Finite Word-Length Effects in Implementation of Distributions for Time–Frequency Signal Analysis

Veselin Ivanović, LJubiša Stanković, and Dušan Petranović

***Abstract*—This correspondence presents an analysis of the finite register length influence on the accuracy of results obtained by the time–frequency distributions (TFD's). In order to measure quality of the obtained results, the variance of the proposed model is found, signal-to-quantization noise ratio (SNR) is defined, and appropriate expressions are derived. Floating- and fixed-point arithmetic are considered, with the analysis of discrete random and discrete deterministic signals. It is shown that commonly used reduced interference distributions (RID's) exhibit similar performance with respect to the SNR. We have also derived the expressions establishing the relationship between the number of bits and the required quality of representation (which is defined by the SNR), which may be used for register-length design in hardware implementation of time–frequency algorithms.**

I. Introduction

Classical Fourier analysis provides the spectrum of the analyzed signals. However, the obtained spectrum does not provide time distribution of the spectral components. As opposed to the classical Fourier analysis, time–frequency signal analysis gives the distribution in time of the signal's spectral components. The quadratic shift-covariant TFD's, which may be treated as special cases of the Cohen class of TFD's (CD), [2], [4], [5], [8], [13], [18], play the central role in this analysis. The most prominent members of this class are the Wigner distribution (WD) and the spectrogram, [3], [5], [8]. Realizations of these TFD's admit both hardware and software implementation. For real-time applications, it is often necessary to use hardware implementation that gives rise to some new issues, one of the most important being the selection of appropriate register length. Shorter register length requires less hardware, but it may produce lower resolution and range. Registers of finite length used to represent signals in time–frequency analysis also introduce quantization errors [11] that may adversely affect the obtained results. Rounding of arithmetic operation results also introduce errors, whose influence to the final results depends on the chosen number representation (fixed point or floating point). Fixed-point arithmetic is characterized by a narrow range and is more sensitive to addition overflow [6], [11]. To overcome this problem, the floating-point representation and arithmetic is used. It significantly extends dynamic range, but for a given register length, it must be done at the expense of the precision. Therefore, a tradeoff between the lengths of mantissa and exponent should be carefully considered in the selection of hardware for implementation.

The effects of finite register length have been analyzed in the case of the WD [6], [14]. This correspondence extends the analysis of [6] and [14] to the TFD's from the general CD [5], [8], [19] for the cases of floating-point and fixed-point representations. Distributions' variance and the signal-to-quantization noise ratio (SNR) have been derived (using results from [1], [7], and [15]) and used as criteria for quantitative comparison of various TFD's from the CD. Deterministic and quasistationary random signals have been analyzed. The analysis and comparison of the most frequently used TFD's, with regard to the finite word-length effects, have been performed. The relationship between dynamic range of used registers and required quality of representation defined by the SNR is derived. The obtained expression can be used for the register lengths selection in the hardware realization of TFD's. The same expression may also be used in determination of mantissa and exponent lengths in hardware designs for floating-point arithmetic.

The paper is organized as follows. After an introduction, in Section II, the variance of the CD of noisy signals is found, and the SNR is defined. In Section III, influence of the finite word-length on the results obtained by the TFD's and the floating-point arithmetic implementation is analyzed, both for random and deterministic signals. In addition, the expressions that may be used for mantissa and exponent length selection are derived. For the sake of completeness, corresponding results and conclusions for the fixed-point case are derived in Section IV.

II. Variance of the Cohen Class of Distributions

Discrete form of the CD of signal $f(n)$ is defined by [7], [9], [19]

$$\begin{aligned} C_f(n,k;\varphi) &= \sum_{i=-L}^{L-1} r_f(n,i)e^{-j\frac{4\pi}{N}ki} \\ r_f(n,i) &= \sum_{m=-L}^{L-1} \varphi(m,i)f(n+m+i)f^*(n+m-i) \end{aligned} \tag{1}$$

where $N = 2L$ is the duration determined by the time-lag kernel $\varphi(m,i)$ width along the time and lag axis, whereas $r_f(n,i)$ is the generalized autocorrelation function of $f(n)$. In order to analyze the influence of the registers' lengths to the accuracy of results obtained by TFD's from the CD, it is necessary to find the variance of the Cohen's estimator when the signal $x(n) = f(n) + \nu(n)$ is used. The analyzed signal is denoted by $f(n)$, whereas $\nu(n)$ denotes additive noise with variance σ_ν^2.

Deterministic Signal: Suppose that the analyzed signal $f(n)$ is deterministic. In that case, it can be shown [1], [14]–[16] that the variance of Cohen's estimator is frequency dependent, and it can be described as $\sigma_{xx}^2(k) = \sigma_{f\nu}^2(k) + \sigma_{\nu\nu}^2(k)$, where $\sigma_{f\nu}^2(k)$ is the variance component depending on both the analyzed signal $f(n)$ and noise $\nu(n)$, whereas $\sigma_{\nu\nu}^2(k)$ depends on the additive noise $\nu(n)$ only. The mean of the total variance of Cohen's estimator, in the case of frequency-modulated deterministic signals $f(n) = Ae^{j\psi(n)}$ and in the case of white Gaussian noise, has the form

$$\overline{\sigma_{xx}^2(k)} = \left(2A^2 + \sigma_\nu^2\right)\sigma_\nu^2 E_\varphi \tag{2}$$

where $E_\varphi = \sum_{m=-L}^{L-1}\sum_{i=-L}^{L-1}|\varphi(m,i)|^2$ is the energy of kernel $\varphi(m,i)$. More details about the derivation of this relation may be found in [1], whereas the forms in the case of whether the signal and noise were real or analytic are studied in [15]. The principal conclusions that may be derived in the real signal and noise case, as well as in the analytic case, remain the same as in the case of complex signal and noise [15]. For this reason, we will very often refer to the complex noise case in the rest of the paper. It is very interesting to emphasize that the variance $\sigma_{\nu\nu}^2(k)$, in the case of white, uniformly

Manuscript received April 1, 1997; revised December 1, 1997. This work was supported by the Alexander von Humboldt Foundation. The associate editor coordinating the review of this paper and approving it for publication was Prof. Moeness Amin.

