Time–Frequency Distributions—A Review

LEON COHEN

Invited Paper

A review and tutorial of the fundamental ideas and methods of joint time-frequency distributions is presented. The objective of the field is to describe how the spectral content of a signal is changing in time, and to develop the physical and mathematical ideas needed to understand what a time-varying spectrum is. The basic goal is to devise a distribution that represents the energy or intensity of a signal simultaneously in time and frequency. Although the basic notions have been developing steadily over the last 40 years, there have recently been significant advances. This review is presented to be understandable to the nonspecialist with emphasis on the diversity of concepts and motivations that have gone into the formation of the field.

I. Introduction

The power of standard Fourier analysis is that it allows the decomposition of a signal into individual frequency components and establishes the relative intensity of each component. The energy spectrum does not, however, tell us *when* those frequencies occurred. During a dramatic sunset, for example, it is clear that the composition of the light reaching us is very different than what it is during most of the day. If we Fourier analyze the light from sunrise to sunset, the energy density spectrum would not tell us that the spectral composition was significantly different in the last 5 minutes. In this situation, where the changes are relatively slow, we may Fourier analyze 5-minute samples of the signal and get a pretty good idea of how the spectrum during sunset differed from a 5-minute strip during noon. This may be refined by sliding the 5-minute intervals along time, that is, by taking the spectrum with a 5-minute time window at each instant in time and getting an energy spectrum as a continuous function of time. As long as the 5-minute intervals themselves do not contain rapid changes, this will give an excellent idea of how the spectral composition of the light has changed during the course of the day. If significant changes occurred considerably faster than over 5 minutes, we may shorten the time window appropriately. This is the basic idea of the short-time Fourier transform, or spectrogram, which is currently the standard method for the study of time-varying signals. However, there exist natural and man-made signals whose spectral content is changing so rapidly that finding an appropriate short-time window is problematic since there may not be any time interval for which the signal is more or less stationary. Also, decreasing the time window so that one may locate events in time reduces the frequency resolution. Hence there is an inherent tradeoff between time and frequency resolution. Perhaps the prime example of signals whose frequency content is changing rapidly and in a complex manner is human speech. Indeed it was the motivation to analyze speech that led to the invention of the sound spectrogram [113], [160] during the 1940s and which, along with subsequent developments, became a standard and powerful tool for the analysis of nonstationary signals [5], [6], [68], [75], [111], [112], [126], [150], [151], [158], [163], [164], [174]. Its possible shortcomings not withstanding, the short-time Fourier transform and its variations remain the prime methods for the analysis of signals whose spectral content is varying.

Starting with the classical works of Gabor [80], Ville [194], and Page [152], there has been an alternative development for the study of time-varying spectra. Although it is now fashionable to say that the motivation for this approach is to improve upon the spectrogram, it is historically clear that the main motivation was for a fundamental analysis and a clarification of the physical and mathematical ideas needed to understand what a time-varying spectrum is. The basic idea is to devise a joint function of time and frequency, a distribution, that will describe the energy density or intensity of a signal simultaneously in time and frequency. In the ideal case such a joint distribution would be used and manipulated in the same manner as any density function of more than one variable. For example, if we had a joint density for the height and weight of humans, we could obtain the distribution of height by integrating out weight. We could obtain the fraction of people weighing more than 150 lb but less than 160 lb with heights between 5 and 6 ft. Similarly, we could obtain the distribution of weight at a particular height, the correlation between height and weight, and so on. The motivation for devising a joint time–frequency distribution is to be able to use it and manipulate it in the same way. If we had such a distribution, we could ask what fraction of the energy is in a certain frequency and time range, we could calculate the distribution of fre-

Manuscript received March 21, 1988; revised March 30, 1989. This work was supported in part by the City University Research Award Program.

The author is with the Department of Physics and Astronomy, Hunter College and Graduate Center, City University of New York, New York, NY 10021, USA.

IEEE Log Number 8928443.

quency at a particular time, we could calculate the global and local moments of the distribution such as the mean frequency and its local spread, and so on. In addition, if we did have a method of relating a joint time-frequency distribution to a signal, it would be a powerful tool for the construction of signals with desirable properties. This would be done by first constructing a joint time-frequency function with the desired attributes and then obtaining the signal that produces that distribution. That is, we could synthesize signals having desirable time-frequency characteristics. Of course, time-frequency analysis has unique features, such as the uncertainty principle, which add to the richness and challenge of the field.

From standard Fourier analysis, recall that the instantaneous energy of a signal $s(t)$ is the absolute value of the signal squared,

$$|s(t)|^2 = \text{intensity per unit time at time } t$$

or

$$|s(t)|^2 \Delta t = \text{fractional energy in time interval } \Delta t \text{ at time } t. \tag{1.1}$$

The intensity per unit frequency,[1] the energy density spectrum, is the absolute value of the Fourier transform squared,

$$|S(\omega)|^2 = \text{intensity per unit frequency at } \omega$$

or

$$|S(\omega)|^2 \Delta\omega = \text{fractional energy in frequency interval } \Delta\omega \text{ at frequency } \omega \tag{1.2}$$

where

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int s(t) e^{-j\omega t}\, dt. \tag{1.3}$$

We have chosen the normalization such that

$$\int |s(t)|^2\, dt = \int |S(\omega)|^2\, d\omega = \text{total energy} = 1 \tag{1.4}$$

where, for convenience, we will always take the total energy to be equal to 1.[2] The fundamental goal is to devise a joint function of time and frequency which represents the energy or intensity per unit time per unit frequency. For a joint distribution $P(t, \omega)$ we have

$$P(t, \omega) = \text{intensity at time } t \text{ and frequency } \omega$$

or

$$P(t, \omega)\Delta t\, \Delta\omega = \text{fractional energy in time-frequency cell } \Delta t\, \Delta\omega \text{ at } t, \omega.$$

Ideally the summing up of the energy distribution for all frequencies at a particular time would give the instantaneous energy, and the summing up over all times at a particular frequency would give the energy density spectrum,

$$\int P(t, \omega)\, d\omega = |s(t)|^2 \tag{1.5}$$

$$\int P(t, \omega)\, dt = |S(\omega)|^2. \tag{1.6}$$

[1]We use angular frequency. All integrals go from $-\infty$ to $+\infty$ unless otherwise stated.

[2]Signals that cannot be normalized may be handled as limiting cases of normalized ones or by using generalized functions.

The total energy E, expressed in terms of the distribution, is given by

$$E = \int P(t, \omega)\, d\omega\, dt \tag{1.7}$$

and will be equal to the total energy of the signal if the marginals are satisfied. However, we note that it is possible for a distribution to give the correct value for the total energy without satisfying the marginals.

Do there exist joint time-frequency distributions that would satisfy our intuitive ideas of a time-varying spectrum? Can their interpretation be as true densities or distributions? How can such functions be constructed? Do they really represent the correlations between time and frequency? What reasonable conditions can be imposed to obtain such functions? The hope is that they do exist, but if they do not in the full sense of true densities, what is the best we can do? Is there one distribution that is the best, or are different distributions to be used in different situations? Are there inherent limitations to a joint time-frequency distribution? This is the scope of time-frequency distribution theory.

Scope of Review, Notation, and Terminology: The basic ideas and methods that have been developed are readily understood by the uninitiated and do not require any specialized mathematics. We shall stress the fundamental ideas, motivations, and unresolved issues. Hopefully our emphasis on the fundamental thinking that has gone into the development of the field will also be of interest to the expert.

We confine our discussion to distributions in the spirit of those proposed by Wigner, Ville, Page, Rihaczek, and others and consider only deterministic signals. There are other qualitatively different approaches for joint time-frequency analysis which are very powerful but will not be discussed here. Of particular note is Priestley's theory of evolutionary spectra [162] and we point out that his discussions of the basic concepts relating to time-varying spectra are among the most profound. Also, we will not consider the Gabor logon approach, although it is related to the spectrogram discussed in Section VI.

As usual, when considering many special cases and situations, one may quickly become embroiled in a morass of subscripts and superscripts. We have chosen to keep the notation simple and, if no confusion arises, we differentiate between cases and situations by context rather than by notation.

Some of the terminology may be unfamiliar or puzzling to some readers. Many words, like distribution in the probability sense, are used because of historical reasons. These distributions first arose in quantum mechanics where the words "probability density" or "distribution" are applied properly. For deterministic signals where no probabilistic considerations enter, the reader should think of distributions as "intensities" or "densities" in the common usage of the words, or simply as how the energy is "distributed" in the time-frequency cells. Of course many probability concepts apply to intensities, such as averages and spreads. When we deal with stochastic signals, probability concepts do properly enter. As we will see, many of the known distributions may become negative or even complex. Hence they are sometimes called quasi or pseudo distributions. Also from probability theory, the word "marginal" is used to indicate the individual distribution. The marginals are derived from the joint distribution by integrating out the other variables. Hence we will say that $|s(t)|^2$ and $S(\omega)|^2$ are the marginals of $P(t, \omega)$, as per Eqs. (1.5) and (1.6).

II. Brief Historical Perspective and Examples

Although we will discuss the particular distributions in detail, it is of value to give a short historical perspective here. The two original papers that addressed the question of a joint distribution function in the sense considered here are those of Gabor [80] and Ville [194]. They were guided by a similar development in quantum mechanics, where there is a partial mathematical resemblance to time-frequency analysis. We discuss this resemblance later, but we emphasize here that the physical interpretations are drastically different and the analogy is only formal. Gabor developed a mathematical method closely connected to so-called coherent states in quantum mechanics. In the same paper Gabor introduced the important concept of the analytic signal. Ville derived a distribution that Wigner [199] gave in 1932 to study quantum statistical mechanics. At the same time as Ville, Moyal [143] used an identical derivation in the quantum mechanical context. Although we use the word "derivation," we emphasize that there is an ambiguity in the method of Ville and Moyal, and many later authors used the same derivation to obtain other distributions. The Wigner-Ville distribution is

$$W(t, \omega) = \frac{1}{2\pi} \int s^*\left(t - \frac{1}{2}\tau\right) e^{-j\tau\omega} s\left(t + \frac{1}{2}\tau\right) d\tau. \qquad (2.1)$$

It satisfies the marginals, but we do not show that now. We shall see later that by simple inspection the properties of a distribution can readily be determined. A year after Wigner, Kirkwood [107] came up with another distribution and argued that it is simpler to use than the Wigner distribution for certain problems. The distribution is simply

$$e(t, \omega) = \frac{1}{\sqrt{2\pi}} s(t)\, S^*(\omega) e^{-j\omega t}. \qquad (2.2)$$

This distribution and its variations have been derived and studied in many ways and independently introduced in signal analysis. A particularly innovative derivation based on physical considerations was given by Rihaczek [167]. Levin [125] derived it by modifying the considerations that led to the Page [152] distribution. Margenau and Hill [133] derived it by the Ville and Moyal methods. Equation (2.2) is complex and is sometimes called the complex energy spectrum. Its real part is also a distribution and satisfies the marginals.

In 1952 Page [152] developed the concept of the running spectrum. He obtained a new distribution from simple conceptual considerations and coined the phrase "instantaneous power spectra." The Page distribution is

$$P^-(t, \omega) = \frac{\partial}{\partial t} \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} s(t') e^{-j\omega t'}\, dt' \right|^2. \qquad (2.3)$$

It was pointed out by Turner [190] and Levin [125] that the Page procedure can be used to obtain other distributions.

A comprehensive and far-reaching study was done by Mark [138] in 1970, where many ideas commonly used today were developed. He pinpointed the difficulty of the spurious values in the Wigner distribution, introduced the "physical" spectrum, which is basically the spectrogram, and showed its relation to the Wigner distribution. Fundamental considerations regarding time-frequency distributions and nonstationary processes were given by Blanc-Lapierre and Picinbono [24], Loynes [128], and Lacoume and Kofman [121].

One of the main stumbling blocks in developing a consistent theory is the fact that the behavior of these few distributions is dramatically different and each has peculiar properties. However, each does satisfy the marginals, has other desirable properties, and presumably is a good candidate for the time-varying spectrum. Furthermore each has been derived from seemingly plausible ideas. It was unclear how many more existed and whether the peculiarities were general features or individual ones. It was subsequently realized [58] that an infinite number can be readily generated from

$$P(t, \omega) = \frac{1}{4\pi^2} \iiint e^{-j\theta t - j\tau\omega + j\theta u} \phi(\theta, \tau) \cdot s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) du\, d\tau\, d\theta \qquad (2.4)$$

where $\phi(\theta, \tau)$ is an arbitrary function called the kernel[3] by Claasen and Mecklenbrauker [56]. By choosing different kernels, different distributions are obtained at will. For example the Wigner, Rihaczek, and Page distributions are obtained by taking $\phi(\theta, \tau) = 1$, $e^{j\theta\tau/2}$, and $e^{j\theta|\tau|/2}$, respectively. Having a simple method to generate all distributions has the advantage of allowing one to prove general results and to study what aspects of a particular distribution are unique or common to all. Equally important is the idea that by placing constraints on the kernel one obtains a subset of the distributions which have a particular property [58]. That is, the properties of the distribution are determined by the corresponding kernel.

There has been a great surge of activity in the past 10 years or so and the initial impetus came from the work of Claasen and Mecklenbrauker [54]-[56], Janse and Kaizer [97], Boashash (aka Bouachache) [35], and others. The importance of their initial contributions is that they developed ideas uniquely suited to the time-frequency situation and demonstrated useful methods for implementation. Moreover, they were innovative in using the similarities and differences with quantum mechanics. In an important set of papers, Claasen and Mecklenbrauker [54]-[56] developed a comprehensive approach and originated many new ideas and procedures for the study of joint distributions. Boashash [35] was perhaps the first to utilize them for real problems and developed a number of new methods. In particular he realized that even though a distribution may not behave properly in all respects or interpretations, it could still be used if a particular property such as instantaneous frequency is well described. He initially applied them to problems in geophysical exploration. Escudie [71], [77] and coworkers transcribed directly some of the early quantum

[3]In general the kernel may depend explicitly on time and frequency and in addition may also be a functional of the signal. To avoid notational complexity we will use $\phi(\theta, \tau)$, and the possible dependence on other variables will be clear form the context. As we will see in Section IV, the time- and frequency-shift invariant distributions are those for which the kernel is independent of time and frequency.

mechanical results, particularly the work on the general class of distributions [58], [132], into signal analysis language. The work by Janse and Kaizer [97] was remarkable in its scope and introduction of new methodologies. They developed innovative theoretical and practical techniques for the use of these distributions.

Many divergent attitudes toward the meaning, interpretation, and use of these distributions have arisen over the years, ranging from the attempt to describe a time-varying spectrum to merely using them as carrying the information of a signal in a convenient way. The divergent viewpoints and interests have led to a better understanding and implementation. We will discuss some of the common attitudes and unresolved issues in the conclusion. The subject is evolving rapidly and most of the issues are open.

Preliminary Illustrative Examples: Before proceeding we present a simple example of the above distributions so that the reader may get a better feeling for the variety and difficulties. We consider the signal illustrated in Fig. 1(a). Initially the sine wave has a frequency ω_1 in the interval $(0, t_1)$, then it is shut off in the interval (t_1, t_2) and turned on again in the interval (t_2, t_3) with a frequency ω_2. This simple signal is an idealization of common situations that we hope these distributions will handle effectively. The signal is highly nonstationary, has intermediate periods of silence common in acoustic signals, and has sudden onsets. Everyone has a sense of what the distribution should be. We expect the distribution to show a peak at ω_1 in the interval $(0, t_1)$ and another peak at ω_2 for the interval (t_2, t_3), and of course to be zero in the interval (t_1, t_2). Fig. 1 illustrates the distributions mentioned thus far and we see that they all imply intensities, that is, nonzero values, at places that are not expected. The Wigner distribution is not zero in the range (t_1, t_2) although the signal is. This is a fundamental property which we discuss later. The Rihaczek distribution has nonzero values at ω_2 at time (t_1, t_2), although we would expect zero intensity at that frequency for those times. Similar statements hold for the interval (t_2, t_3) at frequency ω_1. The distribution is such that all values of the spectrum are reflected at each time. The Page distribution is similar to that of Rihaczek, but it reflects only those frequencies that have already occurred. We also note that while the Wigner distribution peaks in the middle of each portion of the signal, the Rihaczek distribution is flat and the Page distribution gradually increases as more of the signal at that frequency comes through in time.

We emphasize that all three distributions satisfy the instantaneous energy and spectral energy exactly. Although very different in appearance, they are equivalent in the sense that each one can be obtained from the other uniquely and contains the same amount of information. They are very different in their energy concentration, but nonetheless all three have been used with considerable profit. We note that these are just three possibilities out of an infinite number of choices, all with vastly different behavior.

III. The Distributions and Methods for Obtaining Them

One of the remarkable facts regarding time-frequency distributions is that so many plausible derivations and approaches have been suggested, yet the behavior of each distribution is dramatically different. It is therefore important to understand the ideas and arguments that have been given, as variations and insights of them will undoubtedly lead the way to further development. We will not present these approaches in historical order, but rather in a sequence that logically develops the ideas and techniques. However, the different sections may be read independently. With the benefit of hindsight we have streamlined some of the original arguments.

A. Page Distribution and Variations

Page [152] argues that as the signal is evolving, our knowledge of it consists of the signal up to the current time t and we have no information about the future part. Conceptually we may consider a new signal $s_t(t')$,

$$s_t(t') = \begin{cases} s(t'), & t' \le t \\ 0, & t' > t \end{cases} \tag{3.1}$$

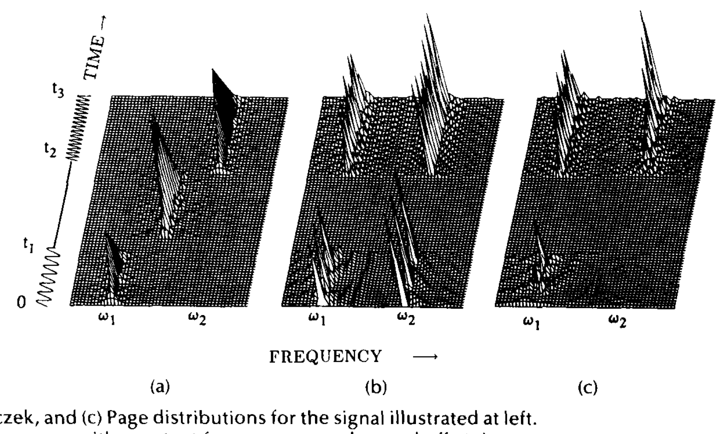


Fig. 1. (a) Wigner, (b) Rihaczek, and (c) Page distributions for the signal illustrated at left. The signal is turned on at time zero with constant frequency ω_1 and turned off at time t_1, turned on again at time t_2 with frequency ω_2 and turned off at time t_3. All three distributions display energy density where one does not expect any. The positive parts of the distributions are plotted. For the Rihaczek distribution we have plotted the real part, which is also a distribution.

where t' is the running time and t is the present instant. The Fourier transform of $s_t(t')$ is

$$S_t^-(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s_t(t') e^{-j\omega t'}\, dt'$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} s(t') e^{-j\omega t'}\, dt' \qquad (3.2)$$

which is called the running spectrum. The (−) notation is to signify that we have observed the signal from $-\infty$. In analogy with Eq. (1.6) we expect the energy observed per unit frequency up to time t to be

$$\int_{-\infty}^{t} P^-(t', \omega)\, dt' = |S_t^-(\omega)|^2. \qquad (3.3)$$

This equation can be used to determine $P^-(t, \omega)$, since differentiation with respect to t yields

$$P^-(t, \omega) = \frac{\partial}{\partial t} |S_t^-(\omega)|^2 \qquad (3.4)$$

which is the Page distribution. It can be obtained from the general class of distributions, Eq. (2.4), by taking

$$\phi(\theta, \tau) = e^{j\theta|\tau|/2}. \qquad (3.5)$$

Substituting Eq. (3.2) into (3.4) and carrying out the differentiation, we also have

$$P^-(t, \omega) = 2 \operatorname{Re} \frac{1}{\sqrt{2\pi}} s^*(t)\, S_t^-(\omega) e^{j\omega t} \qquad (3.6)$$

which is a convenient form for its calculation.

As for the general behavior of the Page distribution, we note that the longer a particular frequency is observed, the larger the intensity is at that frequency. This is illustrated in Fig. 2(a), where we plot the Page distribution for the finite-duration signal,

$$s(t) = e^{j\omega_0 t}, \qquad 0 \le t \le T. \qquad (3.7)$$

The distribution is

$$P^-(t, \omega) = \begin{cases} \dfrac{t}{\pi} \operatorname{sinc}(\omega - \omega_0)t, & 0 \le t \le T \\ 0, & \text{otherwise.} \end{cases} \qquad (3.8)$$

As time increases, the distribution becomes more and more peaked at ω_0. In Fig. 2(b) we have also plotted the Wigner distribution for the same signal for later discussion. We remark here that up to $t = T/2$ the distributions are identical, but after that, their behavior is quite different. The Wigner distribution always goes to zero at the beginning and end of a finite-duration signal. That is not the case with the Page distribution.

It was subsequently realized by Turner [190] that Page's definition and procedure have two arbitrary aspects. First we can add to the Page distribution the function $\rho(t, \omega)$, which Turner called the complementary function, and obtain a new distribution,

$$P_{\text{new}}^-(t, \omega) = \frac{\partial}{\partial t} |S_t^-(\omega)|^2 + \rho(t, \omega). \qquad (3.9)$$

The marginals are still satisfied if the complementary function satisfies

$$\int \rho(t, \omega)\, dt = 0 \quad \text{and} \quad \int \rho(t, \omega)\, d\omega = 0. \qquad (3.10)$$

Turner also pointed out that taking the interval from $-\infty$ to t is not necessary; other intervals can be taken, each producing a different distribution function related to each other by a complementary function satisfying the above conditions.

Levin [125] defined the future running transform by

$$S_t^+(\omega) = \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} s(t') e^{-j\omega t'}\, dt' \qquad (3.11)$$

and using the same argument, we have

$$\int_{t}^{\infty} P^+(t', \omega)\, dt' = |S_t^+(\omega)|^2. \qquad (3.12)$$

Differentiation with respect to time leads to

$$P^+(t, \omega) = -\frac{\partial}{\partial t} |S_t^+(\omega)|^2$$

$$= 2 \operatorname{Re} \frac{1}{\sqrt{2\pi}} s^*(t)\, S_t^+(\omega) e^{j\omega t}. \qquad (3.13)$$

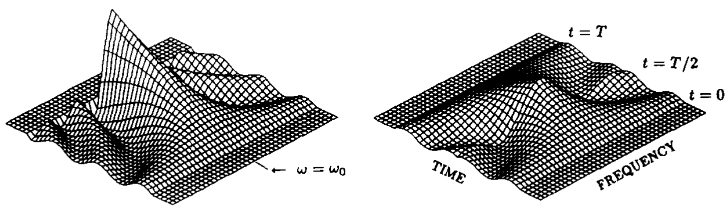
He also argued that we should treat past and future on the same footing, by defining the instantaneous energy spectrum as the average of the two,

$$P(t, \omega) = \frac{1}{2} [P^+(t, \omega) + P^-(t, \omega)] \qquad (3.14)$$

$$= \frac{1}{\sqrt{2\pi}} \operatorname{Re} s^*(t)\, e^{j\omega t} [S_t^-(\omega) + S_t^+(\omega)] \qquad (3.15)$$

$$= \frac{1}{\sqrt{2\pi}} \operatorname{Re} s^*(t)\, e^{j\omega t} S^*(\omega). \qquad (3.16)$$

This distribution is the real part of the distribution given by Eq. (2.2), which corresponds to the kernel $\phi(\theta, \tau) = \cos \frac{1}{2}\theta\tau$.



(a) (b)

Fig. 2. (a) Page and (b) Wigner distributions for the finite-duration signal $s(t) = e^{j\omega_0 t}$, $0 \le t \le T$. As time increases, the Page distribution continues to increase at ω_0. The Wigner distribution increases until $T/2$ and then decreases because the Wigner distribution always goes to zero at the beginning and end of a finite-duration signal.

Because of the symmetry between time and frequency we can also define the running signal transform by

$$s_\omega^-(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\omega} S(\omega')e^{j\omega' t}\, d\omega' \tag{3.17}$$

which yields the distribution

$$\begin{aligned}\hat{P}^-(t, \omega) &= \frac{\partial}{\partial\omega} |s_\omega^-(t)|^2 \\ &= 2\,\mathrm{Re}\, \frac{1}{\sqrt{2\pi}} s_\omega^-(t)\, S(\omega)e^{-j\omega t}.\end{aligned} \tag{3.18}$$

Similarly,

$$\hat{P}^+(t, \omega) = -\frac{\partial}{\partial\omega} \left| \frac{1}{\sqrt{2\pi}} \int_{\omega}^{\infty} s(t')e^{-j\omega t'}\, dt' \right|^2. \tag{3.19}$$

If Eqs. (3.17) and (3.18) are added together, we again obtain Eq. (3.16).

Filterbank Method: Grace [83] has given a interesting derivation of the Page distribution. The signal is passed through a bank of bandpass filters and the envelope of the output is calculated. The squared envelope is given by

$$|G(t, \omega)|^2 = \left| \int_{-\infty}^{t} s(\tau)\, h(t - \tau)e^{j\omega\tau}\, d\tau \right|^2 \tag{3.20}$$

where $h(t)e^{j\omega t}$ is the impulse response of one filter. By choosing the impulse response $h(t)$ to be 1 up to time t and zero afterward, we obtain the frequency density, the right-hand side of Eq. (3.3), and the Page distribution follows as before.

B. Complex Energy Spectrum

As already noted, the Rihaczek distribution was used and derived in many ways, but Rihaczek [167] and Ackroyd [2], [3] derived it from physical considerations. Consider a time-dependent voltage $V(t)$ going through a pure reactance whose admittance function is zero for all frequencies except for a narrow band around ω, where it is 1. If we decompose the voltage into its frequency components,

$$V(t) = \frac{1}{\sqrt{2\pi}} \int V_\omega e^{j\omega t}\, d\omega \tag{3.21}$$

the voltage at each frequency is $(1/\sqrt{2\pi})V_\omega e^{j\omega t}$. The current at that frequency is

$$i_\omega(t) = \frac{1}{\sqrt{2\pi}} V_\omega e^{j\omega t} \tag{3.22}$$

and the total current in the frequency range ω to $\omega + \Delta\omega$ is

$$i(t) = \int_{\omega}^{\omega+\Delta\omega} i_\omega(t)\, d\omega = \frac{1}{\sqrt{2\pi}} \int_{\omega}^{\omega+\Delta\omega} V_\omega e^{j\omega t}\, d\omega. \tag{3.23}$$

The complex power at time t is $V(t)\, i^*(t)$, and hence the energy in the time interval Δt is

$$\begin{aligned}E(t, \omega) &= \int_{t}^{t+\Delta t} V(t)\, i^*(t)\, dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{t}^{t+\Delta t} \int_{\omega}^{\omega+\Delta\omega} V_\omega^* V(t)e^{-j\omega t}\, d\omega\, dt.\end{aligned} \tag{3.24}$$

We now obtain the energy density ω at t by going to the limit,

$$e(t, \omega) = \lim_{\Delta t, \Delta\omega \to 0} \frac{E(t, \omega)}{\Delta t\, \Delta\omega} = \frac{1}{\sqrt{2\pi}} V_\omega^* V(t)e^{-j\omega t}. \tag{3.25}$$

Associating a signal $s(t)$ with the voltage $V(t)$ [in which case V_ω is $S(\omega)$], we have the distribution

$$e(t, \omega) = \frac{1}{\sqrt{2\pi}} s(t)\, S^*(\omega)e^{-j\omega t} \tag{3.26}$$

which is Eq. (2.2).

C. Ville-Moyal Method and Generalization

The Ville [194] approach is conceptually different and relies on traditional methods for constructing distributions from characteristic functions, although with a new twist. Ville used the method to derive the Wigner distribution but did not notice that there was an ambiguity in his presentation and that other distributions can be derived in the same way. As previously mentioned, Moyal [143] used the same approach.

Suppose we have a distribution $P(t, \omega)$ of two variables t and ω. Then the characteristic function is defined as the expectation value of $e^{j\theta t + j\tau\omega}$, that is,

$$M(\theta, \tau) = \langle e^{j\theta t + j\tau\omega} \rangle = \iint e^{j\theta t + j\tau\omega} P(t, \omega)\, dt\, d\omega. \tag{3.27}$$

It has certain manipulative advantages over the distribution. For example, the joint moments can be calculated by differentiation,

$$\langle t^n \omega^m \rangle = \frac{1}{j^n j^m} \frac{\partial^{n+m}}{\partial\theta^n \partial\tau^m} M(\theta, \tau) \Big|_{\theta, \tau = 0}. \tag{3.28}$$

By expanding the exponential in Eq. (3.27) it is straightforward to show

$$M(\theta, \tau) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{(j\theta)^n (j\tau)^m}{n!m!} \langle t^n \omega^m \rangle \tag{3.29}$$

which shows how the characteristic function can be constructed from the joint moments. In general the characteristic function is complex. However, not every complex function is a characteristic function since it must be the Fourier transform of some density. We point out that there are cases where the joint moments do not determine a unique distribution.

