Gaussian Wavelet Features and Their Applications for Analysis of Discretized Signals

G.Ososkov [a,1], A.Shitov [b]

[a] *Laboratory of Computing Techniques and Automation, Joint Institute for Nuclear Research, Dubna, 141980 Russia,*
e-mail: osg@lcta39.jinr.dubna.su

[b] *Department of Theoretical Physics, Ivanovo State University, Ermaka str. 37, 153317 IVANOVO, Russia*
e-mail: shitov@ivgu.polytech.ivanovo.su

Abstract

Problems of analysis of symmetric, bell-shaped signals registered in a discrete form are considered. Fast and direct methods for their processing are proposed on the basis of vanishing momentum wavelets. Unlike previous works wavelets of higher order are used extensively in these methods. A new wavelet feature is observed: the permanence of their relative square. It makes possible to choose an optimal scale coefficient that is common for several wavelet-transforms. Numerical simulations show the high accuracy of proposed algorithms comparable with the more laborious methods of a gaussian fitting to discrete measurements.

Keywords: Wavelets; Shift and dilation; Vanishing momentum wavelet; Discretization; Contamination; Superposed signals; Wavelet spectrum

1 Introduction

Analysis of signals registered in a discrete form is of importance for data processing in almost any field of experimental physics (see, for example [1–3]). In many important cases the signal form is symmetric, bell-shaped, in particular, a gaussian is the most often used approximant due to many physical reasonings [1,2,4]. The problem is to evaluate parameters of such a discrete signal, i.e. its position, amplitude and also its half-width in presence of noise, detector uncertainties and influence of other close signals. Rigid timing requirements inherent in contemporary detectors demand to elaborate a fast

[1] Supported by RFFI, grant N 97-01-01027, Russia.

and direct methods for the signal parameter evaluations. Passing through a detector a particle produces an electronic shower. Its bell-shape surface can be approximated by 2D-gaussian:

$$N(x, y, A, x_0, y_0) = A \exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}\right), \tag{1}$$

where x_0, y_0 – are the shower center coordinates, A - is the maximum amplitude. The shower half-widths σ_x, σ_y by corresponding axes are supposed to be known for the considered detector domain (or can be calculated from the known drift velocity). Due to the factorized wiev of (1) the problem usualy is reduced to handle several 1D-gaussians:

$$g(x; A, x_0) = A \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right). \tag{2}$$

In such 1D presentation (2) a doublet of two overlapping signals can be approximated as

$$G(x; A, x_1; B, x_2) = A \exp\left(-\frac{(x-x_1)^2}{2\sigma^2}\right) + B \exp\left(-\frac{(x-x_2)^2}{2\sigma^2}\right). \tag{3}$$

In the process of registration a signal is to be discretized as a histogramm $\{h_k\}$ on the interval (x_{beg}, x_{end}),

$$h_k = \frac{1}{\tau} \int\limits_{x_{k-1}}^{x_k} G(x; A, x_1; B, x_2)\, dx, \tag{4}$$

where $x_k = x_{beg} + k\tau$, τ – is the bin width. Besides an electronic noise gives a contribution to each histogram bin. Noise values are, in principle, correlated, but according to the earlier study [8] we treat them as independent normal variables with rms equal up to 10% of the mean amplitude. The weak signals are cut off on the threshold depending on the maximum noise level. However, despite of noise thresholding, some background signals contaminating useful signals can also appear above the threshold level.