V. Ivanović is with the Elektrotehnicki Fakultet, University of Montenegro, Montenegro, Yugoslavia.

LJ. Stanković is with the Signal Theory Group, Ruhr University Bochum, Bochum, Germany on leave from the Elektrotehnicki Fakultet, University of Montenegro, Montenegro, Yugoslavia (e-mail: l.stankovic@ieee.org).

D. Petranović is with the Advanced Development Laboratory, LSI Logic, Mountain View, CA 94035 USA.

Publisher Item Identifier S 1053-587X(98)04431-6.

distributed noise $\nu(n)$, is slightly different from the case of white Gaussian noise, but the final result may approximately be described by (2), as well.[1]

Random Signal: Let us assume the complex, quasistationary, stochastic process $f(n)$ with variance $\sigma_f^2(n)$ and complex noise $\nu(n)$ with variance σ_ν^2, both with independent real and imaginary parts with equal variances $E\{f(n_1)f(n_2)\} = E\{\nu(n_1)\nu(n_2)\} = 0$ and $E\{f^*(n_1)f^*(n_2)\} = E\{\nu^*(n_1)\nu^*(n_2)\} = 0$, as well as $E\{f(n+i_1)f^*(n+i_2)\} \cong \sigma_f^2(n)\delta(i_1 - i_2)$ [10] ($\sigma_f^2(n)$ is a slowly varying function). Although this case has limited practical importance (time–frequency analysis takes place for highly nonstationary signals), it allows a derivation of very simple relations that are used for the estimation of a very complicated model for the finite register lengths influence on the CD [11]. Applying analysis from [10] in the case of quasistationary processes and supposing that the signal and noise processes are uncorrelated, it can be shown that the variance of estimator $C_x(n,\omega;\varphi)$, in the case of white Gaussian stochastic processes, takes the final form

$$\sigma_{xx}^2(n,k) = \left(\sigma_f^2(n) + \sigma_\nu^2\right)^2 E_\varphi. \tag{3}$$

III. Analysis of the Quatization Effects with Floating-Point Arithmetic

In the implementations based on the floating-point arithmetic, the quantization only affects the mantissa. Thus, for the floating-point representation, relative multiplicative error appears. In other words, if we denote the quantized value as $Q[x]$ and its value before quantization as x, we can write $Q[x] = x(1+\varsigma(n))$, where $\varsigma(n)$ is the relative error due to the quantization of the arithmetic operation result [11]. In order to make the appropriate analysis, we will assume the following [11], [14].

1) The length of the mantissa is $(b+1)$ bits, and they are organized in the following manner. b bits are used for the absolute value of mantissa and one bit for sign.
2) The random variables of each of the relative quantization error are uncorrelated, i.e., the quantization error is a white-noise process that has the uniform distribution over the range -2^{-b} to 2^{-b}.
3) The error sources are uncorrelated with one another.
4) All the errors are uncorrelated with the input and, consequently, with all signals in the system. Note that the mean and the variance of each assumed relative error $\varsigma(n)$ are $m_\varsigma = 0$ and $\sigma_\varsigma^2 = 2^{-2b}/3 = \sigma_B^2$, where σ_B^2 is the base variance.

According to the assumptions, in the analysis of the finite register length influence on the CD, we will use the model

$$\begin{aligned} C(n,k;\varphi) &= \sum_{i=-L}^{L-1} r(n,i)e^{-j\frac{4\pi}{N}ki}[1+\mu(n,i,k)] \\ &\quad \times \prod_{p=1}^{L_p}[1+g(n,i,k,p)] \\ r(n,i) &= \sum_{m=-L}^{L-1} \Bigg\{ \varphi(m,i)x(n+m+i)x^*(n+m-i) \\ &\quad \times [1+e(n+m,i)][1+\rho(n+m,m,i)] \\ &\quad \times \prod_{q=1}^{L_q}[1+d(n+m,m,i,q)] \Bigg\} \end{aligned} \tag{4}$$

where $x(n) = f(n) + \epsilon(n)$. The following noise sources are introduced in the above equations:

$\epsilon(n)$	noise due to quantization of the complex input $f(n)$;
$e(n+m,i)$	noise due to quantization of the product $x(n+m+i)x^*(n+m-i)$;
$\rho(n+m,m,i)$	noise due to quantization of product of the kernel $\varphi(m,i)$ with $x(n+m+i)x^*(n+m-i)$;
$\mu(n,i,k)$	noise due to quantization of product of the auto-correlation function $r(n,i)$ with the basis functions $e^{-j4\pi ki/N}$.

Considering the definitions and the introduced assumptions, we can conclude that the corresponding variances of these noise sources are $2\sigma_\epsilon^2 = \sigma_e^2 = \sigma_\rho^2 = \sigma_\mu^2 = 4\sigma_B^2$.

The noise sources $g(n,i,k,p)$ and $d(n+m,m,i,q)$, which are produced by the additions, are also included in (4). Namely, the floating-point additions also produce the quantization errors, which are represented by the multiplicative noise. Suppose that the additions in our model are done in the following manner: adding the adjacent elements in the first step, then the adjacent sums in the next step, and so on (what corresponds to the butterflies in the FFT algorithms); then, L_p and L_q belonging to (4) are $L_p = L_q = \log_2 N$. Note that the error due to the quantization of the basic functions $e^{-j4\pi ki/N}$ has not been taken into analysis because it exhibits some deterministic properties, although it can also be modeled as white noise [11]. The same reason is applied to the kernel quantization error.

