DECONVOLUTION OF THE ENERGY SPECTRA

By A. Barcz, B. Cichocki*, A. Turos and L. Wieluński

Institute of Nuclear Research, Nuclear Reactions Department, Warsaw**

(Received July 14, 1977)

Deconvolution of distributions affected by statistical errors is one of the so-called "incorrectly posed problems" in mathematical physics and thus requires special predictions for the resulting functions. In this paper a modified method of statistical regularization has been described which serves as a tool for experimental spectra processing. It enables the deconvolution of the detector and electronic noise to obtain a final resolution significantly improved. Examples of the application of the technique for the spectra characteristic to nuclear physics and nuclear microanalysis have been presented.

1. Introduction

The problem of spectra distorsions caused by the finite energy resolution of the experimental system is common in all fields of experimental physics dealing with measurements and interpretations of the energy spectra. The problem is furthermore complicated by the presence of statistical fluctuations involved in any real experiment.

In this paper we describe the method of deconvolution i.e. the mathematical procedure which permits one to determine the shape of the ideal energy spectrum provided the resolution function of the experimental system is well-known. The method referred to as statistical regularization is based on the works of Phillips [1] and Turchin et al. [2, 3]. It was applied to the analysis of the typical energy spectra obtained in the course of the investigations of solid state structures using nuclear reactions and elastic backscattering [4, 5]. Nevertheless it can be applied, without significant modifications, to the analysis of other kinds of experiments, particularly in low energy nuclear physics.

2. Statement of the problem

The ideal energy spectrum produced by the interaction of the incident particles with a target is given by

$$Y(E_f)dE_f = I\Omega\sigma_\theta[E(x)]C(x)dx, \tag{1}$$

* Permanent address: Instytut Fizyki Teoretycznej UW, Hoża 69, 00-681 Warszawa, Poland.
** Address: Instytut Badań Jądrowych, Hoża 69, 00-681 Warszawa, Poland.

where: I — number of incident particles, Ω — solid angle of detection, $\sigma_\theta(E)$ — differential cross section at an angle of detection θ, $C(x)$ — concentration profile i.e. number of a given kind of atoms per cm^3 at the depth x.

The shape of the $Y(E_f)$ spectrum corresponds to a somewhat fictitious case of an experimental system having δ-function resolution. No statistical fluctuations are taken into account either. The resolution function of a real system is composed of several contributing factors including: detector resolution, incident beam energy spread, energy straggling

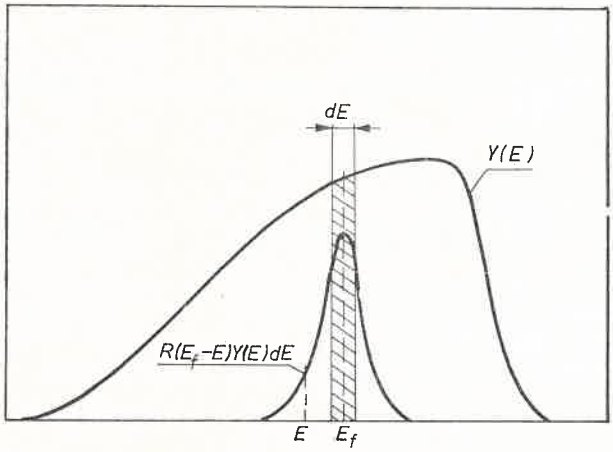


Fig. 1. Convolution of the ideal spectrum $Y(E)$ with the resolution function $R(E_f-E)$

of particles, interaction kinematics, geometry of the experiment etc. [6]. The real spectrum $F(E_f)$ is formed as shown in Fig. 1. Each monoenergetic component of the ideal spectrum $Y(E)dE$ is distributed after detection according to the shape of the resolution function $R(E)$. As a consequence of this fact the number of particles at a given energy E_f is equal to $R(E_f-E)Y(E_f)dE$. The other monoenergetic components of the ideal spectrum will have similar distributions hence the intensity of the spectrum at the energy E_f is given by

$$F(E_f) = \int_{-\infty}^{\infty} R(E_f-E)Y(E)dE \tag{2}$$

i.e. by the convolution of the ideal spectrum and the resolution function. As the energy spectra are usually obtained by the use of a multichannel pulse height analyzer which accumulates information in the finite energy intervals (channels) the integral in Eq. (2) should be transformed into a set of linear algebraic equations

$$\sum_{i=1}^{n} R_{ji}y_i = f_j \quad (j = 1, 2 \ldots n) \tag{3}$$

n being the number of channels. This can be written for simplicity in the operator form

$$\hat{R}\bar{y} = \bar{f}. \tag{4}$$

The first approach to the problem of deconvolution is to solve Eq. (4) by the calculation of inverse matrix $\hat{R}^{-1}$. However, as was shown in Refs. [1, 7], the direct solution

$$\bar{y} = \hat{R}^{-1}\bar{f} \tag{5}$$

is either impossible or does not give a unique result if the real spectrum contains statistical fluctuations. In most cases direct solutions have large oscillating components leading to unphysical results. In fact an additional term, which has been overlooked in Eq. (4), should be taken into account, namely the vector $\bar{\varrho}$ describing the counting statistics in each of the channels

$$\sum_{i=1}^{n} R_{ji}y_i + \varrho_j = f_j. \tag{6}$$

The only way of solving the problem is to introduce some a priori constraints to the unknown function $Y(E)$ or, in the algebraized version, on the vector $\bar{y}$. These constraints should represent the knowledge or anticipation the experimentalist has had before he started the experiment. It may be based on the results of a previous experiment or even on the basic physical features of the investigated phenomenon. We would like to stress that without introducing some information a priori the solution of the problem is impossible.