Well-known restrictions of Fourier analysis motivated our interest to such a modern signal analysis mean as **wavelet-transfom** [5] The wavelet transform of a signal $f(x)$ is determined as

$$W_\psi(a, b)f = \frac{1}{\sqrt{C_\psi}} \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}} \psi\left(\frac{b-x}{a}\right) f(x)\, dx. \tag{5}$$

with the normalizing constant

$$C_\psi = 2\pi \int\limits_{-\infty}^{\infty} \frac{|\tilde{\psi}(\omega)|^2}{|\omega|}\, d\omega \;<\; \infty, \tag{6}$$

where $\tilde{\psi}(\omega)$ – is the Fourier transform of the wavelet $\psi(x)$. The condition $C_\psi < \infty$ is, at the same time, the condition of the wavelet ψ existence. It would be true, in particular, if the first $n-1$ momenta are equal to zero:

$$\int\limits_{-\infty}^{\infty} |x|^m \psi(x) dx = 0, \quad 0 \le m < n. \tag{7}$$

According to the gaussian-like shape of our signals it is natural to choose, as a basic wavelets, the family of **vanishing momentum wavelets** (VMW), since they are generated by gaussian distribution function:

$$g_n(x) = (-1)^{n+1} \frac{d^n}{dx^n} \mathrm{e}^{-x^2/2}, \;\; n > 0 \tag{8}$$

The VMW family is called so because the condition (7) is always holds for it. Two first of VMW are most known [5]:

$$g_1(x) = -x\mathrm{e}^{-\frac{x^2}{2}}, \qquad g_2(x) = (1-x^2)\mathrm{e}^{-\frac{x^2}{2}}.$$

(the second one is also known as "the Mexican hat")
Nevertheless, we found that the power of the wavelet-analysis can be really extended, if we would use the higher order VMW, in particular,

$$g_3(x) = (3x - x^3)\mathrm{e}^{-\frac{x^2}{2}}, \qquad g_4(x) = (6x^2 - x^4 - 3)\mathrm{e}^{-\frac{x^2}{2}}.$$

The normalizing coefficients of these wavelets C_{g_n} are $2\pi(n-1)!$

2 VMW properties

Between useful VMW properties we stress two, which are related to the VMW derivatives and integrals:

$$\frac{d\, g_n(x)}{dx} = -g_{n+1}(x), \qquad \int\limits_{x_1}^{x_2} g_n(x) = g_{n-1}(x_1) - g_{n-1}(x_2). \tag{9}$$

The most important VMW property consists in preserving their relative square, which we define as [6]

$$w(x) = \frac{\int\limits_0^x |g_n(x)|\, dx}{\int\limits_0^\infty |g_n(x)|\, dx}. \tag{10}$$

The first four of VMWs are shown in Fig.1. As we have checked, the relative

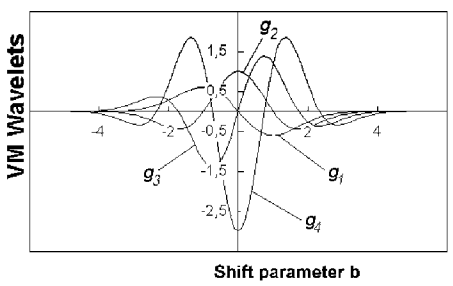

Fig. 1. First four of VMWs

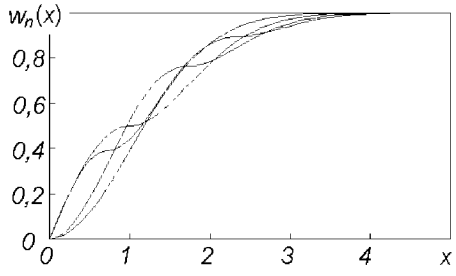

Fig. 2. Their relative squares

VMW squares calculated for the first ten of VMWs are almost equal forming a specific narrow plait. One can see that clearly in Fig.2 for the first four VMWs. That is used below in sect. 4 for the optimal choise of the dilation parameter.

It is a remarkable fact that the wavelet transformation of a gaussian (2) looks as the corresponding wavelet. Therefore the general expression for the n-th wavelet coefficient has the following form:

$$W_{g_n}(a,b)g = \frac{A\sigma a^{n+1/2}}{\sqrt{(n-1)!}s^{n+1}} g_n\left(\frac{b-x_0}{s}\right),$$

where we denote $s = \sqrt{a^2+\sigma^2}$.