Since the quantization errors are small, all higher order error terms can be neglected $\prod_{p=1}^{L_p}[1+g(n,i,k,p)] \cong 1 + \sum_{p=1}^{L_p} g(n,i,k,p)$, and $\prod_{q=1}^{L_q}[1+d(n+m,m,i,q)] \cong 1 + \sum_{q=1}^{L_q} d(n+m,m,i,q)$, and the proposed model (4), reduces to

$$\begin{aligned} C(n,k;\varphi) &\cong \sum_{i=-L}^{L-1} \left\{ r(n,i)e^{-j\frac{4\pi}{N}ki}[1+\eta(n,i,k,p)] \right\} \\ r(n,i) &\cong \sum_{m=-L}^{L-1} \{\varphi(m,i)x(n+m+i)x^*(n+m-i) \\ &\quad \times [1+\epsilon_{\rm eq}(n+m,m,i,q)]\} \end{aligned} \tag{5}$$

where $\eta(n,i,k,p)$ and $\epsilon_{\rm eq}(n+m,m,i,q)$ represent the equivalent noises

$$\begin{aligned} \eta(n,i,k,p) &= \mu(n,i,k) + \sum_{p=1}^{L_p} g(n,i,k,p) \\ \epsilon_{\rm eq}(n+m,m,i,q) &= e(n+m,i) + \rho(n+m,m,i) \\ &\quad + \sum_{q=1}^{L_q} d(n+m,m,i,q) \end{aligned} \tag{6}$$

with the corresponding variances $\sigma_\eta^2 = \sigma_\mu^2 + L_p\sigma_g^2$ and $\sigma_{\rm eq}^2 = \sigma_e^2 + \sigma_\rho^2 + L_q\sigma_d^2$. Based on the central limit theorem, the equivalent noises $\eta(n,i,k,p)$ and $\epsilon_{\rm eq}(n+m,m,i,q)$ behave as Gaussian since they represent sums of the mutually statistically independent small noises. After some straightforward transformations (the same ones as in [1], [7], and [14]–[16]), we obtain the variance of the CD model, given by (4), having in mind (6), in the form

$$\begin{aligned} \sigma^2(n,k) &\cong \sigma_{xx}^2(n,k) + \sigma_{\rm eq}^2 \sum_{i=-L}^{L-1}\sum_{m=-L}^{L-1} |\varphi(m,i)|^2 \\ &\quad \times E\{|x(n+m+i)|^2|x(n+m-i)|^2\} \\ &\quad + \sigma_\eta^2 \sum_{i=-L}^{L-1} E\{|r_x(n,i)|^2\}. \end{aligned} \tag{7}$$

[1] For the case of stationary, white, uniformly distributed complex noise $\nu(n)$, variance $\sigma_{\nu\nu}^2(k)$ has the form $\sigma_{\nu\nu}^2(k) = \sigma_\nu^4(E_\varphi - \frac{6}{5}|\sum_{m=-L}^{L-1}\varphi(m,0)|^2) \cong \sigma_\nu^4 E_\varphi$. In the sequel, it will be proved that for the RID distributions that satisfy the frequency marginal, the second term may be neglected.

TABLE I

FACTORS FOR SOME TIME-FREQUENCY DISTRIBUTIONS: BORN–JORDAN DISTRIBUTION (BJD), OPTIMAL AUTOTERM DISTRIBUTION (OATD), CHOI–WILLIAMS DISTRIBUTION (CWD), BUTTERWORTH DISTRIBUTION (BD), SINC DISTRIBUTION (SINCD), PSEUDO WIGNER DISTRIBUTION WITH THE HANNING WINDOW $w^2(\tau)$ (PWD)

Factors	BJD	OATD	CWD	BD	SINCD	PWD
Kernel $c(\Theta,\tau)$	$\frac{\sin(\frac{\Theta\tau}{2})}{\Theta\tau/2}$	$e^{-\frac{\lvert\Theta\tau\rvert}{\sigma}}$	$e^{-\frac{\Theta^2\tau^2}{\sigma^2}}$	$\frac{1}{1+(\frac{\Theta\tau}{\Theta_1\tau_1})^4}$	$rect(\frac{\Theta\tau}{\alpha})$	$w^2(\tau)$
$E_\varphi = \sum_{m=-L}^{L-1}\sum_{i=-L}^{L-1} \lvert\varphi(m,i)\rvert^2$	12.5358	12.9121	15.4919	19.8952	22.3258	192
$\lvert\varphi(0,0\rvert^2/E_\varphi$	0.0798	0.0774	0.0645	0.0503	0.0448	0.0052
$C=\sqrt{\sum_{i=-L}^{L-1}\sum_{m=-L}^{L-1}\lvert\varphi(m,i)\rvert}$	4.9277	4.7620	4.5476	5.0046	4.8527	16
SNR[dB]	81.3373	81.3462	81.3963	81.4525	81.4742	81.6344

In the last equation $\sigma_{xx}^2(n,k)$ is the variance of the CD of the signal $x(n)=f(n)+\epsilon(n)$ when the arithmetic is ideal, i.e., when only noise $\epsilon(n)$, due to the quantization of input, exists. Its value, which has been obtained from the analysis of random signal $f(n)$, is presented in (3), whereas in the case of deterministic FM signal $f(n)$, its mean value is given by (2) [for $\nu(n)=\epsilon(n)$].