Several authors approached the problem of deconvolution using the method of the Fourier transform [8]. The a priori information is introduced as an arbitrarily chosen high-frequency cut-off which in most cases has no physical meaning. Another way to introduce a priori restrictions is the parametrization of the unknown function. In that case the method of least squares is usually applied.

In this paper we present the method of statistical regularization based on the theory extensively described in [2]. The principle underlying the method consists in the probabilistic way of introducing a priori information. This seems to be the most appropriate way of data processing, for which the introduction of probabilistic concepts is evident because of the random nature of the errors involved.

3. Statistical regularization

Referring to Eq. (6) the deconvolution problem is now as follows: the unknown quantities y_i represent the state of nature, and some random process gives us the quantities f_i. How are we to find the set of y_i given f_i? The general answer to this question is given by the decision theory [9] which states that, on the average, the best estimate on the statistical ensamble is the Bayesian estimate. To calculate the Bayesian estimate of the vector $\bar{y}$ one should find the probability distribution $P(\bar{y}|\bar{f})$ i.e. the probability of $\bar{y}$ given $\bar{f}$, and then calculate the statistical average of $\bar{y}$,

$$\langle \bar{y} \rangle = \int \bar{y} P(\bar{y}|\bar{f}) d\bar{y}. \tag{7}$$

The a posteriori probability $P(\bar{y}|\bar{f})$ is given by the Bayes formula [7]

$$P(\bar{y}|\bar{f}) = \frac{P(\bar{y})P(\bar{f}|\bar{y})}{\int P(\bar{y})P(\bar{f}|\bar{y})}, \tag{8}$$

where: $P(\bar{f}|\bar{y})$ — is the conditional probability of $\bar{f}$ given $\bar{y}$, $P(\bar{y})$ — the probability distribution, which represents the a priori knowledge or degree of belief in the different hypotheses.

The $P(\bar{f}|\bar{y})$ probability distribution can be deduced from the nature of the statistical fluctuations involved in the measurements. If these fluctuations do not appear the ideal spectrum could be easily determined from the experimental one. Therefore the probability distribution of obtaining the real spectrum $\bar{f}$, assuming one knows exactly the ideal spectrum $\bar{y}$ and resolution function R, is given by the probability distribution of vector $\bar{\varrho}$ (cf. Eq. (6))

$$P(\bar{f}|\bar{y}) = P(\bar{f}-\hat{R}\bar{y}) = P(\bar{\varrho}). \tag{9}$$

Each component of vector $\bar{\varrho}$ is therefore given by

$$\varrho_j = (\bar{f}-\hat{R}\bar{y})_j = f_j - \sum_{i=1}^{n} R_{ji}y_i. \tag{10}$$

Assuming that fluctuations in the number of counts in a particular channel are normally distributed about the mean value $(\hat{R}\bar{y})_j$ with dispersion s_j one has

$$P(\varrho_j) = (2\pi s_j^2)^{-1/2} \exp\left(-\frac{\varrho_j^2}{2s_j^2}\right). \tag{11}$$

Since statistical fluctuations in different channels are independent, the probability distribution of vector $\bar{\varrho}$ is given by the product of distributions of all components

$$P(\bar{\varrho}) = \prod_{j=1}^{n} (2\pi s_j^2)^{-1/2} \exp\left(-\frac{\varrho_j^2}{2s_j^2}\right)$$

$$= \prod_{j=1}^{n} (2\pi s_j^2)^{-1/2} \exp\left[-\frac{1}{2s_j^2}\left(f_j - \sum_{i=1}^{n} R_{ji}y_i\right)^2\right]. \tag{12}$$

This expression can be considerably simplified by introducing the following new matrices

$$\hat{W} \text{ where } \left(W_{ij} = \frac{1}{s_j^2}\,\delta_{ij}\right), \qquad \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and

$$\hat{B} = \hat{R}^{+}\hat{W}\hat{R}, \; \hat{R}^{+} \text{ being the transposed matrix.}$$

As is shown in Appendix A, equation (12) can be written in the form

$$P(\bar{f}|\bar{y}) = c_1 \exp\{-\tfrac{1}{2}(\bar{y}, B\bar{y}) + (\bar{y}, \hat{R}^{+}\hat{W}\bar{f}) - \tfrac{1}{2}(\bar{f}, \hat{W}\bar{f})\}, \tag{13}$$

where $(\bar{y}, \hat{B}\bar{y}) = \sum_{i,j} y_i B_{ij} y_j$ being a scalar product. Now we have reached the most critical point of calculating the Bayesian estimate: the introduction of a priori information. Obviously the quantity of such information should be as low as possible. If $P(\bar{y})$ is much more informative than $P(\bar{f}|\bar{y})$, there would be no sense in performing the experiments. For

$P(\bar{y}) = \text{const}$ we obtain an unregularized solution which is the same as the direct solution of Eq. (11).

