A peak-search method based on spectrum convolution

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS D: APPLIED PHYSICS

J. Phys. D: Appl. Phys. **36** (2003) 1903–1909 PII: S0022-3727(03)61606-2

A peak-search method based on spectrum convolution

A Likar[1,2] **and T Vidmar**[2]

[1] Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
[2] Jožef Stefan Institute, Ljubljana, Slovenia

Received 2 April 2003, in final form 13 May 2003
Published 16 July 2003
Online at stacks.iop.org/JPhysD/36/1903

Abstract
A peak-search method for high resolution gamma-ray spectroscopy, based entirely on spectrum convolution, is presented. The detection of the peak and determination of the peak position, width and area can be realized in successive steps, independent of each other. To realize each of the steps, the spectrum is convoluted with an appropriate near-optimal function. Analytical expressions for the uncertainties of positions, areas and widths can be derived. For resolving doublets the standard least-square technique is used, but on the convoluted spectra, avoiding the problems associated with the background component. The performance of the software package based on the proposed method was successfully tested against two standard sets of test spectra of the International Atomic Energy Agency and intercomparison with other peak-search packages proves that the proposed approach is robust and reliable.

1. Introduction

The first step in the analysis of a high resolution gamma-ray spectrum is usually the determination of the peak positions, widths and areas. In routine applications the spectra are analysed using a suitable peak-search routine. There are many such routines available (see [1])[3], all sharing the same basic approach: the data in the region of a peak are viewed as a sum of two features, the 'background' and the 'peak'. Analytical forms for both components are usually assumed and peak location is then performed, followed by the peak area evaluation. The latter always takes on either of the two forms: addition of the counts in the peak within a prescribed region or modelling of the peak shape with an appropriate function. In both cases a separate determination of the background shape is also required.

All but the simplest of these procedures require non-linear least-square fitting of functions of several parameters to the peak itself or the background beneath it. Although well established, the numerical procedures that carry out this fitting are demanding and do not always yield reliable estimates of the uncertainties of the obtained parameters. The problem becomes more difficult when double peaks are to be resolved. In many peak-search techniques, the peak shape is then described as a linear combination of several functions [2]. Amplitudes of these functions can be determined accurately only for the minority of the peaks in an average spectrum and if the spectrum is poor due to a short measurement time, such an approach is not justified. The uncertainty of the peak area is also increased since many of these functions are located outside the interval where the major portion of the peak lies and in the region where the background is dominant.

In this paper we report on a peak-search method which enables maximal control over its performance and builds a specific context of the measurement already at the first stage of the spectrum analysis. The novelty of the proposed method is that we treat at all stages of the peak-search procedure the convoluted spectrum instead of the original one. This approach is usually only used for peak location. In [3] the convolution method was used to determine the peak area, assuming the Gaussian profile of the peaks with known width. It can be shown, however, that working with the convoluted spectrum also makes it possible to determine the widths and areas of the peaks and to do so searching for the optimal value of each of these parameters in turn, rather than at the same time, as it is the case with least-square fitting. Thus, the problem of determining the parameters of the peak shape is deconstructed into individual independent steps for each parameter, yielding reliable estimates of their values and uncertainties. Furthermore, there is no need to fit the shape of the background, since the background is filtered out of the original spectrum through the convolution process.

[3] Intercomparison of gamma-ray analysis software packages, IAEA-TECDOC-1011 (Vienna: IAEA) 1998.

The algorithm is implemented as a multiple pass code. In the first pass one finds the highest peaks and determines the sensitivity level. In the next pass the trend of the widths of the peaks is searched for and the areas of the unambiguous, well separated peaks are determined. In the third pass one uses the information about the widths to accurately determine the areas of weak peaks close to the strong ones and resolve the doublets.

Our approach assumes a simple Gaussian shape for the peak, thus neglecting a possible peak asymmetry and counts in the tails of the peak due to incomplete charge collection, pile-up or other effects. The advantage of this approach is, however, that it strongly reduces the statistical error of the peak area. A strong support for this statement comes from the International Atomic Energy Agency (IAEA) 1998 inter-comparison study[3], which demonstrated that the peak-shape model dependency of the study results was not found. All the programs tested there reported on average the same peak areas, irrespectively of the different peak-shape models used.

2. The method

The use of the convolution method to locate the peaks is a well-established approach and is utilized in many of the available codes. It started with Mariscotti [4, 5] who used a smoothed second derivative of the spectrum to suppress the background and enhance the peaks. The so-called correlation technique emerged later [6], using the second derivative of the Gaussian as the convolution function, called the correlator. Robertson *et al* [7] gave an optimal function, which minimized a specially defined signal-to-noise ratio. These and alternative functions were carefully analysed by Hnatowicz [8]. Since then the convolution method has been widely used for peak position determination, but never for peak area and peak width determination in the manner we propose here.