A. Random Signal

Assume that the analyzed signal $f(n)$ is a complex, quasistationary, white Gaussian stochastic process (with variance $\sigma_f^2(n)$), with independent real and imaginary parts. In this case, the variance of the CD's model has the form

$$\begin{aligned}\sigma^2(n,k) \cong{}& \left(\sigma_f^4(n)+2\sigma_f^2(n)\sigma_\epsilon^2\right)E_\varphi \\ &+\sigma_f^4(n)\sigma_{\rm eq}^2\left[E_\varphi+\sum_{m=-L}^{L-1}|\varphi(m,0)|^2\right] \\ &+\sigma_f^4(n)\sigma_\eta^2\left[E_\varphi+\left|\sum_{m=-L}^{L-1}\varphi(m,0)\right|^2\right]. \qquad (8)\end{aligned}$$

The last equation can be simplified in the case of all RID's, satisfying the frequency marginal property, as well as in the WD case. In these cases, $\varphi(m,i)$ is mainly concentrated at the origin of the (m,i) plane and around the i $(m=0)$ axis [15]; therefore, we have $\sum_{m=-L}^{L-1}|\varphi(m,0)|^2=|\sum_{m=-L}^{L-1}\varphi(m,0)|^2=|\varphi(0,0)|^2$ for all TFD's satisfying the frequency marginal condition, where $\varphi(0,0)$ is a constant $(\varphi(0,0)=1)$. In addition, using the definition of noises $\eta(n,i,k,p)$ and $\epsilon_{\rm eq}(n+m,m,i,q)$ and applying $2\sigma_\epsilon^2=\sigma_e^2=\sigma_\rho^2=\sigma_\mu^2=\sigma_d^2=\sigma_g^2=4\sigma_B^2=\sigma_c^2$, we get the variance of the model $\sigma^2(n,k)$ in the form

$$\begin{aligned}\sigma^2(n,k) \cong{}& \left(\sigma_f^4(n)+\sigma_f^2(n)\sigma_c^2\right)E_\varphi \\ &+\sigma_f^4(n)(3+L_p+L_q)\sigma_c^2[E_\varphi+|\varphi(0,0)|^2]. \qquad (9)\end{aligned}$$

Note that the variance $\sigma^2(n,k)$ takes different values for different TFD's from the CD, depending on the factor E_φ. In [7], it has been shown that this factor is minimized (under the marginal conditions and time-support constraint) with the kernel of the Born–Jordan distribution (BJD), and consequently, it can be concluded that the minimal value of the variance $\sigma^2(n,k)$, in this case, is obtained by the BJD. Almost the same value of E_φ, as in the case of the BJD, is achieved by the autoterm optimal distribution kernel [13], which has been derived in [15].

As a criterion for comparison of the individual TFD, we define the quantization noise-to-signal ratio (NSR) as

$$\text{NSR}=\frac{\sigma^2-\sigma^2_{|\text{without noise}}}{\sigma^2_{|\text{without noise}}} \qquad (10)$$

where $\sigma^2_{|\text{without noise}}$ is the variance of model (4), assuming ideal arithmetic and σ^2 is given by (9)

$$\text{NSR}\cong\frac{\sigma_c^2}{\sigma_f^2(n)}+(3+L_p+L_q)\left[1+\frac{|\varphi(0,0)|^2}{E_\varphi}\right]\sigma_c^2. \qquad (11)$$

Since $|\varphi(0,0)|^2 \ll E_\varphi$, we can approximate the last equation with

$$\text{NSR}\cong\sigma_c^2/\sigma_f^2+(3+L_p+L_q)\sigma_c^2 \qquad (12)$$

where $\sigma_f^2=\min_n\{\sigma_f^2(n)\}$ corresponds to the worst case with respect to the register length design. In this case, all considered TFD's show approximately equal characteristics with respect to the NSR. The degree of the proposed approximation, i.e., the error we introduce with approximation (12), is different for different TFD's and depends on the factor E_φ. This factor has been analyzed in detail in [1], [7], [15], and Table I. The kernels are given in the analogue ambiguity domain (for details, see [2], [4], [5], [8], [9], [13], and [18]). Discretization is done taking the range $|\Theta|\le\sqrt{\pi L}$ and $|\tau|\le\sqrt{\pi L}$, with $L=256$. The kernel $\varphi(m,i)$ is calculated as a Fourier transform $\varphi(m,i)=\text{FT}_\theta[c(\theta,i)]$, where $c(\theta,i)$ are samples of $c(\Theta,\tau)$ along τ, and θ is a discrete-domain frequency $\theta=\Theta\frac{\pi}{\sqrt{\pi L}}$. In order to compare various TFD's, their parameters $(\sigma,\alpha,\theta_1\tau_1)$ are chosen according to the results in [13]. The values of the SNR $=1/\text{NSR}$ [dB] (11) for the commonly used TFD's belonging to the CD and for $b=16$ bits (b is number of bits used to represent mantissa) are given in Table I. The error made by approximation (12) have been calculated, and it has been concluded that it ranges from the case when $|\varphi(0,0)|^2/E_\varphi=0.0798$ for the BJD, to 0 for the Zhao–Atlas–Marks (ZAM) distribution since its kernel [2] $c(\Theta,\tau)=|\tau|\sin(\Theta\tau/2)/(\Theta\tau/2)$ has $c(\theta,0)=0$ (does not satisfy frequency marginal condition), and consequently, $\varphi(m,0)=0$ for every m.

Another interesting distribution, which in the case of multicomponent signals may produce a sum of the Wigner distributions of each signal component separately, is the S method [13], [16], [17]. Its kernel in the time-lag domain is

$$\varphi_{\rm SM}(m,i)=w(m+i)w(m-i)\frac{\sin[2\pi m(2L_d+1)/N]}{(2L_d+1)K\sin(2\pi m/N)}.$$

For $L_d=0$, the spectrogram follows, whereas for $2L_d+1=N$, we get the WD. Note that the kernel $\varphi_{\rm SM}(m,i)$ is not generally a separable function. Factor K is introduced in order to keep the unbiased energy condition for any L_d, $K=\sum_{l=-L_d}^{L_d}W_{w^2}(2l)/(2L_d+1)$, and $W_{w^2}(l)=\text{FT}\{w^2(m)\}$. For example, for the Hanning window and $L_d=4$, we get $E_\varphi=9.1104$ and SNR[dB] $=81.6509$, whereas for the spectrogram $(L_d=0)$, we have SNR[dB] $=81.6559$.