The distribution function may be obtained from $M(\theta, \tau)$ by Fourier inversion,

$$P(t, \omega) = \frac{1}{4\pi^2} \iint M(\theta, \tau)e^{-j\theta t - j\tau\omega}\, d\theta\, d\tau. \tag{3.30}$$

Is it possible to find the characteristic function for the situation we are considering and hence obtain the distribution? Clearly, it seems, we must have the distribution to start with. Recall, however, that the characteristic function is an average. Ville devised the following method for the calculation of averages, a method that is rooted in the quantum mechanical method of associating operators with ordinary variables. If we have a function $g_1(t)$ of time only, then its average value can be calculated in one of two ways, directly using the signal or by way of the spectrum, that is,

$$\langle g_1(t)\rangle = \int |s(t)|^2 g_1(t)\,dt = \int S^*(\omega)\, g_1\left(j\frac{d}{d\omega}\right) S(\omega)\,d\omega. \tag{3.31}$$

This is easy to show by assuming that the function $g_1(t)$ can be expanded in a power series. Therefore in the frequency domain time is "represented" by the operator $jd/d\omega$. Similarly for a function of frequency only, $g_2(\omega)$, its average value can be calculated by

$$\langle g_2(\omega)\rangle = \int |S(\omega)|^2\, g_2(\omega)\,d\omega = \int s^*(t)\, g_2\left(-j\frac{d}{dt}\right) s(t)\,dt. \tag{3.32}$$

and hence frequency becomes the operator $-jd/dt$ in the time domain.

Therefore we can associate time and frequency with the operators $\mathfrak{T}$ and $\mathfrak{W}$ so that

$$\mathfrak{T} \rightarrow t \qquad \mathfrak{W} \rightarrow -j\frac{d}{dt} \qquad \text{in the time domain}$$

$$\mathfrak{T} \rightarrow j\frac{d}{d\omega} \qquad \mathfrak{W} \rightarrow \omega \qquad \text{in the frequency domain.}$$

But what if we have a function $g(t, \omega)$ of time *and* frequency? How do we then calculate its average value? Ville proposed that we do it the same way, namely, by using

$$\langle g(t, \omega)\rangle = \int s^*(t)\, \mathbf{G}(t, \mathfrak{W})\, s(t)\,dt \tag{3.33}$$

in the time domain and

$$\langle g(t, \omega)\rangle = \int S^*(\omega)\, \mathbf{G}(\mathfrak{T}, \omega)\, S(\omega)\,d\omega \tag{3.34}$$

in the frequency domain, where $\mathbf{G}(\mathfrak{T}, \mathfrak{W})$ is the operator "associated" with or "corresponding" to $g(t, \omega)$. Since the characteristic function is an expectation value, we can use Eq. (3.33) to obtain it, and in particular,

$$M(\theta, \tau) = \langle e^{j\theta t + j\tau\omega}\rangle \rightarrow \int s^*(t) e^{j\theta\mathfrak{T} + j\tau\mathfrak{W}} s(t)\,dt. \tag{3.35}$$

We now proceed to evaluate this expression. Because the quantities are noncommuting operators,

$$\mathfrak{T}\mathfrak{W} - \mathfrak{W}\mathfrak{T} = j \tag{3.36}$$

one has to be careful in manipulating them. To break up the exponent, one can use a special case of the Baker–Hausdorf [201] theorem

$$e^{j\theta\mathfrak{T} + j\tau\mathfrak{W}} = e^{-j\theta\tau/2} e^{j\tau\mathfrak{W}} e^{j\theta\mathfrak{T}} = e^{j\theta\tau/2} e^{j\theta\mathfrak{T}} e^{j\tau\mathfrak{W}}. \tag{3.37}$$

The operator $e^{j\tau\mathfrak{W}}$ is the translation operator,

$$e^{j\tau\mathfrak{W}} s(t) = e^{\tau(d/dt)} s(t) = s(t + \tau) \tag{3.38}$$

and substituting into Eq. (3.35) we have

$$M(\theta, \tau) = \int s^*(t)\, e^{j\theta\tau/2} e^{j\theta t} s(t + \tau)\,dt. \tag{3.39}$$

Making the change of variables $u = t + \frac{1}{2}\tau$, $du = dt$, we obtain

$$M(\theta, \tau) = \int s^*\left(u - \frac{1}{2}\tau\right) e^{j\theta u} s\left(u + \frac{1}{2}\tau\right) du. \tag{3.40}$$

Inverting as per Eq. (3.30), we obtain the distribution

$$P(t, \omega) = \frac{1}{4\pi^2} \iiint s^*\left(u - \frac{1}{2}\tau\right) e^{j\theta u} s\left(u + \frac{1}{2}\tau\right) \cdot e^{-j\theta t - j\tau\omega}\,d\theta\,d\tau\,du. \tag{3.41}$$

The θ integration gives a delta function, and hence

$$P(t, \omega) = \frac{1}{2\pi} \iint s^*\left(u - \frac{1}{2}\tau\right) e^{-j\tau\omega} \delta(u - t) \cdot s\left(u + \frac{1}{2}\tau\right) d\tau\,du \tag{3.42}$$

$$= \frac{1}{2\pi} \int s^*\left(t - \frac{1}{2}\tau\right) e^{-j\tau\omega} s\left(t + \frac{1}{2}\tau\right) d\tau \tag{3.43}$$

which is the Wigner distribution.

It was subsequently pointed out [58], [132] that there is an inherent ambiguity in the derivation because the characteristic functions written in terms of the classical variables allow many operator correspondences. The method was generalized by devising a simple method to generate correspondences and distributions [58]. Instead of

$$e^{j\theta t + j\tau\omega} \rightarrow e^{j\theta\mathfrak{T} + j\tau\mathfrak{W}} \tag{3.44}$$

which is called the Weyl correspondence, we could take

$$e^{j\theta t + j\tau\omega} \rightarrow e^{j\theta\mathfrak{T}} e^{j\tau\mathfrak{W}} \tag{3.45}$$

or

$$e^{j\theta t + j\tau\omega} \rightarrow e^{j\tau\mathfrak{W}} e^{j\theta\mathfrak{T}} \tag{3.46}$$

which are called normal ordered correspondences. The symmetrical correspondence is the average of the two,

$$e^{j\theta t + j\tau\omega} \rightarrow \tfrac{1}{2}(e^{j\theta\mathfrak{T}} e^{j\tau\mathfrak{W}} + e^{j\tau\mathfrak{W}} e^{j\theta\mathfrak{T}}). \tag{3.47}$$

There are many other expressions which reduce to the left-hand side when operators $\mathfrak{T}$ and $\mathfrak{W}$ are not considered operators but ordinary functions. The nonunique procedure of going from a classical function to an operator function is called a correspondence rule, and there are an infinite number of such rules [58]. Each different correspondence rule will give a different characteristic function, which in turn will give a different distribution. If we use the correspondence given by Eq. (3.46), we obtain

$$M(\theta, \tau) = \int s^*(t)\, e^{j\tau\mathfrak{W}} e^{j\theta t} s(t) = \int s^*(t)\, e^{j\theta(t+\tau)} s(t + \tau)\,dt = \int s^*(t - \tau)\, e^{j\theta t} s(t)\,dt. \tag{3.48}$$

Inverting to get the distribution we have

$$P(t, \omega) = \frac{1}{4\pi^2} \iiint s^*(u)\, e^{j\theta(u+\tau)} s(u + \tau) \cdot e^{-j\theta t - j\tau\omega}\,d\theta\,d\tau\,du \tag{3.49}$$

$$= \frac{1}{2\pi} s(t) \int s^*(t - \tau) e^{-j\tau\omega}\,d\tau = \frac{1}{\sqrt{2\pi}} s(t)\, e^{-jt\omega} S^*(\omega) \tag{3.50}$$

which is the Rihaczek distribution, Eq. (2.2). If we use Eq. (3.45) instead, we get the complex conjugate of Eq. (3.50), and if we use Eq. (3.47), we get the real part.

Another way to get the correspondence is by associating arbitrary mixed products of time and frequency [58]. That is, we associate

$$t^n \omega^m \rightarrow C(\mathfrak{I}, \mathfrak{W}) \tag{3.51}$$

where $C(\mathfrak{I}, \mathfrak{W})$ indicates a correspondence, and then we calculate the characteristic function from Eq. (3.29). Once the characteristic function is determined, the distribution is obtained as above. Many correspondences of the form given by Eq. (3.51) have been proposed. An early one was that of Born and Jordan [33]. When the above-mentioned procedure is carried out, one obtains a distribution with the kernel [58]

$$\phi(\theta, \tau) = \frac{\sin \frac{1}{2}\theta\tau}{\frac{1}{2}\theta\tau} \tag{3.52}$$

which has some interesting properties [58], [51], [76], [26], [123].[4]

Hence one way to approach the problem of obtaining a joint distribution is to write the totality of possible correspondences for the characteristic function and repeat the above calculation. The reason for the wide choice is that the time and frequency operators do not commute, and hence a number of different rules are possible. A general procedure for associating functions with operators has been developed and has been used in a number of different fields. To an ordinary function $g(t, \omega)$ one associates the operator $\mathbf{G}(\mathfrak{I}, \mathfrak{W})$ in the following manner [58], [132], [181]:

$$\mathbf{G}(\mathfrak{I}, \mathfrak{W}) = \iint \gamma(\theta, \tau)\, \phi(\theta, \tau) e^{j\theta\mathfrak{I} + j\tau\mathfrak{W}}\, d\theta\, d\tau \tag{3.53}$$

where

$$\gamma(\theta, \tau) = \frac{1}{4\pi^2} \iint g(t, \omega) e^{-j\theta t - j\tau\omega}\, dt\, d\omega \tag{3.54}$$

or, equivalently,

$$\mathbf{G}(\mathfrak{I}, \mathfrak{W}) = \frac{1}{4\pi^2} \iint g(t, \omega)\, \phi(\theta, \tau) \cdot e^{j\theta(\mathfrak{I} - t) + j\tau(\mathfrak{W} - \omega)}\, d\theta\, d\tau\, dt\, d\omega \tag{3.55}$$

where $\phi(\theta, \tau)$ is an arbitrary function that satisfies $\phi(\theta, 0) = \phi(0, \tau) = 1$. The reason for imposing this condition is that it assures that functions of t or ω only transform according to

$$g_1(t) \rightarrow g_1(\mathfrak{I}), \qquad g_2(\omega) \rightarrow g_2(\mathfrak{W}). \tag{3.56}$$

Now if this procedure is applied to the characteristic function, we obtain the general correspondence

$$e^{j\theta t + j\tau\omega} \rightarrow \phi(\theta, \tau) e^{j\theta\mathfrak{I} + j\tau\mathfrak{W}} \tag{3.57}$$

Substituting this into Eq. (3.33), we have, as before,

$$M(\theta, \tau) = \phi(\theta, \tau) \int s^*\left(u - \frac{1}{2}\tau\right) e^{j\theta u} s\left(u + \frac{1}{2}\tau\right) du \tag{3.58}$$

[4]Although this kernel and the corresponding distribution are sometimes attributed to Born and Jordan, they never considered joint distributions or kernels. It was derived in [58] using the correspondence of Born and Jordan.

and inverting using Eq. (3.30), we obtain the general class of distributions, Eq. (2.4).

The formalisms possible with this approach and the relation to classical theory has been analyzed by Groblicki [84], Srinivas and Wolf [181], and Ruggeri [171].

D. Local Autocorrelation Methods

A general approach to deriving time-dependent spectra is by generalizing the relationship between the power spectrum and the autocorrelation function. The concept of a local autocorrelation function was developed by Fano [72] and Schroeder and Atal [175], and the relationship of their work to time-varying spectra was considered by Ackroyd [2], [3]. A local autocorrelation method was used by Lampard [122] for deriving the Page distribution, and subsequently other investigators have pointed out the relation to other distributions. The basic idea is to write the energy density spectrum as

$$|S(\omega)|^2 = \left| \frac{1}{\sqrt{2\pi}} \int s(t) e^{-j\omega t}\, dt \right|^2 = \frac{1}{2\pi} \iint s^*(t')\, s(t) e^{j\omega(t' - t)}\, dt'\, dt. \tag{3.59}$$

By making the transformation $\tau = t - t'$, $d\tau = -dt'$, we have

$$|S(\omega)|^2 = \frac{1}{2\pi} \iint s^*(t - \tau)\, s(t) e^{-j\omega\tau}\, d\tau\, dt = \frac{1}{2\pi} \int R(\tau) e^{-j\omega\tau}\, d\tau \tag{3.60}$$

where the autocorrelation function is defined as

$$R(\tau) = \int s^*(t)\, s(t + \tau)\, dt = \int s^*(t - \tau)\, s(t)\, dt = \int s^*\left(t - \frac{1}{2}\tau\right) s\left(t + \frac{1}{2}\tau\right) dt. \tag{3.61}$$

One generalizes the relationship between the energy spectrum and $R(\tau)$ as given by Eq. (3.60) by assuming that we can write a time-dependent power spectrum, that is, a joint time-frequency distribution, as

$$P(t, \omega) = \frac{1}{2\pi} \int R_t(\tau) e^{-j\omega\tau}\, d\tau \tag{3.62}$$

where now $R_t(\tau)$ is a time-dependent or local autocorrelation function. Many expressions for $R_t(\tau)$ have been proposed, and we illustrate some of the possibilities before generalizing. One can simply take

$$R_t(\tau) = s(t)\, s^*(t + \tau) \tag{3.63}$$

which, when substituted into Eq. (3.62), yields the Rihaczek distribution. Mark [138] argued for symmetry,

$$R_t(\tau) = s^*\left(t - \frac{1}{2}\tau\right) s\left(t + \frac{1}{2}\tau\right) \tag{3.64}$$

which gives the Wigner distribution. Mark pointed out that one could consider a more general form,

$$R_t(\tau) = s^*(t - k\tau)\, s[t + (1 - k)\tau]. \tag{3.65}$$

He preferred the value of $k = \frac{1}{2}$ because for the autocorrelation function we have $R(\tau) = R^*(-\tau)$, and if we want the same to hold for the local autocorrelation function

$$R_t(\tau) = R_t^*(-\tau) \tag{3.66}$$

then the value of $k = \frac{1}{2}$ must be chosen. However, there are an infinite number of other forms that can be obtained. A generalized time-dependent autocorrelation function can be defined from the general distribution, Eq. (2.4), as was done by Choi and Williams [51]. Comparing Eq. (2.4) with Eq. (3.62), the generalized time-dependent autocorrelation function is

$$R_t(\tau) = \frac{1}{2\pi} \iint e^{j\theta(u-t)} \phi(\theta, \tau)\, s^*\left(u - \frac{1}{2}\tau\right) \cdot s\left(u + \frac{1}{2}\tau\right) du\, d\theta. \tag{3.67}$$

The above special cases can be obtained by particular choices of the kernel function.

It is convenient to write this as

$$R_t(\tau) = \frac{1}{2\pi} \int r(u - t, \tau)\, s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) du \tag{3.68}$$

where

$$r(u, \tau) = \int e^{j\theta u} \phi(\theta, \tau)\, d\theta. \tag{3.69}$$

We now ask, for what types of kernels does the local autocorrelation function satisfy Eq. (3.66)? Taking the complex conjugate of Eq. (3.67) and substituting $-\tau$ for τ, we have

$$R_t^*(-\tau) = \frac{1}{4\pi^2} \iint e^{j\theta(u-t)}\, \phi^*(-\theta, -\tau)\, s^*\left(u - \frac{1}{2}\tau\right) \cdot s\left(u + \frac{1}{2}\tau\right) du\, d\theta. \tag{3.70}$$

Therefore if we want Eq. (3.66) to hold, we must take

$$\phi(\theta, \tau) = \phi^*(-\theta, -\tau). \tag{3.71}$$

What is interesting about this relation is that it is the requirement for the distribution to be real, as discussed in Section IV.

Choi and Williams [51] devised a very interesting way to understand the effects of the kernel by examining the local autocorrelation function. They point out that since the main interest in these distributions is to study phenomena that are occurring locally, we want to give a relatively large weight to $s^*(u - \frac{1}{2}\tau)\, s(u + \frac{1}{2}\tau)$ when u is close to t; otherwise we would not be emphasizing the events near time t. Choi and Williams used this concept very effectively in devising a new distribution. The importance of their work is that it gives a fresh perspective and a concrete prescription for obtaining distributions with desirable properties. They have unified these concepts by way of the time-dependent autocorrelation function and generalized ambiguity function [63], [58], [76]. We discuss their distribution in Section III-G.

E. Pseudo-Characteristic Function Method and General Bilinear Class

We have seen in Section II-C how Ville's method is generalized to obtain an infinite number of distributions. We now give an alternative derivation which avoids operator concepts and depends on the relationship between the characteristic function of two variables and the characteristic functions of the marginals. Suppose we have a joint distribution of $P(t, \omega)$ and that their marginals are given by

$$P_1(t) = \int P(t, \omega)\, d\omega, \qquad P_2(\omega) = \int P(t, \omega)\, dt. \tag{3.72}$$

The characteristic functions of the marginals are

$$M_1(\theta) = \int e^{j\theta t} P_1(t)\, dt, \qquad M_2(\tau) = \int e^{j\tau\omega} P_2(\omega)\, d\omega \tag{3.73}$$

and comparing with Eq. (3.27), we have

$$M(\theta, 0) = M_1(\theta), \qquad M(0, \tau) = M_2(\tau). \tag{3.74}$$

Now suppose a characteristic function satisfies the marginals, that is, Eqs. (3.74). Then

$$M_{\text{new}}(\theta, \tau) = \phi(\theta, \tau)\, M(\theta, \tau) \tag{3.75}$$

will also satisfy them if we take

$$\phi(0, \tau) = \phi(\theta, 0) = 1. \tag{3.76}$$

Therefore any characteristic function, if multiplied by a function satisfying Eq. (3.76), will produce a new characteristic function which will also satisfy the marginals. If we take Eq. (3.40) as the "original" characteristic function, then a whole class is obtained from

$$M(\theta, \tau) = \phi(\theta, \tau) \int s^*\left(u - \frac{1}{2}\tau\right) e^{j\theta u} s\left(u + \frac{1}{2}\tau\right) du \tag{3.77}$$

which when substituted into Eq. (3.30) yields, as before, the general class of distributions, Eq. (2.4).

We emphasize that even though we have been using the terminology "characteristic function," they are not proper characteristic functions since Eq. (3.75) is not a sufficient condition. They should properly be called quasi- or pseudo-characteristic functions. We also point out that the choice of Eq. (3.40) for $M(\theta, \tau)$ in Eq. (3.75) is arbitrary.

F. Positive Distributions

The question of the existence of manifestly positive distributions which satisfy the marginals has been a central issue in the field. Many "proofs" have been given for their nonexistence, and for a long time it was generally believed that they did not exist. The uncertainty principle was often invoked to make it reasonable that positive distributions cannot exist. Mugur-Schachter [144] has shown where the hidden assumptions in these proofs have crept in. Also, Park and Margenau [154] have made a far-reaching study of the relation of joint measurement, joint distributions, and the existence of positive distributions. Positive distributions do exist, and it is easy to generate an infinite number of them [62]. Choose any positive function $\Omega(u, v)$ of the two variables u, v such that

$$\int_0^1 \Omega(u, v)\, dv = 1, \qquad \int_0^1 \Omega(u, v)\, du = 1 \tag{3.78}$$

and construct

$$P(t, \omega) = |S(\omega)|^2 |s(t)|^2 \Omega(u, v). \tag{3.79}$$

For u and v we now substitute

$$u(t) = \int_{-\infty}^{t} |s(t')|^2 \, dt', \qquad v(\omega) = \int_{-\infty}^{\omega} |S(\omega')|^2 \, d\omega'. \tag{3.80}$$

To show that the marginals are satisfied, we integrate with respect to ω,

$$\int P(t, \omega) \, d\omega = |s(t)|^2 \int |S(\omega)|^2 \Omega(u, v) \, d\omega$$
$$= |s(t)|^2 \int_0^1 \Omega(u, v) \, dv = |s(t)|^2. \tag{3.81}$$

The last step follows since $dv = |S(\omega)|^2 \, d\omega$; similarly for integration with respect to t. Functions satisfying Eq. (3.78) can readily be constructed. It has been shown that this procedure generates all possible positive distributions [73], [176]. We note that the positive distributions are not bilinear in the signal and that in general $\Omega(u, v)$ may be a functional of the signal. The relation of bilinearity to the question of positivity is discussed in the conclusion. The kernel which generates the positive distributions from Eq. (2.4) can be obtained [62].

Whether any of these positive distributions can yield intensities that conform to our expectations has been questioned. Janssen and Claasen [101] have pointed out that no systematic procedure exists for choosing a unique $\Omega(u, v)$. That of course is also true with the bilinear distributions. Janssen [102a] has argued that the positive distributions cannot satisfactorily represent a chirplike signal, although it has been noted [102b] that it can if the kernel is taken to be signal dependent. The problem of constructing a joint distribution satisfying the marginals arises in every field of science and mathematics and is one of the major problems to be resolved. There are in general an infinite number of joint distributions for given marginals, although their construction is far from straightforward. Because the marginals do not contain any correlation information, other conditions are needed to specify a particular joint distribution. That information is entered by way of Ω, although a systematic procedure for doing so has not been developed. In the case of signal analysis and quantum mechanics there is the further issue of dealing with a signal (or wave function) and constructing the marginal from it. The two marginals $|s(t)|^2$, and $|S(\omega)|^2$ do not determine the signal. This was pointed out by Reichenbach [166], who attributes to Bargmann a method for constructing different signals which have the same absolute instantaneous energy and energy density spectrum. Vogt [195] and Altes [7] give similar methods of constructing such functions. Since the signal contains information that the marginals do not, in general $\Omega(u, v)$ must be signal dependent. The question of the amount of information needed to construct a unique joint distribution, and of how much more information the signal contains than the combination of the energy density and spectral energy density, requires considerable further research.

G. Choi-Williams Method

A new and novel approach has recently been presented by Choi and Williams [51] where they address one of the main difficulties with the Wigner distribution. As we have already seen from Fig. 1(a), the Wigner distribution sometimes indicates intensity in regions where one would expect zero values. These spurious values, which are due to the so-called cross terms, are particularly prevalent for multicomponent signals [26], [41], [51], [91], [141]. The cause of these effects, sometimes called "artifacts," is usually attributed to the bilinear nature of the distribution, and it was felt by many that it is something we have to live with. In the case of the Wigner distribution, extensive studies have been made and methods devised to remove them in some way. This usually involves violating some of the desired properties like the marginals. Choi and Williams argue that instead of devising procedures to eliminate them from the Wigner distribution, let us find distributions for which the spurious values are minimal. Choi and Williams succeeded in devising a distribution that behaves remarkably well in satisfying our intuitive notions of where the signal energy should be concentrated, and that reduces to a large extent the spurious cross terms for multicomponent signals. Also, the desirable properties of a distribution are satisfied.

Following Choi and Williams we consider a signal made up of components $s_k(t)$,

$$s(t) = \sum_{k=1}^{N} s_k(t). \tag{3.82}$$

Substituting this into the general equation, Eq. (2.4), we can write the distribution as the sum of self and cross terms,

$$P(t, \omega) = \sum_{k=1}^{N} P_{kk}(t, \omega) + \sum_{\substack{k,l=1 \\ l \neq k}}^{N} P_{kl}(t, \omega) \tag{3.83}$$

where

$$P_{kl}(t, \omega) = \frac{1}{4\pi^2} \iiint e^{-j\theta t - j\tau\omega + j\theta u} \phi(\theta, \tau) \cdot s_k^*\left(u - \frac{1}{2}\tau\right) s_l\left(u + \frac{1}{2}\tau\right) du \, d\tau \, d\theta. \tag{3.84}$$

Choi and Williams [51] realized that by a judicious choice of the kernel, one can minimize the cross terms and still retain the desirable properties of the self terms. This aspect is investigated using a generalized ambiguity concept [63] and autocorrelation function, as in Section III. They found a particularly good choice for the kernel,

$$\phi(\theta, \tau) = e^{-\theta^2\tau^2/\sigma} \tag{3.85}$$

where σ is a constant. Substituting into Eq. (2.4) and integrating over θ, one obtains

$$P_{CW}(t, \omega) = \frac{1}{4\pi^{3/2}} \iint \frac{1}{\sqrt{\tau^2/\sigma}} e^{-[(u-t)^2/(4\tau^2/\sigma)] - j\tau\omega} s^*\left(u - \frac{1}{2}\tau\right) \cdot s\left(u + \frac{1}{2}\tau\right) du \, d\tau. \tag{3.86}$$

The ability to suppress the cross terms comes by way of controlling σ. The $r_{CW}(t - u, \tau)$, as defined by Eq. (3.69), is

$$r(u - t, \tau) = \frac{1}{\sqrt{\tau^2/\sigma}} e^{-[(u-t)^2/(4\tau^2/\sigma)]}. \tag{3.87}$$

From the discussion of Section III-D we note that indeed it is peaked when $u = t$ and σ can be used to control the relative importance of τ.

The kernel given by Eq. (3.85) satisfies Eq. (3.71), which shows that the local autocorrelation function satisfies Eq. (3.66) and that the distribution is real. In Section IV we use

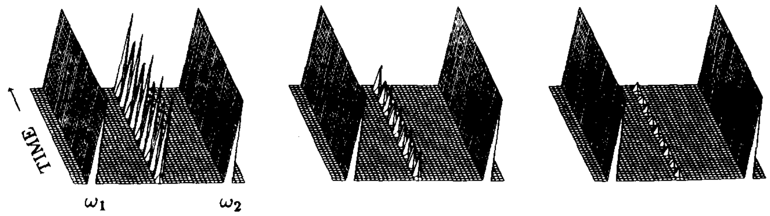


Fig. 3. (a) Wigner and (b), (c) Choi–Williams distributions for the sum of two sine waves, $s(t) = e^{j\omega_1 t} + e^{j\omega_2 t}$. The Wigner distribution is peaked to infinity at the frequencies ω_1, ω_2 and at the spurious value of $\omega = \frac{1}{2}(\omega_1 + \omega_2)$. The middle term oscillates and is due to the cross terms. The Choi–Williams distributions are shown for (b) $\sigma = 10^6$ and (c) $\sigma = 10^5$. Note that all three distributions satisfy the marginals. The values for ω are $\omega_1 = 1$ and $\omega_2 = 9$. The delta functions at ω_1 and ω_2 are symbolically represented and are cut off at the value of 700.

this kernel as an example to demonstrate how the general properties of a distribution can be readily determined by inspection of the kernel.

The importance of the work of Choi and Williams is that they have formulated and implemented effectively a means of choosing distributions that minimize spurious values caused by the cross terms. Moreover they have connected in a very revealing way the properties of a distribution with that of the local autocorrelation function and characteristic function. The kernel given by Eq. (3.85) is a one-parameter family, but their method can be used to find many other kernels having the general desirable properties.

We give three examples to illustrate the considerable clarity in interpretation possible using the distribution of Choi and Williams. We first take the sum of two pure sine waves,

$$s(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t}. \tag{3.88}$$

The Choi–Williams distribution is readily worked out [51],

$$P_{CW}(t, \omega) = A_1^2 \delta(\omega - \omega_1) + A_2^2 \delta(\omega - \omega_2) + 2A_1 A_2 \cdot \cos(\omega_2 - \omega_1)t \, \eta(\omega, \omega_1, \omega_2, \sigma) \tag{3.89}$$

where

$$\eta(\omega, \omega_1, \omega_2, \sigma) = \sqrt{\frac{1}{4\pi(\omega_1 - \omega_2)^2/\sigma}} \cdot \exp\left\{-\frac{[\omega - \frac{1}{2}(\omega_1 + \omega_2)]^2}{4(\omega_1 - \omega_2)^2/\sigma}\right\}. \tag{3.90}$$

We first note that

$$\lim_{\sigma \to \infty} \eta(\omega, \omega_1, \omega_2, \sigma) = \delta[\omega - \tfrac{1}{2}(\omega_1 + \omega_2)] \tag{3.91}$$

and for that case the distribution becomes infinitely peaked at $\omega = \frac{1}{2}(\omega_1 + \omega_2)$. In fact, as $\sigma \to \infty$, it becomes the Wigner distribution, since for that limit the kernel becomes 1. As long as σ is finite, the cross terms will be finite at that point and will increase as $\sqrt{\sigma}$. Note that if σ is small, the cross terms are small and do not obscure the interpretation with spurious values. In Fig. 3 we illustrate the effect of σ. We have represented the delta function symbolically, but the calculation for the cross terms is exact. We see that the cross terms may easily be eliminated for all practical purposes by choosing an appropriate value of σ.

Another revealing example is the sum of two chirplike signals [51],

$$s(t) = A_1\left(\frac{\alpha_1}{\pi}\right)^{1/4} e^{-\alpha_1 t^2/2 + j\beta_1 t^2/2 + j\omega_1 t} + A_2\left(\frac{\alpha_2}{\pi}\right)^{1/4} e^{-\alpha_2 t^2/2 + j\beta_2 t^2/2 + j\omega_2 t} \tag{3.92}$$

which have instantaneous frequencies along $\omega = \beta_1 t + \omega_1$ and $\omega = \beta_2 t + \omega_2$. We expect the concentration of energies to be along the instantaneous frequencies. Fig. 4 is a plot of the Wigner and the Choi–Williams distributions. Note the middle hump in the Wigner distribution. On the other hand, the distribution of Choi and Williams is clear and easily interpretable. We emphasize that the distribution of Choi and Williams satisfies the marginals for any value of σ.

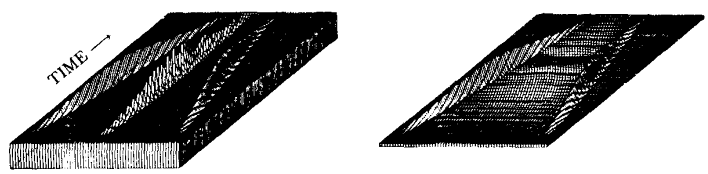


Fig. 4. (a) Wigner and (b) Choi–Williams distributions for the sum of two chirps. The Wigner distribution has artifacts and spurious values in between the two concentrations along the instantaneous frequencies of the chirps. In the Choi–Williams distribution the spurious terms are made negligible by an appropriate choice of σ.