To determine the $P(\bar{y})$ distribution one has to make only one assumption: $y(x)$ is a smooth function. After Turchin et al. [2] we take as a measure of the degree of smoothness of a function the norm of the second derivative in the form $\int \left[\frac{d^2y(x)}{dx^2}\right]^2 dx$. After some algebra is performed the integral should be substituted by a scalar product $(\bar{y}, \hat{\Omega}\bar{y})$, where Ω is the operator of the second derivative of the vector $\bar{y}$. Properties of the matrix $\hat{\Omega}$ are discussed in Appendix B.

In order to take into account all possible values of $\bar{y}$ given by the a priori distribution $P(\bar{y})$ one should calculate the average of $(\bar{y}, \hat{\Omega}\bar{y})$ on this distribution

$$\int (\bar{y}, \hat{\Omega}\bar{y}) P(\bar{y}) d\bar{y} = \omega. \tag{14}$$

Let us assume for a while that the number ω is known. However, Eq. (14) is not sufficient to define $P(\bar{y})$ in a unique way. In order to introduce as little arbitrariness as possible we choose from all distributions $P(\bar{y})$ satisfying Eq. (14) the one that contains the minimum information about $\bar{y}$. According to the information theory [10] the quantitative measure of information is given by the functional

$$I[P(\bar{y})] = \int P(\bar{y}) \ln P(\bar{y}) d\bar{y}. \tag{15}$$

As can be seen in Appendix C the final form of the $P(\bar{y})$ distribution can be obtained by calculating the conditional minimum of the functional (Eq. (15)) for a given value of ω

$$P_\alpha(\bar{y}) = c_2 \exp\left[-\frac{\alpha}{2}(\bar{y}, \hat{\Omega}\bar{y})\right], \tag{16}$$

where $\alpha = \frac{n}{\omega}$ is the smoothness parameter. We can now find from Bayes theorem the a posteriori probability distribution $P(\bar{y}|\bar{f})$ by inserting Eqs. (13) and (16) into Eq. (8)

$$P(\bar{y}|\bar{f}) = c_3 \exp\{-\tfrac{1}{2}(\bar{y}, [\hat{B}+\alpha\hat{\Omega}]\bar{y}) + (\bar{y}, \hat{R}^+\hat{W}\bar{f}) - \tfrac{1}{2}(\bar{f}, \hat{W}\bar{f})\}. \tag{17}$$

As is shown in Appendix D, the $P(\bar{y}|\bar{f})$ distribution has a gaussian form. Hence the regularized solution $\langle\bar{y}\rangle_\alpha$ being the average of the vector $\bar{y}$ for a given value of α-parameter can be obtained by finding the value of $\bar{y}$ corresponding to the zero value of the exponent in Eq. (17)

$$\langle\bar{y}\rangle_\alpha = (\hat{B}+\alpha\hat{\Omega})^{-1}\hat{R}^+\hat{W}\bar{f}. \tag{18}$$

The covariance matrix of the distribution is given by $(\hat{B}+\alpha\hat{\Omega})^{-1}$. The diagonal elements of the covariance matrix are mean square errors of the solution

$$\sigma_i^2 = [(\hat{B}+\alpha\hat{\Omega})^{-1}]_{ii}. \tag{19}$$

To complete the calculation pattern one needs a method to estimate the most probable value of the smoothness parameter α. This can be done on the basis of the experimental

data making once more use of Bayes theorem. The probability of getting different values of α for a given experimental spectrum $\bar{f}$ can be calculated from

$$P(\alpha|\bar{f}) = \frac{P(\alpha)P(\bar{f}|\alpha)}{\int P(\alpha)P(\bar{f}|\alpha)d\alpha}. \tag{20}$$

There is, however, no reason for preferring a certain value of α, therefore we assume $P(\alpha) = \text{const}$.

Then $P(\bar{f}|\alpha)$ can be written as

$$P(\bar{f}|\alpha) = \int P(\bar{f}|\bar{y})P(\bar{y}|\alpha)d\bar{y}, \tag{21}$$

where $P(\bar{f}|\bar{y})$ is given by Eq. (13). $P(\bar{y}|\alpha)$ can be deduced from the solution of Eq. (16) for different values of α. After the distribution $P(\alpha|\bar{f})$ is known, one can compute the mean value $\langle\bar{y}\rangle$

$$\langle\bar{y}\rangle = \int \langle\bar{y}\rangle_\alpha P(\alpha|\bar{f})d\alpha. \tag{22}$$

In most cases of physical interest the distribution $P(\alpha|\bar{f})$ is sufficiently narrow as to replace $\langle\alpha\rangle$ by $\alpha_{\max}$ without introducing significant error. This imposes replacing of $\langle\bar{y}\rangle_\alpha$ by $\bar{y}_{\alpha_{\max}}$ which is much easier to calculate. For a detailed calculation of $\alpha_{\max}$ see Appendix E.

