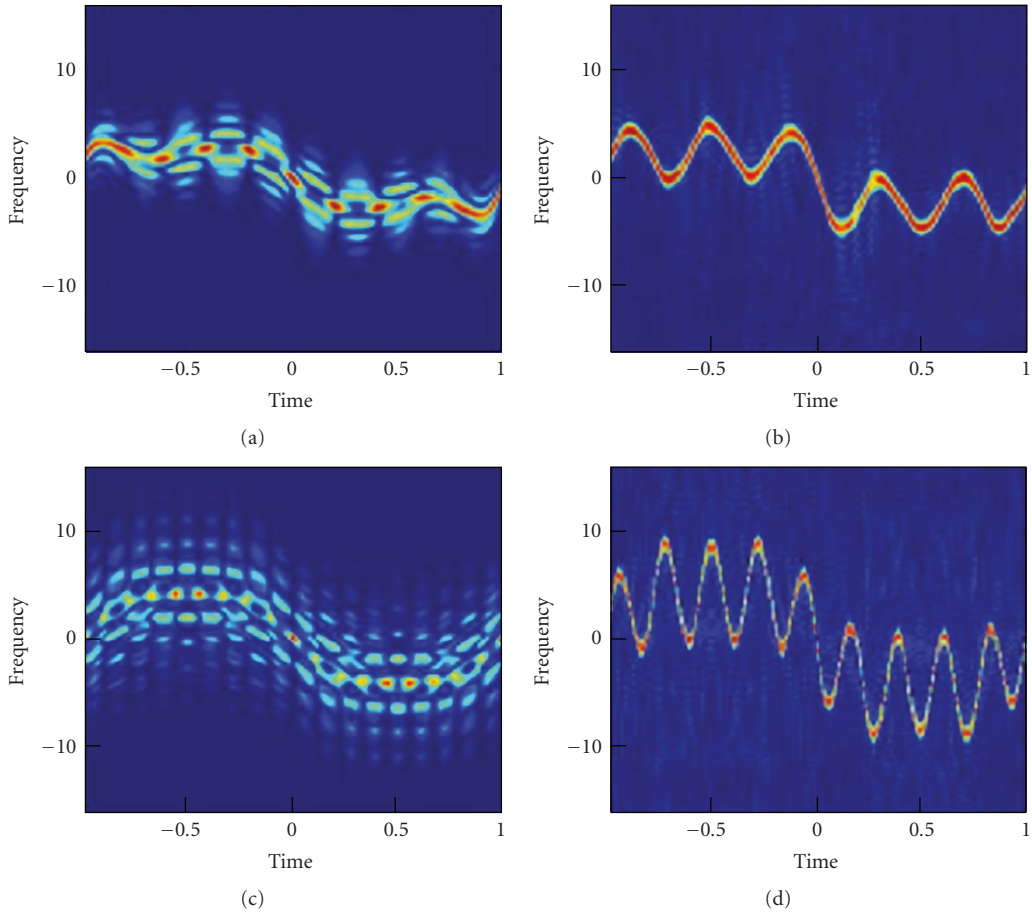

Figure 21: TFDs of test signals, (a), (b) a noisy signal, SNR = 10 dB: (a) WD, and (b) sixth GCAD, (c), (d) a signal with fast-varying IF: (c) WD, and (d) sixth GCAD (all adopted from Cornu et al. [174]).

view that the IF can be computed by averaging frequencies at each time instant, a definition suggested by many researchers [3]. These empirical values are paired with the vectors obtained from target TFDs. The vectors are clustered based on the elbow criterion [182] in relation to underlying image features like edges present in the data. The objective is to divide the input vector space into number of sub-spaces, S_n, described by directional unit vectors, v_n. A vector will lie in the subspace S_n represented by v_n that is most similar to this vector measured by an inner-product similarity measure.

A single layer feed-forward backpropagation NN model with 40 hidden units is used. The Bayesian regularized Lavenberg-Marquardt backpropagation training algorithm is used for training the multiple NNs. The hidden layer of sigmoid neurons followed by an output layer of positive linear neurons, respectively, are fixed as the transfer functions [183, 184]. For each cluster, multiple NNs are trained by the selected training algorithm incorporating the Bayesian regularization. By keeping track of the network error, accessible via the training record, the best network is selected in terms of training performance for each cluster. The best networks for the respective clusters are called the LNNs. The LNNs are then fed with the test image vectors and the resultant data is postprocessed.

Simulation Result. For this paper, a real-life multicomponent signal used by the authors is shown to check the proposed model's performance. The test spectrograms is depicted in Figure 23(a), and the NTFD for this test signal is shown in Figure 23(b). For better comparison, Figures 23(c) and 23(d) depict the WD and reassigned TFD for the same test signal. Visual results indicate the superiority of NTFD as compared to the other TFDs. The NTFDs are deblurred but are not valid energy distributions because they do not observe the signature continuity and marginal characteristics or weak signal mitigation and lack the parametric representations. Due to this reason, the results may not be feasible for certain applications as different applications have different preference and requirement to the TFDs. This problem may be attributed to the discontinuities in the target TFDs and is expected to overcome by better pre- and postprocessing techniques and by using continuous target TFDs.

2.2.7. TFDs Based on Signal Expansions. The wide scope of patterns embedded in complex signals and the precision of their characterization motivate decompositions over large and redundant dictionaries of waveforms. Linear expansions in a single basis, whether it is a Fourier, wavelet, or any other basis, are not flexible enough. In Fourier and wavelet

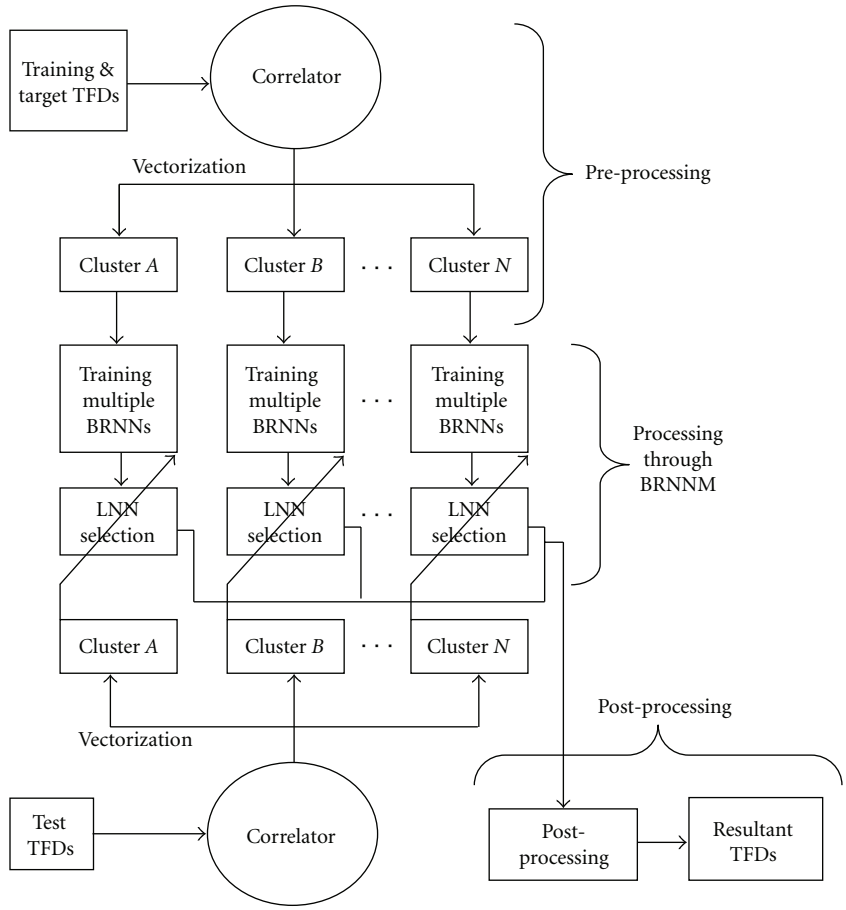

Figure 22: Flow diagram of the NN-based method [178].

basis, it is difficult to detect and identify the signal patterns from their expansion coefficients, because the information is diluted across the whole basis. Due to this reason, there has been an explosion of interest in alternatives to traditional signal representations. Instead of just representing signals as superpositions of sinusoids (the traditional Fourier representation) now there are available alternate dictionaries. Out of such dictionaries, that is, the collections of parameterized waveform, the wavelets dictionary is the best known. Wavelets, steerable wavelets, segmented wavelets, Gabor dictionaries, multiscale Gabor dictionaries, wavelet packets, cosine packets, chirplets, warplets, and a wide range of other dictionaries are now available. Each such dictionary D is a collection of parameterized waveforms $(\varphi_\mu)_{\mu\in\Gamma}$, with μ a parameter. The waveforms φ_μ are discrete-time signals of length n called atoms. Depending on the dictionary, the parameter μ can have the interpretation of indexing frequency, in which case the dictionary is a frequency or Fourier dictionary, of indexing time scale jointly, in which case the dictionary is a time-scale dictionary, or of indexing t-f jointly, in which case the dictionary is a t-f dictionary. A decomposition of a signal x can be envisioned as [185]

$$x = \sum_{i=1}^{m} \gamma_{\mu_i} \varphi_{\mu_i} + \mathbb{R}(m), \tag{53}$$

where γ_{μ_i} are the coefficients and $\mathbb{R}(m)$ is a residual. Depending on the dictionary, such a representation decomposes the signal into pure tones (Fourier dictionary), bumps (wavelet dictionary), chirps (chirplet dictionary), and so forth.

Finding a Suitable Representation Leading to High-Resolution TFDs. The decomposition in (53) is nonunique, because some elements in the dictionary may have representations in terms of other elements. Nonuniqueness gives the possibility of adaptation, that is, of choosing from among many representations one that is most suited to the purposes considered. The advantages sought could be summarized as follows.

(1) *Sparsity*. The sparsest possible representation of the object, that is, the one with the fewest significant coefficients will be obtained.

(2) *Superresolution*. A resolution of sparse objects that is much higher resolution than that possible with traditional nonadaptive approaches will be obtained.

(3) *Speed*. An important constraint which is perhaps in conflict with both the earlier goals. It should be possible to obtain a representation in order $O(n)$ or $O(n\log(n))$ time.

