TIME-FREQUENCY ANALYSIS AND NOISE SUPPRESSION WITH SHIFT-INVARIANT WAVELET PACKETS

Israel Cohen, Shalom Raz and David Malah

Department of Electrical Engineering
Technion — Israel Institute of Technology
Technion City, Haifa 32000, Israel

INTRODUCTION

Time-frequency representations map one-dimensional signals into two-dimensional images that indicate their energy content in the joint time-frequency plane [1, 2]. These representations combine time-domain and frequency-domain analyses to yield a more revealing picture of the temporal localization of spectral components. They have proven indispensable in a wide range of applications, including the study of non-stationary phenomena in high power microwave tubes, such as mode build-up and competition and pulse shortening [3].

A popular time-frequency analysis tool is the spectrogram, a squared magnitude of the short-time Fourier transform. Its major limitation is the inherent trade-off between time and frequency resolution for a particular window function. Since good time resolution requires a narrow window while good frequency resolution requires a wide window, high resolution simultaneously in both directions is unattainable. This limitation promoted the development of bilinear time-frequency representations that attempt to match the window function to the analyzed signal [1]. Unfortunately, the bilinear nature of the latter representations results in a high noise sensitivity and presence of interference terms, which restrict their practical application.

In this paper, we present a wavelet-based method for constructing an efficient time-frequency representation, which is characterized by high time-frequency resolution, noise immunity and reduced interference terms. This method also provides a robust nonlinear technique for estimating a discrete signal from its noisy measurement.

SHIFT-INVARIANT WAVELET PACKET DECOMPOSITION

Overcomplete libraries of waveforms that span redundantly the signal space encourage adaptive signal representations. Instead of representing a prescribed signal in a fixed basis, it is often useful to choose a suitable basis that facilitates a desired application,

Ultra-Wideband, Short-Pulse Electromagnetics 4
Edited by Heyman *et al.*, Kluwer Academic / Plenum Publishers, New York, 1999.

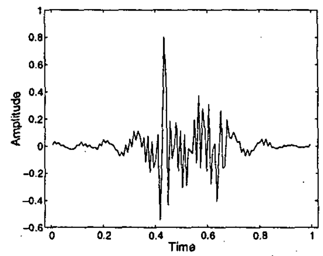


Figure 1: Test signal $g(t)$ consisting of a short pulse, a tone and a nonlinear chirp.

such as compression, identification, classification or noise removal (denoising). The library of wavelet packets is a huge library of bases that consists of translations and dilations of wavelet packets [4, 5]. The basis functions are localized in the time-frequency plane, and organized in a binary tree structure where efficient search algorithms for the best basis can be implemented.

The *shift invariant wavelet packet decomposition* (SIWPD) [6] is an adaptive representation in an *extended* library of wavelet packet bases. The extended library includes an additional degree of freedom that adjusts the time-localization of the basis functions. This degree of freedom is incorporated into the best-basis search algorithm by adaptively selecting between even and odd down-sampling. Specifically, following the low-pass and high-pass filtering, when expanding a parent-node, either all the odd samples or all the even samples are retained, according to the choice which minimizes the cost function.

Let $\{\psi_n(t) : n \in \mathbb{Z}_+\}$ be a wavelet packet family [4] generated by

$$\psi_{2n}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \psi_n(2t-k) \tag{1}$$

$$\psi_{2n+1}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \psi_n(2t-k) \tag{2}$$

where $g_k = (-1)^k h_{1-k}$, and $\psi_0(t) \equiv \varphi(t)$ is an orthonormal scaling function, satisfying

$$\langle \varphi(t-p), \varphi(t-q) \rangle = \delta_{p,q}, \quad p, q \in \mathbb{Z}\,. \tag{3}$$

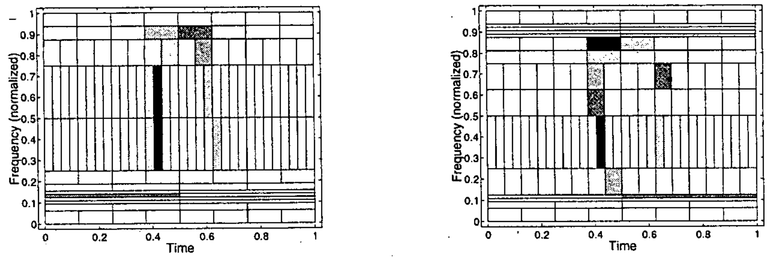
The extended library of wavelet packets is defined as the collection of all the orthonormal bases which are subsets of