In most peak-search routines the peak localization is carried out using convolution techniques, while peak area evaluation is done on the original, non-convoluted spectrum. In our approach, in which a near-optimal convolution function is used, we extract the peak area, position and width from the convoluted spectrum alone. We believe such an approach keeps the statistical errors close down to the optimal level for all the relevant quantities. The transparency of the method is crucial for an easy to follow and robust computer code based on the suggested approach.

Our method of determining the peak parameters from the measured spectrum $p(x)$ is based on the transformation

$$I(y) = \int_{y-\beta}^{y+\beta} p(x) f(x-y)\,\mathrm{d}x \tag{1}$$

The interval of the integration is finite, $[y-\beta, y+\beta]$. A proper choice of the function $f(x-y)$ is the key to the successful implementation of the method. We assume that an ideal spectrum without statistical fluctuations, which we denote by $p_0(x)$, exhibits a line z on a background $\bar{n}(x)$

$$p_0(x) = Az(x, a, \sigma) + \bar{n}(x) \tag{2}$$

where $z(x, a, \sigma)$ represents the shape of the total absorption peak with the scalable width σ and the amplitude A proportional to the intensity of the line.

The function $f(x-y)$ is chosen so as to filter out the background $\bar{n}(x)$. This is achieved by the requirement

$$\int_{y-\beta}^{y+\beta} \bar{n}(x) f(x-y)\,\mathrm{d}x = 0 \tag{3}$$

We next assume that $I(y)$ reaches its maximum value at $y = a$. Since the background is filtered out by the transformation (1), the area of the peak can be deduced from $I(a)$ as

$$P = \alpha I(a) \tag{4}$$

where α has to be calculated knowing the detailed shapes of the functions $f(x-y)$ and $z(x, a, \sigma)$. The shape of the function $f(x-y)$ can be obtained from a restrictive requirement that the statistical uncertainty of the peak position a or the area P be minimal. It is not a difficult task to find such optimal functions. However, although being relatively simple, they are analytically not convenient enough to be used in an actual peak search routine. In the next section we introduce a good approximation to the function $f(x-y)$, which for Gaussian shaped peaks allows the use of symbolic computations.

Let us therefore assume that a line $L_i(x)$ in the spectrum $p(x)$ is perfectly Gaussian-shaped given by

$$L_i(x) = A_i z(x, a, \sigma_i) = A_i \exp\left(-\frac{(x-a_i)^2}{2\sigma_i^2}\right) \tag{5}$$

The parameters A_i and a_i, being the intensity of the line and its position, respectively, are the ones we would like to extract from the spectrum. However, the width of the line, given through the standard deviation σ_i, is also a very informative parameter which makes it possible to judge the quality of the spectrum. The method we elaborate here is heavily based on the proper determination of the standard deviation, which is the key parameter in obtaining the intensity A_i.

In order to filter out from the spectrum any structure which has the desired width and is therefore a good candidate for the total absorption line $L_i(x)$, we convolute the spectrum $p(x)$ with the first order derivative of an unnormalized Gaussian function

$$z(x, y, \sigma_z) = \exp\left(-\frac{(x-y)^2}{2\sigma_z^2}\right)$$

whose maximum value is unity. The width of the function $z(x, \sigma_z)$ is given by the standard deviation σ_z, which we consider as a free parameter for the moment. We later explain how it should be adjusted to the peak shape. Explicitly, we have

$$\frac{\mathrm{d}z(x, y, \sigma_z)}{\mathrm{d}x} = -\frac{x-y}{\sigma_z^2} \exp\left(-\frac{(x-y)^2}{2\sigma_z^2}\right) \tag{6}$$

and adopt this function as a suitable representative of the $f(x-y)$. The convolution

$$I_1(y) = \int_{-\infty}^{\infty} p(x) \frac{\mathrm{d}z}{\mathrm{d}x}\,\mathrm{d}x \tag{7}$$

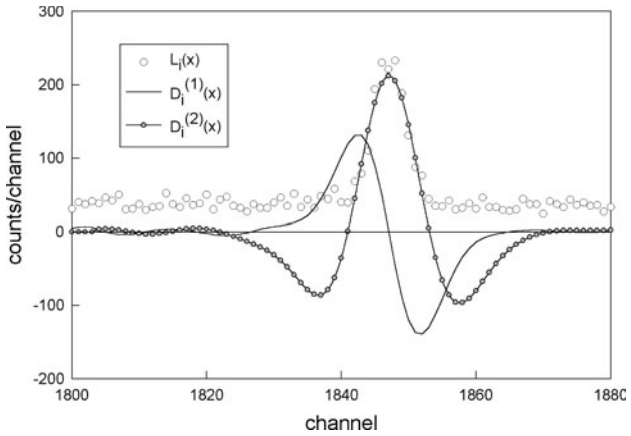

Figure 1. The filtering-out of the background with broad features and the enhancement of the lines $L_i(x)$ (○) is achieved by using the convolution of the spectrum with $\mathrm{d}z/\mathrm{d}x$, defined by equation (6), which yields the derivative-like structure (——). The second convolution of this result with the same function is also shown (solid line with grey dots).