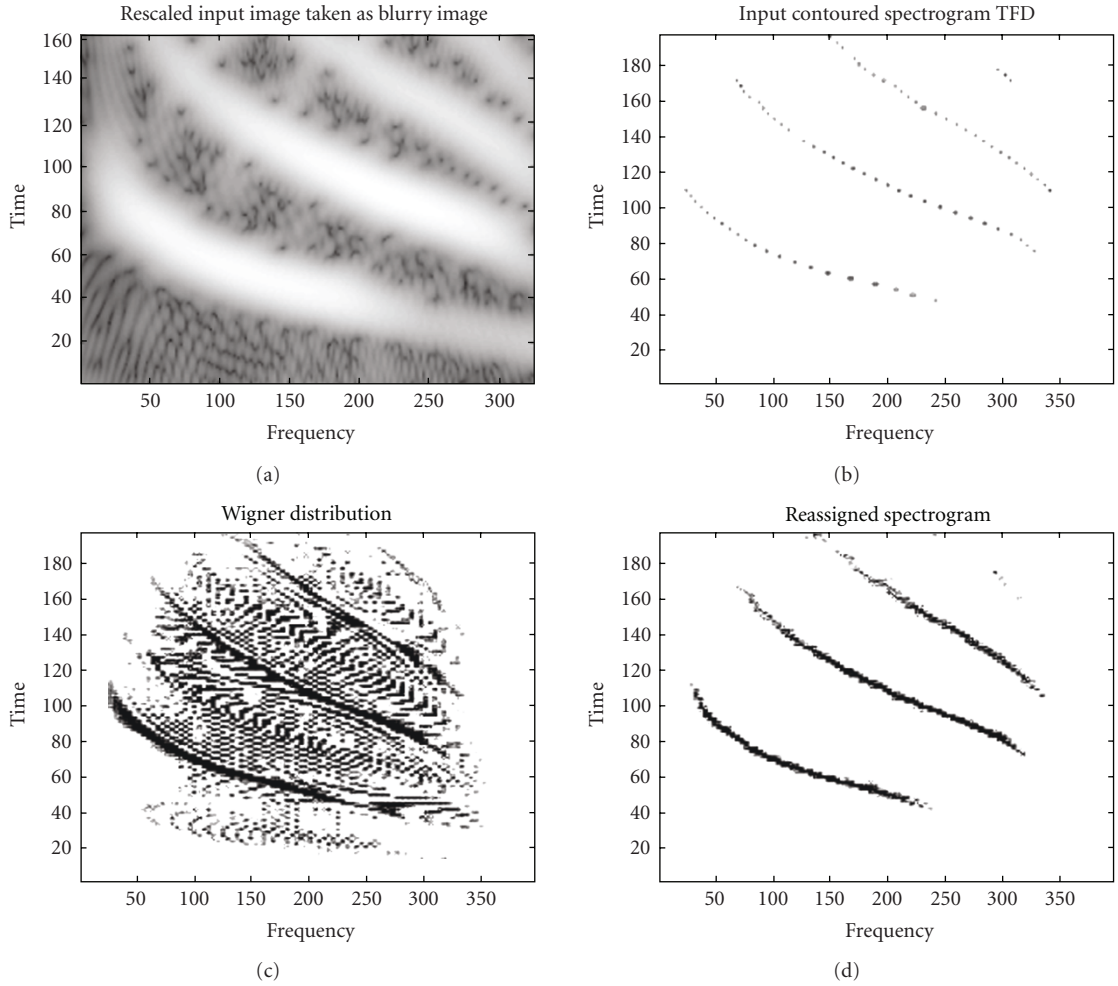

Figure 23: Various TFDs for bat chirps signal, (a) the spectrogram (test TFD), (b) NTFD [178], (c) WD, and (d) reassigned TFD [121].

Several methods have been proposed for obtaining signal representations in *overcomplete dictionaries*. (Because they start out that way or because complete dictionaries are merged, obtaining a new megadictionary consisting of several types of waveforms (e.g., Fourier and wavelets dictionaries.)) These range from general approaches, like the method of frames (MOF) [186] and the method of matching pursuit (MP) [187], to clever schemes derived for specialized dictionaries, like the method of best orthogonal basis (BOB) [188]. These classical methods have both advantages and shortcomings. The principal emphasis of the proposers of these methods is on achieving sufficient computational speed. While the resulting methods are practical to apply to real data, several computational examples reveal that the methods, either quite generally or in important special cases, lack qualities of sparsity preservation and of stable superresolution.

The expansion of the STSC into an infinite number of t-f shifted versions of a weighted elementary atom based on these methods and then applying suitable t-f transform method like WD will result in highly cocentrated and good resolution TFDs. We will discuss some important signal expansion concepts and the resulting TFDs in succeeding paragraphs, from which the t-f research community has specially been benefitted.

Matching Pursuits TFDs with Time-Frequency Dictionaries. Mallat and Zhang [187] introduce an algorithm called MP, that decomposes any signal into waveforms selected among a dictionary of t-f atoms, that are the dilations, translations, and modulations of a single window function. This is achieved using successive approximations of the signal with orthogonal projections on dictionary elements. These waveforms are selected in order to best match the signal structures. Similar algorithms were proposed by Qian and Chen [189] for Gabor dictionaries and by Villemoes [190] for Walsh dictionaries. The MPs provide extremely flexible signal representations since the choice of dictionaries is not limited. Moreover the properties of the signal components are explicitly given by the scale, frequency, time and phase indexes of the selected atoms. This representation is therefore well adapted to information processing. Although an MP is nonlinear, like an orthogonal expansion, it maintains an energy conservation which guaranties its convergence. Mallat

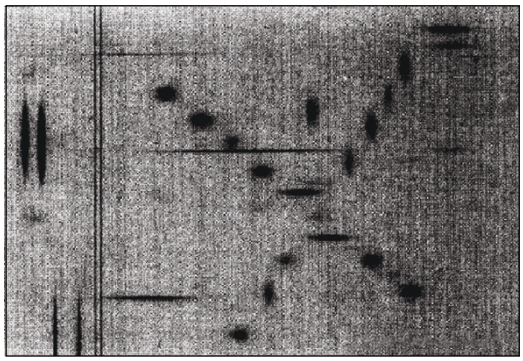

Figure 24: TFD of the example signal. The horizontal axis is time. The vertical axis is frequency. The highest frequencies are on the top. The darkness of this t-f image increases with the value of TFD.

and Zhang then derive a t-f energy distribution, by adding the WD of the selected t-f atoms. Contrarily to the WD or Cohen's class distributions, this energy distribution does not include interference terms and thus provides a clear picture in the t-f plane.

Compact signal coding is another important domain of application of MPs. For a given class of signals, if the dictionary can be adapted to minimize the storage for a given approximation precision, better results are guaranteed than decompositions on orthonormal bases. Indeed, an orthonormal decomposition is a particular case of MP where the dictionary is the orthonormal basis. For dictionaries that are not orthonormal bases, the inner products of the structure book and the indexes of the selected vectors need coding. This requires to quantize the inner product values and use a dictionary of finite size. The MP decomposition is then equivalent to a multistage shape-gain vector quantization in a very high dimensional space. For information processing or compact signal coding, it is important to have strategies to adapt the dictionary to the class of signal, that is, decomposed. If enough prior information is available, the dictionary can be adapted to the probability distribution of the signal class within the signal space. Finding strategies to optimize dictionaries in high dimensions is an open problem that shares similar features with learning problems in NNs.

Numerical Example. Mallat and Zhang formulates the discrete implementation of an MP for a dictionary of Gabor t-f atoms with numerical examples. For example, here the TFD of a signal $s(t)$ that is built by adding chirps, truncated sinusoidal waves and waveforms of different t-f localizations is shown in Figure 24. Each Gabor t-f atom selected by the MP is an elongated Gaussian blob in the t-f plane. Appearance of two chirps that cross each other, with a localized t-f waveform at the top of their crossing point is clearly seen. We can also detect closely spaced Diracs, and truncated sinusoidal waves having close frequencies. Several isolated localized t-f components also appear in this energy distribution.

Basis Pursuit TFDs. The basis pursuit (BP) proposed by Chen et al. [185] finds signal representations in overcomplete dictionaries by convex optimization; it obtains the decomposition that minimizes the η^1 norm of the coefficients occurring in the representation. Because of the nondifferentiability of the η^1 norm, this optimization principle leads to decompositions that can have very different properties from the MOF—in particular, they can be much sparser. Because it is based on global optimization, it can stably superresolve in ways that MP cannot. Moreover BP can be used with noisy data by solving an optimization problem trading off a quadratic misfit measure with an η^1 norm of coefficients. BP is closely connected with linear programming. Recent advances in large-scale linear programming—associated with interior-point methods—can be applied to BP and can make it possible, with certain dictionaries, to nearly solve the BP optimization problem in nearly linear time.

There are important connections between BP and methods like Mallat and Zhong's MP [187] multiscale edge representation and Rudin et al. [133] total variation-based denoising methods, while experimenting with some nonstandard dictionaries, like the stationary wavelet dictionary and the heaviside dictionary.

A. BP Optimization Principle. If it is assumed that the dictionary is overcomplete, then there are in general many representations as in (53). The principle of BP is to find a representation of the signal whose coefficients have minimal η^1 norm. Formally, one solves the problem

$$\min ||\gamma||_1 \text{ subject to } \Phi\gamma = x. \tag{54}$$

From one point of view, (54) is very similar to the MOF where solution to $\min \|\gamma\|_2$ subject to $\Phi\gamma = x$ is sought. Here for BP simply η^1 replaces the η^2 norm as done in (54). However, this apparently slight change has major consequences. The MOF leads to a quadratic optimization problem with linear equality constraints and so involves essentially just the solution of a system of linear equations. In contrast, BP requires the solution of a convex, nonquadratic optimization problem, which involves considerably more effort and sophistication.

The solution of (54) can be obtained by solving an equivalent linear program [191]. The linear programing in so-called standard form [191] is a constrained optimization problem defined in terms of a variable $x \in R^m$ by

$$\min c^T x \text{ subject to } Ax = b, \quad x \geq 0, \tag{55}$$

where $\min c^T x$ is the objective function, $Ax = b$ is a collection of equality constraints, and $x \geq 0$ is a set of bounds. The main question is which variables should be zero. Reformulation of the BP problem is therefore needed by making suitable translations. Thereafter any algorithm from the linear programming literature can be considered as a candidate for solving the BP optimization problem; both the simplex and interior-point algorithms offer interesting insights into BP.