Thus all above-mentioned VMW features are valid also for the gaussian wavelet-transfom. In particular, at the central point $x = x_0$ coefficients of odd VMWs $W_{g_1}(a,x_0)g$ and $W_{g_3}(a,x_0)g$ equal to zero, while coefficients of even, symmetrical wavelets $W_{g_2}(a,x_0)g$ and $W_{g_4}(a,x_0)g$ obtain at this point their maximum (absolute) values

$$W_{g_2}(a,x_0)g = \frac{A\sigma a^{5/2}}{s^3}, \quad W_{g_4}(a,x_0)g = -\sqrt{\frac{3}{2}}\,\frac{A\sigma a^{9/2}}{s^5}. \tag{11}$$

To present the wavelet coefficients of a doublet (3) of two gaussians we use the simplifying normalization:

$$W_n(a,b)g = \frac{W_{g_n}(a,b)g}{w_{g_n}}, \quad w_{g_n} = \frac{A\sigma a^{n+1/2}}{\sqrt{(n-1)!}s^{n+1}}.$$

Then we obtain

$$W_{g_n}(a,b)G = Ag_n\left(\frac{b-x_1}{s}\right) + Bg_n\left(\frac{b-x_2}{s}\right) \tag{12}$$

3 Methods to estimate signal parameters

At this section we derive, first, VMW-based estimates of parameters for "ideal", non-distorted, non-histogrammed signals. These estimates give us the good basis for direct and, therefore, fast algorithms. However in any real registration process various distortions are brought in. So we have then to study, how much these distortions could influence on the wavelet-transfom of such a distorted signal. Only on the basis of this study we can propose recommendations for an optimal choice of shift and dilation VMW parameters, which would guarantee the applicability of proposed algorithms.

Single gaussian signal. For the single gaussian signal we can calculate wavelet transform in a few points and solve the system of corresponding equations. However, applying the ratio of different wavelets we can eliminate the exponent $e^{\left(-\frac{(b-x_0^2)}{2(a^2+\sigma^2)}\right)}$ and obtain the signal position explicitly. For instance, the ratio $W_{g_3}(a,b)g/W_{g_1}(a,b)g$ gives

$$x_0 = b \pm \sqrt{(a^2+\sigma^2)\left[3 + \frac{\sqrt{2}\,(a^2+\sigma^2)}{a^2}\frac{W_{g_3}(a,b)g}{W_{g_1}(a,b)g}\right]}. \tag{13}$$

The true sign in (13) is easy to choose when one would calculate the coefficients $W_{g_3}(a,b)g$ and $W_{g_1}(a,b)g$ in a point, which is far enough from the supposed signal position.

The amplitude value can be evaluated via the value of the half-width of the signal σ (if known) and one of expressions (11). But in the case when the value of σ is unknown it can be also evaluated using $W_{g_2}(a,x_0)g/W_{g_4}(a,x_0)g$:

$$\sigma^2 = -a^2\left(1 + \sqrt{\frac{3}{2}}\frac{W_{g_2}(a,x_0)g}{W_{g_4}(a,x_0)g}\right). \tag{14}$$

Again, the point, in which the ratio $W_{g_2}(a,b)g/W_{g_4}(a,b)g$ is calculated, must be chosen as close to the signal center, as possible.

Doublet of close gaussians. For a doublet of two close signals we can use either

- four first wavelets calculated in one point (method WTS - Wavelet Transform System)
- or one of those wavelets (we choose g_2) calculated in four different points (method g_2-WTS).

The corresponding systems of equations are:

for WTS: $F_n = W_n(a,b)G - W_n(a,b)h = 0, \qquad n = 1,2,3,4;$
for g_2-WTS: $F_n = W_2(a,b_n)G - W_2(a,b_n)h = 0, \qquad n = 1,2,3,4;$

where n-th wavelet coefficient of a histogram h

$$W_n(a,b)h = \frac{s^{n+1}}{\sqrt{2\pi}\,\sigma a^{n+1}} \sum_{k=1}^{N} h_k \int_{x_{k-1}}^{x_k} g_n\left(\frac{x-b}{a}\right) dx$$

is calculated from the source histogram only one time for all iterations. g_2 in g_2-WTS method are taken in four following points: $b_1 = b - h/2, \quad b_2 = b + h/2, \quad b_3 = b_1 - h, \quad b_4 = b_2 + h$ with specially chosen h.