completely filters out any background with rather broad features and strongly enhance the shape of the lines $L_i(x)$ resulting in $D_i^{(1)}(y)$:

$$D_i^{(1)}(y) = \frac{\sqrt{2\pi}(y-a_i)\sigma_z\sigma_i}{(\sigma_i^2+\sigma_z^2)^{3/2}} \exp\left(-\frac{(y-a_i)^2}{2(\sigma_i^2+\sigma_z^2)}\right) \tag{8}$$

The function is shown in figure 1 in solid line along with the corresponding part of the spectrum $p(x)$ (open circles).

The search for the lines is performed by scanning the convoluted spectrum $I_1(y)$ for the characteristic features of the functions $D_i^{(1)}(y)$. Since these functions exhibit clearly detectable positive maxima and negative minima, it is an easy task to locate them. An optimum level, determined from the fluctuation of the baseline due to the background is used, above which the maxima of the $D_i^{(1)}(y)$ are detected, reducing the noise from the background even further. A coarse determination of the positions a_i can be made at this stage by observing the cross-over point and determination of the widths σ_i can be made using the distance between the maximum and the minimum of the functions $D_i^{(1)}(y)$, but the main purpose of this step is to determine the interval where the peak lies.

Once the peak interval has been reliably determined, there are several ways which lead us to the intensity A_i. We follow here the most straightforward one. Intuitive reasoning suggests that the feature $D_i^{(1)}(y)$ in the convoluted spectrum $I_1(y)$ is most selectively detected if the spectrum $I_1(y)$ is convoluted once more with the same function $\mathrm{d}z/\mathrm{d}x$ as above. Following this proposition, we arrive at the spectrum $I_2(y)$, defined as

$$I_2(y) = \int_{-\infty}^{\infty} I_1(x)\frac{\mathrm{d}z}{\mathrm{d}x}\,\mathrm{d}x \tag{9}$$

The $D_i^{(1)}(y)$ in the spectrum $I_1(y)$ is transformed into $D_i^{(2)}(y)$, which has the form

$$D_i^{(2)}(y) = \frac{2\pi\sigma_z^2\sigma_i((y-a_i)^2-\sigma_i^2-2\sigma_z^2)}{(\sigma_i^2+2\sigma_z^2)^{5/2}} \times \exp\left(-\frac{(y-a_i)^2}{2(\sigma_i^2+2\sigma_z^2)}\right) \tag{10}$$

The shape of a function $D_i^{(2)}(y)$ is presented in figure 1 with a line with full circles. We see that it exhibits three extrema, of which the maximum at $y = a_i$ is the most prominent. The value of this maximum depends on σ_z and it is easy to show, by solving the equation $\mathrm{d}D_i^{(2)}/\mathrm{d}\sigma_z = 0$, that it reaches the extreme value when $\sigma_z = \sigma_i$. The study of the functions $\mathrm{d}D_i^{(2)}/\mathrm{d}\sigma_z$ therefore forms the basis for precise determination of the unknown parameters a_i and σ_i of the original lines. The search for the supreme value of the maximum and its position is performed simply by repeating the convolution of the original spectrum twice with the function $\mathrm{d}z/\mathrm{d}x$, always with a different value of the parameter σ_z, and storing the resulting maxima. It is easy to see that when the extreme value of the maximum, denoted by $I_2^{\sup}$, is found, the peak area $P_i = \sqrt{2\pi}\sigma_i A_i$ can be written as

$$P_i = \frac{9I_2^{\sup}\sigma_i}{\sqrt{6\pi}} \tag{11}$$

At the same time the reaching of the supremum yields the exact peak position.

Due to statistical variations of the counts in the spectrum the functions $D_i^{(1)}(y)$ deviate from the ideal shape. Such a deviation can also result from another nearby lying peak. We therefore examine the shape of the functions and note them in the form of a grade attributed to the peak. The highest grade is given to the functions $D_i^{(1)}(y)$ with negligible deviation from the ideal appearance. The decisive criterion is the similarity of the absolute values of the maximum and the minimum. Intense single peaks receive the best mark. The lowest grade reflects the poor symmetry of the function with a barely detectable maximum and minimum. Weak peaks close to the strong ones receive such a mark. Such peaks suffer from interference and the spectrum $I_1(y)$ has to be separately analysed for the peak areas. Since the peak positions are not impaired, the spectrum $I_1(y)$ is analysed by setting

$$I_1(y) = \sum_{i=1}^{2} -\frac{P_i}{2\sqrt{2}}\frac{y-a_i}{\sigma^2}\exp\left(-\frac{(y-a_i)^2}{2\sigma^2}\right) \tag{12}$$

where a_1 and a_2 are approximately known peak positions and σ the width deduced from the single peaks. Only the most intense singlets with very good marks are used to determine this energy dependence of the peak width. In any subsequent analysis of doublets the widths of the peaks comprising them are fixed at the value from this systematics. Alternatively, the user may specify the energy dependence of the FWHM in advance, based on calibration data, and this dependence can be used in the analysis of the doublets. In equation (12) P_1 and P_2 are the peak areas we would like to determine by the best fit to $I_1(y)$ obtained from the experimental spectrum. This is a well posed linear problem which can be solved using standard techniques. In our implementation the positions a_1 and a_2 are varied as well in order to find the best non-linear fit of the parameters to the singly convoluted spectrum. As mentioned above, the peak widths are kept fixed during the fit.