BP-Simplex Algorithm. In standard implementations of the simplex method for linear programming, one first

finds an initial basis B consisting of n linearly independent columns of A for which the corresponding solution $B^{-1}b$ is feasible (nonnegative). Then one iteratively improves the current basis by swapping, at each step, one term in the basis for one term not in the basis, using the swap that best improves the objective function. There always exists a swap that improves or maintains the objective value, except at the optimal solution. Hence the simplex algorithm is explicitly a process of BP; iterative improvement of a basis until no improvement is possible, at which point the solution is achieved.

Translating this linear programming algorithmin to BP terminology, one starts from any linearly independent collection of n atoms from the dictionary. One calls this the current decomposition. Then one iteratively improves the current decomposition by swapping atoms in the current decomposition for new atoms, with the goal of improving the objective function.

BP-Interior Point Algorithm. The collection of feasible points $\{x : Ax = b, x \geq 0\}$ is a convex polyhedron in R^m (a "simplex"). The simplex method, viewed geometrically, works by walking around the boundary of this simplex, jumping from one vertex (extreme point) of the polyhedron to an adjacent vertex at which the objective is better. Interior point methods instead start from a point $x^{(0)}$ well inside the interior of the simplex ($x^{(0)} \gg 0$) and go "through the interior" of the simplex. Since the solution of a linear program is always at an extreme point of the simplex, as the interior-point method converges, the current iterate $x^{(k)}$ approaches the boundary. One may abandon the basic interior-point iteration and invoke a "cross-over" procedure that uses simplex iterations to find the optimizing extreme point.

Translating this linear programming algorithmin to BP terminology, one starts from a solution to the overcomplete representation problem $\Phi y^{(0)} = x$ with $y^{(0)} > 0$. One iteratively modifies the coefficients, maintaining feasibility $\Phi y^{(k)} = x$ and applying a transformation that effectively sparsifies the vector $y^{(k)}$. At some iteration, the vector has $\leq n$ significantly nonzero entries, and it "becomes clear" that those correspond to the atoms appearing in the final solution. One forces all the other coefficients to zero and "jumps" to the decomposition in terms of the $\leq n$ selected atoms.

B. Examples. Chen et al. [185] consider number of practical signals to demonstrate the effectiness of the proposed BP method. Here two synthetic examples are presented including (i) an FM sinusoid superimposed with a pure sinusoid, and (ii) a composite of six atoms: a Dirac, a sinusoid, and four mutually orthogonal wavelet packet atoms.

Figure 25(a) displays the artificial signal consisting of an FM sinusoid superposed with a pure sinusoid: $x = \cos(\delta_0 t) + \cos((\delta_0 t + \alpha\cos(\delta_1 t))t)$. Figure 25(b) shows the ideal phase plane. In Figure 25(c)–25(f), the signal is analyzed using the cosine packet dictionary based on a bell 16 samples wide. It is evident that BOB cannot resolve the nonorthogonality between the sinusoid and the FM signal. Neither can MP. However, BP yields a clean representation of the two structures.

The synthetic signal which is composite of six atoms, adjacent in the t-f plane is depicted in Figure 26. The wavelet packet dictionary of depth $D = \log_2(n)$ is employed, based on filters for symmlets with eight vanishing moments. Figure 26 displays the results in phase-plane form; for comparison, the phase planes obtained using MOF, MP, and BOB are also included. First, note that MOF uses all basis functions that are not orthogonal to the six atoms, that is, all the atoms at times and frequencies that overlap with some atom appearing in the signal. The corresponding phase plane is very diffused or smeared out. Second, MP is able to do a relatively good job on the sinusoid and the Dirac, but it makes mistakes in handling the four close atoms. Third, BOB cannot handle the nonorthogonality between the Dirac and the cosine; it gives a distortion (a coarsening) of the underlying phase plane picture. Finally, BP finds the "exact" decomposition in the sense that the four atoms in the quad, the Dirac, and the sinusoid are all correctly identified.

TFDs Based on Empirical Mode Decomposition. Recently, a new data-driven technique, referred to as empirical mode decomposition (EMD), has been introduced by Huang et al. [192], for analyzing data from nonstationary and nonlinear processes. In their original paper, Huang et al. introduce a general method which requires two steps in analysing the data. The first step is to preprocess the data by the EMD method, with which the data are decomposed into a number of intrinsic mode function (IMF) components. Thus, the data is expanded in a basis derived from the data. The second step is to apply the Hilbert transform to the decomposed IMFs and construct the energy-frequency-time distribution, designated as the Hilbert spectrum, from which the time localities of events are preserved. This construction of TFD is offcourse not limited to any one technique, and the better methods may be used to get TFDs that become highly localized in t-f domain.

The EMD has received more attention in terms of applications [193–205] and interpretations [206, 207]. The major advantage of the EMD is that the basis functions are derived from the signal itself. Hence, the analysis is adaptive in contrast to the traditional methods where the basis functions are fixed. The EMD is based on the sequential extraction of energy associated with various intrinsic time scales of the signal, starting from finer temporal scales (high-frequency modes) to coarser ones (low-frequency modes). The total sum of the IMFs matches the signal very well and, therefore, ensures completeness [192].

The idea is to decompose time series into superposition of components with well-defined Ifs, that is, the IMFs. The components should (approximately) obey earlier requirements of completeness, orthogonality, locality, and adaptiveness. Next construct the Hilbert spectrum of each IMF, representing it in the t-f plane. However the appropriate t-f representation (e.g., reassignment method) of the decomposed IMF result into highly concentrated TFDs as shown in Figure 27 for a synthetic three-component example

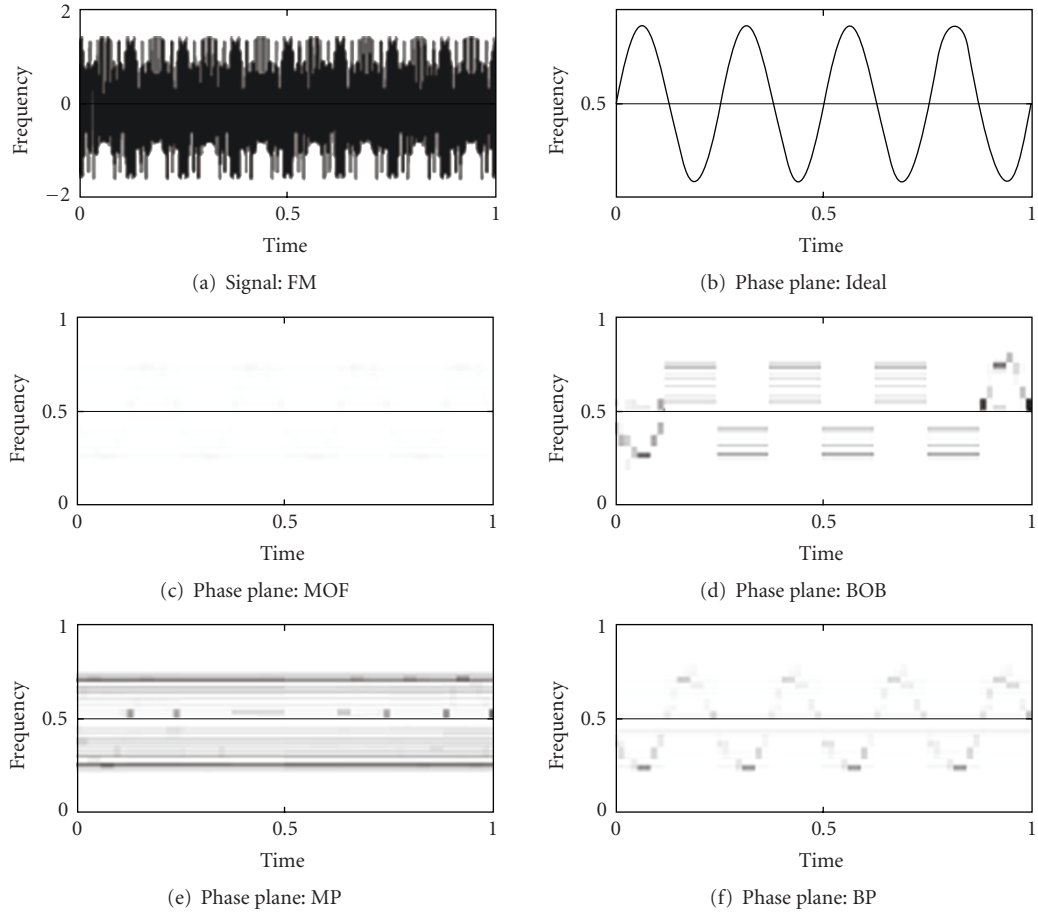

Figure 25: Analyzing the FM cosine signal with a cosine packet dictionary using MOF, BOB, MP, and BP methods (adopted from [185]).

considered by Rilling and Flandrin [208]. The signal in the top row is decomposed by the EMD, resulting in the three IMFs listed below and six others that are not displayed since they are almost zero (they contain less than 0.3% of the total energy). The t-f analysis of the signal (top left of the four bottom diagrams) reveals three t-f signatures that overlap in both time and frequency, thus forbidding the components to be separated by any nonadaptive filtering technique. The t-f signatures of the first three IMFs are extracted by EMD evidence that these modes efficiently capture the three-component structure of the analyzed signal. All TFDs are reassigned spectrograms in this case.

Huang's Algorithm for EMD. The aim of the EMD is to get a representation of the form, given an observation $x(t)$:

$$x(t) = \sum_{k=1}^{K} a_k(t)\psi_k(t), \tag{56}$$

where the $a_k(t)$ measure "amplitude modulations" and the $\psi_k(t)$ "oscillations." The EMD involves the decomposition of $x(t)$ into a series of IMFs through the sifting process, with each one having a distinct time scale [192]. The decomposition is based on the local time scale of the signal and yields adaptive basis functions. The EMD can be seen as a type of wavelet decomposition whose subbands are built up as needed to separate the different components of $x(t)$. Each IMF then replaces the detail signals of $x(t)$ at a certain scale or frequency band [206]. The EMD picks out the highest frequency oscillation that remains in $x(t)$. A function is an IMF if either the number of extrema and the number of zero crossings are equal or differ at most by one, and at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima are zero. Thus, locally, each IMF contains lower frequency oscillations than the one that was extracted before. The EMD does not use any predetermined filter or wavelet function, and thus, it is a fully data-driven method [192]. To be successfully decomposed into IMFs, the signal $x(t)$ must have at least two extrema: one minimum and one maximum.