$$\left\{ B_{\ell,n,m} \;:\; 0 \le \ell \le L,\, 0 \le n, m < 2^{L-\ell} \right\}, \tag{4}$$

where L denotes the finest resolution level, and

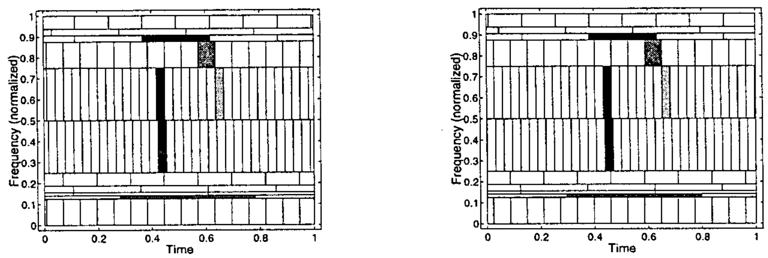
$$B_{\ell,n,m} \equiv \left\{ 2^{\ell/2} \psi_n \left[2^\ell (t - 2^{-L} m) - k \right] \;:\; 0 \le k < 2^\ell \right\}. \tag{5}$$

This library is larger than the standard wavelet packet library by a square power, but can be still structured into a tree configuration which supports fast search algorithms [6]. The additional parameter m provides the crucial degrees of freedom required for the time-adjustment of the basis functions. When an analyzed signal is translated in time by $\tau = q \cdot 2^{-L}$ ($q \in \mathbb{Z}$), a new best-basis is selected whose elements are also translated by τ compared to the former best-basis. Thus the expansion coefficients remain, and the time-frequency representation is shifted in time by the same period.



(a) (b)

Figure 2: Effect of a temporal shift on the time-frequency representation using the WPD with 8-tap Daubechies wavelet filters: (a) $g(t)$ in its best basis, Entropy= 2.69. (b) $g(t-2^{-6})$ in its best basis, Entropy= 2.72.



(a) (b)

Figure 3: Time-frequency representation using the SIWPD with 8-tap Daubechies wavelet filters: (a) $g(t)$ in its best basis, Entropy= 1.72. (b) $g(t-2^{-6})$ in its best basis, Entropy= 1.72. Compared with the WPD (Fig. 2), beneficial properties are shift-invariance and lower information cost.

Compared with the ordinary wavelet packet decomposition (WPD) [4], the SIWPD is determined to be advantageous in the following respects [6]: 1) Shift-invariance; 2) Lower information cost; 3) Improved time-frequency resolution; 4) More stable information cost across a prescribed data set; 5) Controlled computational complexity (at the expense of the information cost down to $O(N log_2 N)$). To illustrate the shift-invariant properties of the SIWPD and its enhanced time-frequency representation compared to the standard WPD, we refer to the expansion of the signal $g(t)$, depicted in Fig. 1. This signal, containing $2^7 = 128$ samples, comprises a short pulse, a tone and a component with nonlinear frequency modulation. For definiteness, we choose D_8 to serve as the scaling function (D_8 corresponds to 8-tap Daubechies least asymmetric wavelet filters [7, page 198]) and the Shannon entropy as the cost function, defined by [4] $\mathcal{M}(\{x_i\}) = -\sum_{i:x_i\neq 0} x_i^2 \log x_i^2$. Figs. 2 and 3 display the best-basis expansions under the WPD and the SIWPD algorithms, respectively, for the signals $g(t)$ and $g(t-2^{-6})$. The sensitivity of WPD to temporal shifts is obvious, while the best-basis SIWPD representation is indeed shift-invariant and characterized by a lower entropy and improved time-frequency resolution.

MODIFIED WIGNER DISTRIBUTION

The Wigner distribution (WD) possesses a number of desirable mathematical properties relevant time-frequency analysis. However, the presence of interference terms renders the WD of multicomponent signals extremely difficult to interpret [1]. Several methods, developed to reduce noise and cross-components at the expense of reduced signal concentration, employ some kind of smoothing kernel or windowing [2]. The choice of the kernel dramatically affects the appearance and quality of the resulting time-frequency representation. Hence adaptive representations often exhibit performance far surpassing that of fixed-kernel representations. However, they are either computationally expensive or have a very limited adaptation capability.