Interference and resolving the doublets is implemented as the second pass of our code. A similar technique is applied to the merged doublets. The candidates are detected through the comparison of the width of the line with the FWHM systematics determined in the first pass from strong singlets.

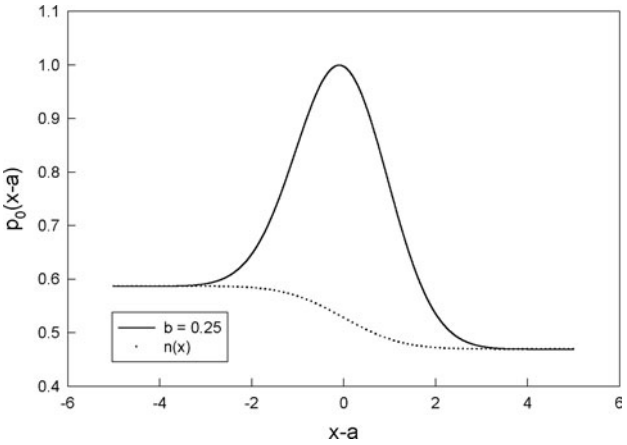

Figure 2. Part of a spectrum with a background enhanced at the low energy part of the peak. The amplitude of the low energy component is $b = 0.25$, the peak-to-background ratio is 1.

A second peak is then added and its position optimized. The widths of the two peaks are taken fixed from the FWHM systematics.

Special marks are given to the structures with only one extreme, either a minimum or a maximum. Such structures are typically a consequence of sharp Compton edges, which we utilize in the analysis in special cases. The grades are used in later passes when the isotope composition of the sample from the extracted peaks is searched for. Usually, however, only peaks with best grades are used in the analysis.

A constant background, which we have assumed in the previous example, is only an approximation. On the low-energy side of real peaks the background is in fact higher due to poor charge collection and almost forward Compton scattering of the photons within the source. We write, as suggested in [9], the background as a sum of a constant and a low energy component in the form

$$\bar{n}(x) = n_0(1 + b(1 - E(x, a))) \tag{13}$$

The function

$$E(x, a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(x-a)^2}{2}\right) \mathrm{d}x$$

is the cumulative normal distribution and b is the amplitude of the low energy component. For the ratio $n_0/A = 1$ and $b = 0.25$ the spectrum p_0 is shown in figure 2. It can be shown that the optimal function $f_0(x-a)$ only weakly depends on the parameter b even if we assume a much exaggerated value of $b = 1$. This suggests that the convolution method, designed for the constant background, correctly eliminates also the background found in actual gamma spectra. This observation is an important one and it will be demonstrated that the method indeed works well when applied to the real spectra.

3. Assessment of uncertainties

The proposed method of the treatment of the spectrum enables one to evaluate statistical uncertainties and write down the results in a closed form. The number of counts $c(x)$ in a channel x is treated as a stochastic variable, which we write in the form

$$c(x) = \bar{n}(x) + w(x)$$

where $w(x)$ is a random variable with zero mean

$$\langle w(x) \rangle = 0$$

and variance

$$\langle w^2(x) \rangle = \bar{n}(x)$$

Moreover, there is no correlation between the fluctuations in different channels of the measured (original) spectrum, so that

$$\langle w(x) w(x') \rangle = 0$$

for all $x \neq x'$. The brackets $\langle\ \rangle$ denote the ensemble average of a quantity. These basic definitions and facts are used throughout this section.

The most important uncertainty which determines the sensitivity of the method to finding weak peaks in the original spectrum concerns the baseline fluctuation of the spectrum $I_1(y)$ given by equation (7). The fluctuation of this spectrum due to statistical fluctuations of the background is calculated from

$$\langle I_1^2(y) \rangle = \int_{-\infty}^{\infty} \langle w^2(x) \rangle \left(\frac{\mathrm{d}z}{\mathrm{d}x}\right)^2 \mathrm{d}x \tag{14}$$

which directly follows from equation (7) and the properties of the noise $w(x)$ listed above. Taking into account the fact that the average number of counts in the background $\bar{n}(x)$ varies slowly across the interval of the width $\sim\sigma_z$, we arrive at the result

$$\langle (I_1^2(y) \rangle = \bar{n}(y) \frac{\sqrt{\pi}}{2\sigma} \tag{15}$$

According to this expression an optimum low discrimination level λ is chosen when searching for the peaks

$$\lambda = \lambda_0 \sqrt{\langle (I_1^2(y) \rangle} \tag{16}$$

A reasonable value for λ_0 is, say, 3.