Implementation. Sifting process involves four major steps [192]. The idea is to identify (locally) the fastest oscillation, subtract it to the initial signal, and iterate it on the residual as follows.

(1) Identify local maxima and minima in the signal.

(2) Deduce an upper and a lower envelope by interpolation (cubic splines)

  (a) Subtract the mean envelope from the signal.

  (b) Iterate until $\#\{extrema\} = \#\{zeroes\} \pm 1$.

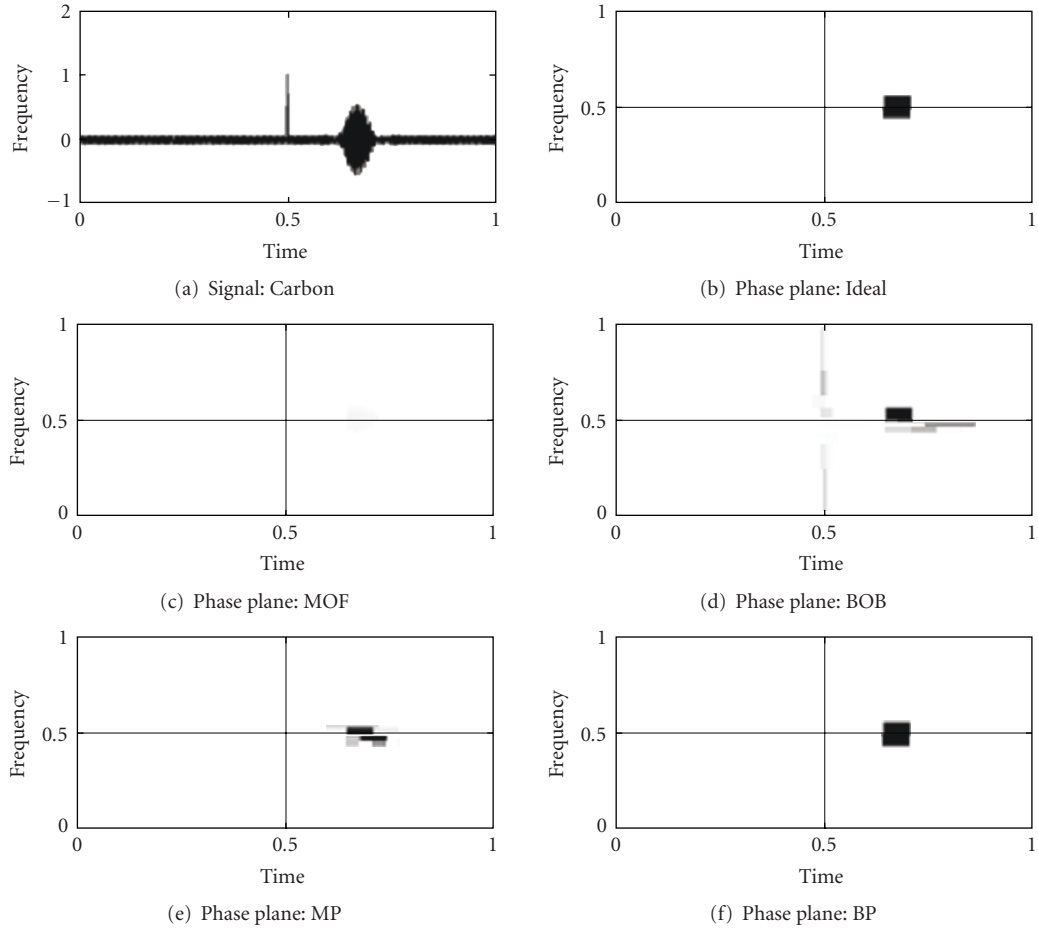

Figure 26: Analyzing the signal which is a composit of six atoms by the MOF, BOB, MP, and BP methods (adopted from [185]).

(3) Subtract the so-obtained mode from the signal.

(4) Iterate on the residual.

The result of the sifting is that $x(t)$ is decomposed into $\mathrm{IMF}_j(t)$, $j = 1, \ldots, C$, and a residual $\mathbb{R}_C(t)$ given by

$$x(t) = \sum_{j=1}^{C} \mathrm{IMF}_j(t) + \mathbb{R}_C(t), \tag{57}$$

where C is the number of modes, which is automatically determined using the stopping criterion. Thus C is signal dependant. The output of the EMD is thus a priori some sort of adaptive multiresolution decomposition [202]. In order to better assess the potential of the method, Flandrin and Gonçalves illustrate its behavior in Figure 27 on a synthetic signal. The results show that the EMD may be very efficient at naturally decomposing signals that are a burden to handle with usual methods based on Fourier or wavelet transform and often necessitate ad hoc solutions. The method proved useful in a variety of applications as diverse as climate variability [209], biomedical engineering [210], or blind-source separation [211]. The technique is, however, faced with the difficulty of being essentially defined by an algorithm, and therefore of not admitting an analytical formulation which would allow for a theoretical analysis and performance evaluation.

Matching Pursuit Adaptive TFDs. A novel approach to extract the IF from its adaptive TFD is proposed recently by Krishnan [212]. The adaptive TFD of a signal is obtained by decomposing the signal into components with reasonable t-f localization and by combining the WD of the components. The adaptive TFD, thus obtained, is free of CTs and is a positive TFD but it does not satisfy the marginal properties. The marginal properties are achieved by applying the MCE optimization to the TFD. Then, IF may be obtained as the first central moment of this adaptive TFD. Krishnan has shown successful extraction of the IF of a set of real-world and synthetic signals of known IF dynamics with the proposed method. In [213], a solution to the multicomponent problem was given by proposing an algorithm to select an optimal TFD from a set of TFDs for a given signal. Krishnan, in his approach, has addressed the same problem by constructing TFDs according to the application in hand, that is, he has tailored the TFD according to the properties of the signal being analyzed. In his method, by using

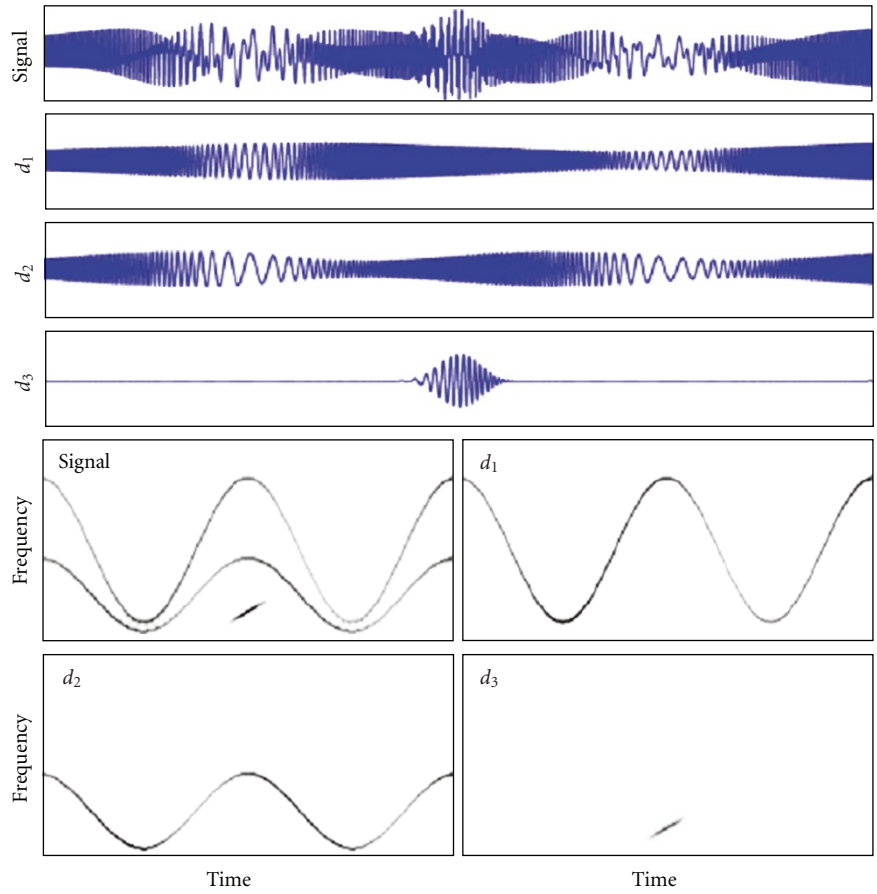

Figure 27: Synthetic three-component example. The signal in the top row is decomposed by the EMD, resulting in the three IMFs listed below. All TFDs are reassigned spectrograms (adopted from [208]).

constraints, the TFDs are modified to satisfy certain specified criteria. It is assumed that the given signal is somehow decomposed into components of a specified mathematical representation. By knowing the components of a signal, the interaction between them can be established and used to remove or prevent CTs. This avoids the main drawback associated with Cohen's class TFDs.

A. *Concept*. The key to successful design of adaptive TFDs lies in the selection of the decomposition algorithm. The components obtained from a decomposition algorithm depend largely on the type of basis functions used. Krishnan makes use of the MP algorithm [187], which decomposes the given signal using basis functions that have excellent t-f properties. The signal $x(t)$ is projected onto a dictionary of t-f atoms obtained by scaling, translating, and modulating a window function $\varsigma(t)$. In [212], the window is selected to be Gaussian type function considered most optimally, that is, $\varsigma(t) = 2^{1/4}\exp(-\pi t^2)$; the t-f atoms are then called Gabor atoms, and they provide the optimal t-f resolution in the t-f plane.

The algorithmic steps followed in [212] are as follows.

(1) The signal is iteratively projected onto a Gabor function dictionary.

(2) This process decomposes the signal into various parts including the inner product of the signal with various t-f atoms and the residue terms.