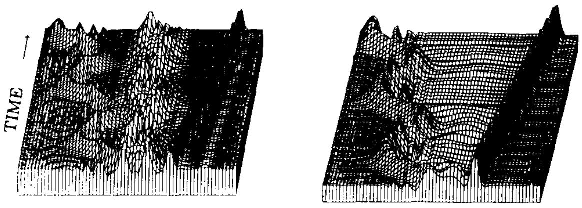


Fig. 5. (a) Wigner and (b) Choi-Williams distributions for signal $s(t) = A_1 e^{j\beta_1 t^2/2 + j\omega_1 t} + A_2 e^{j\omega_2 t + j\beta_2 \sin \omega_m t}$. Both distributions show concentration at the instantaneous frequencies $\omega = \omega_1 + \beta_1 t$ and $\omega = \omega_2 + b_2 \omega_m \cos \omega_m t$. For the Wigner distribution there are spurious terms in between the two frequencies. They are very small in magnitude in the Choi-Williams distribution.

In Fig. 5 we show the Wigner and Choi-Williams distributions for a signal that is the sum of a chirp and a sinusoidal modulated signal,

$$s(t) = A_1 e^{j\beta_1 t^2/2 + j\omega_1 t} + A_2 e^{j\omega_2 t + j\beta_2 \sin \omega_m t}. \tag{3.93}$$

For both distributions we see a concentration along the instantaneous frequencies; however, for the Wigner distribution there are "interference" terms which are very minor in the Choi-Williams distribution.

IV. Unified Approach

As can be seen from the preceding section, there are many distributions with varied functional forms and properties. A unified approach can be formulated in a simple manner with the advantage that all distributions can be studied together in a consistent way. Moveover, a method which readily generates them allows the extraction of those with the desired properties. As we will see, the properties of a distribution are reflected as simple constaints on the kernel. By examination of the kernel one can determine the general properties of the distribution. Also, having a general method to represent all distributions can be used advantageously to develop practical methods for analysis and filtering, as was done by Eichmann and Dong [70]. Excellent reviews relating the properties of the kernel to the properties of the distribution have been given by Janse and Kaizer [97], Janssen [98], Claasen and Mecklenbrauker [56], and Boashash [26].

For convenience we repeat the general form here,

$$P(t, \omega) = \frac{1}{4\pi^2} \iiint e^{-j\theta t - j\tau\omega + j\theta u} \phi(\theta, \tau) \cdot s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) du\, d\tau\, d\theta. \tag{4.1}$$

The kernel can be a function of time and frequency and in principle can be a function of the signal. However, unless otherwise stated, we shall here assume that it is a not a function of time or frequency and is independent of the signal. Independence of the kernel of time and frequency assures that the distribution is time and shift invariant, as indicated below. If the kernel is independent of the signal, then the distributions are said to be bilinear because the signal enters only twice. An important subclass are those distributions for which the kernel is a function of $\theta\tau$, the product kernels,

$$\phi(\theta, \tau) = \phi_{PR}(\theta\tau). \tag{4.2}$$

For notational clarity, we will drop the subscript PR since one can tell whether we are talking about the general case or the product case by the number of variables attributed to $\phi(\theta, \tau)$. In Table 1 we list some distributions and their corresponding kernels.

The general class of distributions can be expressed in terms of the spectrum by substituting Eq. (1.3) into Eq. (2.4) to obtain

$$P(t, \omega) = \frac{1}{4\pi^2} \iiint e^{-j\theta t - j\tau\omega + j\tau u} \phi(\theta, \tau) \cdot S^*\left(u + \frac{1}{2}\theta\right) S\left(u - \frac{1}{2}\theta\right) du\, d\tau\, d\theta. \tag{4.3}$$

Table 1 Some Distributions and Their Kernels

Reference	Kernel $\phi(\theta, \tau)$	Distribution $P(t, \omega)$
Wigner [199], Ville [194]	1	$\frac{1}{2\pi} \int e^{-j\tau\omega} s^*(t - \frac{1}{2}\tau)\, s(t + \frac{1}{2}\tau)\, d\tau$
Margenau and Hill [133]	$\cos \frac{1}{2}\theta\tau$	$\mathrm{Re} \frac{1}{\sqrt{2\pi}} s(t)\, e^{-j\tau\omega} S^*(\omega)$
Kirkwood [107], Rihaczek [167]	$e^{j\theta\tau/2}$	$\frac{1}{\sqrt{2\pi}} s(t)\, e^{-j\tau\omega} S^*(\omega)$
sinc [58]	$\frac{\sin a\theta\tau}{a\theta\tau}$	$\frac{1}{4\pi a} \int \frac{1}{\tau} e^{-j\omega\tau} \int_{t-a\tau}^{t+a\tau} s^*(u - \frac{1}{2}\tau)\, s(u + \frac{1}{2}\tau)\, du\, d\tau$
Page [152]	$e^{j\theta\|\tau\|/2}$	$\frac{\partial}{\partial t} \left\| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} s(t')\, e^{-j\omega t'}\, dt' \right\|^2$
Choi and Williams [51]	$e^{-\theta^2\tau^2/\sigma}$	$\frac{1}{4\pi^{3/2}} \iint \sqrt{\frac{\sigma}{\tau^2}}\, e^{-\sigma(u-t)^2/4\tau^2 - j\tau\omega}\, s^*(u - \frac{1}{2}\tau) \cdot s(u + \frac{1}{2}\tau)\, du\, d\tau$
Spectrogram	$\int h^*(u - \frac{1}{2}\tau)\, e^{-j\theta u} \cdot h(u + \frac{1}{2}\tau)\, du$	$\left\| \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s(\tau)\, h(\tau - t)\, d\tau \right\|^2$

The best practical method to determine the kernel for a given distribution is to put it into the form of Eq. (4.1). Otherwise one can calculate the kernel from

$$\phi(\theta, \tau) = \frac{\iint e^{j\theta t + j\tau\omega} P(t, \omega)\, dt\, d\omega}{\int e^{j\theta u} s^*(u - \tfrac{1}{2}\tau)\, s(u + \tfrac{1}{2}\tau)\, du} \quad (4.4)$$

which is obtained from Eq. (4.1) by Fourier inversion. It is also convenient to define the cross distribution function for two signals, as was done by Eq. (3.84). The main reason for defining it is that the distribution of the sum of two signals

$$s(t) = s_1(t) + s_2(t) \quad (4.5)$$

can be conveniently written as

$$P(t, \omega) = P_{11}(t, \omega) + P_{22}(t, \omega) + P_{12}(t, \omega) + P_{21}(t, \omega). \quad (4.6)$$

If $s_1(t)$ and $s_2(t)$ are each normalized to 1, then an overall normalization may be inserted so that $s(t)$ and the distribution are normalized to 1.

A. Physical Properties Related to Kernel

We now show how the properties of the distribution are related to the properties of the kernel. We shall give only a few derivations to indicate the general approach since the procedures are fairly simple.

Instantaneous Energy and Spectrum: If $P(t, \omega)$ is to be a joint distribution for the intensity, we want it to satisfy the individual intensities of time and frequency. That is, when the frequency variable is integrated out, we expect to have the instantaneous power $|s(t)|^2$, and similarly when time is integrated out, we expect to have the energy density spectrum $|S(\omega)|^2$. Integrating Eq. (4.1) with respect to ω we have

$$\int P(t, \omega)\, d\omega = \frac{1}{2\pi} \iiint \delta(\tau)\, e^{j\theta(u-t)} \phi(\theta, \tau) \cdot s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) d\theta\, du\, d\tau \quad (4.7)$$

$$= \frac{1}{2\pi} \iint e^{j\theta(u-t)} \phi(\theta, 0) |s(u)|^2\, d\theta\, du. \quad (4.8)$$

The only way this can be made equal to $|s(t)|^2$ is if

$$\frac{1}{2\pi} \int e^{j\theta(u-t)} \phi(\theta, 0)\, d\theta = \delta(t - u) \quad (4.9)$$

which forces

$$\phi(\theta, 0) = 1. \quad (4.10)$$

Similarly, if we want

$$\int P(t, \omega)\, dt = |S(\omega)|^2 \quad (4.11)$$

we must take

$$\phi(0, \tau) = 1. \quad (4.12)$$

It also follows that if the total energy is to be preserved, that is,

$$\int P(t, \omega)\, d\omega\, dt = 1 = \text{total energy} \quad (4.13)$$

we must have

$$\phi(0, 0) = 1 \quad (4.14)$$

which is called the normalization condition. We note that this condition is weaker than the conditions given by Eqs. (4.10) and (4.12), that is, it is possible to have a joint distribution whose total energy is the same as that of the signal, but whose marginals are not satisfied. An example of this is the spectrogram discussed in Section VI.

Reality: The bilinear distributions are not in general positive definite, which causes serious interpretive problems. It has been generally argued that at least they should be real. By taking the complex conjugate of Eq. (4.1) and comparing it to the original, it is straightforward to show that a necessary and sufficient condition for a distribution to be real is that the kernel satisfy

$$\phi(\theta, \tau) = \phi^*(-\theta, -\tau). \quad (4.15)$$

Time and Frequency Shifts: If we translate the signal by an amount t_0, we expect the whole distribution to be translated by the same amount. Letting $s(t) \to s_{t_0}(t) = s(t + t_0)$ and substituting in Eq. (4.1), we have

$$P_{t_0}(t, \omega) = \frac{1}{4\pi^2} \iiint e^{-j\theta t - j\tau\omega + j\theta u} \phi(\theta, \tau) \cdot s^*\left(u - \frac{1}{2}\tau + t_0\right) s\left(u + \frac{1}{2}\tau + t_0\right) d\theta\, d\tau\, du \quad (4.16)$$

$$= \frac{1}{4\pi^2} \iiint e^{-j\theta t - j\tau\omega + j\theta(u - t_0)} \phi(\theta, \tau) \cdot s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) d\theta\, d\tau\, du \quad (4.17)$$

$$= \frac{1}{4\pi^2} \iiint e^{-j\theta(t + t_0) - j\tau\omega + j\theta u} \phi(\theta, \tau) \cdot s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) d\theta\, d\tau\, du \quad (4.18)$$

$$= P(t + t_0, \omega). \quad (4.19)$$

Hence a shift in the signal produces a corresponding shift in the distribution. We note that the proof requires that the kernel be independent of time and frequency. A similar result holds in the frequency domain, that is, if we shift the spectrum by a fixed frequency, then the distribution is shifted by the same amount. If

$$S(\omega) \to S(\omega + \omega_0) \quad \text{or} \quad s(t) \to s(t) e^{-j\omega_0 t} \quad (4.20a)$$

then

$$P(t, \omega) \to P(t, \omega + \omega_0). \quad (4.20b)$$

Global and Local Quantities: If we have a function $g(t, \omega)$ of time and frequency, its global average is

$$\langle g(t, \omega) \rangle = \iint g(t, \omega)\, P(t, \omega)\, d\omega\, dt. \quad (4.21)$$

The local or mean conditional value, the average of $g(t, \omega)$ at a particular time, is

$$\langle g(t, \omega) \rangle_t = \frac{1}{P_1(t)} \int g(t, \omega)\, P(t, \omega)\, d\omega \quad (4.22)$$

where $P_1(t)$ is the density in time,

$$P_1(t) = \int P(t, \omega)\, d\omega \tag{4.23}$$

and is equal to $|s(t)|^2$ if Eq. (4.10) is satisfied. Similar equations apply for the expectation value of a function at a particular frequency.

Mean Conditional Frequency and Instantaneous Frequency: The local or mean conditional frequency is given by

$$\langle \omega \rangle_t = \frac{1}{P_1(t)} \int \omega P(t, \omega)\, d\omega. \tag{4.24}$$

We have avoided using the term "instantaneous" frequency for reasons to be discussed shortly. A straightforward calculation leads to

$$\langle \omega \rangle_t P_1(t) = \frac{1}{2\pi} \iint A^2(t) \left[\phi(\theta, 0)\, \varphi'(u) - j \left. \frac{\partial \phi(\theta, \tau)}{\partial \tau} \right|_{\tau=0} \right] \cdot e^{j\theta(u-t)}\, d\theta\, du \tag{4.25}$$

where the signal has been expressed in terms of its amplitude and phase,

$$s(t) = A(t) e^{j\varphi(t)}. \tag{4.26}$$

For the product kernels this becomes

$$\langle \omega \rangle_t P_1(t) = \phi(0)\, A^2(t)\, \varphi'(t) + 2\phi'(0)\, A(t)\, A'(t) \tag{4.27}$$

where the primes denote differentiation.

We must now face the question as to what we want this to be equal to. First we note that if we take

$$\phi(\theta, 0) = 1, \qquad \left. \frac{\partial \phi(\theta, \tau)}{\partial \tau} \right|_{\tau=0} = 0 \tag{4.28}$$

in Eq. (4.25) or

$$\phi(0) = 1, \qquad \phi'(0) = 0 \tag{4.29}$$

in Eq. (4.27), we then have $P_1(t)$ equaling $A^2(t)$ and we obtain

$$\langle \omega \rangle_t = \varphi'(t) \tag{4.30}$$

a pleasing result reminiscent of the usual definition of instantaneous frequency. But it is not. Instantaneous frequency is the derivative of the phase if the analytic signal is used (see Section VIII). Equation (4.30) is true for *any* signal. Moreover, even though instantaneous frequency is meaningfully defined for certain types of signals [79], [153], [168], [191], this result is for all signals. It has been speculated [21], [56] that this indicates a method for a general definition of instantaneous frequency which will hold under all circumstances. However, considerably more work has to be done to fully develop the concept. Conversely, Boashash [26] has argued that since Eq. (4.30) corresponds to the instantaneous frequency only when the analytic signal is used, we should always use the analytic signal in these distributions. These issues are discussed at greater length in Section VIII.

Correlation Coefficient and Covariance: The covariance and the correlation coefficient very often afford much insight into the relationship between two variables. An application of this is given in Section V, where we apply these ideas to the Wigner distribution. For quasi-distributions the covariance was considered first by Cartwright [46]. The covariance is defined as

$$\text{Cov}\,(t\omega) = \langle t\omega \rangle - \langle t \rangle \langle \omega \rangle \tag{4.31}$$

and the correlation coefficient by

$$r = \frac{\text{Cov}\,(t\omega)}{\sigma_t \sigma_\omega} \tag{4.32}$$

where σ_t and σ_ω are the duration and the bandwidth of a signal as usually defined [Eq. (8.10)]. The simplest way to calculate $\langle t\omega \rangle$ is to use Eq. (3.29) with Eq. (3.77),

$$\langle t\omega \rangle = -\left. \frac{\partial^2 M(\theta, \tau)}{\partial\theta\, \partial\tau} \right|_{\theta, \tau = 0} = \int t\varphi'(t)\, A^2(t)\, dt. \tag{4.33}$$

This is an interesting relation because the derivative of the phase is acting as the frequency. We should emphasize that the covariance and the correlation coefficient, as used here, do not always have the same behavior as their standard counterparts because the distribution is not necessarily positive definite.

Spread and Second Conditional Moment: Having obtained a reasonable result for the mean frequency at a particular time, it is natural to ask for the spread or broadness of frequency for that time. This was done by Claasen and Mecklenbrauker [54] for the Wigner distribution case in the signal analysis context, and by others in the quantum mechanical context [82]. Unfortunately difficulties arise, as we shall see. First consider the second conditional moment

$$\langle \omega^2 \rangle_t = \frac{1}{P_1(t)} \int \omega^2 P(t, \omega)\, d\omega. \tag{4.34}$$

The calculation of this quantity is important for many reasons. In quantum mechanics it is particularly relevant because it corresponds to the local kinetic energy. It has been considered by a number of people who have proposed different expressions for it. It was subsequently shown that these different expressions are particular realizations of different distributions. We give the results for the product kernels [123],

$$\langle \omega^2 \rangle_t = \frac{1}{2}[1 + 4\phi''(0)] \left[\frac{A'(t)}{A(t)} \right]^2 - \frac{1}{2}[1 - 4\phi''(0)] \frac{A''(t)}{A(t)} + \varphi'^2(t). \tag{4.35}$$

Even though in general the second conditional moment should be manifestly positive, that is not the case with most of the distributions, including the Wigner distribution. This makes the usual interpretation of these quantities impossible. However, as we will see below, there are distributions for which the second conditional moment and the variance are manifestly positive.

Now consider the spread of the mean frequency at a given time,

$$\langle \sigma_\omega^2 \rangle_t = \int (\omega - \langle \omega \rangle_t)^2 P(t, \omega)\, d\omega = \langle \omega^2 \rangle_t - \langle \omega \rangle_t^2. \tag{4.36}$$

Using Eqs. (4.35) and (4.29) we have

$$\langle \sigma_\omega^2 \rangle_t = \frac{1}{2}[1 + 4\phi''(0)] \left[\frac{A'(t)}{A(t)} \right]^2 - \frac{1}{2}[1 - 4\phi''(0)] \frac{A''(t)}{A(t)}. \tag{4.37}$$

As before, this expression will become negative for most bilinear distributions and therefore cannot be interpreted as a variance. However, consider the choice [123]

$$\phi''(0) = \tfrac{1}{4}. \tag{4.38}$$

Then the spread becomes

$$\langle \sigma_\omega^2 \rangle_t = \left[\frac{A'(t)}{A(t)} \right]^2$$

$$= \frac{1}{|s(t)|^2} \left| \left[\frac{1}{j} \frac{d}{dt} - \varphi'(t) \right] s(t) \right|^2 \tag{4.39}$$

which is manifestly positive. Also, for this case,

$$\langle \omega^2 \rangle_t = \left[\frac{A'(t)}{A(t)} \right]^2 + \varphi'^2(t) \tag{4.40}$$

which is also manifestly positive. There are an infinite number of distributions having this property since there are an infinite number of kernels with the same second derivative at zero.

Group Delay: Suppose we focus on the frequency band around the frequency ω' and assume that the phase of the spectrum is a slowly varying function of frequency so that a good approximation to it, around point ω', is a linear one [153], [168],

$$\psi(\omega) \approx \psi(\omega') + (\omega - \omega')\psi'(\omega') \tag{4.41}$$

where $\psi(\omega)$ is the phase of the spectrum,

$$S(\omega) = |S(\omega)| e^{j\psi(\omega)}. \tag{4.42}$$

If we consider a signal that is made up of the original spectrum concentrated only around the frequencies ω', then we have the corresponding signal

$$s_{\omega'}(t) \approx \frac{1}{\sqrt{2\pi}} \int S(\omega - \omega') e^{j[\psi(\omega') + (\omega - \omega')\psi'(\omega')]} \cdot e^{j\omega t} \, d\omega. \tag{4.43}$$

We now write the spectrum in terms of the original signal as given by Eq. (1.3),

$$s_{\omega'}(t) \approx \frac{1}{2\pi} \iint s(t') e^{j[\psi(\omega') + (\omega - \omega')\psi'(\omega')]} \cdot e^{j\omega t - j(\omega - \omega')t'} \, d\omega \, dt' \tag{4.44a}$$

$$= e^{j[\psi(\omega') - \omega'\psi'(\omega')]} \int s(t') e^{j\omega' t'} \cdot \delta[-t' + \psi'(\omega') + t] \, dt'. \tag{4.44b}$$

Therefore

$$s_{\omega'}(t) \approx s[t + \psi'(\omega')] e^{j\psi(\omega') + j\omega' t}$$

$$= s[t + \psi'(\omega')] e^{j\omega'[t' + \psi(\omega')/\omega']}. \tag{4.45}$$

Hence the envelope of the signal at frequency ω' is delayed by $-\psi(\omega')$, and the phase is delayed by $-\psi(\omega')/\omega'$. The delay of the envelope is called the group delay, which we now write for an arbitrary function ω,

$$t_g = -\psi'(\omega). \tag{4.46}$$

From the point of view of joint time-frequency distributions we may think of the group delay as the mean time at a given frequency. Now by virtue of the symmetry between Eqs. (4.1) and (4.3) everything we have done for the expectation value of frequency at a given time allows us to readily write down the corresponding results for the expectation value of time at a given frequency.

In particular,

$$\langle t \rangle_\omega = -\psi'(\omega) \tag{4.47a}$$

if the kernel is chosen such that

$$\phi(0, \tau) = 1, \qquad \left. \frac{\partial \phi(\theta, \tau)}{\partial \theta} \right|_{\theta = 0} = 0. \tag{4.47b}$$

For the case of product kernels the conditions are the same as that given by Eq. (4.29).

Transformation of Signal and Distribution: In Table 2 we list the transformation properties of the distribution and the characteristic function for simple transformations of the signal.

Range of Distribution: If a signal is zero in a certain range, we would expect the distribution to be zero, but that is not ture for all distributions. It can be seen by inspection that the Rihaczek distribution is always zero when the signal is zero. This is not the case for the Wigner distribution. The general question of when a distribution is zero has not been fully investigated. Claasen and Mecklenbrauker [56] have derived the following condition for determining whether a distribution is zero before a signal starts and after it ends:

$$\int \phi(\theta, \tau) e^{-j\theta t} \, d\theta = 0 \qquad \text{for } |\tau| < 2|t|. \tag{4.48}$$

Even when this holds, it is not necessarily the fact that the distribution is zero in regions where the signal is zero.

Real and Imaginary Parts of Distributions: If a complex distribution satisfies the marginals, then so do the complex conjugate and the real part. The imaginary part indeed must

Table 2 Transformation Properties of Distributions and Characteristic Functions for Transformations of the Signal

Transformation	Signal $s(t)$	Distribution $P(t, \omega)$	Characteristic Function $M(\theta, \tau)$	Kernel $\phi(\theta, \tau)$	Product Kernel $\phi(\theta\tau)$
Time shift	$s(t + t_0)$	$P(t + t_0, \omega)$	$M(\theta, \tau)\, e^{j\theta t_0}$	Any	Any
Frequency shift	$S(\omega + \omega_0)$ or $s(t)\, e^{-j\omega_0 t}$	$P(t, \omega + \omega_0)$	$M(\theta, \tau)\, e^{-j\tau\omega_0}$	Any	Any
Time scaling	$\sqrt{\lvert\alpha\rvert}\, s(\alpha t)$	$P\left(\alpha t, \dfrac{\omega}{\alpha}\right)$	$M\left(\dfrac{\theta}{\alpha}, \alpha\tau\right)$	$\phi\left(\dfrac{\theta}{\alpha}, \alpha\tau\right) = \phi(\theta, \tau)$	Any
Frequency scaling	$\sqrt{\lvert\beta\rvert}\, S(\beta\omega)$ or $\dfrac{1}{\sqrt{\lvert\beta\rvert}}\, s\left(\dfrac{t}{\beta}\right)$	$P\left(\dfrac{t}{\beta}, \beta\omega\right)$	$M\left(\beta\theta, \dfrac{\tau}{\beta}\right)$	$\phi\left(\beta\theta, \dfrac{\tau}{\beta}\right) = \phi(\theta, \tau)$	Any
Time inversion	$s(-t)$	$P(-t, -\omega)$	$M(-\theta, -\tau)$	$\phi(-\theta, -\tau) = \phi(\theta, \tau)$	Any
Complex conjugate	$s^*(t)$	$P(t, -\omega)$	$M(\theta, -\tau)$	$\phi(\theta, -\tau) = \phi(\theta, \tau)$	$\phi(-x) = \phi(x)$

integrate to zero for each variable, that is,

$$\text{Im} \int P(t, \omega)\, dt = 0, \qquad \text{Im} \int P(t, \omega)\, d\omega = 0. \tag{4.49}$$

The complex conjugate distribution

$$P^*(t, \omega) \quad \text{has the kernel} \quad \phi^*(-\theta, -\tau) \tag{4.50}$$

and the real part distribution

$$\text{Re}\, P(t, \omega) \quad \text{has the kernel} \quad \tfrac{1}{2}[\phi(\theta, \tau) + \phi^*(-\theta, -\tau)]. \tag{4.51}$$

For example, using Eq. (4.51) we see that the kernel of the real part of the Rihaczek distribution is $\cos \frac{1}{2}\theta\tau$.

Example: To illustrate how readily one can determine the general properties of a distribution by a simple inspection of the kernel, we use as an example the distribution of Choi and Williams [51]. The Choi–Williams kernel is

$$\phi_{CW}(\theta, \tau) = e^{-\theta^2\tau^2/\sigma} \tag{4.52}$$

and we see that it is a product kernel,

$$\phi_{CW}(x) = e^{-x^2/\sigma} \tag{4.53}$$

where $x = \theta\tau$. From Eqs. (4.10) and (4.12) it is readily seen that the marginals are satisfied, that the mean frequency is the derivative of the phase [verified using Eq. (4.29)], that the shift properties are automatically satisfied, and that all properties and transformations in Table 2 are satisfied since it is a product kernel.

B. Inversion and Representability

To obtain the signal from a distribution we take the inverse Fourier transform of Eq. (4.1) and obtain

$$s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) = \frac{1}{2\pi} \iiint \frac{P(x, \omega)}{\phi(\theta, \tau)} e^{j\theta x + j\tau\omega - j\theta u}\, dx\, d\omega\, d\theta \tag{4.54}$$

or

$$s^*(t')\, s(t) = \frac{1}{2\pi} \iiint \frac{P(x, \omega)}{\phi(\theta, t - t')} \cdot e^{j(t-t')\omega + j\theta[x - (t+t')/2]}\, dx\, d\omega\, d\theta. \tag{4.55}$$

By taking a particular value of t', for instance zero, we have

$$s(t) = \frac{1}{2\pi s^*(0)} \iiint \frac{P(x, \omega)}{\phi(\theta, t)} e^{jt\omega + j\theta(x - t/2)}\, dx\, d\omega\, d\theta. \tag{4.56}$$

Hence the signal can be recovered from the distribution up to a constant phase.

These equations can be written in terms of the generalized characteristic function [58], [63] as defined by Eq. (3.58),

$$s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) = \frac{1}{2\pi} \int \frac{M(\theta, t - t')}{\phi(\theta, \tau)} e^{-j\theta u}\, d\theta \tag{4.57}$$

or

$$s^*(t')\, s(t) = \frac{1}{2\pi} \int \frac{M(\theta, t)}{\phi(\theta, t - t')} e^{-j\theta(t+t')/2}\, d\theta \tag{4.58}$$

or

$$s(t) = \frac{1}{2\pi s^*(0)} \int \frac{M(\theta, \tau)}{\phi(\theta, t)} e^{-j\theta t/2}\, d\theta. \tag{4.59}$$

The preceding relations can be used to determine whether a signal exists that will generate a given $P(t, \omega)$. We call such distributions representable or realizable. A necessary and sufficient [61] condition for representability is that the right-hand side of Eq. (4.55) or Eq. (4.58) result in a product form as indicated by the left-hand side.

Nuttall [148] has made an important contribution regarding the reversibility problem. He has been able to characterize the distributions from which the signal can be recovered uniquely. We note that from a given distribution the characteristic function $M(\theta, \tau)$ can always be determined uniquely since it is the Fourier transform of the distribution as defined by Eq. (3.30). However, to obtain the signal one has to divide the characteristic function by the kernel $\phi(\theta, \tau)$, which may be zero for some values of θ and τ. Nuttall [148] has shown that the signal can be recovered uniquely if the kernel has only isolated zeros. The number of zeros can be infinite, or the kernel may be zero along a line in the θ, τ plane. However, if the kernel is zero in a region of nonzero area, then the signal cannot be recovered. The basic reason is that for isolated zeros the ratio $M(\theta, t)/\phi(\theta, t)$ can be obtained by taking limits at the points where the kernel is zero. However, if the kernel is zero in a region, then the ratio is undefined.

C. Relations Between Distributions

Many special cases relating one particular distribution to another have been given in the literature. A general relationship between any two distributions can be derived readily [61].

Having such a connection allows the derivation of results for a new distribution in a simple way if the answers are already known for another distribution. In addition it clarifies the relation between distributions.

Suppose we have two distributions P_1 and P_2 with corresponding kernels ϕ_1 and ϕ_2. Their characteristic functions are

$$M_1(\theta, \tau) = \phi_1(\theta, \tau) \int e^{j\theta u} s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) du \tag{4.60}$$

$$M_2(\theta, \tau) = \phi_2(\theta, \tau) \int e^{j\theta u} s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) du \tag{4.61}$$

and dividing one by the other we have

$$M_1(\theta, \tau) = \frac{\phi_1(\theta, \tau)}{\phi_2(\theta, \tau)} M_2(\theta, \tau). \tag{4.62}$$

Taking the Fourier transform to obtain the distribution we have

$$P_1(t, \omega) = \frac{1}{4\pi^2} \iiiint \frac{\phi_1(\theta, \tau)}{\phi_2(\theta, \tau)} e^{j\theta(t'-t) + j\tau(\omega'-\omega)} \cdot P_2(t', \omega')\, d\theta\, d\tau\, dt'\, d\omega'. \tag{4.63}$$

It is sometimes convenient to write this as

$$P_1(t, \omega) = \iint g(t' - t, \omega' - \omega) P_2(t', \omega')\, dt'\, d\omega' \quad (4.64)$$

with

$$g(t, \omega) = \frac{1}{4\pi^2} \iint e^{j\theta t + j\tau\omega} \frac{\phi_1(\theta, \tau)}{\phi_2(\theta, \tau)}\, d\theta\, d\tau. \quad (4.65)$$

A very useful way to express Eq. (4.64) is in operator form. We note the general theorem [60]

$$\frac{1}{4\pi^2} \iiiint G(\theta, \tau)\, e^{j\theta(t-t') + j\tau(\omega-\omega')} H(t', \omega')\, d\theta\, d\tau\, dt'\, d\omega' = G\left(-j\frac{\partial}{\partial t}, -j\frac{\partial}{\partial \omega}\right) H(t, \omega) \quad (4.66)$$

which, when applied to Eq. (4.63), gives

$$P_1(t, \omega) = \frac{\phi_1\left(j\frac{\partial}{\partial t}, j\frac{\partial}{\partial \omega}\right)}{\phi_2\left(j\frac{\partial}{\partial t}, j\frac{\partial}{\partial \omega}\right)} P_2(t, \omega). \quad (4.67)$$

D. Other Topics

Mean Values of Time-Frequency Functions: We have already defined and used global and local expectation values. There exists a general relationship between averages and correspondence rules which has theoretical interest and is very often the best way to calculate global averages. One can show [58] that the expectation value calculated in the usual way

$$\langle g(t, \omega) \rangle = \int g(t, \omega)\, P(t, \omega)\, d\omega\, dt \quad (4.68)$$

can be short-circuited by calculating instead

$$\langle g(t, \omega) \rangle = \int s^*(t)\, \mathbf{G}(\mathfrak{T}, \mathfrak{W})\, s(t)\, dt \quad (4.69)$$

if the correspondence between g and $\mathbf{G}$ is given by Eq. (3.55).