Important uncertainties are also the ones of the peak position a_i, its intensity A_i and the width σ_i. The peak function $L_i(x)$ must be treated as a stochastic quantity which can be written as

$$p_i(x) = A_i \exp\left(-\frac{(x-a_i)^2}{2\sigma_i^2}\right) + \bar{n}(x) + w(x) \tag{17}$$

The uncertainty of the width is derived from the doubly convoluted spectrum $I_2(y)$. The peaks appear there in the form of $D_i^{(2)}(y)$, which, for this purpose, we write in a different form as

$$D_i^{(2)}(y) = \int_{-\infty}^{\infty} p_i(x) F(x, y, \sigma_z)\, \mathrm{d}x \tag{18}$$

where the function $F(x, y, \sigma_z)$ is the convolution of the function $\mathrm{d}z/\mathrm{d}x$ with itself. In explicit form this is

$$F(x, y, \sigma_z) = \int_{-\infty}^{\infty} \frac{y-z}{\sigma_z^2} \exp\left(-\frac{(y-z)^2}{2\sigma_z^2}\right) \frac{z-x}{\sigma_z^2} \times \exp\left(-\frac{(z-x)^2}{2\sigma_z^2}\right) \mathrm{d}z \tag{19}$$

which can be integrated analytically with the result

$$F(x, y, \sigma_z) = \frac{\sqrt{\pi}}{4{\sigma_z}^3}((x-y)^2 - 2\sigma_z^2) \times \exp\left(-\frac{(x-y)^2}{4\sigma_z^2}\right) \quad (20)$$

The best estimate of σ_i, which we denote by σ_m, maximizes the expression given in equation (18). The corresponding equation is

$$\int_{-\infty}^{\infty} L_i(x) \frac{\partial F(x, a_m, \sigma_z)}{\partial \sigma_z} \, \mathrm{d}x = 0 \quad (21)$$

To examine the fluctuation $\delta\sigma$ of the actual solution σ_m around the σ_i, that is $\sigma_m = \sigma_i + \delta\sigma$ due to presence of stochastic $w(x)$, we expand the function $\partial F/\partial\sigma$ in powers of $\delta\sigma$ and retain only the linear term.

$$\left.\frac{\partial F(x, a_m, \sigma)}{\partial \sigma}\right|_{\sigma_m} = \left.\frac{\partial F(x, a_m, \sigma)}{\partial \sigma}\right|_{\sigma_i} + \left.\frac{\partial^2 F(x, a_m, \sigma)}{\partial \sigma^2}\right|_{\sigma_i} \delta\sigma \quad (22)$$

The resulting expression for constant background $\bar{n}(x) = \bar{n}$ is

$$\langle(\sigma_m - \sigma_i)^2\rangle = \sigma_{\sigma_m}^2 = \frac{\int_{-\infty}^{\infty}\langle w(x)^2\rangle(\partial F(x, y, \sigma_z)/\partial\sigma_z)^2 \, \mathrm{d}x}{\int_{-\infty}^{\infty} L_i(x)(\partial^2 F(x, y, \sigma_z)/\partial\sigma_z^2) \, \mathrm{d}x} \quad (23)$$

A straightforward calculation yields the following result for this uncertainty

$$\frac{\sigma_{\sigma_m}^2}{\sigma_i^2} = \frac{1.22 + 9.8(\bar{n}\sigma_m/P_m)}{P_m} \quad (24)$$

Here, P_m is our best estimate of the true value of the peak area P_i. The decimal numbers are given in the formula above in order to avoid rather involved fractions of integers and the powers of π, which we obtain executing the integrals.

A similar analysis can be performed to assess the uncertainty of the peak position, which yields

$$\langle(a_m - a_i)^2\rangle = \frac{\sigma_{a_m}^2}{\sigma_i^2} = \frac{1}{P_m}\left(1.54 + 7.1\frac{\bar{n}\sigma_m}{P_m}\right) \quad (25)$$

The calculations of the uncertainty of the peak area requires an assessment of the uncertainty of I_2^{sup} and of the covariance between this quantity and uncertainty of the deduced σ_m. We give here only the final result which is

$$\sigma_{P_m}^2 = \left(1.24 + 11.4\frac{\bar{n}\sigma_m}{P_m}\right) P_m \quad (26)$$

In the case of negligible background this result is close to the ideal uncertainty $\sigma_{P_i}^2 = P_i$.