(3) The process is continued by projecting the residue onto the subsequent functions in the dictionary and after M iteration on simplification:

$$x(t) = \sum_{n=0}^{M-1} \langle R^n x, \varsigma_{\lambda_n}\rangle \varsigma_{\lambda_n} + R^M x(t), \tag{58}$$

where R^M and ς_{λ_n} are Mth residue after approximating in the direction of ς_{λ_n} and the nth Guassina type t-f atom, respectively, with $R^0 x(t) = x(t)$. The author suggests two ways to stop the iterations either by using a prescribed limiting number M of the t-f atoms or by checking the energy of the residue $R^M x(t)$. Doing so, Krishnan takes the WD of the t-f atoms in (58) and subsequently comes up with a signal-decomposition-based TFD after rejecting the CTs, which is given as

$$Q(t,\omega) = \sum_{n=0}^{M-1} |\langle R^n x, \varsigma_{\lambda_n}\rangle|^2 W_{\varsigma_{\lambda_n}}(t,\omega). \tag{59}$$

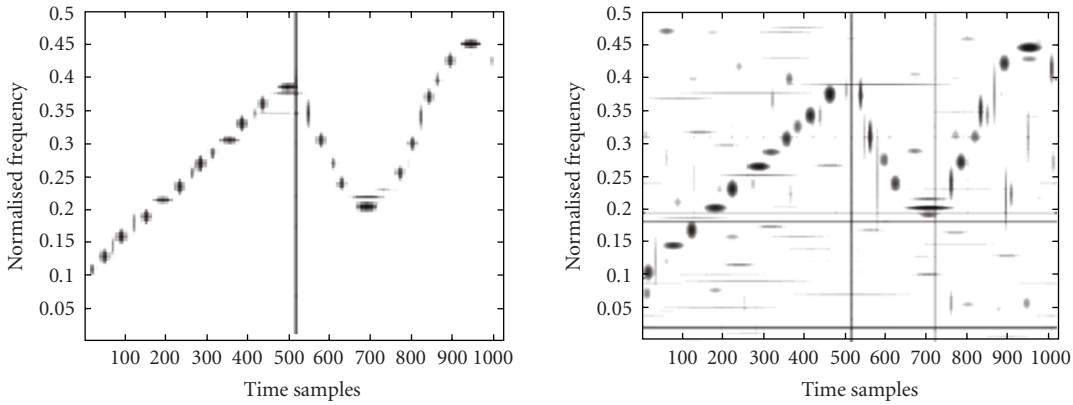

Figure 28: OMP TFDs of a monocomponent, nonstationary, synthetic signal consisting of a chirp, an impulse, and a sinusoidal FM component (SNR = 10 dB) and (SNR = 0 dB), respectively.

B. *Simulation Results*. The TFD in (59) is found free of CTs, termed by author as matching pursuit TFD (MPTFD). He further optimizes it using the cross-entropy minimization method [214, 215] to satify the marginal properties. The resultant TFD is found to have good signal representation and is claimed appropriate for analysis of nonstationary multicomponent signals. The IF of a signal is computed as the first moment of TFD long for each time slice. The method applied to synthetic signal composed of nonoverlapping chirp, transient, and sinusoidal FM components. To simulate noisy signal conditions, the signal is further corrupted by adding random noise of different SNR values. The suggested method by Krishnan gives a clear picture of the IF representation, as we find that the three simulated components are reasonably localized in the TFDs shown in Figure 28.

3. Concluding Remarks

The attempt to clearly understand what a time-varying spectrum is, and to represent the properties of a signal simultaneously in time and frequency without any ambiguity, is one of the most fundamental and challenging aspects of analysis. The t-f processing with regard to improved concentration and resolution is found essential for the ideal and unambiguous characterization of the STSC, a fact authenticated by the large amount of published scientific literature. In this reveiw paper, we attempt to provide a response to the following questions:

(1) why high concentration and good resolution is important?,

(2) what are the motivations of various researchers to propose and implement newer methods for this purpose? and most immportantly,

(3) how different researchers have used new ideas and implemented the techniques to achieve the desired objectives?

Concentrating on various methods and well-tested algorithms, the paper discusses their basic concept, important properties, implementation methods, and simulation results that emphasize the importance and significance of the technique to the analysis signals. Indeed different applications have different preferences and requirements to the TFDs. In general the choice of a TFD in a particular situation depends on many factors such as the relevance of properties satisfied by TFDs, the computational cost and speed of the TFD, and the trade-off in using the TFD. However as this task is achieved by many different types of t-f techniques, it is important to search for the one that is most pertinent to the application. Although the WD and the spectrogram QTFDs are often the easiest ot use, they do not always provide an accurate characterization of the real data. The spectrogram results in a blurred version, and the use of the WD in practical applications has been limited by the presence of CTs and inability to produce ideal concentration for nonlinear IF variations. The spectrogram, for example, could be used to obtain an overall characterization of the STSC structure, and then the information could be used to invest in another QTFD that is well matched to the data for further processing that requires information that is not provided by the spectrogram, an idea conceived and used by Shafi et al. [178].

Here, we barely scratch the surface of the possible ideas and methods that are used to obtain highly concentrated and good resolution distributions to achieve abovementioned objectives due to limitation of space. There are a large number of proposed methods, and only a few have been explored in a sequence with an aim to produce the ideas and techniques in a logical way. Our emphasis has been on the techniques and methodologies that have been developed steadily with stress over the fundamentals. It is important to highlight that all the concepts and techniques developed earlier or in the recent past are truly impressive. However, it is clear that still much more work lies ahead.

References

[1] L. Cohen, "Time-frequency distributions—a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.

[2] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. II. Algorithms and applications," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 540–568, 1992.

[3] L. Cohen, *Time Frequency Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 1995.

[4] S. Erkucuk, S. Krishnan, and M. Zeytinoglu, "Robust audio watermarking using a chirp based technique," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '03)*, vol. 2, pp. 513–516, Baltimore, Md, USA, July 2003.

[5] A. Ramalingam and S. Krishnan, "A novel robust image watermarking using a chirp based technique," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE '04)*, vol. 4, pp. 1889–1892, Ontario, Canada, May 2004.

[6] S. Qian and D. Chen, "Joint time-frequency analysis," *IEEE Signal Processing Magazine*, vol. 16, no. 2, pp. 52–67, 1999.

[7] B. Boashash, *Time-Frequency Signal Analysis and Processing*, Prentice-Hall, Upper Saddle River, NJ, USA, 2003.

[8] B. Boashash, *Time-Frequency Signal Analysis. Methods and Applications*, Longman-Cheshire/Wiley, Melbourne, Australia/New York, NY, USA, 1992.

[9] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution—a tool for time-frequency signal analysis—part I: continuous-time signals," *Philips Journal of Research*, vol. 35, no. 3, pp. 217–250, 1980.

[10] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution-a tool for timefrequency signal analysis—part II: discrete time signals," *Philips Journal of Research*, vol. 35, no. 4-5, pp. 276–300, 1980.

[11] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution-a tool for timefrequency signal analysis—part Ill: relations with other time-frequency signal transformations," *Philips Journal of Research*, vol. 35, pp. 372–389, 1980.

[12] C. P. Janse and A. J. M. Kaizer, "Time-frequency distributions of loudspeakers: the application of the Wigner distribution," *Journal of the Audio Engineering Society*, vol. 31, no. 4, pp. 198–223, 1983.

[13] B. Bouachache, "Representation temps-frequence," Soc. Nat. ELF Aquitaine, Pau, France, Publ. Recherches, no. 373–378, 1978.

[14] P. Flandrin and B. Escudie, "Time and frequency representation of finite energy signals: a physical property as a result of an Hilbertian condition," *Signal Processing*, vol. 2, no. 2, pp. 93–100, 1980.

[15] L. Cohen, "Generalized phase-space distribution functions," *Journal of Mathematical Physics*, vol. 7, no. 5, pp. 781–786, 1966.

[16] H. Margenau and L. Cohen, "Probabilities in quantum mechanics," in *Quantum Theory and Reality*, M. Bunge, Ed., Springer, New York, NY, USA, 1967.

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

[18] R. A. Altes, "Detection, estimation, and classification with spectrograms," *The Journal of the Acoustical Society of America*, vol. 67, no. 4, pp. 1232–1246, 1980.

[19] A. Dziewonski, S. Bloch, and M. Landisman, "A technique for the analysis of transient signals," *Bulletin of the Seismological Society of America*, vol. 59, pp. 427–444, 1969.

[20] J. Flanagan, *Speech Analysis Synthesis and Perception*, Springer, New York, NY, USA, 1972.

[21] A. L. Levshin, V. F. Pisarenko, and G. A. Pogrebinsky, "On a frequency-time analysis of oscillations," *Annales Geophysicae*, vol. 28, pp. 211–218, 1972.

[22] A. V. Oppenheim, "Speech spectrograms using the fast Fourier transform," *IEEE Spectrum*, vol. 7, no. 8, pp. 57–62, 1970.

[23] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 55–59, 1980.

[24] D. Gabor, "Theory of communication," *Journal of the Institute of Electrical Engineers*, vol. 93, no. 11, pp. 429–457, 1946.

[25] L. J. Stanković and S. Stanković, "Analysis of instantaneous frequency representation using time-frequency distributions-generalized Wigner distribution," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 549–552, 1995.

[26] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Physical Review*, vol. 40, no. 5, pp. 749–759, 1932.

[27] J. Ville, "Theorie et applications de la notion de signal analytique," *Cables et Transmission*, vol. 2, no. 1, pp. 61–74, 1946.

[28] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.

[29] H. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 6, pp. 862–871, 1989.

[30] B. Boashash and B. Ristic, "Polynomial WVD's and time-varying polyspectra," in *Higher Order Statistical Processing*, B. Boashash, et al., Ed., Longman Chesihire, London, UK, 1993.

[31] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—part 1: fundamentals," *Proceedings of the IEEE*, vol. 80, pp. 519–538, 1992.

[32] B. Boashash and V. Sucic, "Resolution measure criteria for the objective assessment of the performance of quadratic time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1253–1263, 2003.

[33] L. J. Stanković, "Highly concentrated time-frequency distributions: pseudo quantum signal representation," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 543–551, 1997.

[34] L. Cohen and T. E. Posch, "Positive time-frequency distribution functions," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 31–38, 1985.

[35] H. Margenau and R. N. Hill, "Correlation between measurements in quantum theory," *Progress of Theoretical Physics*, vol. 26, pp. 772–738, 1961.

[36] P. Flandrin and W. Martin, "A general class of estimators for the Wigner-Ville spectrum of nonstationary processes," in *Systems Analysis and Optimization of Systems*, Lecture Notes in Control and Information Sciences, pp. 15–23, Springer, Berlin, Germany, 1984.

[37] R. D. Hippenstiel and P. M. de Oliveira, "Time-varying spectral estimation using the instantaneous power spectrum (IPS)," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1752–1759, 1990.

[38] M. Born and P. Jordan, "Zur Quantenmechanik," *Zeitschrift für Physik*, vol. 34, no. 1, pp. 858–888, 1925.