4. Application to the deconvolution of energy spectra with sharp steps

The method described above has been applied to a number of problems connected with experimental data correction [2]. However, it appeared useful only for slowly varying functions. If rapid changes are present in the spectrum the procedure yields negative data points or oscillations which are physically meaningless solutions. Thus when dealing with the spectra composed of both smooth and sharp parts, the smoothness postulate cannot be fulfilled completely for the whole function. In order to resolve the problem let us redefine the second derivative operator $\hat{\Omega}$:

$$\hat{\Omega}[y(x)] = \int\limits_{X-\bigcup_{j=1}^{l}[x_j,x_j+\Delta_j]} \left(\frac{d^2y}{dx^2}\right)^2 dx + \sum_{j=1}^{l} \beta_j \int\limits_{[x_j,x_j+\Delta_j]} \left(\frac{d^2y}{dx^2}\right)^2 dx, \tag{23}$$

where: l — number of discontinuity points, $[x_j, x_j+\Delta_j]$ — interval with rapid change. The first component corresponds to the smooth part of the spectrum whereas the second one covers the region where rapid changes are suspected to appear. Introducing β_j parameters means that in the regions in question the second derivative is allowed to be $1/\beta_j$ times greater than in the smooth regions. The application of the matrix $\hat{\Omega}$ defined by Eq. (23) has led to satisfactory results for correcting parameters β_j of the order of 0.1—0.001. Figure 2 shows the comparison between the spectra subjected to the deconvolution procedure according to (b) unmodified and (c) modified method. Spectrum (a) presents the raw experimental data.

The parameters $\beta_j, j = 1, \ldots l$ can be found from the experimental spectrum, in a similar way as α, applying once more the Bayes formula. If the values of α and $\beta_j, j \neq i$ are fixed one obtains:

$$P(\beta_i|\bar{f}) = \frac{P(\beta_i)P(\bar{f}|\beta_i)}{\int P(\beta_i)P(\bar{f}|\beta_i)d\beta_i}. \tag{24}$$

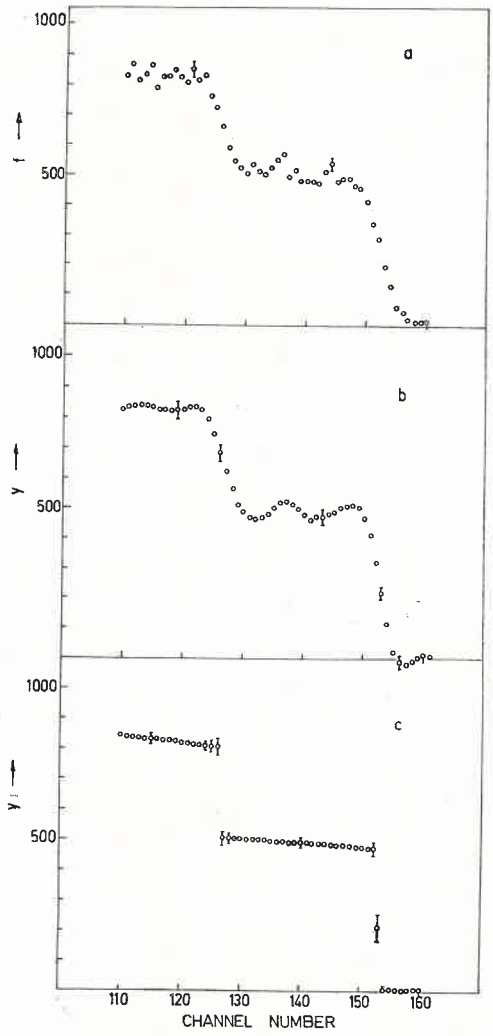


Fig. 2. Comparison of different methods of deconvolution: a) — experimental spectrum, b) — according to Turchin et al. [2], c) — present work

Since there is no reason to prefer any particular β_i value we assume $P(\beta_i) = \text{const}$, which leads to

$$P(\beta_i|\bar{f}) = c_5 \int P(\bar{f}|\bar{y})P_{\beta_i}(\bar{y})d\bar{y} \tag{25}$$

where $P_{\beta_i}(\bar{y})$ is the distribution given by Eq. (16), for $\hat{\Omega} = \hat{\Omega}(\beta_i)$. This expression can be transformed into the final form

$$P(\beta_i|\bar{f}) = c_5 \exp\{\tfrac{1}{2}(\hat{R}^+\hat{W}\bar{f}, [\hat{B}+\alpha\hat{\Omega}]^{-1}\hat{R}^+\hat{W}\bar{f})\}[\det(\hat{B}+\alpha\hat{\Omega})]^{-1/2}. \tag{26}$$

Finding the distribution of the conditional probability density $P(\beta_i|\bar{f})$ requires one to calculate, for each value of β_i, the determinant of n-dimensional matrix, where n is the dimension of vector $\bar{f}$. This is the way to determine the parameter $\beta_{i_{max}}$.