TABLE II
VARIANCE, QUANTIZATION NOISE-TO-SIGNAL RATIO, AND REGISTER LENGTH FOR THE COHEN CLASS OF DISTRIBUTIONS, SUPPOSING SUFFICIENTLY SMALL SIGNAL SO THAT AN OVERFLOW DOES NOT OCCUR

Random signal	**Deterministic signal**
$\sigma^2(k) = [(\sigma_f^2 + \frac{\sigma_c^2}{2})^2 + \sigma_c^2]E_\varphi + (N^2+N)\sigma_c^2$	$\overline{\sigma^2(k)} = [(2A^2 + \frac{\sigma_c^2}{2})\frac{\sigma_c^2}{2} + \sigma_c^2]E_\varphi + (N^2+N)\sigma_c^2$
$NSR \cong \frac{\sigma_c^2}{\sigma_f^2} + \frac{\sigma_c^2}{\sigma_f^4} + \frac{(N^2+N)\sigma_c^2}{\sigma_f^4 E_\varphi} \cong \frac{N^2\sigma_c^2}{\sigma_f^4 E_\varphi}$	$SNR_{\max} \cong \frac{N^2A^4}{(A^2+1)\sigma_c^2 E_\varphi + (N^2+N)\sigma_c^2} \cong \frac{A^4}{\sigma_c^2}$
$b \cong \nu - 0.8 + \frac{1}{6.02}\{SNR[dB] - 10\log(E_\varphi) - 20\log(\sigma_f^2)\}$	$b \cong \frac{1}{6.02}\{SNR_{\max}[dB] - 40\log A\} - 0.8$

Substituting $L_p = L_q = \log_2 N$ and σ_c^2 into (12) and knowing that the duration of a kernel commonly takes an integer power of 2, $N = 2^\nu$, the NSR can be represented in the forme12

$$\text{NSR} \cong \frac{4}{3}\left(3 + 2\nu + 1/\sigma_f^2\right) \cdot 2^{-2b}. \tag{13}$$

Observe that the NSR consists of two parts $\text{NSR} \cong \frac{4}{3}(3 + 1/\sigma_f^2) \cdot 2^{-2b} + \frac{8}{3}\nu \cdot 2^{-2b} = \text{NSR}_1 + \text{NSR}_2$. The first component depends only on the number of bits needed to represent mantissa, whereas the second component depends on both the kernel width (represented by $N = 2^\nu$) and the number of bits b. The value of NSR_1 on a logarithmic scale is given by

$$\text{NSR}_1[\text{dB}] = 10\log(\text{NSR}_1) = 10\log\left(4 + \frac{4}{3\sigma_f^2}\right) - 6.02b. \tag{14}$$

It is clear that NSR_1 [dB] decreases approximately 6 dB for each bit added to the register length.

On the other hand, the second part is proportional to ν, as opposed to the case of the fixed-point arithmetic, where it is proportional to the square of N (see Section IV). At the same time, the NSR_2 is proportional to 2^{-2b}, and consequently, quadrupling ν (i.e., increasing the signal's duration N to the fourth power) results in an increase in the NSR_2, which corresponds to the reduction of b by one bit. Thus, in order to maintain the NSR_2 at the same value, the increase of the distribution duration to the power of four can be compensated with the increase of the register length by one bit.

It is interesting to present (12) as a fundamental dependence of dynamic range of the registers on the SNR

$$b \cong 0.2075 + \frac{1}{6.02}\left\{10\log\left(3 + 2\nu + 1/\sigma_f^2\right) + \text{SNR[dB]}\right\}. \tag{15}$$

Using this expression, the number of bits needed to represent the absolute value of mantissa for a given value of SNR can be easily determined. For example, for $\nu = 10$ and $\sigma_f^2 = 1$, in order to keep SNR[dB] ≥ 80 dB, we have to use $b = 16$ bits to represent the mantissa. Equation (15) is very useful for the design of hardware for implementation of time–frequency algorithms. It can be used to appropriately dimension registers in order to satisfy required quality, as expressed by the SNR, as well as to determine the number of bits necessary to represent the mantissa and exponent in order to find a tradeoff between required accuracy and range.

B. Deterministic Signal

In this section, we will analyze deterministic signal $f(n)$. In finding the variance (7) of the model, we have decided to use its mean value since it requires a lower degree of knowledge of the analyzed deterministic signal $f(n)$ [1], [15].

The mean value of the variance of the CD (7) can be presented in the form

$$\begin{aligned}\overline{\sigma^2(k)} \cong\ & \overline{\sigma_{xx}^2(k)} + \sigma_{\text{eq}}^2 \sum_{i=-L}^{L-1} \sum_{m=-L}^{L-1} |\varphi(m,i)|^2 \\ & \times E\{|x(n+m+i)|^2 |x(n+m-i)|^2\} \\ & + \sigma_\eta^2 \sum_{i=-L}^{L-1} E\{|r_x(n,i)|^2\}\end{aligned} \tag{16}$$

where $\overline{\sigma_{xx}^2(k)}$ is the mean value of the same variance for an ideal arithmetic.

Neglecting all higher order noises and applying a set of straightforward modifications, we get

$$\begin{aligned}\overline{\sigma^2(k)} \cong\ & \overline{\sigma_{xx}^2(k)} + \sigma_{\text{eq}}^2 \sum_{i=-L}^{L-1} \sum_{m=-L}^{L-1} |\varphi(m,i)|^2 \\ & \times |f(n+m+i)|^2 |f(n+m-i)|^2 \\ & + \sigma_\eta^2 \sum_{i=-L}^{L-1} |r_f(n,i)|^2.\end{aligned} \tag{17}$$

Using the definitions of the equivalent noises (6), the last equation can be written as

$$\begin{aligned}\overline{\sigma^2(k)} \cong\ & \overline{\sigma_{xx}^2(k)} + \left(\sigma_e^2 + \sigma_\rho^2 + L_q\sigma_d^2\right) \\ & \times \sum_{i=-L}^{L-1} \sum_{m=-L}^{L-1} |\varphi(m,i)|^2 |f(n+m+i)|^2 \\ & \times |f(n+m-i)|^2 + \left(\sigma_\mu^2 + L_p\sigma_g^2\right) \\ & \times \sum_{i=-L}^{L-1} \left| \sum_{m=-L}^{L-1} \varphi(m,i) f(n+m+i) f^*(n+m-i) \right|^2.\end{aligned} \tag{18}$$