[39] J. Jeong and W. J. William, "Alias-free generalized discrete-time time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2757–2765, 1992.

[40] A. Papandreou and G. F. Boudreaux-Bartels, "Distributions for time-frequency analysis: a generalization of Choi-Williams and the Butterworth distributions," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, vol. 5, pp. 181–184, San Francisco, Calif, USA, March 1992.

[41] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, vol. 23, no. 1, pp. 85–102, 1984.

[42] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[43] O. Rioul and P. Flandrin, "Time-scale energy distributions: a general class extending wavelet transforms," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1746–1757, 1992.

[44] J. Bertrand and P. Bertrand, "Time-frequency representations of broadband signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '88)*, pp. 2196–2199, New York, NY, USA, April 1988.

[45] A. Papandreou-Suppappola, Ed., *Applications in Time-Frequency Signal Processing*, CRC Press LLC, Boca Raton, Fla, USA, 2003.

[46] D. L. Jones and T. W. Parks, "A resolution comparison of several time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 40, no. 2, pp. 413–420, 1992.

[47] A. Dziewonsi, S. Bloch, and M. Landisman, "A technique for the analysis of transient seismic signals," *Bulletin of the Seismological Society of America*, pp. 427–449, 1969.

[48] G. L. Duckworth, *Processing and inversion of Arctic Ocean refraction data*, Sc.D. dissertation, Massachusetts Institute of Technology, Campridge, Mass, USA, 1983.

[49] S. Grassin and R. Garello, "Spectral analysis of the swell using the reassigned Wigner-Ville representation," in *Proceedings of IEEE Oceans Conference*, vol. 3, pp. 1539–1544, Fort Lauderdale, Fla, USA, September 1996.

[50] A. Kayhan, A. El-Jaroudi, and L. Chaparro, "The evolutionary periodogram for non-stationary signals," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, 1994.

[51] A. Kayhan, A. El-Jaroudi, and L. F. Chaparro, "Data-adaptive evolutionary spectral estimation," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 204–213, 1995.

[52] A. Akan and L. F. Chaparro, "Evolutionary chirp representation of non-stationary signals via Gabor transform," *Signal Processing*, vol. 81, no. 11, pp. 2429–2436, 2001.

[53] A. Akan and L. F. Chaparro, "Evolutionary spectral analysis using a warped Gabor expansion," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96)*, vol. 3, pp. 1403–1406, Atlanta, Ga, USA, May 1996.

[54] M. B. Priestley, "Evolutionary spectra and nonstationary processes," *Journal of the Royal Statistical Society B*, vol. 27, no. 2, pp. 204–237, 1965.

[55] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press, London, UK, 1981.

[56] M. B. Priestley, *Nonlinear and Non-Stationary Time Series Analysis*, Academic Press, London, UK, 1988.

[57] Ci. Melard and A. Herteler de Schutter, "Contributions to evolutionary spectral theory," *Journal of Time Series Analysis*, vol. 10, no. 1, pp. 41–63, 1989.

[58] A. M. Yaglom, *An Introduction to the Theory of Stationary Random Functions*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1962.

[59] C. S. Detka, A. El-Jaroudi, and L. F. Chaparro, "Relating the bilinear distribution and the evolutionary spectrum," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '93)*, vol. 4, pp. 496–499, 1993.

[60] J. W. Pitton, "Positive time-frequency distributions via quadratic programming," *Multidimensional Systems and Signal Processing*, vol. 9, no. 4, pp. 439–445, 1998.

[61] P. J. Loughlin, J. W. Pitton, and L. E. Atlas, "Construction of positive time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2697–2705, 1994.

[62] P. Loughlin, J. Pitton, and B. Hannaford, "Approximating time-frequency density functions via optimal combinations of spectrograms," *IEEE Signal Processing Letters*, vol. 1, no. 12, pp. 199–202, 1994.

[63] J. W. Pitton, "An algorithm for weighted least squares positive time-frequency distributions," in *Advanced Signal Processing: Algorithms, Architectures, and Implementations VII*, vol. 3162 of *Proceedings of SPIE*, pp. 90–98, 1997.

[64] J. W. Pitton, "Linear and quadratic methods for positive time-frequency distributions," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, vol. 5, pp. 3649–3652, 1997.

[65] J. W. Pitton, L. E. Atlas, and P. J. Loughlin, "Applications of positive time-frequency distributions to speech processing," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 554–566, 1994.

[66] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[67] D. Thomson and A. Chave, "Jackknifed error estimates for spectra, coherences, and transfer functions," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed., vol. 1, pp. 58–113, Prentice-Hall, Englewood Cliffs, NJ, USA, 1991.

[68] A. Akan, "Signal-adaptive evolutionary spectral analysis using instantaneous frequency estimation," *Frequenz*, vol. 59, no. 7-8, pp. 201–205, 2005.

[69] L. F. Chaparro, R. Suleesathira, A. Akan, and B. Ünsal, "Instantaneous frequency estimation using discrete evolutionary transform for jammer excision," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 6, pp. 3525–3528, Salt Lake, Utah, USA, May 2001.

[70] R. Suleesathira, L. F. Chaparro, and A. Akan, "Discrete evolutionary transform for time-frequency signal analysis," *Journal of the Franklin Institute*, vol. 337, no. 4, pp. 347–364, 2000.

[71] R. Suleesathira and L. F. Chaparro, "Interference mitigation in spread spectrum using discrete evolutionary and Hough transforms," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*, vol. 5, pp. 2821–2824, Istanbul, Turkey, June 2000.

[72] S. Barbarossa, "Analysis of multicomponent LFM signals by a combined Wigner-Hough transform," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1511–1515, 1995.

[73] S. Barbarossa, A. Scaglione, S. Spalletta, and S. Votini, "Adaptive suppression of wideband interferences in spread-spectrum communications using the Wigner-Hough transform," in *Proceedings of the IEEE International Conference on*

Acoustics, Speech, and Signal Processing (ICASSP '97), vol. 5, pp. 3861–3864, Munich, Germany, April 1997.

[74] L. F. Chaparro and A. Alshehri, "Jammer excision in spread spectrum communications via wiener masking and frequency-frequency evolutionary transform," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 4, pp. 473–476, Hong Kong, April 2003.

[75] A. Akan and L. F. Chaparro, "Multi-window Gabor expansion for evolutionary spectral analysis," *Signal Processing*, vol. 63, no. 3, pp. 249–262, 1997.

[76] M. Jachan, G. Matz, and F. Hlawatsch, "Time-frequency ARMA models and parameter estimators for underspread nonstationary random processes," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4366–4381, 2007.

[77] M. Niedźwiecki, *Identification of Time-Varying Processes*, Wiley, New York, NY, USA, 2000.

[78] G. Matz and F. Hlawatsch, "Nonstationary spectral analysis based on time-frequency operator symbols and underspread approximations," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1067–1086, 2006.

[79] G. Matz and F. Hlawatsch, "Time-varying power spectra of nonstationary random processes," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed., chapter 9.4, pp. 400–409, Elsevier, Oxford, UK, 2003.

[80] M. Wax and T. Kailath, "Efficient inversion of Toeplitz-block Toeplitz matrix," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 5, pp. 1218–1221, 1983.

[81] Y. Grenier, "Parametric time-frequency representations," in *Traitement du Signal/Signal Processing, Les Houches, Session XLV*, J. L. Lacoume, T. S. Durrani, and R. Stora, Eds., pp. 338–397, Elsevier, Amsterdam, The Netherlands, 1987.

[82] M. Jachan, G. Matz, and F. Hlawatsch, "Time-frequency-autoregressive random processes: modeling and fast parameter estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 6, pp. 125–128, Hong Kong, April 2003.

[83] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 4, pp. 899–911, 1983.

[84] N. A. Abdrabbo and M. B. Priestley, "On the prediction of nonstationary processes," *Journal of the Royal Statistical Society B*, vol. 29, no. 3, pp. 570–585, 1967.

[85] M. Jachan, F. Hlawatsch, and G. Matz, "Linear methods for tfarma parameter estimation and system approximation," in *Proceedings of the 13th IEEE Workshop on Statistical Signal Processing*, pp. 909–914, Bordeaux, France, July 2005.

[86] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego, Calif, USA, 1999.

[87] F. Hlawatsch and P. Flandrin, "The interference structure of theWigner distribution and related time-frequency signal representations," in *The Wigner Distribution—Theory and Applications in Signal Processing*, W. Mecklenbräuker and F. Hlawatsch, Eds., pp. 59–133, Elsevier, Amsterdam, The Netherlands, 1997.

[88] M. Jachan, G. Matz, and F. Hlawatsch, "TFARMA models: order estimation and stabilization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 4, pp. 301–304, Philadelphia, Pa, USA, March 2005.

[89] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.

[90] S. M. Kay, *Modern Spectral Estimation*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.

[91] S. I. Shah, L. F. Chaparro, and A. El-Jaroudi, "Generalized transfer function estimation using evolutionary spectral deblurring," *IEEE Transactions on Signal Processing*, vol. 47, no. 8, pp. 2335–2339, 1999.

[92] J. Wexler and S. Raz, "Discrete Gabor expansions," *Signal Processing*, vol. 21, no. 3, pp. 207–220, 1990.

[93] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 4, pp. 899–911, 1983.

[94] T. Subba Rao, "The fitting of non-stationary time-series models with time-dependent parameters," *Journal of the Royal Statistical Society B*, vol. 32, no. 2, pp. 312–322, 1970.

[95] M. Kahn, L. F. Chaparro, and E. W. Kamen, "Frequency analysis of nonstationary signal models," in *Proceedings of the Conference on Information Sciences and Systems*, pp. 617–622, Baltimore, March 1989.

[96] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.

[97] S. I. Shah, *Generalized transfer function estimation and informative priors for positive time-frequency distributions*, Ph.D. dissertation, University of Pittsburgh, Pittsburgh, Pa, USA, 1997.

[98] J. Capon, "High-resolution frequency-wave number spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[99] S. I. Shah, P. J. Loughlin, L. F. Chaparro, and A. El-Jaroudi, "Informative priors for minimum cross-entropy positive time-frequency distributions," *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 176–177, 1997.

[100] R. B. Unsal Artan, A. Akan, and L. F. Chaparro, "Higher order evolutionary spectral analysis," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 6, pp. 633–636, Hong Kong, April 2003.