Bilinear Transformations: A general bilinear transformation may be written in the form

$$P(t, \omega) = \iint K(t, \omega; x, x')\, s^*(x)\, s(x')\, dx\, dx' \quad (4.70)$$

as has been done by Wigner [200], Kruger and Poffyn [114], and others [90], [130], [170]. By requiring the distribution to satisfy desirable properties, Wigner obtained the conditions to be imposed on K. If we require that the distribution be time-shift invariant, then it can be shown [114], [200] that K must be a function of $t - x$ and $t - x'$, or equivalently a function of $2t - x - x'$ and $x - x'$. In addition if the distribution is to be frequency-shift invariant, then K must be of the form

$$K(t, \omega; x, x') = e^{j(x-x')\omega} K(t, 0; x, x'). \quad (4.71)$$

Hence kernels which yield time- and frequency-shift invariant distributions are of the form [114], [200], [130]

$$K(t, \omega; x, x') = e^{j(x-x')\omega} K_0(2t - x - x', x - x') \quad (4.72)$$

where the new kernel K_0 is only a function of two variables. By comparing Eq. (4.70) with Eq. (2.4) we have

$$K(t, \omega; x, x') = \frac{e^{-j(x'-x)\omega}}{4\pi^2} \int e^{-j\theta(2t-x-x')/2} \cdot \phi(\theta, x' - x)\, d\theta \quad (4.73)$$

or

$$K_0(t, \tau) = \frac{1}{4\pi^2} \int e^{-j\theta t/2} \phi(\theta, -\tau)\, d\theta. \quad (4.74)$$

We note that $K_0(t, \tau)$ is essentially $r(t, \tau)$, as defined in Eq. (3.69).

Additional conditions imposed on the distribution are reflected as constraints [90], [114], [170], [200], [130] on K in the same manner that we have imposed constraints on $f(\theta, \tau)$. However, as shown by Kruger and Poffyn [114], the contraints $f(\theta, \tau)$ are much simpler to formulate and express as compared to those on K, and that is why Eq. (2.4) is easier to work with than Eq. (4.70). For example, the time- and frequency-shift invariant requirement is imposed by simply requiring that $\phi(\theta, \tau)$ not be a function of time and frequency.

Characteristic Functions and Moments: We have seen in Section III that characteristic functions are a powerful way to derive distributions. The characteristic function is closely related to the ambiguity function (see Section VI). We would like to emphasize that, in addition, characteristic functions are often the most effective way of studying distributions. For example, consider the transformation property of the characteristic function as given by Eq. (4.62) and compare it to the transformation property of the distributions as given by Eq. (4.63)

We also point out the relationship of the generalized characteristic functions and the generalized autocorrelation function. Comparing Eq. (3.67) with Eq. (3.77) we see that

$$R_t(\tau) = \frac{1}{2\pi} \int M(\theta, \tau) e^{-j\theta t}\, d\theta. \quad (4.75)$$

This relation can be used to derive the transformation properties of the autocorrelation function. If $R_t^{(1)}$ and $R_t^{(2)}(\tau)$ are the autocorrelation functions corresponding to two different distributions, then we have

$$R_t^{(1)}(\tau) = \frac{1}{2\pi} \int M_1(\theta, \tau) e^{-j\theta t}\, d\theta \quad (4.76)$$

$$= \frac{1}{2\pi} \int \frac{\phi_1(\theta, \tau)}{\phi_2(\theta, \tau)} M_2(\theta, \tau) e^{-j\theta t}\, d\theta \quad (4.77)$$

where we have used Eq. (4.62). Writing the characteristic function in terms of the autocorrelation function we have

$$R_t^{(1)}(\tau) = \frac{1}{2\pi} \int g_\tau(t - t')\, R_{t'}^{(2)}(\tau)\, dt' \quad (4.78)$$

where

$$g_\tau(t) = \frac{1}{2\pi} \int \frac{\phi_1(\theta, \tau)}{\phi_2(\theta, \tau)} e^{-j\theta t}\, d\theta. \quad (4.79)$$

Using the characteristic function, we can also obtain the transformation of mixed moments. Using Eqs. (3.29) and (4.61),

$$\langle t^n \omega^m \rangle_1 = \frac{1}{j^n j^m} \frac{\partial^{n+m}}{\partial\theta^n \partial\tau^m} \frac{\phi_1(\theta, \tau)}{\phi_2(\theta, \tau)} M_2(\theta, \tau)\bigg|_{\theta,\tau=0}. \quad (4.80)$$

Some straightforward manipulation yields

$$\langle t^n \omega^m \rangle_1 = \sum_{l=0}^{n} \sum_{k=0}^{m} a_{l,k}^{n-l,m-k} \langle t^l \omega^k \rangle_2 \tag{4.81}$$

where

$$a_{l,k}^{n-l,m-k} = \frac{1}{j^{n+m-l-k}} \binom{n}{l} \binom{m}{k} \cdot \frac{\partial^{n+m-l-k}}{\partial \theta^l \partial \tau^k} \frac{\phi_1(\theta, \tau)}{\phi_2(\theta, \tau)} \bigg|_{\theta, \tau = 0}. \tag{4.82}$$

These are very convenient relations for obtaining the moments of a distribution if one has already found them for another one.

V. Wigner Distribution

The Wigner distribution was the first to be proposed and is certainly the most widely studied and applied. The discovery of its strengths and weaknesses has been a major thrust in the development of the field. It can be obtained from the general class by taking

$$\phi_W(\theta, \tau) = 1. \tag{5.1}$$

The Wigner distribution is

$$W(t, \omega) = \frac{1}{2\pi} \int s^*\left(t - \frac{1}{2}\tau\right) e^{-j\tau\omega} s\left(t + \frac{1}{2}\tau\right) d\tau \tag{5.2}$$

and in terms of the spectrum, it is

$$W(t, \omega) = \frac{1}{2\pi} \int S^*\left(\omega + \frac{1}{2}\theta\right) e^{-jt\theta} S\left(\omega - \frac{1}{2}\theta\right) d\theta. \tag{5.3}$$

A. General Properties

Because the kernel is equal to 1, the properties of the Wigner distribution are readily determined. Using the general equations of Section IV, we see that the Wigner distribution satisfies the marginals, that it is real, and that time and frequency shifts in the signal produce corresponding time and frequency shifts in the distribution. Since the kernel is a product kernel, all the transformation properties of Table 2 in Section IV are seen to be satisfied.

The inversion properties are easily obtained by specializing Eqs. (4.54)–(4.56) for the case of $\phi = 1$,

$$s^*\left(t - \frac{1}{2}\tau\right) s\left(t + \frac{1}{2}\tau\right) = \int W(t, \omega) e^{j\tau\omega} \, d\omega \tag{5.4}$$

$$s^*(t') s(t) = \int W\left[\frac{1}{2}(t' + t), \omega\right] e^{j(t-t')\omega} \, d\omega \tag{5.5}$$

$$s(t) = \frac{1}{s^*(0)} \int W\left(\frac{1}{2}t, \omega\right) e^{jt\omega} \, d\omega. \tag{5.6}$$

Mean Local Frequency, Group Delay, and Spread: Since the kernel for the Wigner distribution is 1, we have from Eq. (4.28)

$$\langle \omega \rangle_t = \varphi'(t) \quad \text{if} \quad s(t) = A(t) e^{j\varphi(t)}. \tag{5.7}$$

From Eqs. (4.35) and (4.37), the local mean-squared frequency and the local standard deviation are given by

$$\langle \omega^2 \rangle_t = \frac{1}{2}\left[\frac{A'(t)}{A(t)}\right]^2 - \frac{1}{2}\frac{A''(t)}{A(t)} + \varphi'^2(t) \tag{5.8}$$

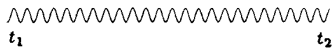
$$\langle \sigma_\omega^2 \rangle_t = \langle \omega^2 \rangle_t - \langle \omega \rangle_t^2 \tag{5.9}$$

$$= \frac{1}{2}\left[\frac{A'(t)}{A(t)}\right]^2 - \frac{1}{2}\frac{A''(t)}{A(t)} \tag{5.10}$$

a result obtained by Claasen and Mecklenbrauker [54]. As they have pointed out, it generally goes negative and cannot be interpreted properly.

B. Range of Wigner Distribution

From the functional relation to the signal one can develop some simple rules of thumb to ascertain the behavior of the Wigner distribution [59]. From Eq. (5.2) we see that for a particular time we are adding up pieces made from the product of the signal at a past time multiplied by the signal at a future time, the time into the past being equal to the time into the future. Therefore to see whether the Wigner distribution is zero at a point, one may mentally fold the part of the signal to the left over to the right and see whether there is any overlap. If so, the Wigner distribution will not be zero, otherwise it will. Now consider a finite-duration signal in the interval t_1 to t_2 as illustrated:



If we are any place left of t_1 and fold over the signal to the right, there will be no overlap since there is no signal to the left of t_1 to fold over. This will remain true up to the start of the signal at time t_1. Hence for finite-duration signals, the Wigner distribution is zero up to the start. This is a desirable feature since we should not have a nonzero value for the distribution if the signal is zero. At any point to the right of t_1 but less than t_2, the folding will result in an overlap. Similar arguments hold for points to the right of t_2. Therefore for a time-limited signal, the Wigner distribution is zero before the signal starts and after the signal ends, that is,

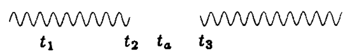
$$W(t, \omega) = 0 \quad \text{for } t \le t_1 \text{ or } t \ge t_2 \text{ if } s(t) \text{ is nonzero only in the range } (t_1, t_2). \tag{5.11}$$

Due to the similar structures of Eqs. (5.2) and (5.3), the same considerations apply to the frequency domain. If we have a band-limited signal, the Wigner distribution will be zero for all frequencies that are not included in that band,

$$W(t, \omega) = 0 \quad \text{for } \omega \le \omega_1 \text{ or } \omega \ge \omega_2 \text{ if } S(\omega) \text{ is nonzero only in the range } (\omega_1, \omega_2). \tag{5.12}$$

These properties are sometimes called the support properties of the Wigner distribution.

Now consider a signal of the following form:



where the signal is zero from t_2 to t_3, and focus on point t_a. Mentally folding the right and left parts of t_a it is clear that there will be an overlap, and hence the Wigner distribution is not zero even though the signal is. In general the Wigner distribution is not zero when the signal is zero, and this causes considerable difficulty in interpretation. In speech, for example, there are silences which are important, but the Wigner distribution masks them. These spurious values can be cleaned up by smoothing, but smoothing destroys some other desirable properties of the Wigner distribution.

Range of Cross Terms: Similar considerations apply to the cross Wigner distribution. In particular, for any two functions $s_1(t)$ and $s_2(t)$ which are zero outside the intervals (t_1, t_2) and (t_3, t_4), respectively, the cross Wigner distribution is zero for the ranges indicated [60], [92a], [92b]

$$W_{12}(t, \omega) = \frac{1}{2\pi} \int s_1^*\left(t - \frac{1}{2}\tau\right) e^{-j\tau\omega} s_2\left(t + \frac{1}{2}\tau\right) d\tau = 0$$

$$\text{if } t \le \frac{1}{2}(t_1 + t_3), \quad t \ge \frac{1}{2}(t_2 + t_4). \quad (5.13)$$

In the frequency domain we have that for two band-limited functions $S_1(\omega)$ and $S_2(\omega)$, which are zero outside the intervals (ω_1, ω_2) and (ω_3, ω_4), respectively,

$$W_{12}(t, \omega) = \frac{1}{2\pi} \int S_1^*\left(\omega - \frac{1}{2}\theta\right) e^{j\theta t} S_2\left(\omega + \frac{1}{2}\theta\right) d\theta = 0$$

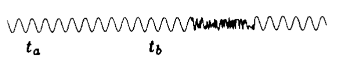
$$\text{if } \omega \le \frac{1}{2}(\omega_1 + \omega_3), \quad \omega \ge \frac{1}{2}(\omega_2 + \omega_4). \quad (5.14)$$

These relationships are useful in calculating the Wigner distribution for finite-duration signals [60], [92a], [92b].

C. Propagation of Characteristics (e.g., Noise)

Almost every worker who has applied the Wigner distribution has noticed that it is "noisy." Indeed we will show that in general if there is noise for a small finite time of the signal, that noise will "appear" at other times, and if the signal is infinite, then it will appear for all time. This effect is a general property of the Wigner distribution that must be fully understood if one is to develop a feeling for its behavior. The important point to realize is that the Wigner distribution at a particular time generally reflects properties that the signal has at other times because the Wigner distribution is *highly* nonlocal.

Consider a finite-duration signal, illustrated below, where we have indicated by the wavy lines any local characteristic, but which for convenience we shall call noise:



Suppose we calculate the Wigner distribution at a time where this characteristic does not appear, say, at point t_a. Folding over the signal, we see that there is no overlap of the left part of the signal with the noise in the signal, and hence noise will not appear in the distribution at time t_a. Now consider point t_b. The overlap will include the noise in the signal and therefore noise will appear in the distribution, even though there is no noise in the signal at that time. Now considering a signal that goes from minus infinity to plus infinity, noise will appear *everywhere* since for any point we choose, folding over the signal about that point will always have an intersection with the noise in the signal. In Fig. 6 we give an example for a finite-duration signal. Noise appears for times in between the arrows, even though the noise in the signal was of shorter duration.

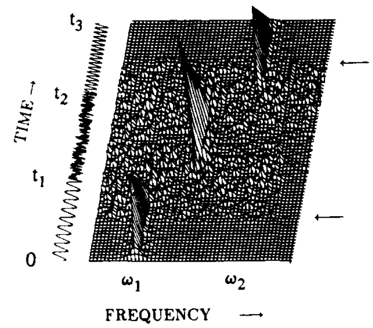


Fig. 6. Example illustrating the "propagation" of noise in the Wigner distribution. Noise (between arrows) appears in the Wigner distribution at times for which there is no noise in the signal. If the signal was of infinite duration, noise would appear for all time in the Wigner distribution, although it was of finite duration in the signal.

D. Examples

Example 1: For signals of the form

$$s(t) = \left(\frac{\alpha}{\pi}\right)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \quad (5.15)$$

the Wigner distribution is

$$W(t, \omega) = \frac{1}{\pi} e^{-\alpha t^2 - (\omega - \beta t - \omega_0)^2/\alpha}. \quad (5.16)$$

The first thing we should note is that the Wigner distribution is positive, and this is the only signal for which it is positive [93], [157], [180]. If α is small, then the distribution is concentrated along the line $\omega = \omega_0 + \beta t$, which is the derivative of the phase and corresponds to the local mean frequency. In the extreme case where we have a chirp (i.e., $\alpha = 0$) the distribution becomes[5]

$$W(t, \omega) = \delta(\omega - \beta t - \omega_0) \quad \text{when } \alpha = 0 \quad (5.17)$$

which shows that the energy is totally concentrated along the instantaneous frequency. If we further take $\beta = 0$, then the distribution is peaked only at the carrier frequency,

$$W(t, \omega) = \delta(\omega - \omega_0) \quad \alpha, \beta = 0. \quad (5.18)$$

In Fig. 7 we plot the Wigner distribution to illustrate how it behaves as the Gaussian becomes more chirplike.

The correlation coefficient for this case gives a revealing answer. We find that

$$\langle t \rangle = 0, \quad \langle \omega \rangle = \omega_0, \quad \langle t\omega \rangle = \frac{\beta}{2\alpha}$$

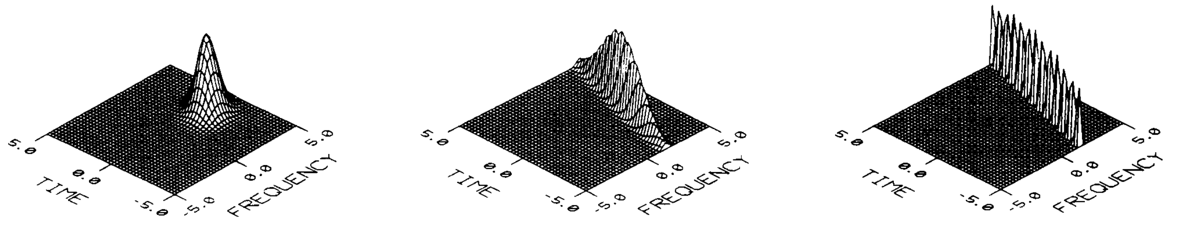
$$\sigma_t = \sqrt{\frac{1}{2\alpha}}, \quad \sigma_\omega = \sqrt{\frac{\alpha^2 + \beta^2}{2\alpha}}. \quad (5.19)$$

The covariance is therefore

$$\text{Cov}(t\omega) = \frac{\beta}{2\alpha}. \quad (5.20)$$

When $\beta \to 0$, the covariance goes to zero, which implies that we have no correlation between time and frequency. That is reasonable because we have a pure sine wave for all time. As $\alpha \to 0$, the covariance goes to infinity and we have total correlation, which is also reasonable since a chirp forces total dependence between time and frequency. Similar considerations apply to the correlation coefficient as

[5] One must be careful in taking the limit because the signal can no longer be normalized. The normalizing factor is omitted when calculating Eq. (5.17).



(a) (b) (c)

Fig. 7. Wigner distribution for Gaussian signal $s(t) = Ae^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t}$. As $\alpha \to 0$, making the signal more chirplike, the Wigner distribution becomes concentrated around the instantaneous frequency $\omega = \omega_0 + \beta t$. $\beta = 0.3$, $\omega_0 = 3$. (a) $\alpha = 1$. (b) $\alpha = 0.5$. (c) $\alpha = 0.1$. Time and frequency variables are plotted from -5 to 5 units.

defined by Eq. (4.23),

$$r = \frac{1}{\sqrt{1 + (\alpha/\beta)^2}}. \tag{5.21}$$

The correlation coefficient goes to 1 as $\alpha \to 0$, which indicates perfect correlation, and it goes to zero as $\alpha \to \infty$, which implies no correlation.

Example 2: We take

$$s(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t}. \tag{5.22}$$

The Wigner distribution is

$$W(t, \omega) = A_1^2 \delta(\omega - \omega_1) + A_2^2 \delta(\omega - \omega_2) + 2A_1A_2 \cdot \delta[\omega - \tfrac{1}{2}(\omega_1 + \omega_2)] \cos(\omega_2 - \omega_1)t. \tag{5.23}$$

Besides the concentration at ω_1 and ω_2 as expected, we also have non zero values at the frequency $\frac{1}{2}(\omega_1 + \omega_2)$. This is an illustration of the cross-term effect discussed previously. The point $\omega = \frac{1}{2}(\omega_1 + \omega_2)$ is the only point for which there is an overlap since the signal is sharp at both ω_1 and ω_2. This distribution is illustrated in Fig. 3(a). For the sum of sine waves, we will always have a spurious value of the distribution at the midway point between any two frequencies. Hence for N sine waves we will have $\frac{1}{2}N(N - 1)$ spurious values. We note that the spurious terms oscillate and can be removed to some extent by smoothing.

Example 3: For the finite-duration signal

$$s(t) = e^{j\omega_0 t}, \qquad 0 \le t \le T \tag{5.24}$$

the Wigner distribution is

$$W(t, \omega) = \begin{cases} \dfrac{t}{\pi} \operatorname{sinc}(\omega - \omega_0)t, & 0 \le t \le \dfrac{T}{2} \\[2ex] \dfrac{T - t}{\pi} \operatorname{sinc}(\omega - \omega_0)(T - t), & \dfrac{T}{2} < t < T. \end{cases} \tag{5.25}$$

This is plotted in Fig. 2(a). As discussed, the Wigner distribution goes to zero at the end point of a finite-duration signal. Also, note the symmetry about the middle time, which is not the case with the Page distribution. The Wigner distribution treats past and present on an equal footing.

Example 4: For the sum of two Gaussians

$$s(t) = A_1\left(\frac{\alpha_1}{\pi}\right)^{1/4} e^{-\alpha_1 t^2/2 + j\beta_1 t^2/2 + j\omega_1 t} + A_2\left(\frac{\alpha_2}{\pi}\right)^{1/4} e^{-\alpha_2 t^2/2 + j\beta_2 t^2/2 + j\omega_2 t} \tag{5.26}$$

a straightforward calculation gives

$$W(t, \omega) = \frac{A_1^2}{\pi} e^{-\alpha_1 t^2 - (\omega - \beta_1 t - \omega_1)^2/\alpha_1} + \frac{A_2^2}{\pi} e^{-\alpha_2 t^2 - (\omega - \beta_2 t - \omega_2)^2/\alpha_2} + A_1A_2(\alpha_1\alpha_2)^{1/4} 4\sqrt{2} \operatorname{Re} \sqrt{\frac{1}{\gamma_2 + \gamma_1}}\, e^{-(\gamma_2 + \gamma_1)t^2/2 + j(\omega_2 - \omega_1)t} \cdot \exp \frac{\{\frac{1}{2}(\gamma_2 - \gamma_1)t + j[\omega - \frac{1}{2}(\omega_2 + \omega_1)]\}^2}{\frac{1}{2}(\gamma_2 + \gamma_1)} \tag{5.27}$$

where

$$\gamma_1 = \alpha_1 + j\beta_1, \qquad \gamma_2 = \alpha_2 - j\beta_2. \tag{5.28}$$

Fig. 4(a) illustrates such a case. The middle hump is again due to the cross-term effect.

E. Discrete Wigner Distribution

A significant advance in the use of the Wigner distribution was its formulation for discrete signals. A number of difficulties arise, but much progress has been made recently. A fundamental result was obtained by Claasen and Mecklenbrauker [53], [54], where they applied the sampling theorem to the Wigner distribution. They showed that the Wigner distribution for a band-limited signal is determined from the samples by

$$W(t, \omega) = \frac{T}{\pi} \sum_{k=-\infty}^{\infty} s^*(t - kT) e^{-2j\omega kT} s(t + kT) \tag{5.29}$$

where $1/T$ is the sampling frequency and must be chosen so that $T \le \pi/2\omega_{max}$, where ω_{max} is the highest frequency in the signal. The sampling frequency must be chosen so that

$$\omega_s \ge 4\omega_{max}. \tag{5.30}$$

For a discrete-time signal $s(n)$, with T equal to 1, this becomes

$$W(n, \theta) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} s^*(n - k) e^{-2j\theta k} s(n + k). \tag{5.31}$$

In the discrete case as given by Eq. (5.29) the distribution is periodic in ω, with a period of π rather than 2π, as in the continuous case. Hence the highest sampling frequency that must be used is twice the Nyquist rate. Chan [48] devised an alternative definition, which is periodic in π. Boashash and Black [25] have also given a discrete version of the Wigner distribution and argued that the use of the analytic signal eliminates the aliasing problem. They devised a real-time method of calculating the Wigner distribution and the analytic signal. Choi and Williams [51] devised a discrete version of the exponential distribution. An alternative approach has been used by Boudreaux-Bartels and Parks [38], [41], who approximated the Fourier integral using spline approximations.

A new and unified approach to the discrete Wigner distribution has recently been given by Peyrin and Prost [155], which naturally reduces to and preserves the properties of the continuous case. Starting with the relation for the sampled signal $\hat{s}(t)$ in terms of the continuous signal $s(t)$,

$$\hat{s}(t) = \sum_n s(nT)\delta(t - nT) \tag{5.32}$$

they substitute into the continuous Wigner distribution and obtain the discrete version for the time variable

$$\hat{W}(n, \omega) = \sum_k s^*((n - 2k)T)s(kT)e^{-j\omega(2k-n)T}. \tag{5.33}$$

Equation (5.33) retains the frequency as a continuous variable. The identical procedure can be used to derive the case when frequency is discrete and time is continuous. They also generalize to the case where both time and frequency are discrete. Their approach is general and can undoubtedly be applied to other distributions.

Amin [9] has given explicit recursion relations for the calculation of the discrete Wigner distribution and discrete smoothed Wigner distribution.

F. Smoothed Distributions

The major motivations for smoothing the Wigner distribution are that for certain types of smoothing, a positive distribution is obtained [34], [47], [66], and that many of the so-called artifacts illustrated before are suppressed. The fundamental idea is to smooth the Wigner distribution by the double convolution

$$W_S(t, \omega) = \int L(t - t', \omega - \omega')W(t', \omega')' \, dt' \, d\omega' \tag{5.34}$$

where L is a smoothing function. It is hoped that a judicious choice of L will result in a new distribution with desirable properties. We stress that if L is taken to be independent of the signal, then the only way to obtain a positive distribution is by sacrificing the marginals. The most common smoothing function used is a Gaussian,

$$L(t, \omega) = \frac{1}{\alpha\beta} e^{-t^2/\alpha - \omega^2/\beta} \tag{5.35}$$

and it has been known for a long time in the quantum literature [34], [47], [66], [118]–[120], [142], [181] that for certain values of α and β, a positive distribution is obtained. The condition is that [47], [66]

$$\alpha\beta \geq 1. \tag{5.36}$$

More general smoothing functions have been considered in Bertrand *et al.* [23], where they showed that a sufficient condition for positivity is that the smoothing function be a Wigner distribution of any normalized signal. Janssen [100] and Janssen and Claasen [101] considered smoothing of the general class of distributions, Eq. (4.1), with a Gaussian. In Section VI we discuss the spectrogram which can be viewed as a smoothing procedure.

Soto and Claverie [179] have given a concise summary of the effects of smoothing with a Gaussian for quantum mechanical distributions, and their results can readily be transcribed into signal analysis language. The global expectation values of frequency and time are the same for the Wigner and the smoothed Wigner distributions. That is not the case for the second global moments where we have $\langle\omega^2\rangle_S = \langle\omega^2\rangle_W + \frac{1}{2}\alpha$ and similarly for $\langle t^2\rangle$. The higher moments are also changed, which is a reflection that the marginals are not preserved. Soto and Claverie show that smoothing very often gives erroneous answers in the quantum mechanical case.

Garudadri *et al.* [81] have made important contributions in understanding the effects and limitations of smoothing. They show that smoothing causes loss of phase information. However, they show that partial smoothing can be of advantage.

Amin [8] has obtained conditions for the selection of time and lag window for smoothing the Wigner distribution and showed the relation with the autocorrelation function.

The recent work of Andrieux *et al.* [10] has made significant strides toward utilizing smoothing effectively. They consider optimal smoothing of the Wigner distribution to be that which preserves as much as possible of the basic characteristics of the Wigner distribution. They argue that the smoothing should involve regions of the time–frequency plane which are as small as possible and yet still lead to a positive distribution. They obtain general conditions for this minimum smoothing in terms of the rate of change of the phase for special signals of the form $s(t) = e^{j\varphi(t)}$.

Nuttall [146], [147] has considered smoothing with a more general Gaussian form,

$$L(t, \omega) = 2\sqrt{Q}e^{-t^2/\alpha - \omega^2/\beta - 2c\omega t} \tag{5.37a}$$

with

$$Q = \frac{1}{\alpha\beta} - c^2 \tag{5.37b}$$

and has shown that the resulting distribution will be positive if $Q \leq 1$, that is, if

$$\alpha\beta \geq \frac{1}{1 + c^2}. \tag{5.38}$$

He points out that this smoothing function need not be a Wigner distribution function of some signal, unless $Q = 1$. He has derived a number of interesting alternative expressions for the smoothed Wigner distribution and showed how smoothing can be optimized in advantageous ways.

Choi and Williams [51], Nuttall [146], [147], and Flandrin [76] have found the ambiguity function plane to be a very effective way of finding kernels.

Although we have discussed smoothing in the section on the Wigner distribution, this is really a distribution-independent process in the sense that smoothing a distribution with one smoothing function is equivalent to smoothing another distribution with a different smoothing function. The end result is the same and is a member of the bilinear class. In particular if P_1 is smoothed with the smoothing function $L_1(t', \omega'; t, \omega)$ to obtain P_{L_1},

$$P_{L_1}(t, \omega) = \iint P_1(t', \omega')L_1(t', \omega'; t, \omega) \, dt' \, d\omega' \tag{5.39a}$$

the same smoothed distribution can be obtained from distribution P_2 with smoothing function $L_2(t', \omega'; t, \omega)$,

$$P_{L_1}(t, \omega) = P_{L_2}(t, \omega) = \iint P_2(t', \omega') L_2(t', \omega'; t, \omega)\, dt'\, d\omega' \tag{5.39b}$$

if the smoothing functions are related by

$$L_2(t'', \omega''; t, \omega) = \iint g(t'' - t', \omega'' - \omega') \cdot L_1(t', \omega'; t, \omega)\, dt'\, d\omega' \tag{5.39c}$$

where $g(t, \omega)$ is defined by Eq. (4.65).

G. Other Properties and Results

Positivity: We have seen that the Wigner distribution is manifestly positive for the Gaussian signal, Eq. (5.15). In fact that is the only signal for which the Wigner distribution is positive, as was shown by Hudson [93] and Piquet [157]. Soto and Claverie [180] have proved it for the multidimensional case.