4. Results

The method was first tested against a standard set of spectra[4], supplied by the IAEA in 1977. The spectra are designed to test a peak search routine's ability to detect single small peaks near the limit of detectability and to detect double peaks with various relative intensities and degrees of overlap. The philosophy of this test closely follows our approach to the spectrum evaluation. Emphasis is on the possibility of the method to detect small peaks on high background using a single parameter to control the detectability limit. The IAEA set of spectra are all 2048 channels long at about 0.5 keV per channel. The end channel thus corresponds to the energy of about 1000 keV.

The IAEA approach is based on the notion that a peak-search algorithm only needs to perform well in a relative sense when it comes to peak positions and areas. That is, the results need to be consistent when the reference spectrum and the spectrum of an unknown sample are both analysed by the same peak-search program, but no absolute accuracy of the peak-search algorithm is required. In this respect, the IAEA sets provide a high-quality reference spectrum, against which the peak-search program is to be calibrated and with the obtained parameters the rest of the spectra are to be analysed in relation to the reference spectrum. The reference spectrum consists of a combination of experimental spectra recorded with the highest possible precision and all the supplied spectra were obtained by subsequent computer manipulation of these experimental spectra, including randomization simulating the effects of counting statistics.

The first of the test spectra consists of small photo-peaks close to the limit of detectability. The aim of the test is to set the sensitivity of the program to a level where the largest number of true peaks and the lowest number of false peaks is reported. The spectrum consists of 22 peaks, just as the reference spectrum does, with some of the peak positions shifted in the test spectrum. The attenuation factors of the peaks with respect to the reference spectrum range from below 100 to about 2600. Optimum performance of our program was found when the low discrimination level of the routine λ was set to the value given in equation (16) and in such configurations 15 true peaks were identified and two false peaks, with the unidentified and false peaks being among the most attenuated ones.

The purpose of the second test, which consists of six identical spectra with different randomizations of the counting statistics, is to check the consistency of the performance of the program. The same 20 peaks as in the reference spectrum are contained in all the six spectra with the same attenuation, ranging from a factor of 5 to a factor of 340, and the same channel shifts. Our program was able to correctly identify all the 20 spectra in all the six cases with only five spurious peaks found altogether. The results for the peak areas, positions and width are summarized in tables 1–3, all of them with respect to the reference spectrum. We have also utilized the six spectra to assess the consistency of our evaluation of the uncertainties of the derived quantities by comparing their *a priori* and *a posteriori* uncertainties. It can be inferred from the tables that our algorithm performs well when it comes to the values of the peak areas, positions and widths and that it assesses the uncertainties of these variables correctly. The relative differences between the quantities for the reference peak and the average of the quantities obtained from the analysis of the six test peaks are small and are well accounted for by the uncertainties of the results.

[4] Information sheet G-1, gamma-ray spectrometry reference service, International atomic energy agency, Vienna, 1977.

Table 1. The results of the peak-search algorithm for the six spectra differing in the counting statistics, relative to the reference spectrum. By c_0 we denoted the position of peak n in the reference spectrum (in channels) and by $\bar{c}$ the average position in the six spectra. The average uncertainty of the deduced average position, as reported by the program, is denoted by $\sigma(\bar{c})$ and the uncertainty computed *a posteriori* from all the six results for the peak position by $\sigma'(\bar{c})$. Peaks 7 and 16 were omitted from the list because they are shifted in the test spectra with respect to the reference spectrum.

n	$(\bar{c}-c_0)/c_0$ (%)	$\bar{c}-c_0$	$\sigma'(\bar{c})$	$\sigma(\bar{c})$
1	0.001	0.001	0.015	0.045
2	−0.013	−0.027	0.077	0.073
3	−0.002	−0.005	0.027	0.025
4	−0.001	−0.003	0.122	0.072
5	−0.002	−0.010	0.019	0.027
6	−0.002	−0.010	0.095	0.094
8	0.000	0.004	0.068	0.103
9	0.001	0.012	0.011	0.025
10	0.002	0.015	0.101	0.102
11	0.004	0.036	0.025	0.036
12	0.004	0.045	0.107	0.137
13	0.001	0.011	0.020	0.040
14	0.004	0.046	0.163	0.145
15	0.001	0.013	0.023	0.031
17	0.000	−0.007	0.024	0.043
18	0.000	0.004	0.040	0.049
19	−0.007	−0.126	0.164	0.245
20	0.002	0.030	0.282	0.172
21	0.000	−0.002	0.030	0.054
22	−0.001	−0.020	0.059	0.051

Table 2. The results of the peak-search algorithm for the six spectra differing in the counting statistics, relative to the reference spectrum. By A_0 we denote the area of peak n in the reference spectrum and by $\bar{A}$ the average peak area in the six spectra. The average uncertainty of the deduced average peak area, as reported by the program, is denoted by $\sigma(\bar{A})$ and the uncertainty computed *a posteriori* from all the six results for the peak area by $\sigma'(\bar{A})$.