For the deterministic FM signals, $\overline{\sigma_{xx}^2(k)}$ is given by (2). The application of the Cauchy inequality [12] on the last component of the mean variance (18) (i.e., using the form $\sum_{i=-L}^{L-1} |\sum_{m=-L}^{L-1} \varphi(m,i) f(n+m+i) f^*(n+m-i)|^2 \leq NA^4 E_\varphi$) for the given values of the individual quantization noises results in

$$\overline{\sigma^2(k)} \leq [A^2 + A^4(2 + L_q) + NA^4(1 + L_p)]\sigma_c^2 E_\varphi. \tag{19}$$

For a reasonable choice of the duration of the analyzed signal $f(n)$ ($N \gg 1$), the second term in (19) can be neglected, and we can obtain the limit for the mean of variance $\overline{\sigma^2(k)}$ as

$$\max\{\overline{\sigma^2(k)}\} \cong [A^2 + NA^4(1 + L_p)]\sigma_c^2 E_\varphi. \tag{20}$$

Equation (20) represents a general maximized expression for the mean variance in the case of FM signals. However, this expression cannot be used as a close approximation for many special signal forms since in the Cauchy's inequality, the equivalence very seldom occurs. Consequently, we shall consider the specific signal forms because, in

TABLE III
VARIANCE, QUANTIZATION NOISE-TO-SIGNAL RATIO, AND REGISTER LENGTH FOR THE COHEN CLASS OF DISTRIBUTIONS WHEN THE ANALYZED SIGNAL IS APPROPRIATELY SCALED AT THE INPUT

Random signal	**Deterministic signal**
$\sigma^2(k)=[\frac{1}{C^4}(\sigma_f^2+\frac{\sigma_c^2}{2})^2+\sigma_c^2]E_\varphi+(N^2+N)\sigma_c^2$	$\overline{\sigma^2(k)}=[\frac{1}{C^4}(2A^2+\frac{\sigma_c^2}{2})\frac{\sigma_c^2}{2}+\sigma_c^2]E_\varphi+(N^2+N)\sigma_c^2$
$NSR\cong\frac{\sigma_c^2}{\sigma_f^2}+\frac{C^4\sigma_c^2}{\sigma_f^4}+\frac{C^4(N^2+N)\sigma_c^2}{\sigma_f^4E_\varphi}\cong\frac{C^4N^2\sigma_c^2}{\sigma_f^4E_\varphi}$	$SNR_{\max}\cong\frac{N^2A^4/C^4}{(A^2/C^4+1)\sigma_c^2E_\varphi+(N^2+N)\sigma_c^2}\cong\frac{A^4}{C^4\sigma_c^2}$
$b\cong\nu-0.8+\frac{1}{6.02}\{SNR[dB]-10\log(E_\varphi/C^4)-20\log(\sigma_f^2)\}$	$b\cong\frac{1}{6.02}\{SNR_{\max}[dB]-40\log A+40\log C\}-0.8$

TABLE IV
VARIANCE, QUANTIZATION NOISE-TO-SIGNAL RATIO, AND REGISTER LENGTH FOR THE COHEN CLASS OF DISTRIBUTIONS IMPLEMENTED USING THE FFT ALGORITHMS WITH SCALE FACTORS OF 1/2

Random signal	**Deterministic signal**
$\sigma^2(k)=\frac{1}{N^2}[(\sigma_f^2+\sigma_c^2/2)^2+\sigma_c^2]E_\varphi+5\sigma_c^2$	$\overline{\sigma^2(k)}=\frac{1}{N^2}[(2A^2+\sigma_c^2/2)\sigma_c^2/2+\sigma_c^2]E_\varphi+5\sigma_c^2$
$NSR\cong\frac{\sigma_c^2}{\sigma_f^2}+\frac{\sigma_c^2}{\sigma_f^4}+\frac{5N^2\sigma_c^2}{\sigma_f^4E_\varphi}\cong\frac{5N^2\sigma_c^2}{\sigma_f^4E_\varphi}$	$SNR_{\max}\cong\frac{N^2A^4}{(A^2+1)\sigma_c^2E_\varphi+5N^2\sigma_c^2}\cong\frac{A^4}{5\sigma_c^2}$
$b\cong\nu+0.3685+\frac{1}{6.02}\{SNR[dB]-10\log(E_\varphi)-20\log(\sigma_f^2)\}$	$b\cong\frac{1}{6.02}\{SNR_{\max}[dB]-40\log A\}+0.3685$

the special cases, the approximation closer to the exact value can be found. It makes the analysis of the mean variance $\overline{\sigma^2(k)}$ more reliable.

Example: Let us consider the special form of an FM signal $f(n) = Ae^{j\psi(n)}$ with a slowly varying frequency $\omega(n)$ such that $f(n+m\pm i)\varphi(m,i)\cong A\varphi(m,i)e^{j(\psi(n)+\omega(n)(m\pm i))}$ within the considered lag interval. In this case, we have $|r_f(n,i)|^2 = A^4|c(0,i)|^2$ so that (17), for the TFD's satisfying the time marginal condition, may be reduced to

$$\overline{\sigma^2(k)}\cong\left(2A^2\sigma_\epsilon^2+A^4\sigma_{\mathrm{eq}}^2\right)E_\varphi+NA^4\sigma_\eta^2. \tag{21}$$

Observe that the mean variance of the model of CD is directly proportional to the factor E_φ, as in the case of the maximal value of $\overline{\sigma^2(k)}$ for the FM signals (20).