Modulation and Convolution: Consider the Wigner distribution of the product of two signals,

$$s(t) = s_1(t) s_2(t). \tag{5.40}$$

To write it in terms of the Wigner distribution of the individual signals, substitute the signal into Eq. (5.1) and use the inverse relations of Eq. (5.5) to obtain [54]

$$W(t, \omega) = \int W_1(t, \omega') W_2(t, \omega - \omega')\, d\omega'. \tag{5.41}$$

For the case of convolution where

$$s(t) = \int s_1(t') s_2(t - t')\, dt' \quad \text{or} \quad S(\omega) = S_1(\omega) S_2(\omega) \tag{5.42}$$

we immediately get (by symmetry) that

$$W(t, \omega) = \int W_1(t', \omega) W_2(t' - t, \omega)\, dt'. \tag{5.43}$$

Moyal Formula: An interesting relation exists between the overlap of two signals and the overlap of their Wigner distributions,

$$\left| \int s_1(t) s_2^*(t)\, dt \right|^2 = 2\pi \iint W_1(t, \omega) W_2(t, \omega)\, dt\, d\omega. \tag{5.44}$$

This was first shown by Moyal [143]. Janssen [98] showed that there are an infinite number of other distributions with this property. The requirement is that the kernel satisfy $|\phi(\theta, \tau)|^2 = 1$. In the case where this is not so we have

$$\left| \int s_1(t) s_2^*(t)\, dt \right|^2 = 2\pi \iiiint P_1(t, \omega) P_2^*(t', \omega') \cdot K(t - t', \omega - \omega')\, dt\, d\omega\, dt'\, d\omega' \tag{5.45}$$

where

$$K(t, \omega) = \frac{1}{4\pi^2} \iint \frac{e^{j\theta t + j\tau\omega}}{|\phi(\theta, \tau)|^2}\, d\theta\, d\tau. \tag{5.46}$$

Some have made Moyal's formula a requirement of distribution, but it is not clear why that should be so. As Janssen [98] pointed out, it has a certain appeal in quantum mechanics but is "perhaps not really necessary for signal analysis." In fact it is not really used in quantum mechanics either. Of course, the inner product is a fundamental quantity in signal analysis and quantum mechanics, and what one needs is a way to relate it to the respective distributions. Equation (5.45) does so and there is no particular reason why the relation has to be of the form given by Eq. (5.44). We note that Moyal's formula has been found to be useful in detection problems [78], [117], [31].

Performance in Noise: The behavior of the estimate of the Wigner distribution for a deterministic signal in zero mean additive stationary noise has been analyzed by Nuttall [146]. He obtains explicit expressions for the mean Wigner distribution (ensemble average over all possible realizations of the noise) and its variance. Since it is assumed that the noise is additive, the signal and noise process is

$$x(t) = s(t) + n(t). \tag{5.47}$$

As the noise term is stationary and does not decay to zero at infinity, $x(t)$ is weighted by a known deterministic function $v(t)$, which may be chosen advantageously depending on the circumstances. The weighted process is

$$y(t) = v(t) x(t) = v(t)[s(t) + n(t)]. \tag{5.48}$$

The Wigner distribution can therefore be written as the convolution, with respect to frequency, of the distribution of $v(t)$ with the distribution of $s(t) + n(t)$, as per Eq. (5.41). The Wigner distribution of $s(t) + n(t)$ consists of four terms, namely, the distribution of the signal, the distribution of the noise, and the two cross terms which are linear in the noise. The linear noise terms ensemble average to zero because we are dealing with zero mean noise. Therefore the mean Wigner distribution is

$$\overline{W}_y(t, \omega) = \int W_v(t, \omega')[W_s(t, \omega - \omega') + \overline{W}_n(t, \omega - \omega')]\, d\omega' \tag{5.49}$$

where overbars denote an ensemble average over all possible realizations of the noise.

The mean Wigner distribution of the noise can be simplified for stationary noise. Namely, the noise covariance $C(t)$ is a function of the difference of the times,

$$C(t' - t) = \overline{n(t) n^*(t')} \tag{5.50}$$

and hence

$$\overline{W}_n(t, \omega - \omega') = \frac{1}{2\pi} \int \overline{n^*\left(t - \frac{1}{2}\tau\right) n\left(t + \frac{1}{2}\tau\right)}\, e^{-j\tau(\omega - \omega')}\, d\tau \tag{5.51}$$

$$= \frac{1}{2\pi} \int C(\tau) e^{-j\tau(\omega - \omega')}\, d\tau$$

$$= G(\omega - \omega') \tag{5.52}$$

where $G(\omega)$ is the power density spectrum. The mean Wigner distribution of $y(t)$ is therefore

$$\overline{W}_y(t, \omega) = \int W_v(t, \omega')[W_s(t, \omega - \omega') + G(\omega - \omega')]\, d\omega'. \tag{5.53}$$

With the further assumption that $n(t)$ is a Gaussian with $\overline{n(t) n(t')} = 0$, Nuttall [146] obtains an explicit expression for the variance of W_y. He shows that for any noise spectrum, the variance will be infinite if the signal is not weighted, that is, if $v(t) = 1$ for all time. Moreover, he shows that for the case of white noise (power density constant for all frequencies) the variance is infinite for *any* weighting function. To have finite variance, the frequencies outside the band of the signal must be filtered out.

Other Derivations and Properties: The Wigner distribution can be derived by different methods. A particularly interesting one using the Radon transform was given by Bertrand and Bertrand [22], who also studied the behavior of these distributions for broad-band signals. Kobayashi and Suzuki [110] have shown that for mono-component signals the Wigner distribution may give rise to side lobes.

General reviews of the Wigner distribution have been given by Hillery *et al.* [89], Mecklenbrauker [141], Boashash [26], [29], and Boudreaux-Bartels [41].

VI. Spectrogram and Ambiguity Function

A. Short-Time Fourier Spectrum and Spectrogram

The spectrogram [5], [6], [68], [75], [111], [112], [126], [150], [151], [158], [163], [164], [174] has been the most widely used tool for the analysis of time-varying spectra. The concept behind it is simple and powerful. If we want to analyze what is happening at a particular time, then we just use a small portion of the signal centered around that time, calculate its energy spectrum, and do it for each instant of time. Specifically, for a window function $h(t)$ centered at t, we calculate the spectrum of $s(t')h(t' - t)$,

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}} \int e^{-j\omega t'} s(t')h(t' - t)\, dt' \tag{6.1}$$

which is the short-time Fourier transform. The energy density spectrum or spectrogram is

$$P_S(t, \omega) = |S_t(\omega)|^2 \tag{6.2}$$

which can be considered as the energy density at points t and ω. The window function controls the relative weight imposed on different parts of the signal. By choosing a window that weighs the interval near the observation point a greater amount than other points, the spectrogram can be used to estimate local quantities. Depending on the application and field, different forms of display have been used. The most common display is a two-dimensional projection where the intensity is represented by different shades of gray. This is possible for the spectrogram, because it is manifestly positive. The earliest application was used to discover the fundamental aspects of speech. The mathematical description of the spectrogram is closely connected to the work of Fano [72] and Schroeder and Atal [175], although their approach was from the correlation point of view. There have been many modifications [111], [112] of the spectrogram, and an excellent comparison of the different approaches is presented by Kodera, Gendrin, and De-Villedary [112]. Altes [6] has given a comprehensive analysis of the spectrogram and derived a number of interesting relations pertinent to the issues we are addressing in this review.

Another perspective is gained if we express the short-time Fourier transform in terms of the Fourier transforms of the signal $S(\omega)$ and window $H(\omega)$,

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}} e^{-j\omega t} \int e^{j\omega' t} S(\omega')H(\omega - \omega')\, d\omega' \tag{6.3}$$

which by analogy with the preceding discussion can be used to study the behavior of the properties around the frequency point ω. This is done by choosing a time window function whose transform is weighted relatively higher at the frequency ω.

The more compact or peaked we make the window in the time domain, the more time resolution is achieved. Similarly, if we choose a window peaked in the frequency domain, high-frequency resolution is obtained. Because of the uncertainty principle, both $h(t)$ and $H(\omega)$ cannot be made arbitrarily narrow; hence there is an inherent tradeoff between time and frequency resolutions in the spectrogram for a *particular* window. However, different windows can be used for estimating different properties.

The basic properties [68], [126], [111], [112] and effectiveness of the spectrogram for a particular signal depend on the functional form of the window, although we expect that the estimated properties are not too sensitive to the details of the window. Indeed one would hope that the results are in some sense window independent. As an illustration consider calculating the first conditional moment of frequency by using Eq. (4.24) in order to estimate the instantaneous frequency. If we write the signal in terms of its amplitude and phase as in Eq. (4.26), and similarly for the window

$$h(t) = A_h(t)e^{j\varphi_h(t)} \tag{6.4}$$

then the first conditional moment is calculated to be

$$\langle\omega\rangle_t = \frac{1}{P_1(t)} \int A^2(t')A_h^2(t' - t)[\varphi'(t') + \varphi_h'(t' - t)]\, dt' \tag{6.5}$$

where $P_1(t)$ is the marginal distribution in time,

$$P_1(t) = \int |S_t(\omega)|^2\, d\omega = \int A^2(t')A_h^2(t' - t)\, dt'. \tag{6.6}$$

This may be derived directly or by using Eq. (4.25), with the kernel of the spectrogram to be given by Eq. (6.17). If different windows are used, different results are obtained for $\langle\omega\rangle_t$. If the window is narrowed in such a way as to approach a delta function [123], $A_h^2(t) \rightarrow \delta(t)$, then $P(t) \rightarrow A^2(t)$. Using Eq. (6.5) for real windows, the estimated instantaneous frequency approaches the derivative of the phase

$$\langle\omega\rangle_t \rightarrow \varphi'(t). \tag{6.7}$$

We note, however, that although the average approaches the derivative of the phase, its standard deviation approaches infinity [123]. This is due to the fact that as $A_h^2(t) \rightarrow \delta(t)$, the modified signal $s(t')h(t' - t)$ is very narrow as a function of t' and hence has a large spread in the frequency domain.

The energy concentration of the spectrogram in the time-frequency plane is illustrated effectively by the following example [112], where we have been able to put the final result in a revealing analytic form. For the signal we take an amplitude-modulated linear FM as given by Eq. (5.15) (we take $\omega_0 = 0$ for convenience) and choose the window to be

$$h(t) = \left(\frac{a}{\pi}\right)^{1/4} e^{-at^2/2 + jbt^2/2}. \tag{6.8}$$

The short-time Fourier transform can be calculated analytically and, after some algebra, the energy density spectrum can be written in either of two forms,

$$P_S(t, \omega) = \frac{P_1(t)}{\sqrt{2\pi\sigma_2^2}} e^{-[\omega - \langle\omega\rangle_t]^2/2\sigma_2^2} \tag{6.9}$$

$$= \frac{P_2(\omega)}{\sqrt{2\pi\sigma_1^2}} e^{-[t - \langle t\rangle_\omega]^2/2\sigma_1^2} \tag{6.10}$$

where $P_1(t)$ and $P_2(\omega)$ are the marginal distributions of time and frequency, respectively,

$$P_1(t) = \sqrt{\frac{a\alpha}{\pi(\alpha + a)}} e^{-[a\alpha/(\alpha + a)]t^2} \tag{6.11}$$

$$P_2(\omega) = \sqrt{\frac{a\alpha/\pi}{\alpha(a^2 + b^2) + a(\alpha^2 + \beta^2)}} e^{-\{a\alpha/[\alpha(a^2+b^2)+a(\alpha^2+\beta^2)]\}\omega^2} \tag{6.12}$$

and where

$$\langle\omega\rangle_t = \frac{a\beta - b\alpha}{\alpha + a} t \tag{6.13}$$

$$\langle t\rangle_\omega = \frac{a\beta - b\alpha}{\alpha(a^2 + b^2) + a(\alpha^2 + \beta^2)} \omega \tag{6.14}$$

$$\sigma_2^2 = \frac{1}{2}(\alpha + a) + \frac{1}{2}\frac{(\beta + b)^2}{\alpha + a} \tag{6.15}$$

$$\sigma_1^2 = \frac{1}{2}\frac{(\alpha + a)^2 + (\beta + b)^2}{\alpha(a^2 + b^2) + a(\alpha^2 + \beta^2)}. \tag{6.16}$$

From Eqs. (6.9) and (6.10) we see that for a given time, the maximum concentration is along the *estimated* instantaneous frequency and that for a given frequency, the concentration is along the *estimated* time delay. If we want high time resolution, which will give a good estimate of the instantaneous frequency, we must take a narrow window which is accomplished by making a large. From Eq. (6.5) we see that then $\langle\omega\rangle_t \rightarrow \beta t$, as expected from our previous discussion.

Properties of the Spectrogram and Kernel: As previously mentioned, the spectrogram is a member of the class given by Eq. (2.4). Expanding Eq. (6.2) and comparing with Eq. (2.4), we see that the kernel that produces the spectrogram is [56]

$$\phi_S(\theta, \tau) = \int h^*\left(t - \frac{1}{2}\tau\right)h\left(t + \frac{1}{2}\tau\right) e^{-j\theta t}\, dt \tag{6.17}$$

which is related to the symmetrical ambiguity function (see next section) of the window. It can also be expressed in terms of the Fourier transform of the window,

$$\phi_S(\theta, \tau) = \int H^*\left(\omega - \frac{1}{2}\theta\right)H\left(\omega + \frac{1}{2}\theta\right)e^{j\tau\omega}\, d\omega. \tag{6.18}$$

Equation (6.17) is a very convenient way to study and derive the properties of the spectrogram. For example, to see whether energy can be preserved, we examine the kernel at $\theta, \tau = 0$, as per Eq. (4.14),

$$\phi_S(0, 0) = \int |h(t)|^2\, dt. \tag{6.19}$$

This should be equal to 1 if we want the total energy to be preserved, and that can be achieved if the window is normalized to 1. To examine whether the marginals are satisfied we examine the conditions given by Eqs. (4.10) and (4.12),

$$\phi_S(\theta, 0) = \int |h(t)|^2 e^{-j\theta t}\, dt \tag{6.20}$$

$$\phi_S(0, \tau) = \int |H(\omega)|^2 e^{j\tau\omega}\, d\omega. \tag{6.21}$$

To have the correct marginals, both of these quantities should be equal to 1. The only way we can make $\phi_s(\theta, 0)$ equal to 1 is if we choose a window whose square approaches a delta function. The closer it approaches a delta function, the closer the time marginal of the spectrogram will approach the instantaneous energy. However, for a narrow window, the Fourier transform will be very broad and the spectral energy density will be represented poorly.

We have already given the first conditional moment of frequency for a given time, Eq. (6.5), and showed its relationship to the instantaneous frequency. A similar result holds for the time delay. If the Fourier transforms of the signal and the window are written as

$$S(\omega) = B(\omega)e^{j\psi(\omega)}, \qquad H(\omega) = B_H(\omega)e^{j\psi_H(\omega)} \tag{6.22}$$

the conditional expected value of time for a given frequency is

$$\langle t\rangle_\omega = -\frac{1}{P_2(\omega)} \int B^2(\omega')B_H^2(\omega - \omega')[\psi'(\omega') - \psi_h'(\omega - \omega')]\, d\omega' \tag{6.23}$$

where $P_2(\omega)$ is the marginal distribution in time,

$$P_2(\omega) = \int B^2(\omega')B_H^2(\omega - \omega')\, d\omega'. \tag{6.24}$$

If the window is narrowed in the frequency domain, then a similar argument as before shows that the estimated group delay goes as $\langle t\rangle_\omega \rightarrow -\psi'(\omega)$ for real windows that are narrow in the frequency domain.

One can perform a further integration of the conditional moments to get the mean time and mean frequency of the signal. They are given by

$$\langle t\rangle|_S = \langle t\rangle|_s - \langle t\rangle|_h \tag{6.25}$$

$$\langle\omega\rangle|_S = \langle\omega\rangle|_s + \langle\omega\rangle|_h = \langle\varphi'(t)\rangle|_s + \langle\varphi_h'(t)\rangle|_h \tag{6.26}$$

where the subscript S signifies that we are using the spectrogram for the calculation and the subscripts s and h indicate calculation with the signal or window only, that is, with $A^2(t)$ or with $A_h^2(t)$. The mean value of the window has a very determined effect on these global quantities. If the window is chosen so that its mean time and frequency are zero (e.g., by choosing a window symmetrical in time and frequency), then the average of the spectrogram will be identical to that of the window, irrespective of its shape. However, the second global moment and the standard deviation will depend on the window characteristics.

Similar considerations apply to the covariance. The first joint moment is

$$\langle t\omega\rangle|_S = \langle t\varphi'(t)\rangle|_s - \langle t\varphi_h'(t)\rangle|_h - \langle t\rangle|_h\langle\varphi'(t)\rangle|_s + \langle t\rangle|_s\langle\varphi_h'(t)\rangle|_h \tag{6.27}$$

and using Eqs. (6.25) and (6.26) we see that the covariance can be written in the form

$$\mathrm{Cov}\,(t\omega)|_S = \mathrm{Cov}\,(t\omega)|_s - \mathrm{Cov}\,(t\omega)|_h. \tag{6.28}$$

This shows that the covariance of the energy density spectrum is the difference between the covariance of the signal and that of the window.

Inversion and Representability: The inversion problem for the spectrogram poses no new difficulties and the discussion of Section IV regarding inversion applies. To determine whether a particular window function leads to a spectrogram from which a signal can be recovered, one calculates the kernel by way of Eq. (6.17) and applies the general criteria of Nuttall [148] as discussed in Section IV-B.

Comparison with Bilinear Distributions Satisfying the Marginals: The spectrogram has a simple intuitive interpretation and by choosing appropriate windows, the physical parameters of the signals can be measured or estimated. However, one must manipulate the window depending on the quantities being estimated. For example, if one wants to obtain accurate results for both the instantaneous frequency and the time delay, different windows must be used. Also the optimum window to use will in general depend on time [14], [183]. On the other hand, as we have seen, for some of the bilinear distributions, the instantaneous frequency and time delay are obtained *exactly* by calculating the conditional averages, and no decisions with respect to the windows have to be made. This is an important advantage afforded by distributions like the Wigner, which has been used with considerable profit to estimate the instantaneous frequency of a signal. However, the spectrogram has the advantage that it is always positive. The bilinear distributions which give the proper marginals are never positive for an arbitrary signal. Also, the results they give for other conditional moments can very often not be interpreted [56]. The relative merits and the usefulness of these distributions are developing subjects and as we gain experience with a variety of distributions, their advantages and drawbacks will be clarified. It is very likely that different distributions should be used for different signals and for obtaining different properties of a signal.

For comparison we list in Table 3 some of the important properties of the spectrogram and compare them to the bilinear distributions which satisfy the marginals and satisfy Eqs. (4.28) and (4.47b).

Relation to Other Distributions: Historically the development of the spectrogram and bilinear distributions of the type discussed in Section I evolved separately and were motivated by different approaches and physical arguments. Only recently has the connection between them been appreciated. In the quantum context Bopp [34] and Kuryshkin *et al.* [118]–[120] developed the theory of spectrograms as an alternative approach to the Wigner distribution. Perhaps the earliest connection between the spectrogram and other distributions, pointed out by Ackroyd [2], [3] and later used by Altes [6], is the relation with the Rihaczek distribution,

$$P_S(t, \omega) = \iint e_s(t', \omega') e_h(t' - t, \omega - \omega')\, dt'\, d\omega' \tag{6.29}$$

where $e_s(t, \omega)$ and $e_h(t, \omega)$ are the Rihaczek distributions of the signal and window functions, respectively. The spectrogram can be thought of as the time-frequency distribution of the signal smoothed with the time-frequency distribution of the window. Mark [138] and Claasen and Mecklenbrauker [56] pointed out a similar relation with the Wigner distribution,

$$P_S(t, \omega) = \iint W_s(t', \omega') W_h(t' - t, \omega - \omega')\, dt'\, d\omega'. \tag{6.30}$$

Many researchers working with the Wigner distribution, who have been unaware of Eq. (6.29), have implied that this shows some unique and important connnection between the spectrogram and the Wigner distribution. We now know that these relations are just special cases of the general relation which connects any two different bilinear distributions.

In fact these two special cases can be generalized for other distributions,

$$P_S(t, \omega) = \iint P_s(t', \omega') P_h(t' - t, \omega - \omega')\, dt'\, d\omega' \tag{6.31}$$

for all kernels such that $\phi(-\theta, \tau)\phi(\theta, \tau) = 1$, where P_s and P_h are the distribution functions of the signal and the window, respectively. To show this, suppose M_s and M_h are the characteristic functions of the signal and the window. Using Eq. (3.77) we have that

$$\phi_S(\theta, \tau) = \frac{M_h(-\theta, \tau)}{\phi(-\theta, \tau)}. \tag{6.32}$$

Table 3 Comparison of Spectrogram with Distributions Satisfying Marginals and Eqs. (4.28) and (4.47b)*

Property	Spectrogram	"Distribution"
Total energy	1	1
Time marginal $P_1(t)$	$\int A^2(t')\, A_h^2(t' - t)\, dt'$	$A^2(t)$
Frequency marginal $P_2(\omega)$	$\int B^2(\omega')\, B_H^2(\omega - \omega')\, d\omega'$	$B^2(\omega)$
Mean time	$\langle t \rangle\|_s - \langle t \rangle\|_h$	$\langle t \rangle\|_s$
Mean frequency	$\langle \omega \rangle\|_s + \langle \omega \rangle\|_h$	$\langle \omega \rangle\|_s$
Mean time for given ω	$-\frac{1}{P_2(\omega)} \int B^2(\omega')\, B_H^2(\omega - \omega')\, [\psi'(\omega') - \psi_H'(\omega - \omega')]\, d\omega'$	$-\psi'(\omega)$
Mean frequency for given t	$\frac{1}{P_1(t)} \int A^2(t')\, A_h^2(t' - t)\, [\varphi'(t') + \varphi_h'(t' - t)]\, dt'$	$\varphi'(t)$

*The signal and window are written as $A(t)\, e^{j\varphi(t)}$ and $A_h(t)\, e^{j\varphi_h(t)}$, respectively, and both are normalized to 1. Their Fourier transforms are expressed as $B(\omega)\, e^{j\psi(\omega)}$ and $B_H(\omega)\, e^{j\psi_H(\omega)}$, respectively. The sybmols $|_s$ and $|_h$ indicate that the calculation is done with the signal or the window only.

The characteristic function of the spectrogram is

$$M_S(\theta, \tau) = \iint |S_t(\omega)|^2 e^{j\theta t + j\tau\omega} \, dt \, d\omega$$
$$= \frac{M_s(\theta, \tau) M_h(-\theta, \tau)}{\phi(\theta, \tau)\phi(-\theta, \tau)}. \tag{6.33}$$

If we take the Fourier transform of both sides according to Eq. (3.30) to obtain the distribution, then Eq. (6.31) follows. If $\phi(-\theta, \tau)\phi(\theta, t)$ does not equal 1, then

$$P_S(t, \omega) = \iiiint G(t'', \omega'') \, P_s(t', \omega') \cdot P_h(t'' + t' - t, -\omega'' + \omega - \omega') \, dt' \, dt'' \, d\omega' \, d\omega'' \tag{6.34}$$

where

$$G(t, \omega) = \frac{1}{4\pi^2} \iint \frac{e^{-j\theta t - j\tau\omega}}{\phi(-\theta, \tau)\phi(\theta, \tau)} \, d\theta \, d\tau. \tag{6.35}$$

B. Ambiguity Function

The Woodward [202] ambiguity function has been an important tool in analyzing and constructing signals associated with radar. It relates range and velocity resolution, and the performance characteristics of a waveform can be formulated in terms of it. By constructing signals having a particular ambiguity function, desired performance characteristics are achieved, at least in theory. A comprehensive discussion of the ambiguity function can be found in [168], and shorter reviews of its properties and applications are found in [67] and [177].

The connection between the ambiguity function and time-frequency distribution functions as discussed here has been recognized for a long time [108], [109]. Indeed Woodward [202] noted the connection with the Ville characteristic function. The similarities between the ambiguity function and pseudo-characteristic functions as discussed in Section III are many. Having a connection between the two often helps to clarify relations.

There are a number of minor differences in terminology regarding the ambiguity function. We shall use the definition of [168],

$$\chi(\theta, \tau) = \int s^*(t - \tau) s(t) e^{j\theta t} \, dt. \tag{6.36}$$

The symmetrical ambiguity function is defined [168] by

$$\chi_S(\theta, \tau) = \int s^*\left(t - \frac{1}{2}\tau\right) e^{j\theta t} s\left(t + \frac{1}{2}\tau\right) dt. \tag{6.37}$$

We note that very often the complex conjugate of (6.36) or its absolute value, or the absolute value squared, are called the ambiguity function.

Comparing with Eq. (3.19) it is seen that the ambiguity function is the characteristic function of the Rihaczek distribution, and comparing with Eq. (3.40) we see that the symmetrical ambiguity function is the characteristic function of the Wigner distribution. The mathematical and possible physical analogy between the two enhances the interpretation of the properties of the ambiguity function. For example, the condition that $\chi(0, 0) = 1$ is easily understood from a characteristic function point of view since it is a reflection of the fact that the distribution is normalized to the total energy, which has been chosen to be 1. The relation that $\chi(\theta, \tau) = \chi^*(-\theta, -\tau)$ implies that the distribution is real. If we look at the column labeled characteristic function in Table 2, we recognize that those are properties usually associated with the ambiguity function, and that in many cases the interpretation in terms of distributions is more transparent. The analogy can be extended by defining a generalized ambiguity function through Eq. (3.77). Choi and Williams [51] have used them with profit to analyze the effects of the kernel on the behavior of the distribution. There are many more analogies than we have indicated here. A number of excellent articles exploring the relationship between the ambiguity function and time-frequency distributions can be found in [56], [69], [187].

Some have argued that a particular distribution, such as the Wigner, is "better" than the ambiguity function. A number of reasons are usually given, among them that the Wigner distribution is real while the ambiguity function is complex. This is a mistaken view for the following reasons. Characteristic functions are very often much more revealing than the distribution. Furthermore they are very useful in calculation as, for example, to calculate the mixed moments. The properties of a distribution are often easier to determine from the characteristic function than from the manipulation of the distribution. Also, the ambiguity function plane is a very effective means for choosing kernels [51], [146], [147], [76]. Finally we point out that the characteristic function has been a main tool for obtaining these distributions.

VII. Time-Frequency Filtering and Synthesis

If the concepts and methods of filter theory could be generalized to the time-frequency plane, it would offer a powerful tool for the construction of signals with desirable time-frequency properties. However, time-frequency filtering presents unique difficulties which have not been fully overcome. Perhaps the first attempt to obtain input-output relations for a joint quasi-distribution was by Liu [127], who used the Page distribution. He calculated the output relations for a number of causal linear systems and obtained interesting bounds on the output distribution. Subsequently Bastiaans [17], [18] and Claasen and Mecklenbrauker [56] have obtained the transformation properties for the Wigner distribution. Eichmann and Dong [70] formulated a general optical method for time-frequency filtering and produced methods that may be applied to many distributions.

As Saleh and Subotic [173] have pointed out, unlike a standard transfer function, the output for these bilinear distributions is not a simple multiplicative function of the input distribution. As a matter of fact, the output distribution will almost always not be representable, that is, no signal will exist that will produce it. There are two qualitatively different reasons why distributions are not representable. They can be categorized into distribution-independent and distribution-dependent reasons. For the sake of simplicity we restrict ourselves to distributions that satisfy the marginals.

1) Distribution-Independent Conditions: From a potential candidate for a distribution $P(t, \omega)$, one can calculate the two marginals

$$P_1(t) = \int P(t, \omega) \, d\omega, \qquad P_2(\omega) = \int P(t, \omega) \, dt. \tag{7.1}$$

Now for the distribution to be representable, P_1 and P_2 must be the absolute squares of functions that are Fourier transform pairs of each other, that is, there must exist a signal $s(t)$ whose Fourier transform is $S(\omega)$, such that $P_1(t) = |s(t)|^2$ and $P_2(\omega) = |S(\omega)|^2$. An example of nonrepresentability, as pointed out by Saleh and Subotic [173], would be a distribution that produced marginals which are nonzero only in finite regions. Such marginals could not produce Fourier pairs since they are time and band limited.

2) Distribution-Dependent Conditions: The above requirements are clear and within the experience of working with Fourier transforms. The second set of reasons depend on the functional relationship between the distribution and the signal, and reflect the peculiarities of the distribution. Thus the design of time-variant transfer functions cannot be based solely on physical grounds but must take into account the peculiarities of the distribution, and unfortunately each distribution has its own peculiarities. To illustrate these difficulties, we give some examples. Suppose that for an input Wigner distribution function the transfer function cuts a strip parallel to the frequency axis for a finite time interval, indicating silence at all frequencies. The resulting distribution is never representable, although what we have done to it, namely, asked for some silence, is certainly reasonable. Is the nonrepresentability an indication of some violation of physical impossibility? No, it is merely a peculiarity of the Wigner distribution. For a more dramatic example, suppose we have the Wigner distribution for a Gaussian signal. If we multiply the distribution by ω^2, do we get a representable distribution? No. In fact if we multiply it by any positive function other than a Gaussian, we can be certain that the resulting distribution is not proper. The reason is that the Gaussian signal is the only one that gives a positive Wigner distribution, and multiplying the distribution with another positive function which is not Gaussian cannot result in *any* Wigner distribution. Now if we use the Rihaczek distribution for the silence example, the resulting distribution is a proper Rihaczek distribution. However, if we multiply a Rihaczek distribution by a function of time and frequency which is not a product of functions of time and frequency, the output will never be a Rihaczek distribution. Again, this is a peculiarity of the distribution and not a reflection of some inherent physical impossibility. Hence procedures that appear reasonable, as reflected by reasonable time-variant transfer functions, often do not work for a particular distribution, but may work for another. The failure is not due to any violation of physical law, but just a reflection of the peculiarities of the distribution. How to recognize and deal with these peculiarities is one of the major stumbling blocks. The above difficulties have been investigated for only a few distributions, and it is possible that there may be distributions for which the difficulties do not arise.