After putting the $\beta_{i_{max}}$ value to the operator $\hat{\Omega}$, one finds the distribution $P(\beta_{i+1}|\bar{f})$. This has to be continued for all indices "i". Note that each $P(\beta_i|\bar{f})$ distribution depends on the values of the remaining β_j parameters, thus the procedure is to be repeated until the required convergence is achieved. Another difficulty arises from the dependence of the distribution $P(\beta_i|\bar{f})$ (Eq. (26)) on the smoothness parameter α. Therefore one has to include in the iteration process the determination of α_{max}, for any set of $\beta_{i_{max}}$. This yields finally an estimate of the maximum of the many-dimensional function $P(\beta_1, \ldots \beta_l, \alpha|\bar{f})$ depending on $l+1$ variables.

To test the method, the following step-shaped function has been taken:

$$y^s(x) = \begin{cases} 1000 & 1 \leqslant x_i \leqslant 24 \\ 2000 & 25 \leqslant x_i \leqslant 50 \end{cases} \tag{27}$$

where x_i is a channel number. The convolution of this function with the gaussian distribution R (FWHM = 5) gives:

$$f^s(x) = \int R(x-x')y^s(x')dx'. \tag{28}$$

Then we have to add the statistic errors to each bin so as to obtain the "experimental" spectrum

$$f(x_i) = f^s(x_i)+\delta_i. \tag{29}$$

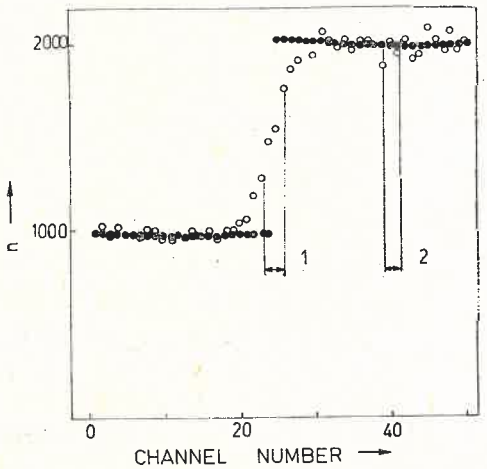


Fig. 3. Test spectrum – ○ and that deconvoluted with $\alpha = 0.16$ and $1/\beta = 390$ – ●

Such an artificial spectrum is shown in Fig. 3. The result of deconvolution differs from the original function $y^s(x)$ by less than 1% which is well below the statistical errors. Table I gives the values β_{max} and α_{max} for the successive steps of determination of the distribution $P(\beta, \alpha|\bar{f})$.

TABLE I

$1/\beta$	α
200	0.06
240	0.077
270	0.1
330	0.13
390	0.16
390	

Figure 4 shows two curves of the probability density $P(1/\beta)$. Since the distribution corresponding to region *1* (in Fig. 4) is a broad one, the integration of $\bar{y}$ over the whole distribution ought to be made to find $\langle\bar{y}\rangle$. Fortunately, it can be proved that the β-parameter only

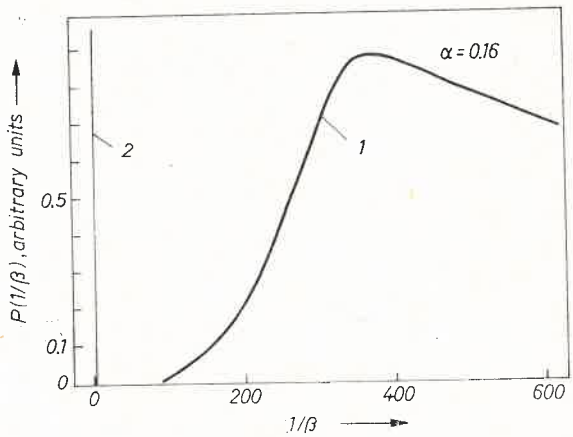


Fig. 4. Probability distribution of the β-parameter for: *1* — sharp and *2* — flat part of test spectrum

weakly influences the vector $\bar{y}$ (in terms of deviation from $\bar{y}^s$), which enables one to replace the integral $\langle\bar{y}\rangle$ with $\bar{y}_{\alpha_{max},\beta_{max}}$. This simplifies calculations significantly. For the region *2* the distribution has a strong, narrow maximum about $\beta = 1$, as it would have to be expected for any region which covers the "smooth" part of the spectrum.

5. Results and discussion

The method has been applied to the deconvolution of spectra characteristic to the nuclear microanalysis [4, 5] and therefore containing sharp steps due to the rapid changes of the concentration distribution of a given element in the investigated sample.