n	$(\bar{A}-A_0)/A_0$ (%)	$\bar{A}-A_0$	$\sigma'(\bar{A})$	$\sigma(\bar{A})$
1	0.303	70.820	473.201	477.333
2	2.060	100.298	479.522	400.667
3	−0.549	−104.090	249.589	467.167
4	0.049	2.701	472.598	408.000
5	−0.283	−51.801	418.543	446.500
6	3.897	193.956	466.696	423.333
7	1.475	326.152	295.353	478.500
8	0.953	44.178	433.591	414.500
9	−0.147	−37.980	411.517	505.500
10	0.832	44.173	268.885	432.667
11	4.067	750.307	583.813	501.333
12	1.393	15.015	124.587	101.833
13	0.260	15.738	102.284	164.833
14	0.169	1.575	71.923	91.333
15	0.470	47.861	74.269	191.667
16	1.957	19.690	79.041	92.000
17	−1.226	−86.929	95.324	174.167
18	−1.643	−87.632	91.146	152.333
19	−5.583	−35.561	30.898	82.500
20	−2.713	−24.451	96.667	87.667
21	0.389	22.692	112.728	164.333
22	0.915	50.987	74.401	147.167

Table 3. The results of the peak-search algorithm for the six spectra differing in the counting statistics, relative to the reference spectrum. By w_0 we denoted the FWHM of peak n in the reference spectrum (in channels) and by $\bar{w}$ the average FWHM in the six spectra. The average uncertainty of the deduced average FWHM, as reported by the program, is denoted by $\sigma(\bar{w})$ and the uncertainty computed *a posteriori* from all the six results for the peak position by $\sigma'(\bar{w})$.

n	$(\bar{w}-w_0)/w_0$ (%)	$\bar{w}-w_0$	$\sigma'(\bar{w})$	$\sigma(\bar{w})$
1	0.967	0.010	0.018	0.045
2	0.924	0.010	0.133	0.073
3	0.321	0.004	0.016	0.025
4	−0.290	−0.003	0.116	0.072
5	−0.013	0.000	0.024	0.027
6	5.746	0.074	0.071	0.094
7	0.762	0.010	0.028	0.026
8	−0.738	−0.010	0.114	0.103
9	0.035	0.001	0.024	0.025
10	0.769	0.011	0.097	0.102
11	2.903	0.045	0.047	0.036
12	4.429	0.070	0.127	0.137
13	−0.441	−0.007	0.039	0.040
14	−2.186	−0.038	0.123	0.145
15	−0.115	−0.002	0.009	0.031
16	2.403	0.042	0.092	0.142
17	0.228	0.004	0.033	0.043
18	−0.688	−0.013	0.020	0.049
19	1.635	0.033	0.180	0.245
20	−2.063	−0.041	0.248	0.172
21	0.501	0.010	0.050	0.054
22	0.829	0.017	0.051	0.051

The third test deals with the ability of the method to resolve the doublets. The test spectrum contains nine double peaks with components of various relative intensities and degrees of overlap. The peak shapes of the components are the same as in the original reference spectrum. The peak positions are, however, not the same as in the reference spectrum so the peak search method under test has to resolve the doublets without prior knowledge of the positions of the components. This difficult task has been successfully performed for the peaks where the components were three channels apart if the relative intensities were less or equal to 1 : 3. The FWHM of the components was 3.6 channels. For separation of the components of six channels all the doublets in the spectrum were resolved. The FWHM of the components was 4.6 channels and the maximum ratio between the components 1 : 10. This complies with the ANSI standard. The ability of the method to resolve merged doublets has also been tested.

The second set of test spectra considered was the one made available by the IAEA peak search intercomparison exercise carried out in 1998[3]. These spectra span 8192 channels at about 0.4 keV per channel, yielding the highest energy in the spectra to be about 3200 keV. For the description of scope of the exercise and the different types of spectra, statistical criteria used for the performance evaluation and the general philosophy behind the exercise the reader is kindly referred to the report itself, since an appropriate account of it would be too lengthy. The results of the fully automatic mode for our analysis are given in table 4. As mentioned, in our implementation doublets are detected by comparing the deduced width of the considered peak with the width obtained from the systematic

Table 4. The results of the peak-search algorithm for the IAEA 1998 test spectra. Structure of the table is the same as in the publication[2]. Capital N denotes the number of events reported by the program. The χ^2 values refer to the differences between reported and 'true' values for the peak areas, pondered with reported uncertainties or reference uncertainties where the reported ones were not available (missed peaks). The normalization is such that in an ideal case the χ^2 value should be equal to 1.