Finally, define the maximal signal-to-quantization noise ratio $(\mathrm{SNR}_{\max})$ for deterministic signals as a ratio between maximal square absolute value of the considered distribution and the model mean variance[2] [16]

$$\mathrm{SNR}_{\max}=\max\{|C_f(n,k;\varphi)|^2\}/\overline{\sigma^2(k)}. \tag{22}$$

Analysis of the sinusoidal signal $f(n)$ in the case of TFD's satisfying the time-marginal condition, relatively easily shows that $C_f(n,k;\varphi)=NA^2\delta(k-k_0)$; therefore, the ratio $\mathrm{SNR}_{\max}$ may be represented as:

$$\mathrm{SNR}_{\max}=\frac{N^2A^4}{\left(2A^2\sigma_\epsilon^2+A^4\sigma_{\mathrm{eq}}^2\right)E_\varphi+NA^4\sigma_\eta^2}. \tag{23}$$

[2] Another possible definition of the SNR is the local ratio of distribution and its model mean variance:

$$\mathrm{SNR}=\frac{\left|C_f(n,k;\varphi)\right|^2}{\overline{\sigma^2(k)}}.$$

However, we preferred definition (22) since it produces simpler results. It also compares the peak value of the TFD with the quantization noise in the time–frequency plane. This is very reasonable in many practical applications, where a TFD [its peak value(s)] is used to estimate the instantaneous frequency of a signal. In this case, we are not interested in the local ratio, especially at the points where the TFD is equal to zero. For that point, it is better to compare the mean variance, due to quantization effects, with the maximum value of the TFD since this ratio represents the measure of a possible false peak detection (i.e., wrong frequency detection).

Let us, finally represent the above analysis for deterministic sinusoidal signal in the form of a relationship between the dynamic register range (i.e., number of bits used for representation of the mantissa's absolute value) and the induced quantization error

$$b\cong0.2075-\frac{\nu}{2}+\frac{1}{6.02}\left\{10\log\left[(2+\nu+1/A^2)\frac{E_\varphi}{N}+1+\nu\right]+\mathrm{SNR}_{\max}[\mathrm{dB}]\right\} \tag{24}$$

where the values of E_φ for the considered TFD's may be found in Table I. The above equation may be used in hardware design to determine registers' lengths necessary to keep $\mathrm{SNR}_{\max}$ at an acceptable level. For example, for $A=1$, $N=512$ and $\mathrm{SNR}_{\max}[\mathrm{dB}]\geq 80$ dB, we get mantissa length $b=11$ for all considered RID's satisfying marginal properties.

IV. ANALYSIS OF THE QUANTIZATION EFFECTS WITH FIXED-POINT ARITHMETIC

When the numbers are represented using fixed-point notation, the quantization errors occur only for multiplication. However, it is possible to cause an overflow when preforming an addition operation. In the analysis of the influence of the finite register length in fixed-point arithmetic, we will use the model

$$\begin{aligned}C(n,k;\varphi)&=\sum_{i=-L}^{L-1}\left\{r(n,i)e^{-j\frac{4\pi}{N}ki}+\mu(n,i,k)\right\}\\ r(n,i)&=\sum_{m=-L}^{L-1}\{\varphi(m,i)[x(n+m+i)x^*(n+m-i)\\&\quad+e(n+m,i)]+\rho(n+m,m,i)\}.\end{aligned} \tag{25}$$

Quantization errors stemming from this model are analogous to the ones induced by the floating-point arithmetic (Section III) [6], [11], [14] with the corresponding variances $2\sigma_\epsilon^2=\sigma_e^2=\sigma_\rho^2=\sigma_\mu^2=4\sigma_B^2=\sigma_c^2$ ($\sigma_B^2=2^{-2b}/12$—basic variance value). However, as opposed to the case of floating-point arithmetic, these errors are additive [11].

1) Assume first that the analyzed signal is small enough so that an overflow cannot occur. After several appropriate transformations (the same with ones presented in Section III), it may

be shown that the model variance takes the form

$$\sigma^2(k) = \sigma_{xx}^2(k) + \sigma_e^2 E_\varphi + N^2\sigma_\rho^2 + N\sigma_\mu^2. \tag{26}$$

The above result is obtained by assuming calculations based on the conventional DFT arithmetic. Calculations are usually performed by the FFT algorithms. The results, however, remain the same by using, for example, the "decimation-in-time" algorithm. Namely, in that case, the last component from (26) is $(N-1)\sigma_\mu^2 \cong N\sigma_\mu^2$ [11].

When signal $f(n)$ is not small enough, we should take care to prevent the overflow effects. Assuming that the samples $f(n)$ are located within the interval $[0, 1)$, we may use one of the following two methods to account for possible overflow.

2) If the signal is divided by $C = \sqrt{\sum_{i=-L}^{L-1}\sum_{m=-L}^{L-1}|\varphi(m,i)|}$, an overflow cannot occur. In this case, the variance $\sigma^2(k)$ is

$$\sigma^2(k) = \sigma_{xx}^2(k)/C^4 + \sigma_e^2 E_\varphi + N^2\sigma_\rho^2 + N\sigma_\mu^2. \tag{27}$$

3) Using the scaling with factors of 1/2 [11], in the FFT algorithm, we may avoid the overflow as well. All the signals at the input of an FFT block—generalized auto-correlation function $r_x(n,i)$ and the noises $e(n)$ and $\rho(n)$—get lowered by the factor of N at its output. At the same time, we should prevent an overflow in the calculation of the generalized autocorrelation function given by (25) so that the analyzed signal is scaled by the factor $C_1 = \sqrt{\max_i \sum_{m=-L}^{L-1}|\varphi(m,i)|}$. For the considered RID's, which satisfy the frequency marginal, and for the WD, we have $C_1 = \sqrt{\sum_{m=-L}^{L-1}|\varphi(m,0)|} = 1$. Thus, in this case, variance (26) may be represented by

$$\sigma^2(k) = \frac{1}{N^2}\left(\sigma_{xx}^2(k) + \sigma_e^2 E_\varphi\right) + \sigma_\rho^2 + 4\sigma_\mu^2. \tag{28}$$

As before (Section III), we have considered both random and deterministic signals $f(n)$.