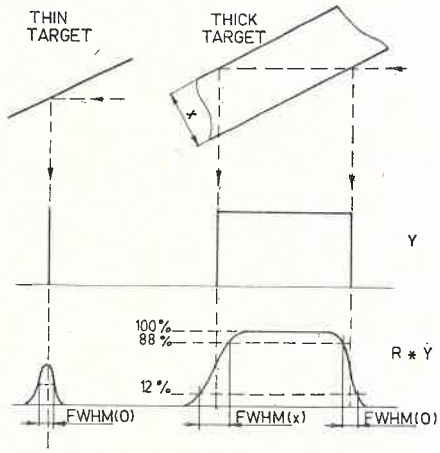


Fig. 5. Determination of resolution function from the experimental spectra

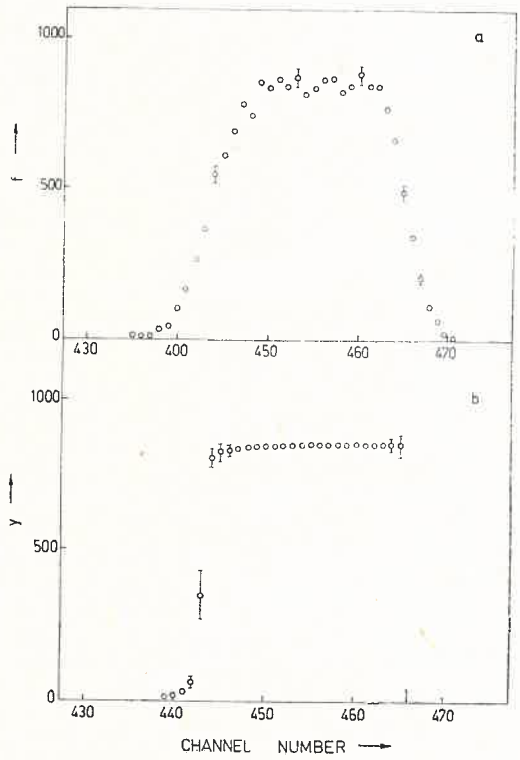


Fig. 6. Energy spectra of α-particles emitted in the $^{16}O(d, \alpha)^{14}N$ reaction from 900 keV deuteron bombardment of a 2100 Å thick SiO_2 surface layer, a) – experimental, b) – after deconvolution

Performing the procedure of deconvolution we have to assume that the resolution distribution R is measurable. In practice this can be done by using a radioisotopic source of monoenergetic α-particles which are registered directly by a detector. Another way to determine the resolution function is illustrated schematically in Fig. 5. Applying a thin and/or thick target and analyzing the spectra of backscattered or nuclear reaction-produced particles we may obtain the R distribution or all its required momenta. Fig. 6a shows the energy spectrum of α-particles from $^{16}O(d, \alpha)^{14}N$ reaction induced by 900 keV deuteron

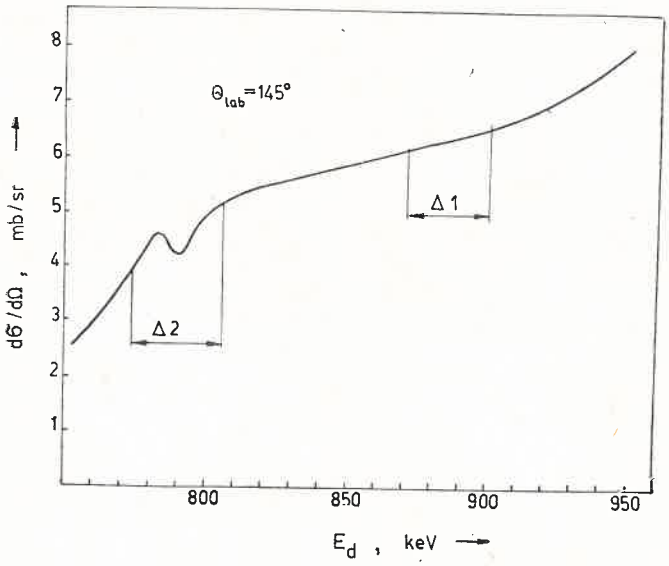
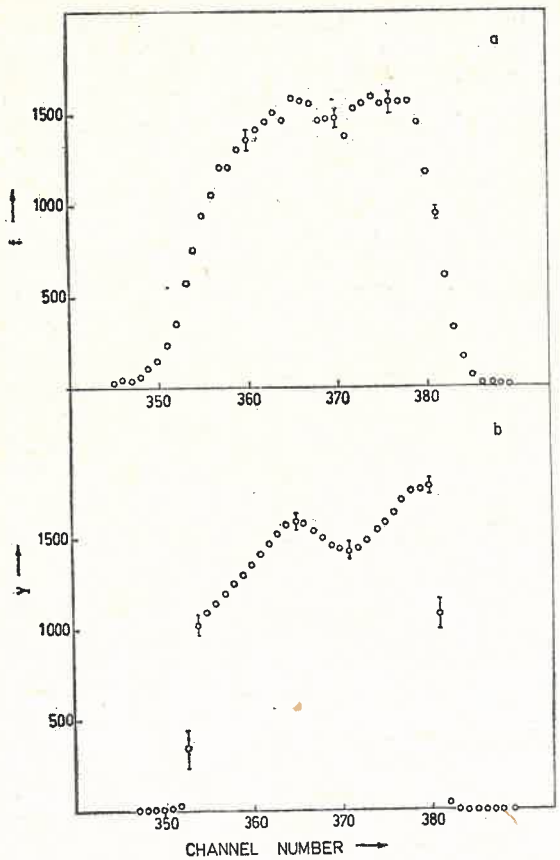


Fig. 7. Excitation function of the $^{16}O(d, \alpha)^{14}N$ reaction

bombardment of 2100 Å thick SiO_2 surface layer. The incident deuterons energy has been chosen as 900 keV so as to cover the flat region $\Delta 1$ (Fig. 7) in the excitation curve. Since the oxygen concentration has been found to be uniform, the slope of the spectrum is due to the variation of the reaction cross-section with the incident deuteron energy.