These problems not withstanding, Saleh and Subotic [173] simplified matters considerably. By analogy with the standard transfer function method they multiply the input distribution by a time-variant transfer function to obtain the output. Conceptually this is an ideal method as it is simple and direct. Specifically, they write

$$P_O(t, \omega) = H(t, \omega) P_I(t, \omega) \tag{7.2}$$

where P_O and P_I are the output and input distributions, respectively, and H is the time-varying transfer function. In general the output distribution will not be representable, and they present two methods to synthesize the signal from the output distribution. One technique is based on using Eq. (5.6), irrespective of whether or not the signal is representable, and the other finds a signal that reproduces a distribution as closely as possible, in the least-square sense, to the output distribution. The method of Saleh and Subotic is appealing because it conforms as much as possible with our current intuitive notions of what we would want time-frequency filtering to do. As they point out, their approach applies to other time-frequency distributions. It would be of interest to investigate for which distributions their procedure can be implemented in an optimal manner.

Other innovative methods have been devised for the synthesis problem. Boudreaux-Bartels and Parks [39]–[41] have devised a number of efficient methods for the synthesis of the Wigner distribution, and other methods have been given by Yu and Chang [203], [204] and Boashash *et al.* [26].

Input-Output Relations: We now summarize the input-output relations for a general time-variant linear transformation of the signal,

$$s_O(t) = \int h(t, t') s_I(t')\, dt' \tag{7.3}$$

where $h(t, t')$ is the impulse response [172], [205]. In such a case the relation between the input distribution and the output distribution can always be written as

$$P_O(t, \omega) = \iint K(t, \omega; t', \omega') P_I(t', \omega')\, dt'\, d\omega'. \tag{7.4}$$

A straightforward calculation yields

$$\begin{aligned} K(t, \omega, t', \omega') &= \frac{1}{8\pi^3} \int e^{-j\tau\omega + j\tau'\omega' + j\theta(u - t) - j\theta'(u' - t')} \frac{\phi(\theta, \tau)}{\phi(\theta', \tau')} \\ &\quad \cdot h^*\left(u - \frac{1}{2}\tau, u' - \frac{1}{2}\tau'\right) h\left(u + \frac{1}{2}\tau, u' + \frac{1}{2}\tau'\right) \\ &\quad \cdot du\, d\tau\, d\theta\, du'\, d\tau'\, d\theta'. \end{aligned} \tag{7.5}$$

This simplifies considerably when particular kernels are considered. For the Rihaczek distribution we have

$$K(t, \omega, t', \omega') = \frac{1}{2\pi} h(t, t') S^*(\omega, \omega') e^{-it\omega + it'\omega'} \tag{7.6}$$

where

$$S(\omega, \omega') = \frac{1}{2\pi} \iint h(t, t') e^{j\omega t - j\omega' t'}\, dt\, dt'. \tag{7.7}$$

We note that K is a two-dimensional Rihaczek distribution. For the Wigner distribution we have

$$\begin{aligned} K(t, \omega, t', \omega') = \frac{1}{2\pi} \iint e^{-j\tau\omega + j\tau'\omega'} h^*\left(t - \frac{1}{2}\tau, t' - \frac{1}{2}\tau'\right) \\ \cdot h\left(t + \frac{1}{2}\tau, t' + \frac{1}{2}\tau'\right) d\tau\, d\tau' \end{aligned} \tag{7.8}$$

which is a two-dimensional Wigner distribution [17], [18].

VIII. Other Topics: Instantaneous Frequency, Quantum Mechanics, and Uncertainty Principle

A. Instantaneous Frequency and Analytic Signal

The concept of "instantaneous" frequency has a long history in physics and astronomy. Historically the methodology and description of instantaneous frequency has not always been associated with time-frequency distributions or a time-varying spectrum. A comprehensive theory of joint time-frequency distributions would be able to encompass and clarify the concept of instantaneous frequency, so it is important to appreciate the work that has been done along these lines. It was Armstrong's [11] discovery that frequency modulation for radio transmission reduces noise significantly, which produced a concerted effort to understand and describe the mathematical and conceptual description of frequency modulation and instantaneous frequency. Early comprehensive works on the analysis of frequency modulation were those of Carson and Fry [45] and Van der Pol [192], who defined instantaneous frequency as the rate of change of the phase of the signal. This definition implies that we have some procedure for forming a complex signal from a real one. In general there are an infinite number of complex signals whose real part is a given real signal. A major step was made by Gabor [80], who from the observation that both sin ωt and cos ωt transform into an exponential $e^{j\omega t}$ if we use only their positive spectrum, generalized to the arbitrary case with the prescription to "suppress the amplitudes belonging to negative frequencies and multiply the amplitudes of positive frequency by 2."He noted that this procedure is equivalent to adding to the signal an imaginary part, which is the Hilbert transform of the signal. The positive frequencies are multiplied by 2 to preserve the total energy of the original signal. To see how the Hilbert transform arises from the above prescription, suppose the signal is $s(t)$ and its Fourier transform is $S(\omega)$. The signal $z(t)$ whose spectrum is composed of the positive frequencies of $S(\omega)$ is given by the inverse transform of $S(\omega)$ using only the positive frequencies,

$$z(t) = 2\frac{1}{\sqrt{2\pi}}\int_0^\infty S(\omega)e^{j\omega t}\,d\omega. \tag{8.1}$$

Expressing $S(\omega)$ in terms of the signal $s(t)$ as per Eq. (1.3),

$$z(t) = 2\frac{1}{2\pi}\int_0^\infty \int s(t')e^{-j\omega t'}e^{j\omega t}\,dt'\,d\omega \tag{8.2}$$

and using the fact that

$$\int_0^\infty e^{j\omega x}\,d\omega = \pi\delta(x) + \frac{j}{x} \tag{8.3}$$

we have

$$z(t) = s(t) + \frac{j}{\pi}\int \frac{s(t')}{t-t'}\,dt'. \tag{8.4}$$

The second part of Eq. (8.4) is the Hilbert transform of the signal, and $z(t)$ is called the analytic signal. The derivative of the phase of the analytic signal conforms to our expectations of instantaneous frequency for a wide variety of cases, particularly narrow-band signals. There has been considerable controversy over whether this represents the proper mathematical expression of instantaneous frequency, and a number of other definitions have been given [85], [178]. For example, one can define it in terms of the average number of zeros that a function crosses per unit time.

For a real signal of the form $A(t)\cos[\omega_0 t + \phi(t)]$ the complex signal is often taken to be $A(t)e^{j\omega_0 t + j\phi(t)}$, which is called the quadrature, or exponential, model. The conditions under which this complex signal is a good approximation to the analytic signal have been investigated [169]. Nuttall [145] resolved the issue by defining the error between the exponential and the analytic signal to be the energy of the difference of the two signals. He showed that the error will be zero if the spectrum of $A(t)e^{j\phi(t)}$ is single sided. It is not necessary for the signal to be narrow band. A convenient and useful theorem for the study of Hilbert transforms was given by Bedrosian [20]. It relates the Hilbert transform of a product of two signals to the Hilbert transform of each signal.

"Instantaneous" frequency implies that we are dealing with a local concept, but to calculate the Hilbert transform, the signal for all time must be used. This paradoxical situation was analyzed by Vakman [191], who set up mild and reasonable conditions for the formation of a complex signal and showed that these conditions lead to the analytic signal. He points out that in reality only a small band around the instantaneous frequency, the "active band," is needed to approximate the analytic signal [4], [106], [191].

He makes the interesting observation that, very often, quantities defined globally can, under certain circumstances, be described advantageously by local concepts as, for example, is the case with electromagnetic waves, where in principle we are dealing with waves spread out through space, but under certain conditions, the light ray method is appropriate and useful.

The identification of the derivative of the phase with the concept of instantaneous frequency must not be taken too literally. The relation of the derivative of the phase and the "frequency" that appears in the spectrum has been investigated by Ville [194], Fink [74], and Mandel [131]. Consider at each instant keeping track of the instantaneous frequency and asking for the average frequency. That would be given by the time average

$$\langle\varphi'(t)\rangle_T = \int \varphi'(t)|s(t)|^2\,dt \qquad \text{(time average)}. \tag{8.5}$$

If we compare this to the mean frequency as defined by the spectrum,

$$\langle\omega\rangle_S = \int \omega|S(\omega)|^2\,d\omega \qquad \text{(spectral average)} \tag{8.6}$$

it is easy to prove that they are identical,

$$\langle\omega\rangle_S = \langle\varphi'(t)\rangle_T \tag{8.7}$$

which argues for the identification of the derivative of the phase as the instantaneous frequency. However, if one calculates the second moments, the identification is no longer compatible because they are not equal. In fact [74], [131], [194],

$$\langle\omega^2\rangle_S = \langle\varphi'^2(t)\rangle_T + \int A'^2(t)\,dt. \tag{8.8}$$

Also, the standard deviation of the spectral average is $\sigma_S^2 = \langle\omega^2\rangle_S - \langle\omega\rangle_S^2$, and by a similar expression for the time average, we have

$$\sigma_S^2 = \sigma_T^2 + \int A'^2(t)\, dt. \tag{8.9}$$

Mandel [131] has emphasized that the derivative of the phase does not always coincide with the frequencies that appear in the spectrum, although the averages are equal, as per Eq. (8.7). In fact it is very easy to construct examples where the derivative of the phase of an analytic signal (whose spectrum consists of only positive frequencies) may be negative at certain times. Of course one can *define* instantaneous frequency to be the derivative of the phase, but the conceptual notion embodied in the phrase will not always be reflected in the definition. We note that if there is no amplitude modulation, the two coincide, and we further note that the second term of Eq. (8.9) is identical to the expected value of the local spread, as defined by Eq. (4.39).

From the perspective of joint time-frequency distributions, instantaneous frequency is defined as the *average* frequency at a particular time, that is, the mean conditional moment of frequency. We have already seen that there are an infinite number of distributions which give the derivative of the phase for the mean conditional value of frequency. The condition for this to hold is given by Eq. (4.28) and is usually considered an important and desirable attribute of a distribution. However, we note that the result is true for any complex signal, not just the analytic signal. One may argue that the condition on these distributions should be that the conditional moment of frequency be the derivative of the phase for only certain types of signals, or it should be the derivative of the phase in some approximate sense. It can also be argued that the result should be the derivative of the phase of the analytic signal, even if the actual signal is used in the calculation of the distribution. More fundamentally, a theory of time-frequency distributions should predict the proper expression for instantaneous frequency and the "answer" should not have to be imposed. In addition we point out that some of the distributions which give the derivative of the phase for the first conditional moment give improper results for the second conditional moment, as discussed in Section IV. This indicates that we do not have a fully consistent theory. Considerable further research is needed to clarify the relations between the instantaneous frequency, joint time-frequency distributions, and the result implied by the works of Ville, Fink, and Mandel, Eq. (8.9).

From the practical point of view the question arises as to whether the actual signal or the analytic signal should be used when calculating a joint time-frequency distribution. Many have advocated using the analytic signal. There are three basic reasons for this advocacy. First, as we have discussed, some distributions give the derivative of the phase for the conditional first moment. Hence it is argued that we should use the analytic signal because the instantaneous frequency is defined in terms of the analytic signal. If one wants to use these distributions for the estimation of the instantaneous frequencies, and that is certainly an important application, then the analytic signal should be used. While it is true that a distribution of the analytic signal dramatizes the instantaneous frequency concentration for many signals, the question of what it hides of the original signal has not been fully addressed. We point out that when the analytic signal is used, the marginals of the *original* signal are not properly given. Second, the analytic signal does not have negative frequencies and therefore cannot cause interference terms with the positive frequencies. Although eliminating the negative frequencies does eliminate their overlap, it does not eliminate the interference of the positive frequencies with other positive frequencies. There will always be interference terms, no matter what part of the signal is eliminated, since that is an inherent property of the bilinear distributions. The third argument for using the analytic signal is that aliasing is eliminated and the sampling rate is reduced to the standard Nyquist rate [26].

B. Relation to Quantum Mechanics

The fundamental notion of classical mechanics is that from a knowledge of the initial positions and velocities of a particle, and the knowledge of the forces, one can predict exactly what the position and velocity of the particle will be at a later time. The equation of evolution in classical mechanics is Newton's second law of motion. The breakdown of classical mechanics and the realization that the deterministic viewpoint is incorrect because the laws of nature only predict the probability where a particle will be, is one of the greatest intellectual achievements of humankind. In addition it has had profound practical consequences as evidenced by the modern devices based on quantum effects. The fundamental idea of modern physics is that we can only predict probabilities for observables such as position and velocity, and that this is not a reflection of human ignorance but rather the way that nature operates. The probabilities are predicted by solving Schrödinger's equation of motion, which gives the wave function of position at time t. The probability of finding the particle at position q at time t is then the absolute square of the wave function. Another dramatic departure of quantum mechanics from classical mechanics is that physical observables are represented by operators and not functions. The noncommutation of operators has profound consequences regarding the simultaneous measurability of observables. We should point out that in quantum mechanics we may have an additional level of description. That is the case where we do not know, because of human ignorance, what the wave function is and assign a probability to the possible wave functions. This is done in quantum statistical mechanics and is similar to the treatment of stochastic signal in signal theory. We emphasize that in quantum mechanics we are starting with a probability description, but in signal analysis we are starting with a deterministic description.

There is a partial formal mathematical correspondence between quantum mechanics and signal analysis. Historically work on joint time-frequency distributions has often been guided by corresponding developments in quantum mechanics. Indeed the original papers of Gabor and Ville continuously evoked the quantum analogy. However, the analogy is formal only and because the interpretation is dramatically different, one must be particularly cautious in transposing and interpreting results from one field to another. What may be reasonable in quantum mechanics does not necessarily make it resonable in signal theory. Indeed it is often preposterous in signal theory, as will be illustrated with a concrete example.

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 Fourier transform 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. The first fundamental difference is that quantum mechanics is an inherently probabilistic theory. Its probabilistic interpretation is not a question of ignorance but of the fundamental character of the physical world. In signal theory, on the other hand, the signal is inherently deterministic, and the absolute square of the signal is an intensity with no probability connotations. We now come to the most important distinction. In quantum mechanics, physical quantities are always associated with operators. It is the fundamental tenet of quantum mechanics that what can be measured for an observable are the eigenvalues of its operator. This produces some seemingly bizarre results, which are nonetheless true and have been verified experimentally. It is the basis for the quantization of physical quantities and has no counterpart in signal theory. For a dramatic example, consider the sum of two continuous quantities. In quantum mechanics the sum is not necessarily continuous. Specifically consider the position q and momentum p, which are continuous variables; in quantum mechanics $q^2 + p^2$ (appropriately dimensioned) is never continuous under any circumstances, for any particle. It is always quantized, that is, it can have only certain values. The corresponding statement in signal analysis would be that time and frequency are continuous but that $t^2 + \omega^2$ (appropriately dimensioned) is never so, and of course that would be a ludicrous statement to make in signal analysis. Hence, even though there is a mathematical analogy with quantum mechanics, we cannot take the results of quantum mechanics over to joint time-frequency distributions indiscriminately. In Table 4 we outline the formal mathematical correspondence between signal analysis and quantum mechanics.

C. Uncertainty Principle and Joint Distributions

The uncertainty principle expresses a fundamental relationship between the standard deviation of a function and the standard deviation of its Fourier transform. In particular, the standard deviations are defined by

$$(\Delta t)^2 = \int (t - \bar{t})^2 |s(t)|^2 \, dt$$

$$(\Delta\omega)^2 = \int (\omega - \bar{\omega})^2 |S(\omega)|^2 \, d\omega \tag{8.10}$$

where $\bar{t}$ and $\bar{\omega}$ are the mean time and frequency. The uncertainty principle is

$$\Delta t \Delta\omega \geq \tfrac{1}{2} \tag{8.11}$$

for *any* signal. In common usage Δt and $\Delta\omega$ are called the duration and the bandwidth of a signal.

We would like to clarify the role of the uncertainty principle and its significance with regard to joint distributions. We will show that the uncertainty principle is a relationship concerning the marginals *only* and has no bearing on the existence of joint distributions. The phrase "uncertainty" was coined in quantum mechanics, where its connotation is appropriate since quantum mechanics is an inherently probabilistic theory. In quantum mechanics the standard deviations involve the measurement of physical observables. However, in nonprobabilistic contexts the uncertainly principle should be thought of as expressing the fact that a function and its Fourier transform cannot be made arbitrarily narrow.

The proper interpretation of the uncertainty relation in signal analysis has been emphasized by many. In his paper on the representation of signals, for example, Lerner [124] states that the uncertainty principle "... has tempted some individuals to draw unwarranted parallels to the uncertainty principle in quantum mechanics... The analogy is formal only." Equally to the point is Skolnik [177]: "The use of the word 'uncertainty' is a misnomer, for there is nothing uncertain about the 'uncertainty relation.'... It states the well-known mathematical fact that a narrow waveform yields a wide spectrum and a wide waveform yields a narrow spectrum and both the time waveform and the frequency spectrum cannot be made arbitrarily small simultaneously."

In both signal theory and quantum theory we have an uncertainty principle. In quantum mechanics it refers to the probabilistic aspects of measuring quantities, and the word

Table 4 Relationship Between Quantum Mechanics and Signal Analysis*

Quantum Mechanics (Inherently Probabilistic)		Signal Analysis (Deterministic)	
Position	q (random)	Time	t
Momentum	p (random)	Frequency	ω
Time	t	No correspondence	
Wave function	$\psi(q, t)$	Signal	$s(t)$
Momentum wave function	$\phi(p, t) = \frac{1}{\sqrt{2\pi\hbar}} \int \psi(q, t)\, e^{-jqp/\hbar}\, dq$	Spectrum	$S(\omega) = \frac{1}{\sqrt{2\pi}} \int s(t)\, e^{-j\omega t}\, dt$
Probability of position at time t	$\|\psi(q, t)\|^2$	Energy density	$\|s(t)\|^2$
Probability of momentum at time t	$\|\phi(p, t)\|^2$	Energy density spectrum	$\|S(\omega)\|^2$
Expected value of position	$\langle q \rangle = \int q\|\psi(q, t)\|^2\, dq$	Mean time	$\langle t \rangle = \int t\|s(t)\|^2\, dt$
Expected value of momentum	$\langle p \rangle = \int p\|\phi(p, t)\|^2\, dp$	Mean frequency	$\langle \omega \rangle = \int \omega\|S(\omega)\|^2\, d\omega$
Standard deviation of position	$\sigma_q = \sqrt{\langle q^2 \rangle - \langle q \rangle^2}$	Duration	$T = \sqrt{\langle t^2 \rangle - \langle t \rangle^2}$
Standard deviation of momentum	$\sigma_p = \sqrt{\langle p^2 \rangle - \langle p \rangle^2}$	Bandwidth	$B = \sqrt{\langle \omega^2 \rangle - \langle \omega \rangle^2}$
Uncertainty principle	$\sigma_p \sigma_q \geq \hbar/2$	Time-bandwidth relation	$BT \geq \frac{1}{2}$

*The formal mathematical correspondence is (position, momentum) ⇔ (time, frequency). The wave function in quantum mechanics depends on time, but this has no formal correspondence in signal analysis. Planck's constant $\hbar$ may be taken equal to 1. Quantum mechanics is an inherently probabilistic theory in contrast to signal analysis, which is deterministic. Hence while there is the formal mathematical correspondence, the interpretation of results is very different. Both quantum mechanics and signal theory have another level of indeterminism where the wave function or the signal is ensemble averaged.

"uncertainty" connotes the right meaning. It is one of the most profound discoveries and refers to the measurement of physical quantities represented by operators that do not commute, such as position and momentum. In signal analysis it applies only to the broadness of signals, which are related to each other by Fourier transforms, and it does not relate to measurement in the quantum mechanical sense. As Ackroyd [3] has emphasized, "There is a misconception that it is not possible to measure the $t - f$ energy density of a given waveform and that this is a consequence of Gabor's uncertainty relation. However, the uncertainty principle of waveform analysis is not concerned with the measurement of $t - f$ energy density distributions: instead it states that if the effective bandwidth of a signal is W then the effective duration cannot be less than about $1/W$ (and conversely)..."

Additional confusion arises when the Δt and $\Delta\omega$, which are used in the uncertainty principle to connote standard deviations or broadness, are misconstrued with the differential elements of calculus. They are not the same, and the uncertainty principle does not say that we cannot make the differential elements as small as we like. The two uses of Δ should not be confused.

We now address the question of the relationship of the uncertainty principle and joint distributions. Our point is that it has no bearing on the question of joint distributions and relates to the product of the standard deviations of the *marginals*. To understand this, suppose we have a joint distribution and wish to calculate the product $(\Delta t)^2(\Delta\omega)^2$. It would be

$$(\Delta t)^2(\Delta\omega)^2 = \iint (t - \bar{t})^2 P(t, \omega)\, dt\, d\omega \cdot \iint (\omega - \bar{\omega})^2 P(t, \omega)\, dt\, d\omega \quad (8.12)$$

$$= \int (t - \bar{t})^2 |s(t)|^2\, dt \cdot \int (\omega - \bar{\omega})^2 |S(\omega)|^2\, d\omega. \quad (8.13)$$

This is the usual starting point in the derivation of the uncertainty principle, and hence the uncertainty principle follows. This demonstration is rather trivial; however, since there is a general sense that the uncertainty principle has to do with correlations between measurements of time and frequency, the preceding steps force the reader to see that this is *not* the case. The uncertainty principle is calculated *only* from the marginals. Hence any joint distribution that yields the *marginals* will give the uncertainty principle. It has nothing to do with correlations between time and frequency or the measurement for small times and frequencies. What it does say is that the marginals are functionally dependent. But the fact that marginals are related does not imply correlation between the variables and has nothing to do with the existence or nonexistence of a joint distribution.

It is often stated that one cannot have proper positive distributions because that would violate the uncertainty principle. But it is well known that the Wigner distribution *is* positive for some signals. If positivity and the uncertainty principle were incompatible, it must be so for all cases. Furthermore it is possible to generate an infinite number of positive distributions which satisfy the marginals. Also, it is sometimes stated that if one averages the Wigner distribution over an area greater than that given by the uncertainty principle, we will get a positive answer. That is not the case. Many counterexamples have been given.

IX. Applications

There has been a considerable effort to apply these distributions to almost every field where nonstationary signals arise. The purpose of the applications has varied considerably, from the simple graphic presentation of the results, with the expectation that they will reveal more than other methods, to sophisticated manipulation of the distribution. We will emphasize how these distributions have been applied, but we will not go into detail about the particular numerical techniques. The applications can be broadly categorized according to three methodologies. First is calculating the distribution to see whether it does reveal more information than other tools such as the spectrogram. An example is the application to speech, where one hopes more of the fine points of speech such as transients and transitions will be revealed. Second is to use a particular property of the distribution which clearly and robustly represents the time-frequency content for that property, for example, correlating instantaneous frequency with physical quantities one is trying to obtain. Third is to use the distribution as a carrier of the information of a signal and without concern as to whether the distribution truly represents the time-frequency energy density. Many applications do not fall clearly into the above categories, but it is nevertheless useful to keep them in mind because the success or failure of a distribution in a particular application does not necessarily imply success or failure in a different application. For example, the Wigner distribution may be hard to interpret in the analysis of speech, but may be useful for recognition. For applications where the interpretation as true densities is not necessary, the violation of certain properties, such as the marginals, may be acceptable.

Perhaps the earliest application that took advantage of the Wigner distribution was the work of Boashash [35]. His method is based on correlating a physical quantity of the problem at hand with a feature in the Wigner distribution, usually the instantaneous frequency. The importance of Boashash's idea is that one does not have to rely on a full interpretation of the distribution as a joint density but only that some of its predictions need be correct. For example, as long as one is confident that the instantaneous frequency is well described by the distribution, the fact that other properties may not be is unimportant. His first application of this was to geophysical exploration. The basic idea is to send a signal through the ground, measure the resulting signal, and calculate the Wigner distribution. From the distribution one determines the instantaneous frequency, and from the instantaneous frequency one calculates the attenuation and dispersion. This method has been used to study many diverse problems. Boashash *et al.* [27], [28] studied the absorption and dispersion effects in the earth. Imberger and Boashash [94], [95] have applied the method to analyze the temperature gradient microstructure in the ocean by relating the instantaneous frequency to the dissipation of kinetic energy. Bazelaire and Viallix [19] have also used the Wigner distribution to obtain data to measure the absorption and dispersion coefficients of the ground and have formulated a new understanding of seismic noise.

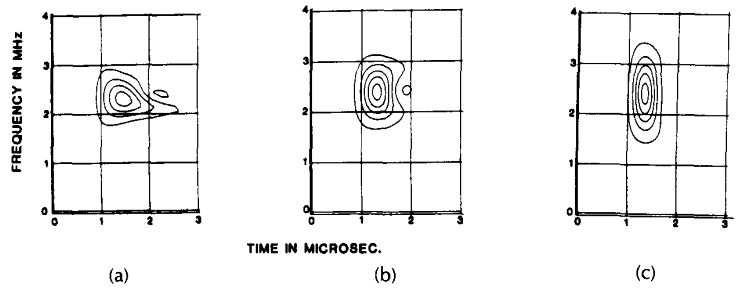


Fig. 8. Contour plots of Wigner distribution for the output of a simulated ultrasonic transducer for various design parameters. The advantage of using a time-frequency distribution is that one can readily see the main features of the output.

Of particular significance was the work of Janse and Kaizer [97]. They developed innovative techniques and laid the foundation for the use of these distributions as practical tools. They calculated the Wigner distribution for a number of standard filters, and found it to be a particularly powerful means of handling the inherently nonstationary signals encountered in loudspeaker design.

In the design of devices to produce waves, time-frequency distributions are an effective and comprehensive indicator of the characteristics of the output. We illustrate this with the work of Marinovic and Smith [137], who used the Wigner distribution as an aid in the design and analysis of ultrasonic transducers. An ultrasonic transducer is a device for producing sound waves and is typically used as the source of waves in medical imaging, sonar, and so on. The most common means of producing high-frequency sound is by exciting a piezoelectric crystal such as a quartz. When mechanical stress is applied to these crystals, the polarization is changed, producing an electric field. Conversely when an electric field is applied, the crystal is strained. By using an oscillating electric field the crystal is made to vibrate, producing acoustic waves in the medium. The output of a transducer depends on many factors, such as shape, thickness, the electronic fields driving it, and the coupling with the medium. In designing a transducer one is interested in the output having certain desirable characteristics. Typically what is required is, for the output, to have a uniform time spread over the frequencies. Uniformity is desired so that there are no aberrations with respect to different frequencies acting in different ways. A number of simulation models have been devised which predict the general characteristics of the output with the various parameters under the designer's control. The advantage of using a time-frequency distribution is that one can quickly and effectively see the effects of varying the parameters. In Fig. 8 we show various contour plots of the Wigner distribution of the output of a simulation program for designing transducers for three choices of design parameters. In Fig. 8(a) the output has a very poor uniformity of energy in the various frequencies. In Fig. 8(b) there is considerable improvement, but no uniformity yet, and in Fig. 8(c) we have an ideal case, the output being quite uniform for the frequencies produced. The advantage of using a joint time-frequency distribution is that within one picture the characteristics of the transducer are readily discerned and one does not have to do various independent time and frequency analyses.

An example that illustrates the use of these distributions for discovery and classification is the work of Barry and Cole [15] on muscle sounds. When a muscle contracts, it produces sounds that can be picked up readily by a microphone. It has been discovered that these sounds are not due to the muscle vibrating as a simple string. The work of Barry and Cole [15] and others is aimed at correlating the properties of the sound with the characteristics of the muscle. If one had a good understanding of the different mechanisms that are producing significant changes in the distribution, one would potentially have an excellent diagnostic tool since these acoustic waves would provide a noninvasive diagnostic tool. Fig. 9 shows the distribution of Choi and Williams for two different sounds produced by a muscle. The first is produced during an isometric con-

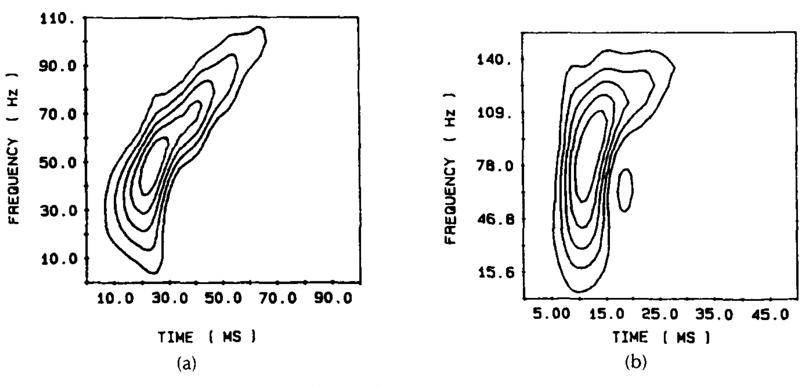


Fig. 9. Choi–Williams distribution for two different sounds produced by a muscle. (a) During isometric contraction. (b) When muscle is twitched.

traction and the second when the muscle is twitched. The general features are quite reproducible from muscle to muscle and hence reflect some general characteristics of muscle contraction. On a fundamental level these figures dramatically show that average frequency or instantaneous frequency changes significantly during a muscle contraction and gives an indication of the spread around the average. The sounds have been correlated with characteristics of the muscle such as its stiffness. The important point from our perspective is that, as Barry and Cole [15] point out, these "time-dependent frequency changes in the acoustical signals would be hard to discern with standard frequency domain analysis."

The propagation of a signal through different media is a very common occurrence in nature, and since it is generally accompanied by time delays and changes in frequencies, a combined time-frequency description is natural. The simplest propagation of a disturbance is one governed by the equation $u_{xx}(x, t) = v^2 u_{tt}(x, t)$, where u is the physical quantity that is changing (e.g., pressure or electric field), x and t are position and time, respectively, and v is the velocity of propagation. If we start with a disturbance at $t = 0$ given by $u(x, 0)$, then the disturbance at a later time will be given by the same function, displaced by vt, that is, the disturbance will propagate with the same shape. Mathematically this is explained by observing that *any* function of the form $u(x - vt)$ is a solution. Examples are electromagnetic waves in free space and sound waves in air (to a large extent). Indeed the reason that a person standing 10 ft away from an object sees and hears the same thing as a person 20 ft away is that the shape of the disturbance has not changed.