[101] A. J. E. M. Janssen, "On the locus and spread of pseudo-density functions in the time-frequency plane," *Philips Journal of Research*, vol. 37, no. 3, pp. 79–110, 1982.

[102] W. Rihaczek, *Principles of High-Resolution Radar*, McGraw-Hill, New York, NY, 1969.

[103] P. M. Woodward, *Probability and Information Theory with Application to Radar*, Pergamon, London, UK, 1953.

[104] G. W. Deley, "Waveform design," in *Radar Handbook*, M. I. Skolnik, Ed., McGraw-Hill, New York, NY, USA, 1970.

[105] M. I. Scolnik, *lntroduction to Radar Systems*, McCraw-Hill, New York, NY, USA, 1980.

[106] H. H. Szu and J. Blodgett, "Wigner distribution and ambiguity functions," in *Optics in Four Dimensions*, L. M. Narducci, Ed., pp. 355–381, American Institute of Physics, New York, NY, USA, 1981.

[107] G. Eichmann and N. M. Marinovic, "Scale-invariant Wigner distribution and ambiguity functions," in *Analog Optical Processing and Computing*, vol. 519 of *Proceedings of SPIE*, pp. 18–24, Cambridge, Mass, USA, 1984.

[108] L. J. Stanković, "A time-frequency distribution concentrated along the instantaneous frequency," *IEEE Signal Processing Letters*, vol. 3, no. 3, pp. 89–91, 1996.

[109] L. J. Stanković, "L-class of time-frequency distributions," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 22–25, 1996.

[110] L. J. Stanković, "On the realization of the highly concentrated time-frequency distributions," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (TFTSA '96)*, pp. 461–464, Paris, France, June 1996.

[111] B. Boashash and P. O'Shea, "Polynomial Wigner-Ville distributions and their relationship to time-varying higher order spectra," *IEEE Transactions on Signal Processing*, vol. 42, no. 1, pp. 216–220, 1994.

[112] L. J. Stankovic, "An analysis of some time-frequency and time-scale distributions," *Annales des Telecommunications*, vol. 49, no. 9-10, pp. 505–517, 1994.

[113] L. J. Stanković, "A method for improved distribution concentration in the time-frequency signal analysis using the L-Wigner distribution," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, 1995.

[114] L. J. Stanković, "A multitime definition of the Wigner higher order distribution: L-Wigner distribution," *IEEE Signal Processing Letters*, vol. 1, no. 7, pp. 106–109, 1994.

[115] L. J. Stanković, "An analysis of the Wigner higher order spectra of multicomponent signals," *Annals of Telecommunications*, vol. 49, no. 3-4, pp. 132–136, 1994.

[116] J. R. Fonolosa and C. L. Nikias, "Wigner higher order moment spectra: definitions, properties, computation and application to the transient signal analysis," *IEEE Transactions on Signal Processing*, vol. 41, no. 1, pp. 245–266, 1993.

[117] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2127–2135, 1990.

[118] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation. Optimal kernel design," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1589–1602, 1993.

[119] D. L. Jones and R. G. Baraniuk, "Adaptive optimal-kernel time-frequency representation," *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2361–2371, 1995.

[120] R. G. Baraniuk and D. L. Jones, "Signal-dependent time-frequency analysis using a radially Gaussian kernel," *Signal Processing*, vol. 32, no. 3, pp. 263–284, 1993.

[121] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995.

[122] P. Flandrin, F. Auger, and E. Chassande-Mottin, "Time-frequency reassignment: from principles to algorithms," in *Applications in Time-Frequency Signal Processing*, A. Papandreou-Suppappola, Ed., chapter 5, pp. 179–203, CRC Press, Boca Raton, Fla, USA, 2003.

[123] K. Kodera, C. de Villedary, and R. Gendrin, "A new method for the numerical analysis of non-stationary signals," *Physics of the Earth and Planetary Interiors*, vol. 12, no. 2-3, pp. 142–150, 1976.

[124] K. Kodera, R. Gendrin, and C. de Villedary, "Analysis of time-varying signals with small BT values," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 64–76, 1978.

[125] D. J. Nelson, "Cross-spectral methods for processing speech," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2575–2592, 2001.

[126] S. A. Fulop and K. Fitz, "A spectrogram for the twenty-first century," *Acoustics Today*, vol. 2, no. 3, pp. 26–33, 2006.

[127] S. A. Fulop and K. Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 360–371, 2006.

[128] http://en.wikipedia.org/wiki/Reassignment_method.

[129] G. Eichmann and B. Z. Dong, "Two-dimensional optical filtering of 1-D signal," *Applied Optics*, vol. 21, no. 17, pp. 3152–3156, 1982.

[130] M. G. Amin and W. J. Williams, "High spectral resolution time-frequency distribution kernels," *IEEE Transactions on Signal Processing*, vol. 46, no. 10, pp. 2796–2804, 1998.

[131] B. Boashash, B. Lovell, and H. Whitehouse, "High resolution time frequency signal analysis by parametric modeling of the Wigner-Ville distribution," in *Proceedings of the International Symposium on Signal Processing and Its Applications (ISSPA '87)*, Brisbane, Australia, August 1987.

[132] P. A. Ramamoorthy, V. K. Iyer, and Y. Ploysongsang, "Autoregressive modeling of the Wigner spectrum," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '87)*, pp. 1509–1512, Dallas, Tex, USA, April 1987.

[133] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.

[134] E. F. Velez and R. G. Absher, "Smoothed Wigner-Ville parametric modeling for the analysis of nonstationary signals," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 507–510, May 1989.

[135] E. F. Velez and R. G. Absher, "Parametric modeling of the Wigner half-kernel and its application to spectral estimation," *Signal Processing*, vol. 26, no. 2, pp. 161–175, 1992.

[136] R. Kumaresan, "On the zeros of the linear prediction error filter for deterministic signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, pp. 217–220, 1983.

[137] S. L. Marple Jr., *Digital Spectral Analysis with Applications*, chapter 11, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.

[138] B. Barkat and B. Boashash, "A high-resolution quadratic time-frequency distribution for multicomponent signals analysis," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2232–2239, 2001.

[139] R. G. Baraniuk, *Shear madness: signal-dependent and metaplectic time-frequency representations*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, August 1992.

[140] R. M. Fano, "Short-time autocorrelation functions and power spectra," *The Journal of the Acoustical Society of America*, vol. 22, pp. 546–550, 1950.

[141] M. R. Schroeder and B. S. Atal, "Generalized short-time power spectra and autocorrelation functions," *The Journal of the Acoustical Society of America*, vol. 34, pp. 1679–1683, 1962.

[142] M. H. Ackroyd, "Instantaneous and time-varying spectra: an introduction," *Radio and Electronic Engineer*, vol. 39, no. 3, pp. 145–152, 1970.

[143] M. H. Ackroyd, "Short-time spectra and time-frequency energy distributions," *The Journal of the Acoustical Society of America*, vol. 50, pp. 1229–1231, 1970.

[144] D. G. Lampard, "Generalization of the wiener-khintchine theorem to nonstationary processes," *Journal of Applied Physics*, vol. 25, no. 6, pp. 802–803, 1954.

[145] W. D. Mark, "Spectral analysis of the convolution and filtering of non-stationary stochastic processes," *Journal of Sound and Vibration*, vol. 11, no. 1, pp. 19–63, 1970.

[146] Y. Zhao, L. E. Atlas, and R. J. Marks, "Use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1084–1091, 1990.

[147] Y. Zhang, M. G. Amin, and G. J. Frazer, "High resolution time-frequency distributions for maneuvering target detection in over-the-horizon radars," *IEE Proceedings: Radar, Sonar and Navigation*, vol. 150, no. 4, pp. 299–304, 2003.

[148] A. Papandreou-Suppappola, "Generalized time-shift covariant quadratic time-frequency representations with arbitrary group delays," in *Proceedings of the 29th Asilomar Conference Signals, Systems and Computers*, pp. 553–557, Pacific Grove, Calif, USA, October 1995.

[149] A. Papandreou-Suppappola, F. Hlawatsch, and G. F. Boudreaux-Bartels, "Quadratic time-frequency representations with scale covariance and generalized time-shift covariance: a unified framework for the affine, hyperbolic, and power classes," *Digital Signal Processing*, vol. 8, no. 1, pp. 3–48, 1998.

[150] A. Papandreou and G. F. Boudreaux-Bartels, "The exponential class and generalized timeshift covariant quadratic time-frequency representations," in *Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (TFTSA '96)*, pp. 429–432, Paris, France, June 1996.

[151] A. Papandreou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "A unified framework for the scale covariant affine, hyperbolic, and power class time-frequenc representations using generalized time-shifts," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '95)*, Detroit, Mich, USA, May 1995.

[152] A. Papandreou-Suppappola, *New classes of quadratic time-frequency representations with scale covariance and generalized time-shift covariance: analysis, detection, and estimation*, Ph.D. thesis, University of Rhode Island, Kingston, RI, USA, 1995.

[153] A. Papandreou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "Hyperbolic class of quadratic time-frequency representations—part I: constant-Q warping, the hyperbolic paradigm, properties, and members," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3425–3444, 1993.

[154] F. Hlawatsch, A. Papandreou, and G. F. Boudreaux-Bartels, "The power classes of quadratic time-frequency representations: a generalization of the affine and hyperbolic classes," in *Proceedings of the 27th Asilomer Conference of Signals, Systems and Computers*, vol. 2, pp. 1265–1270, Pacific Grove, Calif, USA, November 1993.

[155] A. Papandreou-Suppappola, "Time-frequency representations covariant to frequency-dependant time shifts," in *Time-Frequency Signal Analysis and Processing*, B. Boashash, Ed., Prentice-Hall, New York, NY, USA, 2002.

[156] A. Papandreou-Suppappola, R. L. Murray, B.-G. Iem, and G. F. Boudreaux-Bartels, "Group delay shift covariant quadratic time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2549–2564, 2001.

[157] V. Katkovnic and L. Stankovic, "Instantaneous frequency esimation using the Wigner distribution with varying and data-driven window length," *IEEE Transactions on Signal Processing*, vol. 46, pp. 2315–2325, 1998.

[158] Z. M. Hussain and B. Boashash, "Adaptive instantaneous frequency esitmation of multi-component FM signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*, vol. 5, pp. 657–660, Istanbul, Turkey, June 2000.