Recently, we have introduced a *modified Wigner distribution* (MWD) [8], which obtains high resolution and concentration in time-frequency, and is superior in eliminating interference terms. We showed that using extended libraries of orthonormal bases, interference terms can be reduced by adaptively thresholding the cross WD of pairs of basis functions.

Let $g = \sum_\lambda c_\lambda \varphi_\lambda$ be the SIWPD of the signal g. Then its MWD is defined by

$$T_g(t,\omega) = \sum_{\lambda\in\Lambda} |c_\lambda|^2 W_{\varphi_\lambda}(t,\omega) + 2 \sum_{\{\lambda,\lambda'\}\in\Gamma} \mathrm{Re}\{c_\lambda c^*_{\lambda'} W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega)\} \tag{6}$$

where W_{φ_λ} is the auto WD of φ_λ and $W_{\varphi_\lambda,\varphi_{\lambda'}}$ is the cross WD of φ_λ:

$$W_{\varphi_\lambda}(t,\omega) = \int \varphi_\lambda(t+\tau/2)\varphi^*_\lambda(t-\tau/2)e^{-j\omega\tau}d\tau \tag{7}$$

$$W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega) = \int \varphi_\lambda(t+\tau/2)\varphi^*_{\lambda'}(t-\tau/2)e^{-j\omega\tau}d\tau \tag{8}$$

The summations in (6) are limited to basis-functions whose coefficients are large enough, and to pairs which are "close" in time-frequency plane. Let ϵ and D denote respectively thresholds of relative amplitude and time-frequency distance. Then the sets Λ and Γ are given by

$$\Lambda = \{\lambda \mid |c_\lambda| \geq \epsilon M\}, \quad M \equiv \max_\lambda\{|c_\lambda|\} \tag{9}$$

$$\Gamma = \{\{\lambda,\lambda'\} \mid 0 < d(\varphi_\lambda,\varphi_{\lambda'}) \leq D,\ |c_\lambda c_{\lambda'}| \geq \epsilon^2 M^2\}, \tag{10}$$

The distance d between a pair of basis-functions is measured by their degree of adjacency:

$$d(\varphi_\lambda,\varphi_{\lambda'}) = \left[\frac{(\bar{t}_\lambda - \bar{t}_{\lambda'})^2}{\Delta t_\lambda \Delta t_{\lambda'}} + \frac{(\bar{\omega}_\lambda - \bar{\omega}_{\lambda'})^2}{\Delta\omega_\lambda \Delta\omega_{\lambda'}}\right]^{1/2} \tag{11}$$

where $(\bar{t}_\lambda, \bar{\omega}_\lambda)$ is the time-frequency position of the basis-function φ_λ, and Δt_λ and $\Delta\omega_\lambda$ are the corresponding time and frequency uncertainties. By adjusting the distance threshold D and amplitude threshold ϵ, one can effectively balance the cross-term interference, the useful properties of the distribution (time/frequency marginals, energy conservation, instantaneous frequency, *etc.*), and the computational complexity [9].

Here, the basis-functions are of the form

$$\psi_{\ell,n,m,k}(t) = 2^{\ell/2}\psi_n\left[2^\ell(t - 2^{-L}m) - k\right] \tag{12}$$

where ℓ is the resolution-level index ($0 \leq \ell \leq L$), n is the frequency index ($0 \leq n < 2^{L-\ell}$), m is the shift index ($0 \leq m < 2^{L-\ell}$) and k is the position index ($0 \leq k < 2^\ell$). Each basis-function is associated with a rectangular tile in the time-frequency plane which is positioned about

$$\bar{t} = 2^{-\ell}k + 2^{-L}m + (2^{L-\ell} - 1)C_h + (C_h - C_g)R(n), \tag{13}$$