Newton's method is applied to solve these non-linear systems

$$\mathbf{D}\Delta\mathbf{X} = \mathbf{F}, \;\; \mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, \mathbf{A}, \mathbf{B})^{\mathbf{T}}, \;\; \mathbf{F} = (\mathbf{F_1}, \mathbf{F_2}, \mathbf{F_3}, \mathbf{F_4})^{\mathbf{T}},$$

where $\mathbf{D}$ is the matrix of partial derivatives of F_n with respect to components of $\mathbf{X}$. The first approach $\mathbf{X}^{(0)}$ is obtained by a rough estimation of the signal parameters from the source histogram h (see, for example, [8]). The next approach is obtained as $\mathbf{X}^{(1)} = \mathbf{X}^{(0)} + \Delta\mathbf{X}$ and so on iteratively, unless a wanted accuracy is reached. Calculations of the partial derivatives matrix $\mathbf{D}$ can be considerably simplified due to the above mentioned VMW-feature (9).

4 Optimal choise of shift and dilation parameters

As we assume, the choice of VMW parameters should minimize the influence of various signal distortions. Some results of a study of this influence are described below. Here we focused ourselves especially on the signal distortion due to digitizing and its contamination by additive noise. A histogrammed signal can be considered as

$$\tilde{h}_k = h_k + \varepsilon_k, \quad k = \overline{1,N}, \tag{15}$$

where h_k is non-distorted signal obtained according to (4) and ε_k is the noise addition. Therefore wavelet-coefficients should also consists of two parts

$$W_{g_n}(a,b)\tilde{h} = W_{g_n}(a,b)h + W_{g_n}(a,b)\varepsilon. \tag{16}$$

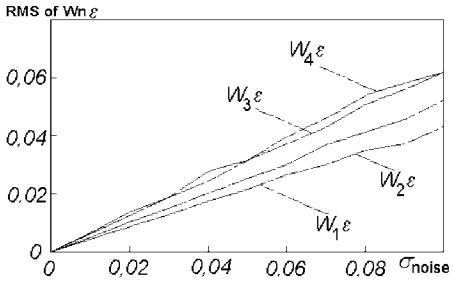

Fig.3. RMS of the wavelet-transforms $W_{g_n}(a,0)\varepsilon$ versus of the noise level

To estimate quantitatively the effect of the second, noise component a numerical experiment was done with 1000 events imitating the contamination of a digitized signal by normally distributed random noise ε with the standard deviation σ_{noise}. The corresponding values of the RMS of $W_{g_n}(a,0)\varepsilon$ are shown in Fig.3.

The non-distorted gaussian signal with the amplitude $A = 1$, zero center and $\sigma = 1.25$ was then histogrammed for different shift values b. The wavelet-transforms $W_{g_n}(a,b)h, \quad n = 1,2,3,4$ of this signal as the functions of b are

depicted in Fig.1. Comparing them with corresponding rms values from Fig.3, one can see that even the maximum wavelet coefficients of noise are two orders of magnitude less than wavelet-coefficients $W_{g_n}(a,b)h$ for almost all values of b except small, clearly distinctive areas for odd and even wavelets. That gives us the following rule for optimal choosing of the shift parameter b:

- odd wavelets g_1 and g_3 must be calculated in points on signal tails;

- even wavelets g_2 and g_4 must be calculated in points cituated as close to the signal center as possible.