Test type	Criterion	Straight	1n1	3n1	1n3
High peak	N	48	86	78	80
	X1	0.8	2.7	3.8	2.2
Small on high	N	18	24	22	22
	X1	2.1	1.6	4.1	1.1
Small on low	N	12	11	13	14
	X1	0.9	1.0	1.3	2.6
All matches	N	78	121	113	116
	X1	1.1	2.4	3.6	2.0
Position	X	1.1	5.0	7.0	4.7
Misses	N	90	206	208	207
	X	6.2	15.8	29.7	19.3
False hits	N	2	9	7	12
	X	5.1	40.2	3.9	1.3
Total	N	170	336	328	335
	X	3.8	12.4	21.4	12.7

deduced from the reference spectrum. In the spectra named ADD10N1 and ADD1N100 the uncertainty of the extracted widths obscures the broadening due to satellite peaks. In such cases intervention from the user is required or the satellite-peak positions should be known and we therefore do not give the results for the last mentioned spectra.

A glance at table 4 otherwise tells us that the present implementation of our approach is highly competitive with other packages tested in the 1998 IAEA intercomparison. Very good results for small peaks detected on either high or low background were obtained. Only one other package (HypermetPC[5]) in the exercise performed comparably for the singlet spectrum. Slightly smaller number of hits for small peaks on low background compared with the small hits on high background is in our case caused by fixing the lowest background to 1 count per channel. The sensitivity was in our case fixed to a relatively high value ($\lambda_0 = 3.2$) to suppress false hits at the level close to zero.

The results for doublet (ADD*) spectra indicate high resolving power of our method which proves that our approach to detect unresolved doublets by width comparison is appropriate in the automatic mode of operation. The criterion for promoting a peak as a doublet is chosen to be on the level of three standard deviations of the extracted width, meaning that the absolute difference between actual width σ_i and systematic singlet one should be greater than three standard deviations of the width uncertainty σ_{σ_i}. Relatively high value of 40.2 for χ^2 of 'false hits' for the ADD1N1 spectrum stems from an attempt to resolve very weak doublet in the presence of a very strong one in one specific case with the ratio of areas of 1–270. Nevertheless, the process of resolving merged multiplets is generally stable and represents an alternative method to those mentioned in [10, 11].

The method was additionally thoroughly tested on many spectra measured with actual detector sets and with spectra prepared with the use of Monte Carlo method together with the GEANT system [12]. The peak-search method proposed in this paper was tested and compared with the results obtained with other commercially available routines. The measured spectra contained 8192 channels that spanned the energy range from a few kilovolts to about 2600 keV and the artificial spectra produced with the Monte Carlo method spanned 8192 channels, corresponding to the energy range of about 3000 keV. The artificial Monte Carlo spectra have in general an advantage over the spectra from the actual detectors in that one knows exactly the right answer regarding the position of the peaks and their intensity. The spectra are nevertheless as complex as the real ones and the width and tailing can be accurately kept under control. In all cases we found adequate performance of the method regarding the determination of the peak positions, widths and the absolute areas.

[5] HypermetPC 4.0, User's Manual, Nuclear Physics Department, Institute of Isotopes KFKI, Budapest, 1995.

5. Conclusion

We present an efficient approach to the peak-search problem for complex spectra from high resolution Ge spectrometers. The method is based on a linear transformation of the spectrum which eliminates the background. Individual peak shape parameters are obtained one at a time without any requirement for non-linear fitting in case of single peaks. Analytical expressions are given for uncertainties of all of the deduced peak parameters. Doublets are detected through deviation of the extracted widths from the systematic dependence assumed for singlets. Their analysis does require non-linear fitting, but on a convoluted spectrum instead of on the original one, which eliminates the need for the modelling of the background. The method is not difficult to implement and appears to be robust and reliable.

The assumed Gaussian shape of the peak allows complete control of the uncertainties of the vital parameters of a peak. Different peak shape models, as proved before[3], yield the same peak areas when confronted with high quality spectra. Our experience with the analysis where absolute areas are important, as they enter non-linear equations, proves the above conclusion.

References

[1] Debertin K and Helmer R 1988 *Gamma- and X-Ray Spectrometry with Semiconductor Detectors* (Amsterdam: North-Holland)
[2] Sasamoto N, Koyama K and Tanaka S-I 1975 *Nucl. Instrum. Methods* **125** 507
[3] Brouwer G and Jansen J A J 1973 *Anal. Chem.* **45** 2239
[4] Mariscotti M A 1967 *Nucl. Instrum. Methods* **50** 309
[5] Routti J T and Prussin S G 1969 *Nucl. Instrum. Methods* **72** 124
[6] Black W W 1969 *Nucl. Instrum. Methods* **71** 317
[7] Robertson A, Prestwich W V and Keneth T J 1972 *Nucl. Instrum. Methods* **100** 317
[8] Hnatowicz V 1976 *Nucl. Instrum. Methods* **133** 137
[9] Winn W G 2000 *Nucl. Instrum. Methods* A **450** 430
[10] Blaauw M, Keyser R M and Fazekas B 1999 *Nucl. Instrum. Methods* A **432** 77
[11] Keyser R M 1990 *Nucl. Instrum. Methods* A **286** 403
[12] Brun R, Bruyant F, Maire M, McPherson A C and Zanarini P 1987 GEANT3, CERN Data Handling Division, Geneva