$$\bar{\omega} = 2^{\ell-L}[GC^{-1}(n) + 0.5], \tag{14}$$

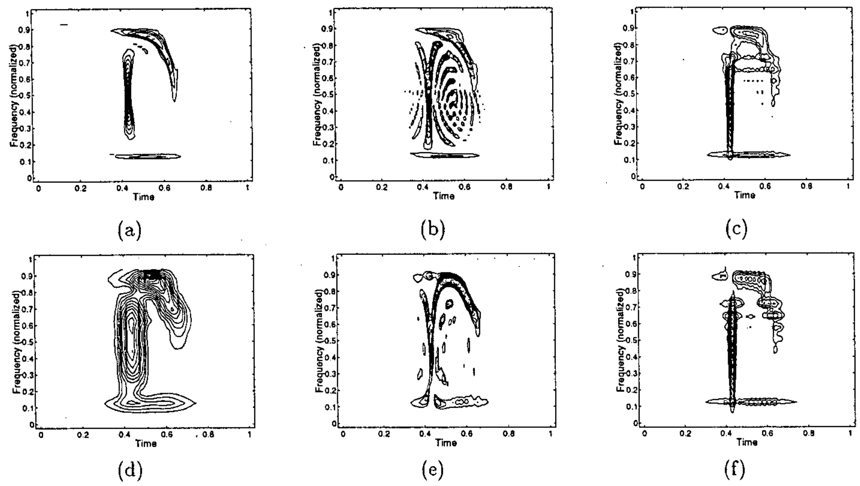


Figure 4: Contour plots for the signal $g(t)$: (a) Modified Wigner distribution; (b) Wigner distribution; (c) Choi-Williams distribution; (d) Spectrogram; (e) Cone-kernel distribution; (f) Reduced interference distribution. The modified Wigner distribution eliminates the interference terms while retaining high energy concentration of signal components.

where C_h and C_g are respectively the centers of energy of the low-pass and high-pass quadrature filters h and g [5], defined by

$$C_h = \frac{1}{\|h\|^2} \sum_{k \in \mathbb{Z}} k|h_k|^2 , \quad C_g = \frac{1}{\|g\|^2} \sum_{k \in \mathbb{Z}} k|g_k|^2 , \tag{15}$$

$R(n)$ is an integer obtained by bit reversal of n in a $L - \ell$ bits binary representation, and GC^{-1} is the inverse Gray code permutation. The width and height of the tile are given by

$$\Delta t = 2^{-\ell} , \quad \Delta\omega = 2^{\ell - L} . \tag{16}$$

Fig. 4(a) shows the MWD for the signal $g(t)$, attained by utilizing expression (6) combined with the thresholds $D = 1.5$ and $\epsilon = 0.1$. Figs. 4(b)–(f) describe respectively the WD, the Choi-Williams distribution, the spectrogram, the cone-kernel distribution and the reduced interference distribution [2]. Clearly, the MWD achieves high resolution and concentration in time-frequency, and is superior in eliminating interference terms associated with the WD.

TRANSLATION-INVARIANT DENOISING

The use of wavelet bases for estimation of signals embedded in noise has been the object of considerable recent research. While traditional methods often remove noise by low-pass filtering, thus blurring the sharp features in the signal, wavelet-based methods show good performance for a wide diversity of signals, including those with jumps, spikes and other nonsmooth features [10, 11, 12]. The transform-based thresholding method consists of three steps: transformation of the noisy data into a time-scale domain, soft or hard thresholding to the resulting coefficients, and transformation back into the original space. This scheme necessitates determination of the "best" basis and threshold value, leading to the best signal estimate. It is useful to employ the library of wavelet-packet bases as a collection of competing models, and select the best model according to the *Minimum Description Length* (MDL) criterion [11, 13]. However, denoising based on the conventional WPD may exhibit visual artifacts, attributable to the lack of shift-invariance [12].

One approach to attaining shift-invariance is to average the translation dependence: applying a range of shifts to the noisy data, denoising the shifted versions with the wavelet transform, then unshifting and averaging the denoised data [12]. This procedure, termed *Cycle-Spinning*, generally yields better visual performance on smooth parts of the signal. However, transitory features may be significantly attenuated [14].

In this section, we present a translation-invariant signal estimator, which is based on the SIWPD and the MDL criterion. A collection of signal models is generated using the *extended* library of orthonormal wavelet-packet bases, and an additive cost function, approximately representing the MDL principle, is derived. We show that the minimum description length of the noisy observed data is achieved by utilizing the SIWPD and thresholding the resulting coefficients. This estimator is efficiently combined with the MWD, yielding robust time-frequency representations that are characterized by high resolution and suppressed interference-terms.