The rule to determine the optimal delation parameter a can be derived from the expression (10) for the invariant relative square w

$$a = \frac{v}{2}\sqrt{\frac{-1}{2\,\ln(1-w)}},\ \ v = x_2 - x_1, \quad 0 < w < 1. \tag{17}$$

5 Simulations and results

The amazing insensitivity of wavelets to various signal distortions is widely known (see, for example, D.Donoho's article in [5]). However it was necessary to test more in details the accuracy and efficiency of proposed methods and the dependences of their results on such factors as (i) distance $d = |x_2 - x_1|/\sigma$ between two components of the doublet signal (3), (ii) noise level and bin shedding, (iii) detector granularity degree. We use Monte-Carlo simulations to fulfil this study. The data were simulated as follows. The gaussian doublet (3) with $\sigma = 4$ was simulated and histogrammed according to (4) with bin number N and bin-size 1 as it depicted in Fig.4-a (N=10) and Fig.4-b (N=32). Both peak positions and amplitudes are randomly distributed (exponential distribution with the mean $\bar{A}$ was used to generate amplitudes. The single gaussian signal (2) needed to test the accuracy of the formula (14) was simulated analogically. Then discretized signals were distorted by adding to the value in each histogram bin a noise value distributed by the normal or (optionally) by the uniform distribution with $\sigma_{noise} = 0.1\overline{A}$). Then weak signals with the amplitude less than 10% of A_{max} were truncated.

Both methods WTS and g_2-WTS were applied to estimate doublet parameters x_1, x_2, A, B. Calculating wavelet-coefficients, we set up the dilation parameter a according to (17) with $w = 0.9$ and $v = 2|\hat{x}_2 - \hat{x}_1|$, where $\hat{x}_1, \hat{x}_2$ are the estimates of position parameters obtained on the current iteration. The maximum number of iteration was set to 10, but it was usually not more than 3-4.

Each point on the dependence plot was obtained by repeating the whole procedure 1000 times. This procedure was implemented as a special WINDOWS-95 software package allowing to vary all simulation parameters and to visualize calculated signals and spectra. Two examples of applying this package to study the influence of various signal distorsions on its VMW spectrum are pre-