However, a typical wave equation governing the propagation of a wave in a medium contains extra terms and does not admit solutions of the form $u(x - vt)$ for an arbitrary u. The equation is a wave equation because it does admit propagating waves of the form $e^{j(\pm\omega t - kx)}$. Only sinusoidal waves propagate without change. One of the most important effects in wave propagation is that the phase velocity may depend on the wavelength. Since an arbitrary disturbance can be decomposed into sinusoids by way of the Fourier transform and each will move at a different velocity, the recombination of them at a later time will not preserve the shape of the signal. This phenomenon, that sinusoidal waves of different frequencies propagate with different velocities, is called dispersion. The reason for this term is that the earliest discovered ramification was that a prism can "disperse" white light into different colors since they travel with different speeds in glass. If the velocity decreases with frequency, one says the dispersion is "normal;" otherwise it is termed "anomalous" dispersion. Another important effect in the propagation of a signal is the attenuation of a wave, the dying out or absorption. The energy is typically dissipated into heat. Again, the amount of attenuation depends on the medium and the frequency. In the case of sound in normal conditions there is almost no attenuation, and that is why we are able to hear from far away. In contrast high-frequency electromagnetic waves are damped within a short distance of entering the surface of a conductor. Also, as a wave propagates from one medium to another, part of it gets reflected and part transmitted.

We are now in a position to understand why a time-frequency analysis offers an effective description of a signal that has propagated through different media. At each instant of time the signal we measure will be the superposition of a number of waves. We may still be measuring the initial wave more or less as it left the source. Superimposed on that will be a delayed signal from a reflection boundary, and that signal may have the same shape as the original one, but delayed in time. The superposition of these two signals may look quite complicated, but in the time-frequency plane we will simply see the distribution of the original signal and the same distribution translated upward with respect to time. This would be immediately recognizable. If in addition we have a wave that was delayed and dispersed, this will be seen in a time-frequency plot as an image similar to the original, displaced upward with a certain amount of relative bending in those frequencies were the dispersion occurred.

To illustrate we use the work of Boashash and Bazelaire [36] and Boles and Boashash [32] on geophysical exploration. Seismic signals are particularly rich and varied due to layers of different media. Not only are there layers of various solids with different properties (e.g., shale, sandstone), but one has layers of water and gas beneath the surface. In addition if one is exploring off shore, the signal obtained from sources beneath the seabed are concealed by the reverberations of the water. What is typically done in exploration is to produce a wave at the surface and measure the resulting wave at one or more places down the field. The initial wave is generated by different means, such as by an explosion caused by dynamite or by vibrating a metal plate coupled to the earth. The resulting acoustic wave travels through the different layers and is reflected upward with possible multiple reflections. The velocity of different layers varies considerably. For air it is about 1000 ft/s and for solid rock it may be as high as 20 000 ft/s. The resulting signal will be a multicomponent signal, and if we did have a good time-frequency distribution, each component would stand out in the time-frequency plane. A schematic diagram of what such a distribution would look like is presented in Fig. 10, which is adapted from Boles and Boashash

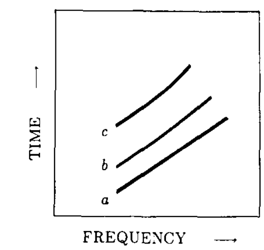


Fig. 10. Schematic diagram of a time-frequency distribution for a wave reflected from media with different characteristics. Curve *a*—initial signal; curve *b*—distribution of a signal that has gone through a layer with little dispersion; curve *c*—from layer with normal dispersion. The high frequencies travel at a slower speed and arrive at a relatively later time. Also the high frequencies are cut off for *b* and *c*, indicating attenuation. (Adopted from Boles and Boashash [32] and Boashash and Bazelaire [36].)

[32]. The different components are labeled *a*–*c*, where *a* is the original signal at the source. The delays are due to reflection from deeper layers which arrive at a later time. The bending upward at the high frequencies is a reflection

of the normal dispersion because the high frequencies travel more slowly and arrive at a relatively later time. For the components that are delayed longest and have traveled furthest, the high frequencies are cut off because of attenuation. The advantage of using a time-frequency description is that one can see all these effects in one picture. Potentially one may obtain the parameters by direct measurement of the delay, attenuation, and dispersion and thereby identify the media through which it propagated. Synthesis methods can be used to decompose the signal into its components. We have gone into some length in describing this type of situation because it is very typical of a variety of phenomena and will most likely be one of the common uses of time-frequency distributions. The implementation of the idealized picture described is currently complicated by various factors. In particular, if the Wigner distribution is used, in addition to the "real components" we have the cross terms, and for more than a few components the number of cross terms is very large. We also have noise. Also, because the layers are not uniform, the simple picture illustrated above becomes considerably more complicated. Boles and Boashash [32] have devised a number of methods to overcome these difficulties and have applied their analysis to real and simulated data. We have already seen that the new distribution of Choi and Williams [51] reduces the cross terms dramatically. It would certainly be interesting to apply that distribution to such a situation.

An innovative use of these distributions has been the work of Marinovic and Eichmann [134], [135], who developed a novel approach that does not depend on interpreting them as true distributions. The Wigner distribution is regarded as the kernel of an integral equation, and the corresponding eigenvalues and eigenfunctions are found. The expansion in terms of the eigenfunctions has been used for pattern recognition as the eigenvalues have been found to be effective classifiers of shapes. This decomposition may also be used to suppress the effects of noise in the Wigner distribution [136]. The noise is spread through all the terms of the decomposition and, retaining only the first few terms of the expansion, suppresses the noise considerably. The expansion method was applied by Marinovic and Smith [136] to show how the reconstructed distribution, which retains only the first few terms of the singular value expansion, allows one to extract the local frequency information in the case of echo signals which get corrupted by interference and noise.

Speech is one of the most complex nonstationary signals and a natural application for these time-frequency distributions. Chester and coworkers [49], [50] were the first to apply the Wigner distribution for the analysis and recognition of speech and constructed hardware for its calculation. They pointed out that the Wigner distribution has considerable noise sensitivity and that the interpretation is not straightforward, as with the spectrogram. However they found it useful for analysis and recognition. Their use of the Wigner distribution is based on the possibility that the characteristic for speech signals is very robust and hence may be used for recognition. Pickover and Cohen [156] used a number of distributions to study speech and found them difficult to interpret compared to the standard spectrogram. They pinpointed the difficulty with each distribution they considered. Riley [170] has considered the question of what kinds of distributions would be desirable to use in speech analysis and has used smoothed distributions to study formant structure. He devised a means of detecting and extracting the relevant speech parameters. Velez and Absher [193] used the smoothed Wigner distribution to display formant structure in speech and found it to be an effective clarifier of speech sounds. We have already mentioned the work of Janse and Kaizer [97] in regard to loudspeaker design. Preis [161] used the Wigner distribution to study various audio signals and found that the combined representation presents a clearer view of the various time-frequency quantities. Szu [189] has given a comprehensive discussion of the applications of bilinear distributions for the study of various acoustical signals, and in particular to questions of the signal processing involved in hearing.

Pattern recognition schemes use the distribution as a two-dimensional representation, not necessarily time and frequency [1], [30], [37], [65], [115]. Particular interpretations of a joint distribution as representing the energy is not required for this type of analysis.

Kumar and Carroll [116], [117] considered the use of the Wigner distribution function for the binary detection problem where one has to make a yes-no decision as to the existence of a signal in the presence of additive noise. They used the integral along the instantaneous frequency of the Wigner distribution as a statistic. The Wigner distribution method performed comparably to cross-correlation methods. They point out possible advantages for nonstationary signals of more complexity. A detection scheme for determining the instantaneous frequency of a chirp in additive noise was considered by Kay and Boudreaux-Bartels [104]. They used the likelihood ratio test to show that it is optimal to integrate the distribution along all straight lines in the time-frequency plane and choose the maximum value to compare to a threshold. This reflects the fact that for chirp-like signals the concentration of the distribution is maximum along the instantaneous frequency. Szu [188] has devised a number of methods to use these distributions for passive surveillance. He has taken advantage of the unique symmetry properties of the Wigner distribution. The synthesis method developed by Boudreaux-Bartels and Parks [39], [41] has been used to separate two signals with different characteristics. A general approach to the detection problem in the time-frequency plane has been formulated by Flandrin [78]. Harris and Abu Salem [88] have compared the performance of the Wigner distribution with other methods for the case of a sinusoid in the presence of additive white noise. They found that for the estimation of amplitude and frequency the Wigner distribution behaves poorly in noise but has the advantage that a priori knowledge of some of the characteristics of the signal is not required as is the case with some other methods. Cohen, Boudreaux-Bartels, and Kadambe [57] have devised a time-frequency approach to tracking mono- and multicomponent chirp signals in noise. Boashash and O'Shea [31] have extended the work of Kumar and Carroll and applied their method to the identification of underwater acoustic transients, in particular machine noise. They developed a general methodology for use of the Wigner distribution and cross Wigner distribution for detection problems.

The Wigner distribution has been used extensively as a tool to study radiance functions and coherence in optics, and methods for its production and optical filtering and display have been proposed and studied [12], [13], [16], [42], [44], [64], [86], [96], [103], [105], [149], [182], [184].

Linear predictive and autoregressive methods have been considered by Ramamoorthy *et al.* [165]. They showed that it results in good time and frequency resolutions, although they find that the interpretation is difficult. Boashash and coworkers [30], [129] showed that autoregressive methods can improve resolution if a careful choice is made of the parameters, otherwise spurious peaks occur which have no significance.

In a unique application Choi, Williams, and Zaveri [52] used the distribution discussed in Section III-G to evaluate the classify "event-related potentials" where certain words were used to induce brain wave responses in patients. The signal obtained is represented by the distribution and used to classify the signal in terms of the types of stimuli. They found the distribution given by Eq. (3.86) to be very effective as it reduces the masking effects of the cross terms.

Breed and Posch [43] have used the Wigner distribution to study an array of receivers and have formulated it in terms of the spatial parameters. They show that is provides a useful range and azimuth estimator. This approach works very well because for moderate ranges the signal has a quadratic phase spatial variation. These are precisely the cases that the Wigner distribution is well suited to handle, as we have seen in Section V. Swindlehurst and Kailath [185] have also devised a method for using the Wigner distribution for source localization for an array of receivers in the near-field approximation.

The relationship between the evolutionary spectrum of Priestley [162] and the Wigner distribution has been made by Hammond and Harrison [87].

The theory of these distributions has been applied to stochastic signals, and many of the original papers in the field addressed this aspect of the problem. This involves a further averaging to take into account the distribution of signals. Comprehensive work has been done by Janssen [99], Martin [139], Martin and Flandrin [140], and White [196]. Martin [139] has coined the phrase Wigner-Ville *spectrum* to indicate the Wigner distribution that has been ensemble averaged over the possible realizations of the signal. White [196] and White and Boashash [197], [198] have devised specific methods for obtaining the important parameters of a random process and have given expressions for the errors involved in estimating the parameters. Posch [159] has shown that if Eq. (4.10) is satisfied by the kernel, then the distribution will be the power spectrum when the input is a random stationary signal.

In concluding this summary of the applications, we emphasize that these distributions have not only been useful to study old ideas, but have also led to new concepts. An innovative concept has been introduced by Szu and Coulfield [186], where they address the question of how to compare the frequency contents of two signals. They devised a four-dimensional Rihaczek distribution, the variables being time and frequency for each signal. From this correlated distribution they compare the frequency contents of two signals.

X. Conclusion

In conclusion we discuss some general attitudes that have arisen in regard to these time-frequency distributions. The enigma of these distributions is that they sometimes give very reasonable results and sometimes absurd ones. For example, the Wigner distribution gives a very reasonable result for the first conditional moment of frequency, but an unreasonable one for the second conditional moment. A common attitude is that when we do get unacceptable results, we will know that the theory does not apply, and we will not use it for those situations. The problem with that point of view is how do you know when the results are absurd? Sometimes it is obvious, but not always. The fact that these distributions cannot be used in a consistent manner is one of the main areas that needs much further theoretical development.

One of the major issues in the field has always been which distribution, if any, is the absolute "best." There has been a general 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 is advocating. Also these lists very often contain requirements that are questionable and are obviously put in to force an issue. An example is the requirement that Moyal's formula should hold, but it is unclear why. If we found a distribution for which the Moyal formula did not hold but nevertheless behaved well, would we reject it on that basis? Clearly not. Witness the recent discovery of the Choi-Williams distribution. Another requirement commonly imposed is the finite support property, that is, for a finite-duration signal the distribution should be zero before the signal stops and after the signal ends. That certainly seems like a desirable condition for if there is no signal, we expect the distribution to be zero. (However, we know from Section V that there are distributions, for example, the Wigner, which have the finite-support property but are not necessarily zero in regions where the signal is zero.) But then the condition should simply be that the distribution should be zero if the signal is zero and not just the finite-support property. The positivity condition is also usually left out, although everyone concerned with choosing a best distribution mentions the advantage of having a positive distribution. We also point out that even plausible sounding conditions have to be applied carefully. An often stated requirement is that the first conditional moment of frequency be the derivative of the phase of the signal because it corresponds to instantaneous frequency. This may sound reasonable, but we already discussed in Section VIII the difficulties of making a total identification of instantaneous frequency, first conditional moment, and derivative of the phase. As has been pointed out in Section VIII, there is the theoretical difficulty that the derivative of the phase does not always correspond to the frequencies in the Fourier spectrum [74], [131]. This indicates that there may be a possible inherent inconsistency with the marginal requirement. This problem requires considerable further investigation. Also the requirement should be in terms of the analytic signal because that is how instantaneous frequency is defined. Moreover it is well known that the usefulness of the definition is meaningful only for certain types of signals, and therefore it is questionable whether we should insist that this hold for all signals. Given all these issues, it is not straightforward to set up the conditions for the satisfaction of the concept of instantaneous frequency.

Indeed, a comprehensive theory of a time-varying spectrum should predict what instantaneous frequency is.

Another approach is to argue that the performance of a distribution is best for a particular property that is deemed desirable. In a penetrating work Janssen [98] considered the performance of distributions for signals of the form $s(t) = e^{j\varphi(t)}$ and attempted to determine which distribution is more concentrated along the line $\omega = \varphi'(t)$. Toward that end he needed a method to determine the spread along that line. As we have seen, the concept of spread using these distributions is far from clear, so Janssen squared the distribution to avoid the fact that the distributions may go negative. Some have assumed that Janssen showed that the Wigner distribution has the least amount of spread around the derivative of the phase. However, Janssen proved this only for the class of distributions that have kernels of the form $\phi(\theta, \tau) = e^{ja\theta\tau}$. Also, we have seen that for multicomponent signals there are distributions that behave better than the Wigner distribution in the sense that the cross terms are smaller in magnitude. Hence it is far from clear whether "optimality" should be set up for mono or multicomponent signals, or perhaps neither.

Another common argument for elevating a particular distribution is to argue, for example, that all time- and shift-invariant distributions can be expressed as a "smoothed" version of it and therefore degraded in some sense. In particular it is often stated that the time- and shift-invariant distributions may be written in the form

$$P(t, \omega) = \iint g(t' - t, \omega' - \omega) W(t', \omega')\, dt'\, d\omega' \quad (10.1)$$

where $W(t, \omega)$ is the Wigner distribution, making it the "fundamental" one. However, we have seen in Section IV that we can equally well express the distributions in terms of, for example, the Rihaczek distribution.

Another view is that the choice of distribution should depend on the application and possibly the class of signals used, much in the same spirit as different window functions are chosen in various applications of the spectrogram or different sets of functions are used to expand the electrostatic potential depending on the geometry of the problem. As with expansions in terms of a complete set of functions, the choice is a matter of convenience, insight, and mathematical simplicity, which depends on the situation. Perhaps the proper attitude should be that the choice of distribution should be signal or application dependent. Indeed the recent work of Choi and Williams [51] and Nuttall [146], [147], which uses the ambiguity plane to choose the kernel, is an indication that different kernels may be appropriate for different signals. This makes kernels signal dependent, and hence the distributions are not necessarily bilinear any longer. Given these exciting developments it appears that at this stage of our knowledge, trying to prove which function is "best" is premature to say the least.

We now turn to what has been a fundamental issue with these joint distributions, and that is the positivity question. Everyone agrees that ideally a distribution should be positive since they are to be interpreted as densities. Many proofs have been given that positive distributions satisfying the marginals do not exist. The common plausibility argument relied on the uncertainty principle as discussed in Section VIII. A thorough and profound analysis of the positivity question has been given by Mugur-Schachter [144], who has identified many of the questionable and hidden assumptions that have gone into the proofs to show that they do not exist. Even before positive distributions were constructed, as explained in Section III, it was clear that there could not be any inherent reason for their nonexistence since the Wigner distribution is positive for certain signals. Park and Margenau [154], in their work on joint measurability, also analyzed the various arguments that have been given for the nonexistence of positive distributions and were able to construct a simple counterexample. Of course we now know that positive distributions are easily constructed, as in Section III-F, and that they do yield the correct marginals and the uncertainty principle.

It is a fact that distributions which are bilinear functionals of the signal cannot be positive for all signals [199], [200]. Joint densities and marginals appear in every field of science and engineering, and certainly bilinearity is never imposed on a distribution. Note that in time-frequency analysis the marginals themselves are bilinear in the signal, and hence bilinear distributions in the sense discussed here are joint distributions which are bilinear to the square root of the marginals. Now even the simplest proper joint distribution, the correlationless one $[P(x, y) = P_1(x) P_2(y)]$, is a product of the marginals and hence quartilinear to the square root of the marginals. In general, proper joint distributions are *highly* nonlinear functionals of the marginals. The possibility of using distributions that are not bilinear in the signal needs considerably more research. That is not to say that the class of bilinear distributions are not useful or desirable. However, we should be clear about the conceptual assumptions and interpretations.

Current knowledge has barely scratched the surface of the possible distributions and methodologies that may be used to describe a time-varying spectrum. There are an infinite number of distributions, and only a few have been explored. Although the concepts and techniques that have been developed in the past 40 years are truly impressive, it is clear that much more work lies ahead. The attempt to understand what a time-varying spectrum is, and to represent the properties of a signal simultaneously in time and frequency, is one of the most fundamental and challenging aspects of analysis.

Additional Comments

I would like to mention some recent results and some additions and omissions in the text.

Elimination of Aliasing in the Discrete Wigner Distribution. For a band-limited signal, the value of the signal at an arbitrary time can be obtained from discrete sampled values if the sampling is done at the Nyquist rate or higher, that is, at a sampling frequency $\omega_s \geq 2\omega_{max}$, where ω_{max} is the highest frequency in the signal. As mentioned in Section V-E, it has generally been believed that to reconstruct the Wigner distribution from discrete samples, one must sample the signal at twice this rate or higher; otherwise aliasing will occur. Nuttall [206] has recently shown that the higher sampling rate is unnecessary and has devised an efficient alias-free method for the computation of the Wigner distribution from a signal sampled at the Nyquist rate. The key to his approach is to take into account *all* the available information in the local autocorrelation function $R_t(\tau)$ as given by Eq. (3.64). By doubly Fourier transforming the local autocorrelation function, Nuttall has shown that nonoverlapping diamond-shaped regions exist in the transformed plane, each containing the fundmental information, provided that the signal was sampled at the Nyquist rate or higher. The end result of his analysis for obtaining the Wigner distribution at an arbitrary time-frequency pair is to construct first

$$\hat{S}(\omega) = \begin{cases} \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} s(kT)\, e^{-j\omega kT}, & \text{if } |\omega| \leq \frac{\pi}{T} \\ 0, & \text{otherwise} \end{cases}$$

where $s(kT)$ are the sampled signal values. The Wigner distribution is then constructed from $\hat{S}(\omega)$ according to

$$W(t, \omega) = \frac{1}{2\pi} \int \hat{S}^*\left(\omega - \frac{1}{2}\theta\right) e^{jt\theta} \hat{S}\left(\omega + \frac{1}{2}\theta\right) d\theta.$$

Nuttall has shown that this is the Wigner distribution of the original continuous signal $s(t)$ and is hence free of aliasing. In practice, both $\hat{S}(\omega)$ and $W(t, \omega)$ will be discretized in time and frequency and accomplished by fast Fourier transforms. We note that interpolation of the sampled values or reconstitution of the continuous signal from the sampled values is not necessary or utilized.

"Data-Adaptive" Distributions: As discussed in the Conclusion, the most common view point regarding time-frequency distributions is to find a "best" one, which will be used for all signals, although, as we mentioned, there are recent indications that different distributions may be appropriate for different signals. Recently Jones and Parks [207], [208] have made an important contribution in developing and implementing a "data-adaptive" method for devising a time-frequency distribution of a certain form. They consider the short-time Fourier transform with a Gaussian window where the parameters of the Gaussian window are varied for each different point in the time-frequency plane. The parameters are chosen so that the time-frequency concentration of the locally dominant component is maximized. They have applied this approach to both idealized and real data with considerable effectiveness in constructing distributions with high resolution.

Resolution Comparison: Jones and Parks [209] have made an interesting comparative study of the resolution properties of the Wigner distribution, spectrogram, and smoothed Wigner distribution. They used a signal composed of two Gaussian components with different time and frequency centers. They showed that for this case the best resolution (defined by the ability of the distribution to separate the two centers) was obtained by the spectrogram with an optimally matched window.

Bilinear Distributions: The approach discussed in Section III-D was extended in O'Connell and Wigner [210], where they considered the question of uniqueness in a distribution.

Local Second Moment: It was mentioned in the text that the second local moment of frequency corresponds to the local kinetic energy in quantum mechanics and that different expressions have been considered. The unified approach where previous expressions are special cases was presented in [211], and the references to previously known expressions are given therein.

Applications: Forrester [212] has applied the Wigner distribution to the study of vibrations of helicopter components with the aim of detecting developing failure of machine parts, in particular gear failure. By examining the vibrations of gears with and without faults he has shown that the Wigner distribution is an excellent discriminator and can be used to detect both the type and the extent of faults.

White and Boashash [213] developed a method for estimating the Wigner distribution for a nonstationary random process. They use a recursive procedure for the specification of estimators having desired characteristics in the particular regions of the time-frequency plane.

Acknowledgment

The author would like to express his appreciation to Dr. Albert Nuttall for the many conversations and valuable insights he has given him on a very wide variety of subjects. He also appreciates the help rendered by Dr. Chongmoon Lee in the preparation of the manuscript and he wishes to thank Profs. D. T. Barry, N. M. Marinovic, and W. J. Williams for supplying some of the graphs relating to their and their coworkers' respective research.

References

[1] R. M. S. S. Abeysekera and B. Boashash, "Time frequency domain features of ECG signals: their application in *P* wave detection using the cross Wigner-Ville distribution," in *Proc. IEEE ICASSP 89*, 1989 (to be published).

[2] M. H. Ackroyd, "Instantaneous and time-varying spectra—an introduction," *Radio Electron. Eng.*, vol. 239, pp. 145-152, 1970.

[3] —, "Short-time spectra and time-frequency energy distributions," *J. Acoust. Soc. Am.*, vol. 50, pp. 1229-1231, 1970.

[4] D. V. Ageyev, "The active band of the frequency spectrum of a function of time," *Trud. Gor'k. Politekh. In-ta*, vol. 11, p. 5, 1955.

[5] J. B. Allen and L. R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, pp. 1558-1564, 1977.

[6] R. Altes, "Detection, estimation and classification with spectrograms," *J. Acoust. Soc. Am.*, vol. 67, pp. 1232-1246, 1980.

[7] —, Tech. Memo. 322, Orincon, personal communication, Apr. 1984.

[8] M. G. Amin, "Time and lag window selection in Wigner-Ville distribution," in *Proc. IEEE ICASSP 87*, pp. 1529-1532, 1987.

[9] —, "Recursion in Wigner distribution," in *Advanced Algorithms and Architectures for Signal Processing III*, F. T. Luk, Ed., *Proc. SPIE*, vol. 975, pp. 221-231, 1989.

[10] J. C. Andrieux, M. R. Feix, G. Mourgues, P. Bertrand, B. Izrar, and V. T. Nguyen, "Optimum smoothing of the Wigner-Ville distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 764-769, 1987.

[11] E. H. Armstrong, "A method of reducing disturbances in radio signaling by a method of frequency-modulation," *Proc. IRE*, vol. 24, pp. 669-740, 1936.

[12] R. A. Athale, J. N. Lee, E. L. Robinson, and H. H. Szu, "Acousto-optic processors for real-time generation of time-frequency representations," *Opt. Lett.*, vol. 8, pp. 166-168, 1983.

[13] R. Bamler and H. Glunder, "The Wigner distribution function of two-dimensional signals coherent-optical generation and display," *Opt. Acta*, vol. 30, pp. 1789-1803, 1983.

[14] N. F. Barber and F. Ursell, "The response of a resonant system to a gliding tone," *Phil. Mag.*, vol. 39, pp. 345-361, 1948.

[15] D. T. Barry and N. M. Cole, "Muscle sounds are emitted at the resonant frequency of skeletal muscle," *IEEE Trans. Biomed. Eng.* (submitted for publication).

[16] H. O. Bartelt, K.-H. Brenner, and W. Lohmann, "The Wigner distribution function and its optical production," *Opt. Commun.*, vol. 32, pp. 32-38, 1980.

[17] M. J. Bastiaans, "Wigner distribution function and its application to first order optics," *J. Opt. Soc. Am.*, vol. 69, pp. 1710-1716, 1979.

[18] —, "Application of the Wigner distribution function to partially coherent light," *J. Opt. Soc. Am.*, vol. 3, pp. 1227-1238, 1986.