[159] V. Katkovnik, "Nonparametric estimation of instantaneous frequency," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 183–189, 1997.

[160] F. Hlawatsch, A. Papandreou, and G. F. Boudreaux-Bartels, "The power classes-quadratic time-frequency representations with scale covariance and dispersive time-shift covariance," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 3067–3083, 1999.

[161] A. H. Costa and G. F. Boudreaux-Bartels, "Design of time-frequency representations using a multiform, tiltable exponential kernel," *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2283–2301, 1995.

[162] A. Papandreou-Suppappola, F. Hlawatsch, and G. F. Boudreaux-Bartels, "Power class time-frequency representations: interference geometry, smoothing, and implementation," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (TFTSA '96)*, pp. 193–196, Paris, France, June 1996.

[163] A. Papandreou and G. F. Boudreaux-Bartels, "Distortion that occurs when the signal group delay does not match the Time-Shift Covariance of a Time-Frequency Representation," in *Proceedings of the 30th Annual Conference on Information Sciences and Systems*, pp. 520–525, Princeton, NJ, USA, March 1996.

[164] A. Papandreou-Suppappola, B. G. Iem, and G. F. Boudreaux-Bartels, "Time-frequency symbols for statistical signal processing," in *Time-Frequency Signal Analysis and Processing*, B. Boashash, Ed., Prentice-Hall, New York, NY, USA, 2002.

[165] L. Stankovic, "Time-frequency distributions with complex argument," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 475–486, 2002.

[166] A. Papandreou-Suppappola and G. F. Boudreaux-Bartels, "Effect of mismatching analysis signals and time-frequency representations," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (TFTSA '96)*, pp. 149–152, Paris, France, June 1996.

[167] P. Guillemain and P. White, "Wavelet transform for the analysis of dispersive systems," in *Proceedings of IEEE UK Symposium on Applications of Time-Frequency and Time-Scale Methods*, pp. 32–39, University of Warwick, Coventry, UK, August 1995.

[168] D. E. Newland, "Time-frequency and time-scale analysis by harmonic wavelets," in *Signal Analysis and Prediction*, A. Prochazka, Ed., chapter 1, Birkhauser, Boston, Mass, USA, 1998.

[169] M. J. Freeman, M. E. Dunham, and S. Qian, "Transionospheric signal detection by time-scale representation," in *Proceedings of IEEE UK Symposium on Applications of Time-Frequency and Time-Scale Methods*, pp. 152–158, University of Warwick, Coventry, UK, August 1995.

[170] J. P. Sessarego, J. Sageloli, P. Flandrin, and M. Zakharia, "Time-frequency Wigner-Ville analysis of echoes scattered by a spherical shell," in *Wavelets, Time-Frequency Methods and Phase Space*, J. M. Combes, A. Grossman, and P. T. Chamitchian, Eds., pp. 147–153, Springer, Heidelberg, Germany, 1989.

[171] A. Papandreou and L. T. Antonelli, "Use of quadratic time-frequency representations to analyze Cetacean mammal sounds," Tech. Rep. 11, Naval Undersea Warfare Centre, Newport, RI, USA, December 2001.

[172] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill, New York, NY, USA, 1953.

[173] V. Szekely, "Distributed RC networks," in *The Circuits and Filters Handbook*, W. K. Chen, Ed., pp. 1203–1221, CRC Press/IEEE Press, Boca Raton, Fla, USA, 1995.

[174] C. Cornu, S. Stankovic, C. Ioana, A. Quinquis, and L. Stankovic, "Generalized representation of phase derivatives

for regular signals," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4831–4838, 2007.

[175] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, NY, USA, 1987.

[176] I. Djurović and L. J. Stanković, "Influence of high noise on the instantaneous frequency estimation using quadratic time-frequency distributions," *IEEE Signal Processing Letters*, vol. 7, no. 11, pp. 317–319, 2000.

[177] P. O'Shea, "A new technique for instantaneous frequency rate estimation," *IEEE Signal Processing Letters*, vol. 9, no. 8, pp. 251–252, 2002.

[178] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif, "Evolutionary time-frequency distributions using Bayesian regularised neural network model," *IET Signal Processing*, vol. 1, no. 2, pp. 97–106, 2007.

[179] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif, "Computing deblurred time-frequency distributions using artificial neural networks," *Circuits, Systems, and Signal Processing*, vol. 27, no. 3, pp. 277–294, 2008.

[180] M. T. Hagan, H. B. Demuth, and M. Beale, *Neural Network Design*, Thomson Learning, Boston, Mass, USA, 1996.

[181] S.-C. Pei and J.-J. Ding, "Relations between Gabor transforms and fractional fourier transforms and their applications for signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4839–4850, 2007.

[182] http://en.wikipedia.org/wiki/Data_clustering.

[183] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif, "Impact of varying neurons and hidden layers in neural network architecture for a time frequency application," in *Proceedings of the 10th IEEE International Multitopic Conference (INMIC '06)*, pp. 188–193, Islamabad, Pakistan, December 2006.

[184] J. Ahmad, I. Shafi, S. I. Shah, and F. M. Kashif, "Analysis and comparison of neural network training algorithms for the joint time-frequency analysis," in *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA '06)*, pp. 193–198, Austria, February 2006.

[185] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[186] I. Daubechies, "Time-frequency localization operators: a geometric phase space approach," *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 605–612, 1988.

[187] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[188] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713–718, 1992.

[189] S. Qian and D. Chen, "Signal representation using adaptive normalized Gaussian functions," *Signal Processing*, vol. 36, no. 1, pp. 1–11, 1994.

[190] L. F. Villemoes, "Best approximation with Walsh atoms," *Constructive Approximation*, vol. 13, no. 3, pp. 329–355, 1997.

[191] P. E. Gill, W. Murray, and M. H. Wright, *Numerical Linear Algebra and Optimization*, Addison-Wesley, Redwood City, Calif, USA, 1991.

[192] N. E. Huang, Z. Shen, S. R. Long, et al., "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, no. 1971, pp. 903–995, 1998.

[193] A.-O. Boudraa, J.-C. Cexus, F. Salzenstein, and L. Guillon, "If estimation using empirical mode decomposition and nonlinear Teager energy operator," in *Proceedings of the 1st International Symposium on Control, Communications, and Signal Processing (ISCCSP '04)*, pp. 45–48, Hammamet, Tunisia, March 2004.

[194] J. C. Cexus and A. O. Boudraa, "Nonstationary signals analysis by Teager-Huang transform (THT)," in *Proceedings of the 13th European Signal Processing Conference (EUSIPCO '06)*, Florence, Italy, 2006.

[195] A.-O. Boudraa and J.-C. Cexus, "EMD-based signal filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 6, pp. 2196–2202, 2007.

[196] J. C. Cexus and A. O. Boudraa, "Teager-Huang analysis applied to sonar target recognition," *International Journal of Signal Processing*, vol. 1, no. 1, pp. 23–27, 2004.

[197] A. O. Boudraa, J. C. Cexus, and Z. Saidi, "EMD-based signal noise reduction," *International Journal of Signal Processing*, vol. 1, no. 1, pp. 33–37, 2004.

[198] Z. Liu and S. Peng, "Boundary processing of bidimensional EMD using texture synthesis," *IEEE Signal Processing Letters*, vol. 12, no. 1, pp. 33–36, 2005.

[199] A. O. Boudraa, J. C. Cexus, F. Salzenstein, and A. Beghdadi, "EMD-based multibeam echosounder images segmentation," in *Proceedings of the 2nd International Symposium on Control, Communications, and Signal Processing (ISCCSP '06)*, Marrakech, Morocco, March 2006.

[200] K. Zeng and M.-X. He, "A simple boundary process technique for empirical mode decomposition," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS '04)*, vol. 6, pp. 4258–4261, 2004.

[201] P. Flandrin, P. Gonçalves, and G. Rilling, "Detrending and denoising with empirical mode decompositions," in *Proceedings of 7th European Signal Processing Conference (EUSIPCO '04)*, pp. 1581–1584, Vienna, Austria, 2004.

[202] P. Flandrin and P. Gonçalves, "Empirical mode decompositions as a data-driven wavelet-like expansions," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 4, pp. 477–496, 2004.

[203] G. Rilling, P. Flandrin, and P. Gonçalves, "Empirical mode decomposition, fractional Gaussian noise and hurst exponent estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 4, pp. 489–492, Philadelphia, Pa, USA, 2005.

[204] R. Deering and J. F. Kaiser, "The use of a masking signal to improve empirical mode decomposition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 4, pp. 485–488, Philadelphia, Pa, USA, 2005.

[205] S. Benramdane, J. C. Cexus, A. O. Boudraa, and J. A. Astolfi, "Transient turbulent pressure signal processing using empirical mode decomposition," in *Proceedings of the Physical Signal and Image Processing Conference*, Mulhouse, France, 2007.

[206] P. Flandrin, G. Rilling, and P. Gonçalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.

[207] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proceedings of the Royal Society A*, vol. 460, no. 2046, pp. 1597–1611, 2004.

[208] G. Rilling and P. Flandrin, "One or two frequencies? The empirical mode decomposition answers," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 85–95, 2008.

[209] K. T. Coughlin and K. K. Tung, "11-year solar cycle in the stratosphere extracted by the empirical mode decomposition method," *Advances in Space Research*, vol. 34, no. 2, pp. 323–329, 2004.

[210] M. Chavez, C. Adam, V. Navarro, S. Boccaletti, and J. Martinerie, "On the intrinsic time scales involved in synchronization: a data-driven approach," *Chaos*, vol. 15, no. 2, Article ID 023904, 2005.

[211] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, "Underdetermined blind audio source separation using modal decomposition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, 15 pages, 2007.

[212] S. Krishnan, "A new approach for estimation of instantaneous mean frequency of a time-varying signal," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 17, pp. 2848–2855, 2005.

[213] V. Sucic and B. Boashash, "Optimisation algorithm for selecting quadratic time-frequency distributions: performance results and calibration," in *Proceedings of the 6th International Symposium on Signal Processing and Its Applications (ISSPA '01)*, vol. 1, pp. 331–334, Kuala Lumpur, Malaysia, August 2001.

[214] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Transactions on Information Theory*, vol. 26, no. 1, pp. 26–37, 1980.

[215] J. E. Shore and R. W. Johnson, "Properties of cross-entropy minimization," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 472–482, 1981.