Let $y(t) = f(t) + z(t)$ represent the noisy observed data, where $f(t)$ is the unknown signal to be estimated, and $z(t)$ is a white Gaussian noise with zero mean and a known power spectral density σ^2. Denote by $\mathcal{B}$ the extended library of wavelet packet bases. The description length of y represented on a basis $B \in \mathcal{B}$ is given by [9]

$$\mathcal{L}(By) = \sum_{(\ell,n,m)\in E} \mathcal{L}(B_{\ell,n,m}y) \tag{17}$$

where

$$\mathcal{L}(B_{\ell,n,m}y) = 3 + \frac{1}{2\sigma^2 \ln 2} \sum_{k=1}^{2^\ell N} \min\left\{C^2_{\ell,n,m,k}(y)\,,\; 3\sigma^2 \ln N\right\} \tag{18}$$

is the codelength associated with a terminal node (ℓ, n, m), and

$$B_{\ell,n,m}y = \left\{C_{\ell,n,m,k}(y) = \langle y\,,\ \psi_{\ell,n,m,k}\rangle\ :\ 1 \le k \le 2^\ell N\right\} \tag{19}$$

are the expansion coefficients of the observed data. Since the codelength in Eq. (17) constitutes an additive cost function, the SIWPD gives the optimal basis according to the MDL principle. The optimal estimate of $f(t)$ is obtained by expanding the observed data $y(t)$ on the optimal basis $A = \left\{\hat{\phi}_k\right\}_{1\le k\le N}$ and hard-thresholding the coefficients by $\tau \equiv \sigma\sqrt{3\ln N}$. Specifically,

$$\hat{f}(t) = \sum_{k=1}^{N} \eta_\tau(y_k)\hat{\phi}_k(t) \tag{20}$$

where $y_k = \left\langle y, \hat{\phi}_k\right\rangle$, and $\eta_\tau(c) \equiv c\mathbf{1}_{\{|c|>\tau\}}$ is the *hard-threshold* function. The time-frequency distribution estimate of y is obtained by computing the MWD for the signal estimate:

$$\hat{T}_y(t,\omega) = T_{\hat{f}}(t,\omega) = \sum_{k\in\Lambda} |y_k|^2 W_{\hat{\phi}_k}(t,\omega) + 2\sum_{\{k,k'\}\in\Gamma} \mathrm{Re}\{y_k y^*_{k'} W_{\hat{\phi}_k,\hat{\phi}_{k'}}(t,\omega)\} \tag{21}$$

where

$$\Lambda = \left\{k\ :\ |y_k| > \sigma\sqrt{3\ln N},\ 1 \le k \le N\right\}, \tag{22}$$

$$\Gamma = \left\{\{k,k'\}\ :\ k,k' \in \Lambda,\ 0 < d(\hat{\phi}_k, \hat{\phi}_{k'}) \le D\right\}. \tag{23}$$

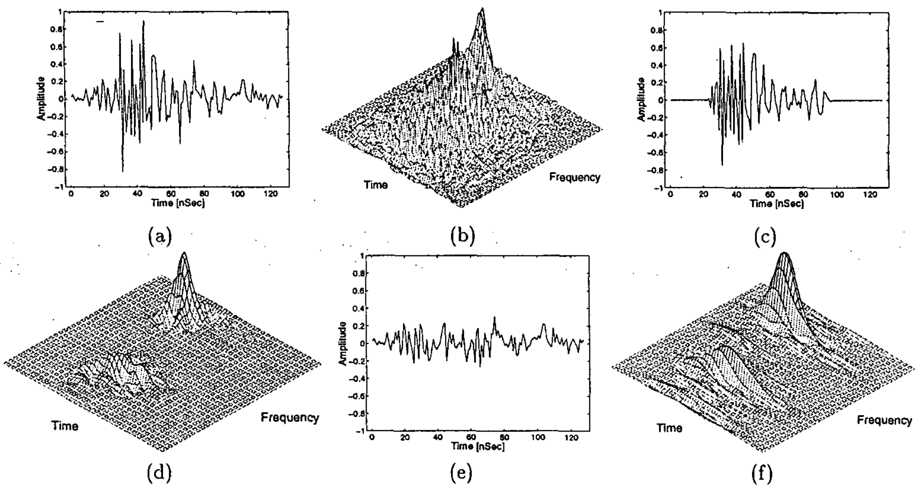


Figure 5: Electromagnetic pulse in a relativistic magnetron (heterodyne detection; local oscillator= 2.6GHz): (a) Noisy measurement $y(t)$. (b) Wigner distribution for $y(t)$. (c) The signal estimate $\hat{f}(t)$ by the MDL principle. (d) The estimate of the modified Wigner distribution for $y(t)$. (e) Residual between $y(t)$ and $\hat{f}(t)$. (f) Smoothed pseudo Wigner distribution for $y(t)$.