sented in Fig.4. 2D wavelet-spectra are depicted as gray-level images ranging

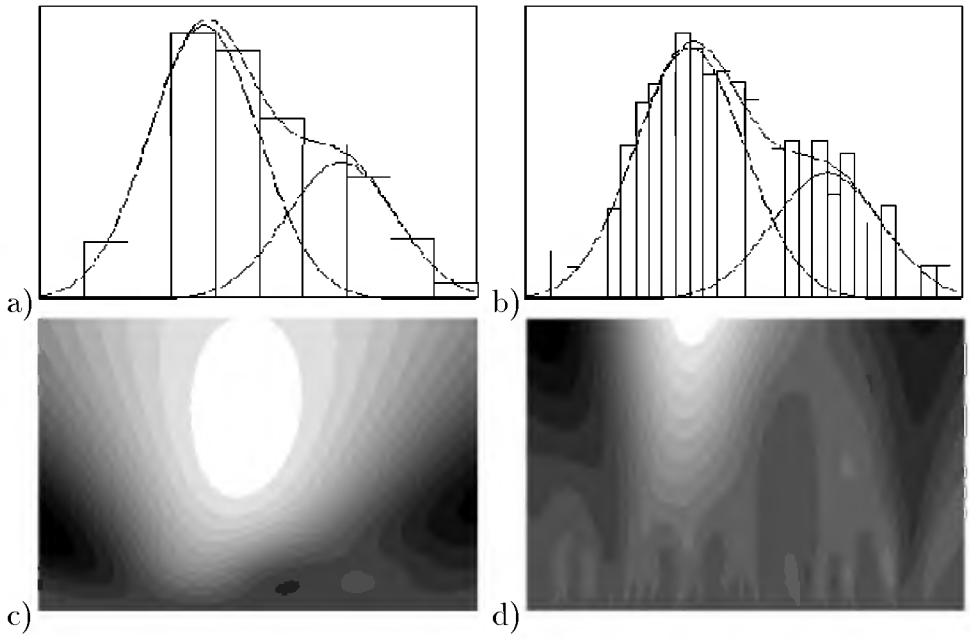

Fig. 4. Effects of various granularity, contamination and thresholding. (a) Histogram of a doublet with low granulatrity (10 bins). (b) The same signal, but discretized for 32 bins, then noise with 30% amplitude from the signal maximum is added to each histogram bin and cut-off is done on the 10% level. (c) An overal view of the wavelet-spectrum of the discretized signal (a). It stays the same for higher granular, distorted and cutoff signal (b) (d) High frequency part of spectrum for signal (b)

from black for minimum to white for maximum values. One can clearly see the striking robustness of the VMW to the various signal distortions: only a relatively thin high-frequency layer of the spectrum is violated while the rest of it looks as a spectrum of the non-distorted signal. As one can see in Fig.5, a similar robustness was developed when three of bins of the histogram depicted in Fig.4(b) were set to zero imitating a malfunctions of data channels.

Results of RMS-error dependence of the signal position estimation on the distance between two signal components determined by WTS and g_2-WTS methods are shown in Fig.6, where distances are given in the signal half-width units common for both components. Error values are given in bin-width. The first method has better accuracy when the distance between two peaks in a doublet is less than 2σ, although when it approaches 1σ-distance, the accuracy increases considerably for both methods. Relative RMS-error of the estimate $\hat{\sigma}_g$ of the single gaussian half-width grows linearly with random noise (in percents of $\bar{A}$) as $\hat{\sigma}_g = 0.2\sigma_{noise} + 0.01$. Dependence of $\hat{\sigma}_g$ on detector granularity is presented in Fig.7.

As one can see from the given examples, wavelet transform can develop even fine effects of asymmetry and other signal deviations from an ideal gaussian shape. We do not touch here questions of a quantitative estimations of such deviations, since it would bring us to the different topic related to wavelet series expansion. One can also indicate these deviations by calculating the

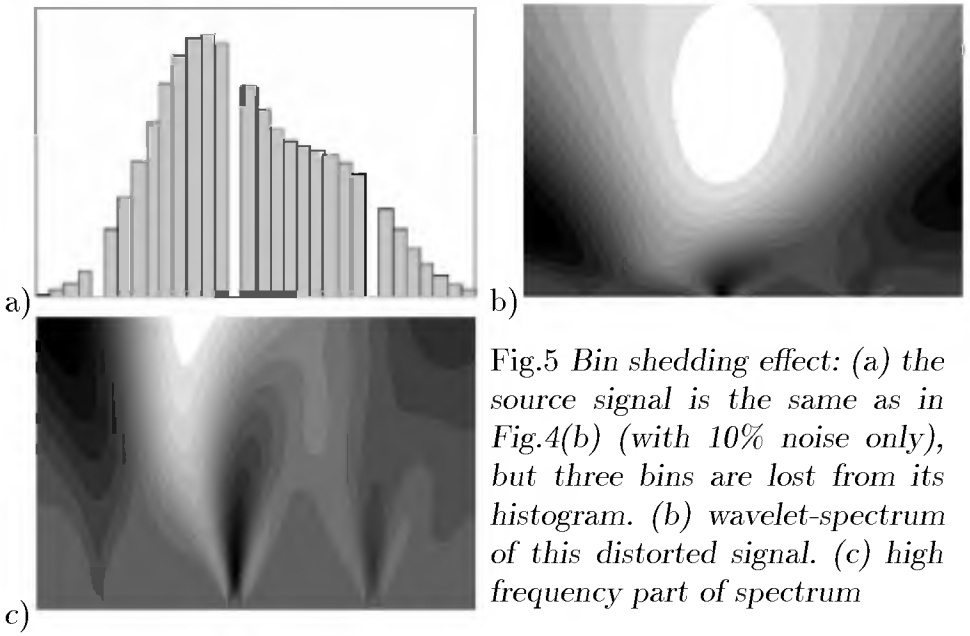

Fig.5 *Bin shedding effect: (a) the source signal is the same as in Fig.4(b) (with 10% noise only), but three bins are lost from its histogram. (b) wavelet-spectrum of this distorted signal. (c) high frequency part of spectrum*