[19] E. Bazelaire and J. R. Viallix, "Theory of seismic noise," in *Proc. 49th Eur. Ass. Explor. Geophys. Mtg.* (Belgrade, Yugoslavia, 1987), pp. 1-2.
[20] E. Bedrosian, "A product theorem for Hilbert transforms," *Proc. IEEE*, vol. 51, pp. 686-689, 1963.
[21] P. Bertrand and J. Bertrand, "Time-frequency representations of broad band signals," in *The Physics of Phase Space*, Y. S. Kim and W. W. Zachary, Eds. New York, NY: Springer, 1987, pp. 250-252.
[22] J. Bertrand and P. Bertrand, "Tomographic procedure for constructing phase space representations," in *The Physics of Phase Space*, Y. S. Kim and W. W. Zachary, Eds. New York, NY: Springer, 1987, pp. 208-210.
[23] P. Bertrand, B. Izrar, V. T. Nguyen, and M. R. Feix, "Obtaining non-negative quantum distribution functions," *Phys. Lett.*, vol. 94, pp. 415-417, 1983.
[24] A. Blanc-Lapierre and B. Picinbono, "Remarques sur la notion de spectre instantané de puissance," *Pub. Sci. Univ. d'Alger*, s. B1, pp. 2-32, 1955.
[25] B. Boashash and P. J. Black, "An efficient real-time implementation of the Wigner-Ville distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1611-1618, 1987.
[26] B. Boashash, "Time-frequency signal analysis, the choice of a method and its applications," in *Advances in Spectral Analysis*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall (to be published).
[27] B. Boashash, B. Lovell, and H. J. Whitehouse, "High-resolution time-frequency signal analysis by parametric modeling of the Wigner-Ville distribution," in *Proc. IASTED Int. Symp. Signal Processing and Its Applications* (Brisbane, Australia, 1987), pp. 297-301.
[28] B. Boashash and H. J. Whitehouse, "Seismic applications of the Wigner-Ville distribution," in *Proc. IEEE Int. Conf. Systems and Circuits*, pp. 34-37, 1986.
[29] B. Boashash, "Non-stationary signals and the Wigner-Ville distribution," in *Proc. IASTED Int. Symp. Signal Processing and Its Applications* (Brisbane, Australia, 1987), pp. 261-268.
[30] B. Boashash, B. Lovell, and L. White, "Time frequency analysis and pattern recognition using singular value decomposition of the Wigner-Ville distribution," in *Advanced Algorithms and Architecture for Signal Processing, Proc. SPIE*, vol. 828, pp. 104-114, 1987.
[31] B. Boashash and P. O'Shea, "A methodology for detection and classification of underwater acoustic signals using time-frequency analysis techniques," preprint, 1989 (to be published).
[32] P. Boles and B. Boashash, "The cross Wigner-Ville distribution—a two-dimensional analysis method for the processing of vibrosis seismic signals," in *Proc. IEEE ICASSP 87*, pp. 904-907, 1988.
[33] M. Born and P. Jordan, "Zur Quantenmechanik, " *Z. Phys.*, vol. 34, pp. 858-888, 1925.
[34] F. Bopp, "La méchanique quantique est-elle une méchanique statistique classique particulière?," *Ann. Inst. H. Poincaré*, vol. 15, pp. 81-112, 1956.
[35] B. Bouachache, "Représentation temps-fréquence," Soc. Nat. ELF Aquitaine, Pau, France, Publ. Recherches, no. 373-378, 1978.
[36] B. Bouachache and E. de Bazelaire, "Pattern recognition in the time-frequency domain by means of the Wigner-Ville distribution," in *Proc. GRETSI 9th Colloq.*, pp. 879-884, 1983.
[37] B. Bouachache and F. Rodriguez, "Recognition of time-varying signals in the time-frequency domain by means of the Wigner distribution," in *Proc. IEEE ICASSP 84*, pp. 22.5/1-4, 1984.
[38] G. F. Boudreaux-Bartels and T. W. Parks, "Reducing aliasing in the Wigner distribution using implicit spline interpolation," in *Proc. IEEE ICASSP 83*, pp. 1438-1441, 1983.
[39] G. F. Boudreaux-Bartels and T. W. Parks, "Signal estimation using modified Wigner distributions," in *Proc. IEEE ICASSP 84*, pp. 22.3/1-4, 1984.
[40] G. F. Boudreaux-Bartels and T. W. Parks, "Time-varying filtering and signal estimation using Wigner distribution synthesis techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 442-451, 1986.
[41] G. F. Boudreaux-Bartels, "Time varying signal processing using the Wigner-distribution time-frequency signal representation," in *Advances in Geophysical Data Processing*, M. Simaan, Ed. Greenwich, CT: JAI Press, 1985, pp. 33-79.
[42] H. Brand, R. Graham, and A. Schenzle, "Wigner distribution of bistable steady states in optical subharmonic generation," *Opt. Commun.*, vol. 32, pp. 359-363, 1980.
[43] B. R. Breed and T. E. Posch, "A range and azimuth estimator based on forming the spatial Wigner distribution," in *Proc. IEEE ICASSP 84*, pp. 41B/9.1-9.2, 1984.
[44] K. H. Brenner and A. W. Lohmann, "Wigner distribution function display of complex 1D signals," *Opt. Commun.*, vol. 42, pp. 310-314, 1982.
[45] J. R. Carson and T. C. Fry, "Variable frequency electric circuit theory with application to the theory of frequency modulation," *Bell Sys. Tech. J.*, vol. 16, pp. 513-540, 1937.
[46] N. Cartwright, "Correlations without joint distributions in quantum mechanics," *Found. Phys.*, 4, pp. 127-136, 1974.
[47] —, "A non-negative quantum mechanical distribution function," *Physica*, vol. 83A, pp. 210-212, 1976.
[48] D. S. K. Chan, "A non-aliased discrete-time Wigner distribution for time-frequency signal analysis," in *Proc. IEEE ICASSP 82*, pp. 1333-1336, 1982.
[49] D. Chester, F. J. Taylor, and M. Doyle, "On the Wigner distribution," in *Proc. IEEE ICASSP 83*, pp. 491-494, 1983.
[50] D. Chester, F. J. Taylor, and M. Doyle, "The Wigner distribution in speech processing applications," *J. Franklin Inst.*, vol. 318, pp. 415-430, 1984.
[51] H. I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, 1989 (in press).
[52] H. I. Choi, W. J. Williams, and H. Zaveri, "Analysis of event related potentials: time frequency energy distribution," in *Proc. Rocky Mountain Bioengineering Symp.*, pp. 251-258, 1987.
[53] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The aliasing problem in discrete-time Wigner distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1067-1072, 1983.
[54] 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 J. Res.*, vol. 35, pp. 217-250, 1980.
[55] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution—a tool for time-frequency signal analysis; part II: discrete time signals," *Philips J. Res.*, vol. 35, pp. 276-300, 1980.
[56] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution—a tool for time-frequency signal analysis; part III: relations with other time-frequency signal transformations," *Philips J. Res.*, vol. 35, pp. 372-389, 1980.
[57] F. S. Cohen, G. F. Boudreaux-Bartels, and S. Kadambe, "Tracking of unknown non-stationary chirp signal using unsupervised clustering in the Wigner distribution space," in *Proc. IEEE ICASSP 88*, pp. 2180-2183, 1988.
[58] L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7, pp. 781-786, 1966.
[59] —, "On a fundamental property of the Wigner distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 559-561, 1987.
[60] —, "Wigner distribution for finite duration or band-limited signals and limiting cases," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 796-806, 1987.
[61] —, "Quantization problem and variational principle in the phase space formulation of quantum mechanics," *J. Math. Phys.*, vol. 17, pp. 1863-1866, 1976.
[62] L. Cohen and T. Posch, "Positive time-frequency distribution functions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 31-38, 1985.
[63] L. Cohen and T. Posch, "Generalized ambiguity functions," in *Proc. IEEE ICASSP 85*, pp. 1025-1028, 1985.
[64] M. Conner and L. Yao, "Optical generation of the Wigner distribution of 2-D real signals," *Appl. Opt.*, vol. 24, pp. 3825-3829, 1985.
[65] G. Cristobal, J. Bescos, and J. Santamaria, "Application of

the Wigner distribution for image representation and analysis," in *Proc. IEEE 8th Int. Conf. Pattern Recognition*, pp. 998–1000, 1986.

[66] N. G. DeBruijn, "Uncertainty principles in Fourier analysis," in *Inequalities*, O. Shisha, Ed. New York, NY: Academic Press, 1967, pp. 57–71.

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

[68] A. Dziewonski, S. Bloch, and M. Landisman, "A technique for the analysis of transient signals," *Bull. Seism. Soc. Am.*, vol. 59, pp. 427–444, 1969.

[69] G. Eichmann and N. M. Marinovic, "Scale-invariant Wigner distribution and ambiguity functions," in *Proc. Int. Soc. Opt. Eng., Proc. SPIE*, vol. 519, pp. 18–24, 1985.

[70] G. Eichmann and B. Z. Dong, "Two-dimensional optical filtering of 1-D signals," *Appl. Opt.*, vol. 21, pp. 3152–3156, 1982.

[71] B. Escudie and J. Grea, "Sur une formulation générale de la représentation en temps et fréquence dans l'analyse de signaux d'énergie finie," *C. R. Acad. Sci. Paris*, vol. A283, pp. 1049–1051, 1976.

[72] R. M. Fano, "Short-time autocorrelation functions and power spectra," *J. Acoust. Soc. Am.*, vol. 22, pp. 546–550, 1950.

[73] P. D. Finch and R. Groblicki, "Bivariate probability densities with given margins," *Found. Phys.*, vol. 14, pp. 549–552, 1984.

[74] L. M. Fink, "Relations between the spectrum and instantaneous frequency of a signal," transl. in *Probl. Inform. Transm.*, vol. 2, pp. 11–21, 1966.

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

[76] P. Flandrin, "Some features of time-frequency representations of multi-component signals," in *Proc. IEEE ICASSP 84*, pp. 41B.4.1–4.4, 1984.

[77] P. Flandrin and B. Escudie, "Time and frequency representation of finite energy signals: a physical property as a result of a Hilbertian condition," *Signal Processing*, vol. 2, pp. 93–100, 1980.

[78] P. Flandrin, "A time-frequency formulation of optimum detection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 1377–1384, 1988.

[79] L. E. Franks, *Signal Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1969.

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

[81] H. Garudadri, M. P. Beddoes, A.-P. Benguerel, and J. H. V. Gilbert, "On computing the smoothed Wigner distribution," in *Proc. IEEE ICASSP 87*, pp. 1521–1524, 1987.

[82] S. K. Ghosh, M. Berkowitz, and R. G. Parr, "Transcription of ground-state density-functional theory into a local thermodynamics," in *Proc. Nat. Acad. Sci.*, vol. 81, pp. 8028–8031, 1984.

[83] O. D. Grace, "Instantaneous power spectra," *J. Acoust. Soc. Am.*, vol. 69, pp. 191–198, 1981.

[84] R. Groblicki, "A generalization of Cohen's theorem," *J. Math. Phys.*, vol. 24, pp. 841–843, 1983.

[85] M. S. Gupta, "Definitions of instantaneous frequency and frequency measurability," *Am. J. Phys.*, vol. 43, pp. 1087–1088, 1975.

[86] A. K. Gupta and T. Askura, "New optical system for the efficient display of Wigner distribution functions using a single object transparency," *Opt. Commun.*, vol. 60, pp. 265–268, 1986.

[87] J. K. Hammond and R. F. Harrison, "Wigner-Ville and evolutionary spectra for covariance equivalent nonstationary random process," in *Proc. IEEE ICASSP 85*, pp. 1025–1027, 1985.

[88] F. J. Harris and A. Abu Salem, "Performance comparison of Wigner-Ville based techniques to standard FM-discriminators for estimating instantaneous frequency of a rapidly slewing FM sinusoid in the presence of noise," in *Advanced Algorithms and Architectures for Signal Processing III*, F. T. Luk, Ed., *Proc. SPIE*, vol. 975, pp. 232–244, 1989.

[89] M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, "Distribution functions in physics: fundamentals," *Phys. Rep.*, vol. 106, pp. 121–167, 1984.

[90] F. Hlawatsch, "Transformation, inversion and conversion of bilinear signal representations," in *Proc. IEEE ICASSP 85*, pp. 1029–1032, 1985.

[91] ——, "Interference terms in the Wigner distribution," in *Digital Signal Processing 84*, V. Cappellini and A. G. Constantinides, Eds. Amsterdam, The Netherlands: Elsevier (North Holland), 1984.

[92] (a) ——, "A note on 'Wigner distribution for finite duration or band-limited signals and limiting cases'," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 927–929, 1988; (b) author's reply, *ibid.*

[93] R. L. Hudson, "When is the Wigner quasi-probability density non-negative?," *Rep. Math. Phys.*, vol. 6, pp. 249–252, 1974.

[94] J. Imberger and B. Boashash, "Signal processing in oceanography," in *Proc. IASTED Int. Symp. Signal Processing and Its Applications* (Brisbane, Australia, 1987), pp. 8–9.

[95] J. Imberger and B. Boashash, "Application of the Wigner-Ville distribution to temperature gradient microstructure: a new technique to study small-scale variations," *J. Phys. Oceanog.*, vol. 16, pp. 1997–2012, 1986.

[96] T. Iwai, A. K. Gupta, and T. Asakura, "Simultaneous optical production of the sectional Wigner distribution function for a two-dimensional object," *Opt. Commun.*, vol. 58, pp. 15–19, 1986.

[97] C. P. Janse and J. M. Kaizer, "Time-frequency distributions of loudspeakers: the application of the Wigner distribution," *J. Audio Eng. Soc.*, vol. 31, pp. 198–223, 1983.

[98] A. J. E. M. Janssen, "On the locus and spread of psuedo-density functions in the time-frequency plane," *Philips J. Res.*, vol. 37, pp. 79–110, 1982.

[99] ——, "Application of Wigner distributions to harmonic analysis of generalized stochastic processes," Math.-Centrum, Amsterdam, The Netherlands, MC-tract 114, 1979.

[100] ——, "Positivity properties of phase-plane distribution functions," *J. Math. Phys.*, vol. 25, pp. 2240–2252, 1984.

[101] A. J. E. M. Janssen and T. C. M. Claasen, "On positivity of time-frequency distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1029–1032, 1985.

[102] (a) A. J. E. M. Janssen, "A note on 'Positive time-frequency distributions'," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 701–704, 1987. (b) Author's reply, *ibid.*

[103] J.-Z. Jiao, B. Wang, and L. Hong, "Wigner distribution function and optical geometrical transformation," *Appl. Opt.*, vol. 23, pp. 1249–1254, 1984.

[104] S. Kay and G. F. Boudreaux-Bartels, "On the optimality of the Wigner distribution for detection," in *Proc. IEEE ICASSP 85*, pp. 1017–1020, 1985.

[105] O. Kenny and B. Boashash, "An optical signal processor for time-frequency signal analysis using the Wigner-Ville distribution," *J. Elec. Electron. Eng.* (Australia), pp. 152–158, 1988.

[106] E. L. Key, E. M. Fowle, and R. D. Haggarty, "A method of designing signals of large time-bandwidth product," in *IRE Int. Conv. Rec.*, vol. 4, pp. 146–154, 1961.

[107] J. G. Kirkwood, "Quantum statistics of almost classical ensembles," *Phys. Rev.*, vol. 44, pp. 31–37, 1933.

[108] J. R. Klauder, "The design of radar signals having both high range resolution and high velocity resolution," *Bell Sys. Tech. J.*, vol. 39, pp. 809–820, 1960.

[109] J. R. Klauder, A. C. Price, S. Darlington, and W. J. Albershein, "The theory and design of chirp radars," *Bell Sys. Tech. J.*, vol. 39, pp. 745–808, 1960.

[110] F. Kobayashi and H. Suzuki, "Time-varying signal analysis using squared analytic signals," in *Proc. IEEE ICASSP 87*, pp. 1525–1528, 1987.

[111] K. Kodera, C. de Villedary, and R. Gendrin, R, "A new method for the numerical analysis of non-stationary signals," *Phys. Earth and Planetary Interiors*, vol. 12, pp. 142–150, 1976.

[112] K. Kodera, R. Gendrin, and C. de Villedary, "Analysis of time-varying signals with small BT values," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 64–76, 1978.

[113] R. Koenig, H. K. Dunn, and L. Y. Lacy, "The sound spectrograph," *J. Acoust. Soc. Am.*, vol. 18, pp. 19–49, 1946.

[114] J. G. Kruger and A. Poffyn, "Quantum mechanics in phase space," *Physica*, vol. 85, pp. 84–100, 1976.

[115] B. V. K. V. Kumar and W. C. Carroll, "Pattern recognition using Wigner distribution function," in *IEEE Proc 10th Int. Optical Computing Conf.*, pp. 130–135, 1983.

[116] B. V. K. V. Kumar and W. C. Carroll, "Effects of sampling on signal detection using the cross-Wigner distribution function," *Appl. Opt.*, vol. 23, pp. 4090–4094, 1984.
[117] B. V. K. V. Kumar and W. C. Carroll, "Performance of Wigner distribution function based detection methods," *Opt. Eng.*, vol. 23, pp. 732–737, 1984.
[118] V. V. Kuryshkin, I. Lyabis, and Y. I. Zaparovanny, "Sur le problème de la règle de correspondance en théorie quantique," *Ann. Fondation Louis de Broglie*, vol. 3, pp. 45–61, 1978.
[119] V. V. Kuryshkin, "Some problems of quantum mechanics possessing a non-negative phase-space distribution function," *Int. J. Theor. Phys.*, vol. 7, pp. 451–466, 1973.
[120] —, "Uncertainty principle and the problems of joint coordinate-momentum probability density in quantum mechanics," in *The Uncertainty Principle and the Foundations of Quantum Mechanics*. New York, NY: Wiley, 1977, pp. 61–83.
[121] J.-L. Lacoume and W. Kofman, "Etude des signaux non stationnaires par la représentation en temps et en fréquence," *A. Telec.*, vol. 30, pp. 231–238, 1975.
[122] D. G. Lampard, "Generalization of the Wiener-Khintchine theorem to nonstationary processes," *J. Appl. Phys.*, vol. 25, p. 802, 1954.
[123] C. Lee and L. Cohen, "Instantaneous frequency, its standard deviation and multicomponent signals," in *Advanced Algorithms and Architectures for Signal Processing III*, F. T. Luk, Ed., *Proc. SPIE*, vol. 975, pp. 186–208, 1989.
[124] R. M. Lerner, "Representation of signals," in *Lectures on Communications System Theory*, E. J. Baghdady, Ed. New York, NY: McGraw-Hill, 1961, pp. 203–242.
[125] M. J. Levin, "Instantaneous spectra and ambiguity functions," *IEEE Trans. Informat. Theory*, vol. IT-13, pp. 95–97, 1967.
[126] A. L. Levshin, V. F. Pisarenko, and G. A. Pogrebinsky, "On a frequency-time analysis of oscillations," *Ann. Geophys.*, vol. 28, pp. 211–218, 1972.
[127] S. C. Liu, "Time-varying spectra and linear transformation," *Bell Sys. Tech. J.*, vol. 50, pp. 2365–2374, 1971.
[128] R. M. Loynes, "On the concept of spectrum for nonstationary processes," *J. R. Stat. Soc.*, ser. B, vol. 30, pp. 1–30, 1968.
[129] B. Lovell and B. Boashash, "Evaluation of criteria for detection of change in nonstationary signals," in *Proc. IASTED Int. Symp. Signal Processing and Its Applications* (Brisbane, Australia, 1987), pp. 291–298.
[130] D. Lowe, "Joint representations in quantum mechanics and signal processing theory: why a probability function of time and frequency is disallowed," Royal Signals and Radar Establishment, Memo. 4017, 1986.
[131] L. Mandel, "Interpretation of instantaneous frequency," *Am. J. Phys.*, vol. 42, pp. 840–846, 1974.
[132] H. Margenau and L. Cohen, "Probabilities in quantum mechanics," in *Quantum Theory and Reality*, M. Bunge, Ed. New York, NY: Springer, 1967.
[133] H. Margenau and R. N. Hill, "Correlation between measurements in quantum theory," *Prog. Theor. Phys.*, vol. 26, pp. 722–738, 1961.
[134] N. M. Marinovic and G. Eichmann, "Feature extraction and pattern classification in spatial frequency domain," in *Proc. Conf. Intelligent Robots and Computer Vision, Proc. SPIE*, vol. 579, pp. 15–20, 1985.
[135] N. M. Marinovic and G. Eichmann, "An expansion of Wigner distribution and its applications," in *Proc. IEEE ICASSP 85*, pp. 1021–1024, 1985.
[136] N. M. Marinovic and W. A. Smith, "Suppression of noise to extract the intrinsic frequency variation from an ultrasonic echo," in *Proc. IEEE 1985 Ultrasonics Symp.*, pp. 841–846, 1985.
[137] N. M. Marinovic and W. A. Smith, "Application of joint time-frequency distributions to ultrasonic transducers," in *Proc. 1986 IEEE Int. Symp. Circuits and Systems*, pp. 50–54, 1986.
[138] W. D. Mark, "Spectral analysis of the convolution and filtering of non-stationary stochastic processes," *J. Sound Vib.*, vol. 11, pp. 19–63, 1970.
[139] W. Martin, "Time-frequency analysis of random signals," in *Proc. IEEE ICASSP 82*, pp. 1325–1328, 1982.
[140] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 1461–1470, 1985.
[141] W. F. G. Mecklenbrauker, "A tutorial on non-parametric bilinear time-frequency signal representations," in *Signal Processing*, J. L. Lacoume, T. S. Durrani, and R. Stora, Eds. Amsterdam, The Netherlands: North Holland, 1985.
[142] G. Mourgues, M. R. Feix, J. C. Andrieux, and P. Bertrand, "Not necessary but sufficient conditions for the positivity of generalized Wigner functions," *J. Math. Phys.*, vol. 26, pp. 2554–2555, 1985.
[143] J. E. Moyal, "Quantum mechanics as a statistical theory," *Proc. Cambridge Phil. Soc.*, vol. 45, pp. 99–124, 1949.
[144] M. Mugur-Schachter, "A study of Wigner's theorem on joint probabilities," *Found. Phys.*, vol. 9, pp. 389–404, 1979.
[145] A. H. Nuttall, "On the quadrature approximation to the Hilbert transform of modulated signals," *Proc. IEEE*, vol. 54, p. 1458, 1966; E. Bedrosian, author's reply, *ibid.*
[146] —, "Wigner distribution function: relation to short-term spectral estimation, smoothing, and performance in noise," Naval Underwater Systems Center, NUSC Tech. Rep. 8225, Feb. 16, 1988.
[147] —, "The Wigner distribution function with minimum spread," Naval Underwater Systems Center, NUSC Tech. Rep. 8317, June 1, 1988.
[148] —, private communication.
[149] J. Ojeda-Castaneda and E. E. Sicre, "Bilinear optical systems, Wigner distribution function and ambiguity function representations," *Opt. Acta*, vol. 31, pp. 255–260, 1984.
[150] A. V. Oppenheim, "Speech spectrograms using the fast Fourier transform," *IEEE Spectrum*, vol. 7, pp. 57–62, 1970.
[151] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[152] C. H. Page, "Instantaneous power spectra," *J. Appl. Phys.*, vol. 23, pp. 103–106, 1952.
[153] A. Papoulis, *Signal Analysis*. New York, NY: McGraw-Hill, 1977.
[154] J. L. Park and H. Margenau, "Simultaneous measurability in quantum theory," *Int. J. Theor. Phys.*, vol. 1, pp. 211–283, 1968.
[155] F. Peyrin and R. Prost, "A unified definition for the discrete-time discrete-frequency, and discrete-time/frequency Wigner distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 858–867, 1986.
[156] C. Pickover and L. Cohen, "A comparison of joint time-frequency distributions for speech signals," *Proc. IEEE Int. Conf. Systems and Circuits*, vol. 1, pp. 42–45, 1986.
[157] C. Piquet, *C. R. Acad. Sci. Paris*, vol. A279, pp. 107–109, 1974.
[158] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 55–69, 1980.
[159] T. E. Posch, "Time frequency distributions for a wide sense stationary random signal," in *Proc. IEEE Int. Conf. Systems and Circuits*, 1989 (to be published).
[160] R. K. Potter, G. Kopp, and H. C. Green, *Visible Speech*. New York, NY: Van Nostrand, 1947.
[161] D. Preis, "Phase distortion and phase equalization in audio signal processing—a tutorial review," *J. Audio Eng. Soc.*, vol. 30, pp. 774–794, 1982.
[162] M. B. Priestley, *Spectral Analysis and Time Series*. New York: NY: Academic Press, 1981.
[163] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[164] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
[165] P. A. Ramamoorthy, V. K. Iyer, and Y. Ploysongsang, "Autoregressive modeling of the Wigner spectrum," in *Proc. IEEE ICASSP 87*, pp. 1509–1512, 1987.
[166] H. Reichenbach, *Philosophical Foundations of Quantum Mechanics*. Berkeley, CA: Univ. California Press, 1944.
[167] W. Rihaczek, "Signal energy distribution in time and frequency," *IEEE Trans. Informat. Theory*, vol. IT-14, pp. 369–374, 1968.
[168] —, *Principles of High-Resolution Radar*. New York, NY: McGraw-Hill, 1969.
[169] —, "Hilbert transforms and the complex representation of

real signals," *Proc. IEEE*, vol. 54, pp. 434–435, 1966; E. Bedrosian, authors' reply, *ibid.*
[170] M. D. Riley, "Beyond quasi-stationarity: designing time-frequency representations for speech signals," in *Proc. IEEE ICASSP 87*, pp. 657–660, 1987.
[171] G. J. Ruggeri, "On phase space description of quantum mechanics," *Prog. Theo. Phys.*, vol. 46, pp. 1703–1712, 1971.
[172] B. E. A. Saleh, "Optical bilinear transformations—general properties," *Opt. Acta*, vol. 26, pp. 777–799, 1979.
[173] B. E. A. Saleh and N. S. Subotic, "Time-variant filtering of signals in the mixed time-frequency domain," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1479–1485, 1985.
[174] R. W. Schafer and L. R. Rabiner, "Design and simulation of a speech analysis-synthesis system based on short-time Fourier analysis," *IEEE Trans. Audio Electrocoust.*, vol. AU-21, pp. 165–174, 1973.
[175] M. R. Schroeder and B. S. Atal, "Generalized short-time power spectra and autocorrelation functions," *J. Acoust. Soc. Am.*, vol. 34, pp. 1679–1683, 1962.
[176] B. Schweizer and A. Sklar, "Probability distributions with given margins: note on a paper by Finch and Groblicki," *Found. Phys.*, vol. 16, pp. 1061–1064, 1986.
[177] M. I. Scolnik, *Introduction to Radar Systems.* New York, NY: McGraw-Hill, 1980.
[178] J. Shekel, "'Instantaneous' frequency," *Proc. IRE*, vol. 41, p. 548, 1953.
[179] F. Soto and P. Claverie, "Some properties of the smoothed Wigner function," *Physica*, vol. 109A, pp. 193–207, 1981.
[180] F. Soto and P. Claverie, "When is the Wigner distribution function of multidimensional systems nonnegative?," *J. Math. Phys.*, vol. 24, pp. 97–100, 1983.
[181] M. D. Srinivas and E. Wolf, "Some nonclassical features of phase-space representations of quantum mechanics," *Phys. Rev.*, vol. D11, pp. 1477–1485, 1975.
[182] S. Stein and H. H. Szu, "Two dimensional optical processing of one-dimensional acoustic data," *Opt. Eng.*, vol. 21, pp. 804–813, 1982.
[183] L. R. O. Storey, "An investigation of whistling atmospherics," *Phil. Trans. R. Soc.*, vol. A246, pp. 113–141, 1953.
[184] N. S. Subotic and B. E. A. Saleh, "Optical time-variant processing of signs in the mixed time-frequency domain," *Opt. Commun.*, vol. 52, pp. 259–264, 1984.
[185] A. L. Swindlehurst and T. Kailath, "Near-field source parameter estimation using a spatial Wigner distribution approach," in *Advanced Algorithms and Architectures for Signal Processing III*, T. Luk, Ed., *Proc. SPIE*, vol. 975, pp. 86–92, 1989.
[186] H. H. Szu and H. J. Coulfield, "The mutual time-frequency content of two signals," *Proc. IEEE*, vol. 72, pp. 902–908, 1984.
[187] H. H. Szu and J. Blodgett, "Wigner distribution and ambiguity functions," in *Optics in Four Dimensions*, L. M. Narducci, Ed. New York, NY: Am. Inst. of Physics, 1981, pp. 355–381.
[188] H. H. Szu, "Two dimensional optical processing of one-dimensional acoustic data," *Opt. Eng.*, vol. 21, pp. 804–813, 1982.
[189] —, "Signal processing using bilinear and nonlinear time-frequency joint distributions," in *The Physics of Phase Space*, Y. S. Kim and W. W. Zachary, Eds. New York, NY: Springer, 1987, pp. 179–199.
[190] C. H. M. Turner, "On the concept of an instantaneous spectrum and its relationship to the autocorrelation function," *J. Appl. Phys.*, vol. 25, pp. 1347–1351, 1954.
[191] D. Y. Vakman, "Do we know what are the instantaneous frequency and instantaneous amplitude of a signal?," transl. in *Radio Eng. Electron. Phys.*, vol. 21, pp. 95–100, June 1976.
[192] B. Van der Pol, "The fundamental principles of frequency modulation," *J. IEE (London)*, vol. 93, pp. 153–158, 1946.
[193] E. F. Valez and R. G. Absher, "Transient analysis of speech signals using the Wigner time-frequency representation," in *Proc. IEEE ICASSP 89*, 1989 (to be published).
[194] J. Ville, "Théorie et applications de la notion de signal analytique," *Cables et Transmission*, vol. 2A, pp. 61–74, 1948.
[195] A. Vogt, "Position and momentum distribuions do not determine the quantum mechanical state," in *Mathematical Foundations of Quantum Theory*, A. R. Marlow, Ed. New York, NY: Academic Press, 1978.
[196] L. B. White, "Some aspects of the time-frequency analysis of random processes," thesis, University of Queensland, Australia, 1989.
[197] L. B. White and B. Boashash, "On estimating the instantaneous frequency of a Gaussian random signal by the use of the Wigner-Ville distribution," *IEEE Trans. Acoust, Speech, Signal Processing*, vol. ASSP-36, pp. 417–420, 1988.
[198] L. B. White and B. Boashash, "Time-frequency coherence—a theoretical basis for cross spectral analysis of non stationary signals," in *Proc. IASTED Int. Symp. Signal Processing and Its Applications* (Brisbane, Australia, 1987), pp. 18–23.
[199] E. P. Wigner, "On the quantum correction for thermodynamic equilibrium," *Phys. Rev.*, vol. 40, pp. 749–759, 1932.
[200] —, "Quantum-mechanical distribution functions revisited," in *Perspectives in Quantum Theory*, W. Yourgrau and A. van der Merwe, Eds. Cambridge, MA: M.I.T. Press, 1971.
[201] R. M. Wilcox, "Exponential operators and parameter differentiation in quantum physics," *J. Math. Phys.*, vol. 8, pp. 962–982, 1967.
[202] P. M. Woodward, *Probability and Information Theory with Application to Radar.* London, England: Pergamon, 1953.
[203] K.-B. Yu and S. Cheng, "Signal synthesis from Wigner distribution," in *Proc. IEEE ICASSP 85*, pp. 1037–1040, 1985.
[204] K.-B. Yu, "Signal representation and processing in the mixed time-frequency domain," in *Proc. IEEE ICASSP 87*, pp. 1513–1516, 1987.
[205] L. A. Zadeh, "Frequency analysis of variable networks," *Proc. IRE*, vol. 38, p. 291, 1950.
[206] A. H. Nuttall, "Alias-free discrete-time Wigner distribution function and complex ambiguity function," Naval Underwater Systems Center, NUSC Tech. Rep. 8533, April 14, 1989.
[207] D. L. Jones, "A high-resolution data-adaptive time-frequency representation," thesis, Rice University, Houston, TX, 1988.
[208] D. L. Jones and T. W. Parks, "A high-resolution data-adaptive time-frequency representation," in *Proc. IEEE ICASSP 87*, pp. 681–683.
[209] D. L. Jones and T. W. Park, "A resolution comparison of several time-frequency representations," *IEEE Trans. Acoust., Speech, Signal Processing*, 1989 (submitted for publication).
[210] R. F. O'Connell and E. P. Wigner, "Quantum-mechanical distribution functions: conditions for uniqueness," *Phys. Lett.*, vol. 83A, pp. 145–148, 1981.
[211] L. Cohen, "Local kinetic energy in quantum mechanics," *J. Chem. Phys.*, vol. 70, pp. 788–789, 1979.
[212] B. D. Forrester, "Use of the Wigner-Ville distribution in helicopter fault detection," in *Proc. ASSPA 89*, pp. 78–82, 1989.
[213] L. B. White and B. Boashash, "A design methodology for the optimal estimation of the spectral characteristics of non-stationary random signals," in *Proc. ASSP Workshop on Spectral Estimation and Modeling*, pp. 65–70, 1988.

Leon Cohen received the B.S. degree from City College, New York, in 1962 and the Ph.D. degree from Yale University, New Haven, CT, in 1966, both in physics.

He is currently Professor of Physics at Hunter College and Graduate Center of the City University of New York. His research interests are in astronomy, quantum mechanics, chemical physics, numerical methods, and signal analysis. In the past few years he has held visiting research positions at the IBM T. J. Watson Research Center, the Center for Advanced Studies at the University of New Mexico, and the Naval Underwater Systems Center.