Fig. 5(a) shows a noisy measurement of an electromagnetic pulse ($\approx$ 100 nanoseconds long) generated by high power ($\approx$ 100 MegaWatts) relativistic magnetron. The measurement involves heterodyning at 2.6GHz, filtering at 500kHz and sampling at 1GHz [3]. The Wigner distribution, depicted in Fig. 5(b), is clearly ineffective as a time-frequency analysis tool, for its high noise sensitivity. Yet, the estimates of the signal and the MWD, as shown in Figs. 5(c) and (d), are potentially valuable when analyzing the measurements and studying the non-stationary phenomena, such as mode build-up and competition and pulse shortening [15], which are common in such high power microwave tubes. The residual between the noisy measurement and the signal estimate is depicted in Fig. 5(e). To ascertain that this residual is actually the noise component, we compare the estimate of the MWD with the smoothed pseudo Wigner distribution of the noisy measurement (Fig. 5(f)). Since these two distributions are similar, in view of the fact that smoothing in the Wigner domain reduces the noise at the expense of smearing the signal components, it is reasonable to assume that the signal estimate contains all the signal components and the residual is mostly noise. A detailed derivation and implementation of the proposed estimator, examples illustrating its performance, and a discussion of the relation to other work is given in [9].

CONCLUSION

Cross terms associated with bilinear distributions are not necessarily interpretable as interference terms. Any signal can be broken up in an infinite number of ways, each of which generates different cross terms. Therefore, it is important to choose an appropriate decomposition that separates the parts which are well delineated in the time-frequency plane. We have presented a modified Wigner distribution, where undesirable interference-terms can be eliminated while still retaining high energy concentration.

A prescribed signal is expanded into a redundant library of orthonormal wavelet-packet bases, from which the best decomposition is selected, and subsequently transformed into the Wigner domain. The discrimination between beneficial cross terms, which primarily enhance the useful properties of the time-frequency representation, and undesirable interference terms is determined according to the degree of adjacency and relative amplitudes of the interacting basis functions; Only adjacent pairs whose coefficients are large enough are related to the same component of the signal. The balance between interference terms, concentration and computational complexity is achieved by adjusting the distance and amplitude thresholds.

A translation-invariant denoising method, which uses the SIWPD and the MDL criterion has been described . The MDL principle is applied for approximating the description length of the noisy observed data and for choosing the optimal wavelet-packet basis. The proposed signal estimator, combined with the modified Wigner distribution, generates robust time-frequency representations.

REFERENCES

[1] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall Inc., 1995.

[2] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations", *IEEE SP Magazine*, Apr. 1992, pp. 21–67.

[3] I. Schnitzer, A. Rosenberg, C. Leibovitch, M. Botton, I. Cohen and J. Leopold, "Evolution of spectral power density in grounded cathode relativistic magnetron", *Proc. SPIE, Intense Microwave Pulses IV*, Vol. 2843, Aug. 1996.

[4] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection", *IEEE Trans. Inform. Theory*, Vol. 38, No. 2, Mar. 1992, pp. 713–718.

[5] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, AK Peters, Ltd, Wellesley, Massachusetts, 1994.

[6] I. Cohen, S. Raz and D. Malah, "Orthonormal shift-invariant wavelet packet decomposition and representation", *Signal Processing*, Vol. 57, No. 3, Mar. 1997, pp. 251–270.

[7] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM Press, Philadelphia, Pennsylvania, 1992

[8] I. Cohen, S. Raz and D. Malah, "Eliminating interference terms in the Wigner distribution using extended libraries of bases", *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-97*, Munich, Germany, 20–24 Apr. 1997, pp. 2133–2136.

[9] I. Cohen, *Shift-Invariant Adaptive Wavelet Decompositions and Applications*, D.Sc. Dissertation, Technion - Israel Institute of Technology, Haifa, Israel, May 1998.

[10] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", *Biometrica*, Vol. 81, 1994, pp. 425–455.

[11] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion", in: E. Foufoula and P. Kumar, eds., *Wavelets in Geophysics*, Academic Press, 1994, pp. 299–324.

[12] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 125–150.

[13] H. Krim, and J.-C. Pesquet, "On the statistics of best bases criteria", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 193–207.

[14] N. A. Whitmal, J. C. Rutledge and J. Cohen, "Wavelet-based noise reduction", *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95*, Detroit, Michigan, 8–12 May 1995, pp. 3003–3006.

[15] J. Benford and J. Swegle, *High Power Microwaves*, Artech House, Norwood, 1992.