Random Signal: If we assume a random, white, and uniformly distributed signal $f(n)$, then variances (26)–(28), the NSR coefficient, and register length b as a function of SNR and $N = 2^\nu$ assume the forms presented in Tables II–IV. For the considered TFD's, the error done by approximation in NSR (with $\sigma_f^2 = 1$, $N = 512$, and $b = 16$) is of order 0.1% and 0.001%, respectively. The relation for the register length b may be used for the hardware realization of TFD's. For example, assuming $N = 512$ (and consequently $\nu = 9$), $\sigma_f^2 = 1$, and SNR[dB] ≥ 80 dB, we get the registers length defined by $b = 20$, $b = 25$, and $b = 21$, respectively. Obtained results are valid for all considered RID's satisfying the frequency marginal property. Note that as in the case of floating-point arithmetic, the dynamic range of registers is represented by b. However, in this case, b is the number of bits necessary for the representation of the signal's sample absolute value.

Deterministic Signal: The means of expressions (26)–(28), assuming a deterministic FM signal $f(n) = Ae^{j\psi(n)}$ according to (2), are given in Tables II–IV, along with the $\mathrm{SNR}_{\max}$ (for the sinusoidal signal and TFD's satisfying the time-marginal condition) and the expressions for the interrelationship between the dynamic range of the used registers (described by b) and the required ratio $\mathrm{SNR}_{\max}$. Relations are presented for the case of conventional DFT (and approximately the FFT) and scaled FFT algorithms using 1/2 factors. The approximation error of the $\mathrm{SNR}_{\max}$ for the analyzed TFD's (Table I) with $A = 1/2$, $N = 512$, and $b = 16$, is of order 0.1% and 0.001%, respectively. From the expression for $\mathrm{SNR}_{\max}$, we can determine the necessary register's word length for the required quality representation. For example, for $A = 1$, $N = 512$, and $\mathrm{SNR}_{\max}$[dB] ≥ 80 dB, we have $b = 13$, $b = 18$, and $b = 14$, respectively, for all considered TFD's from the RID class.

V. Conclusion

We have analyzed finite register length influence on the accuracy of results obtained by the time–frequency analysis for the cases of floating-point and fixed-point arithmetic as well as for the random and deterministic FM signals. It has been shown that commonly used representations from the RID class exhibit similar performance, with respect to the SNR, in all analyzed cases. We have derived the expressions that can be used in making the hardware design decisions related to the conflict between the desire to obtain fine quantization and wide dynamic range while holding the register length fixed. The obtained results may be used in the optimization of register length, which is an important factor in hardware implementations of TFD's.

References

[1] M. G. Amin, "Minimum variance time-frequency distribution kernels for signals in additive noise," *IEEE Trans. Signal Processing*, vol. 44, pp. 2352–2356, Sept. 1996.

[2] L. E. Atlas, Y. Zhao, and R. J. Marks, II, "The use of cone shape kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1084–1091, 1990.

[3] B. Boashash and P. J. Black, "An efficient real-time implementation of the Wigner-Vile distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1611–1618, Nov. 1987.

[4] H. Choi and W. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 862–871, June 1989.

[5] L. Cohen, "Time-frequency distributions—A review," *Proc. IEEE*, vol. 77, pp. 941–981, July 1989.

[6] C. Griffin, P. Rao, and F. Taylor, "Roundoff error analysis of the discrete Wigner distribution using fixed-point arithmetic," *IEEE Trans. Signal Processing*, vol. 39, pp. 2096–2098, Sept. 1991.

[7] S. B. Hearon and M. G. Amin, "Minimum-variance time-frequency distributions kernels," *IEEE Trans. Signal Processing*, vol. 43, pp. 1258–1262, May 1995.

[8] F. Hlawatsch and G. F. Broudreaux-Bartels, "Linear and quadratic time-frequency signal representation," *IEEE Signal Processing Mag.*, pp. 21–67, Apr. 1992.

[9] J. Jeong and W. J. Williams, "Alias-free generalized discrete-time time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 40, pp. 2757–2765, Nov. 1992.

[10] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis os nonstationary processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1461–1470, Dec. 1985.

[11] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1975, pp. 404–479.

[12] A. Papoulis, *Signal Analysis*. New York: McGraw-Hill, 1977.

[13] LJ. Stanković, "Auto-term representation by the reduced interference distributions; A procedure for kernel design," *IEEE Trans. Signal Processing*, vol. 44, pp. 1557–1564, June 1996.

[14] LJ. Stanković and S. Stanković, "On the Wigner distribution on discrete-time noisy signals with application to the study of quantization effect," *IEEE Trans. Signal Processing*, vol. 42, pp. 1863–1867, July 1994.

[15] LJ. Stanković and V. Ivanović, "Further results on the minimum variance time-frequency distributions kernels," *IEEE Trans. Signal Processing*, vol. 45, pp. 1650–1655, June 1997.

[16] LJ. Stanković, V. Ivanović, and Z. Petrović, "Unified approach to the noise analysis in the Wigner distribution and spectrogram using the S-method," *Ann. des Telecomm.*, nos. 11/12, pp. 585–594, Nov./Dec. 1996.

[17] S. Stanković and LJ. Stanković, "Architecture for the realization of a system for time-frequency analysis," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 600–604, July 1997.

[18] D. Wu and J. M. Morris, "Time-frequency representation using a radial Butterworth kernel," in *Proc. IEEE IS-TSTFA*, Philadelphia, PA, Oct. 1994, pp. 60–63.

[19] D. Wu and J. M. Morris, "Discrete Cohen's class of distributions," in *Proc. IEEE IS-TSTFA*, Philadelphia, PA, Oct. 1994, pp. 532–535.