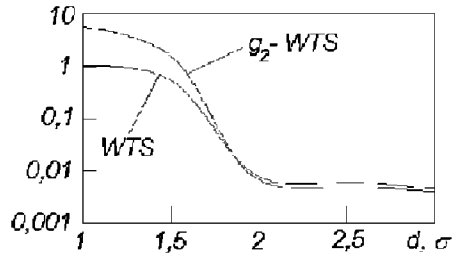

Fig. 6. RMS-error dependences of signal position estimates on the distance between two signal components determined by WTS and g_2-WTS methods.

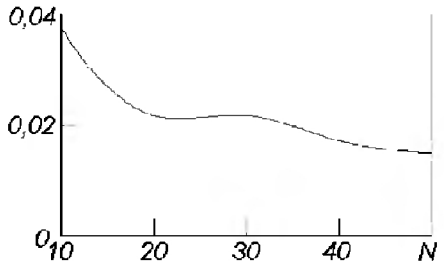

Fig. 7. RMS-error dependences of the single gaussian half-width estimate on detector granularity (bin number)

third and the fourth momenta of the histogrammed signal. We did not also compare here accuracies of the wavelet analysis and the Fourier approach, which is more familiar for experimentalists, since it was already done in [7] for the problem of resolving doublets of close gaussian discretized signals. As it was shown there in a simplified condition, when the position of one of signals is fixed, the rms of the wavelet estimation of the distance between both peaks is 20% better than of the Fourier method.

6 Conclusion

The direct formulae are derived to calculate location, amplitude and scale parameters for a single signal and doublets of overlapped signals. Unlike our previous work [7] wavelets of higher order are used extensively in these al-

gorithms. The observed VMW features allows to choose the optimal wavelet parameters. In particular, the stability of the VMW relative square was found. It makes possible to choose the dilation parameter that is common for several wavelet-transforms. Numerical simulations show the high accuracy of proposed algorithms is comparable with the more laborious methods of a gaussian fitting to discrete measurements [8]. This study was very facilitated by developing a special user friendly software package for the visualization of simulated signals and their 2D-wavelet spectra for any of the first six gaussian wavelets.

Approaches published here and in our previos work [6] are already successfully used by some of the JINR physicists [9], although for different purposes.

Authors would like to thank drs M.Altaiski and M.Khvastunov for the fruitful discussions and some useful remarks.

References

[1] H.Agakishiev et. al., Nucl.Instr. and Methods, **A371** (1996), 243-247.

[2] J.P.Wurm et. al. IEEE TRANS. On Nuclear Science, v. 39, No. 4, (1992), 619-628.

[3] E.Conti et. al., Nucl.Instr. and Meth. **A356** (1995), 286-296.

[4] E.Gatti, P.Rehak, M.Sampietro, Nucl.Instr. and Meth. in Phys. Res. **A274** (1989), 469-476.

[5] I. Daubechies Editor, Proc.of Symp. in Appl. Math, **V 47**, 1993

[6] G.Ososkov, A.Shitov, Commun. of JINR P11-97-347, Dubna, 1997 (in Russian).

[7] M.V.Altaisky, E.A.Kolganova V.E.Kovalenko, G.A.Ososkov, Proc. of the International conference of SPIE -The International Society for Optical Engineering, Proc. SPIE'96, v.2847 (1996) 656-664.

[8] H.Agakishiev et. al., Commun. of JINR E10-97-105, Dubna, 1997.

[9] M.Khvastunov, *Introduction to wavelet analysis: applications to signals with the gaussian shape*, JINR Rapid Commun. 5[91]-98, Dubna, 1998 (in Russian).