In order to check the sensitivity of the method, a more spectacular example of the energy spectrum has been taken. The 6000 Å thick, uniform SiO_2 surface layer has been bombarded with 805 keV deuterons. The α-particles spectrum of that case is shown in Fig. 8a. The spectrum after deconvolution (Fig. 8b) follows exactly the shape of the excitation function in the energy range $\Delta 2$ (Fig. 7) with a characteristic small resonance at 780 keV. Another example of the application of the deconvolution method is shown in Fig. 9. It presents the backscattering energy spectrum of 1.8 MeV α-particles from a 1950 Å thick silicon dioxide surface layer. The deconvoluted spectrum (Fig. 9b) yields the result which is free of experimental system distorsions and thus making further data processing much more convenient.

Figure 10 shows the comparison between two spectra of the same origin but differing with counting statistics. The purpose of this experiment was to test the influence of statistical errors on the resulting functions $\bar{y}$. After normalization both curves are equal to each



Fig, 8

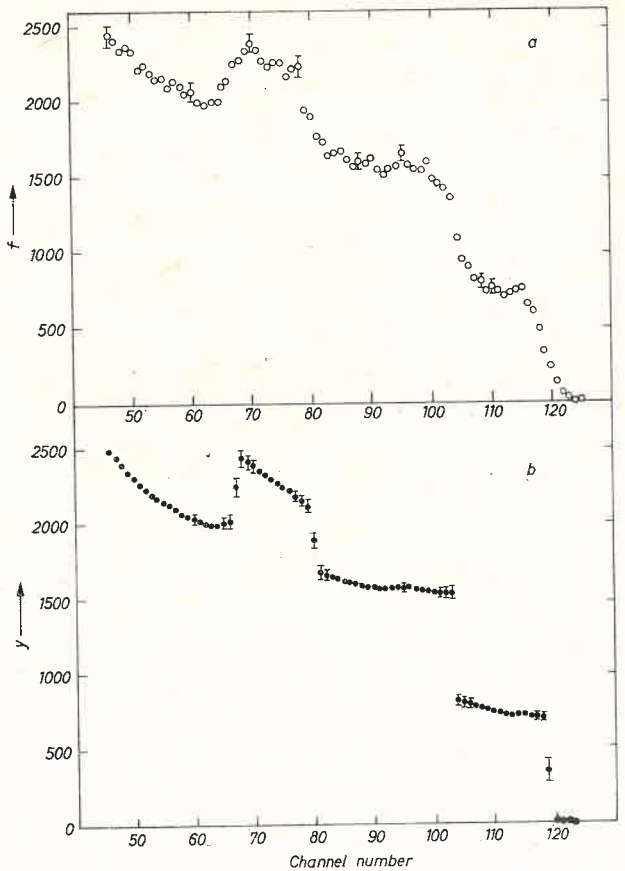


Fig. 9

Fig. 8. Energy spectra of α-particles emitted in the $^{16}O(d, \alpha)^{14}N$ reaction from 805 keV deuteron bombardment of a 6000 Å thick SiO_2 surface layer, a) – experimental, b) – after deconvolution

Fig. 9. Backscattering energy spectra from 1.8 MeV α-particle bombardment of a 4760 Å thick silicon oxide surface layer on silicon, a) – experimental, b) – after deconvolution

other to within an accuracy of 3%, which reveals weak dependence of statistical errors on the vector $\bar{y}$.

As is seen from the above results, the deconvolution method makes it possible to analyse more accurately the energy spectra. The statistical regularization has some advantages over other deconvolution methods:

— it refers to the statistical properties of detection processes involved,

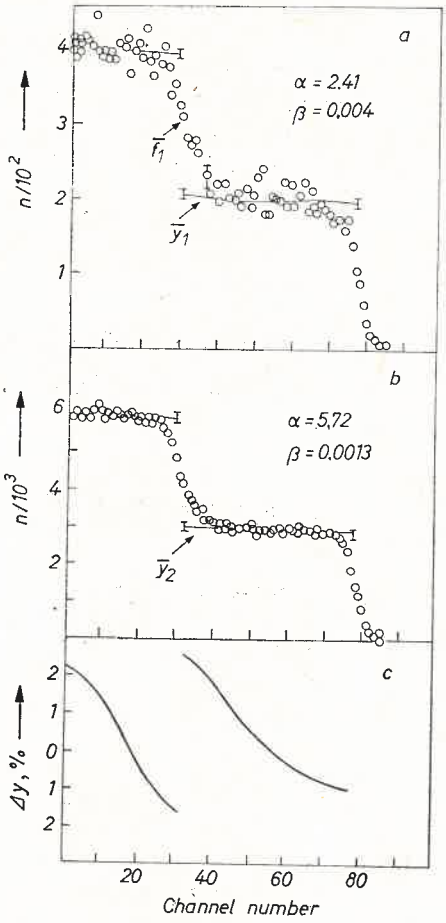


Fig. 10. Comparison of the backscattering spectra from Ag–Au alloy obtained with different counting statistics (a and b); c – the relative difference between the vectors $\bar{y}_i$; $\Delta y = (y_1 I_2 - y_2 I_1)/y_2 I_1$; I_1, I_2 being numbers of incident particles for spectrum a and b, respectively.

— applies the nontrivial a priori information, the amount of which is minimized,

— the errors in solution can be estimated,

— enables one, after adequate modification of matrix $\hat{\Omega}$, to process spectra of any shape.

The assumption of the smoothness of the solutions seems to be reasonable in most cases in which nuclear microanalysis is involved. It is difficult to determine exactly what

is the uncertainty in the energy scale of the deconvoluted spectra. This uncertainty depends on the accuracy of the resolution function estimation and on the statistical errors of the measured spectrum. In the limiting case i.e. if the resolution function is known with absolute accuracy and for the very small statistical errors involved in the measurement one may expect that the fine structure of one channel width could be determined.

Numerical calculations have been performed by the use of CDC CYBER-72 computer of Institute of Nuclear Research. The program has been divided into 10 subroutines to make easy operations like inversion of a matrix, calculation of a determinant etc., during the iteration process. The following input data are required: coordinates of the vector $\bar{f}$, initial values of α and β coefficients, parameters of R distribution, regions where $\beta_{i_{max}}$ values will be estimated. The program, for the dimension of the vector $\bar{f}$ equal 90, occupies the whole computer memory (64k). Time consumed depends on the number of β_i's and ranges between 10 min and 1 hour.

APPENDIX A

$P(\bar{f}\,|\bar{y})$ probability distribution

The exponent in Eq. (12) describing the conditional probability $P(\bar{f}\,|\bar{y})$ can be transformed as follows:

$$\sum_{j=1}^{n} \frac{1}{s_j^2}\left(f_j - \sum_{i=1}^{n} R_{ji}y_i\right)^2 = (\bar{f}-\hat{R}\bar{y},\ \hat{W}\bar{f}-\hat{W}\hat{R}\bar{y})$$

$$= (\bar{f},\ \hat{W}\bar{f})-(\bar{f},\ \hat{W}\hat{R}\bar{y})-(\hat{R}\bar{y},\ \hat{W}\bar{f})+(\hat{R}\bar{y},\ \hat{W}\hat{R}\bar{y})$$

$$= (\bar{f},\ \hat{W}\bar{f})-(\bar{y},\ \hat{R}^+\hat{W}\bar{f})-(\bar{y},\ \hat{R}^+\hat{W}\bar{f})+(\bar{y},\ \hat{R}^+\hat{W}\hat{R}\bar{y})$$

$$= (\bar{f},\ \hat{W}\bar{f})-2(\bar{y},\ \hat{R}^+\hat{W}\bar{f})+(\bar{y},\ \hat{B}\bar{y}).$$

Hence

$$P(\bar{f}\,|\bar{y}) = c_1 \exp\{-\tfrac{1}{2}(\bar{y},\ \hat{B}\bar{y})+(\bar{y},\ \hat{R}^+\hat{W}\bar{f})-\tfrac{1}{2}(\bar{f},\ \hat{W}\bar{f})\}. \tag{A1}$$

APPENDIX B

Properties of the matrix $\hat{\Omega}$

The matrix $\hat{\Omega}$ of the quadratic form $(\bar{y},\ \hat{\Omega}\bar{y})$ should assure the identity of the scalar product and the integral $\int\left(\frac{d^2y}{dx^2}\right)^2 dx$. The numerical value of the second derivative of the spectrum in the channel "j" is given by

$$\Delta_j'' = \frac{1}{h^2}\left[\left(\frac{y_j-y_{j-1}}{h}\right)-\left(\frac{y_{j+1}-y_{j-1}}{h}\right)\right] \tag{B1}$$

and

$$(\Delta_j'')^2 = \frac{1}{h^2}(y_{j-1}^2+2y_{j+1}y_{j-1}-4y_jy_{j-1}+4y_j^2-4y_jy_{j+1}+y_{j+1}^2) \tag{B2}$$

where h is the channel width which is usually assumed to be equal to 1. The scalar product $(\bar{y}, \hat{\Omega}\bar{y})$ should be therefore equal to

$$\sum_{j=1}^{n}(\Delta_j'')^2 = \sum_{j=1}^{n} a_{ij}y_iy_j, \tag{B3}$$

hence

$$\sum_{ij} a_{ij}y_iy_j = \sum_{ij} y_j\Omega_{ij}y_i,$$

$$\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}y_iy_j - \sum_{j=1}^{n} y_j(\sum_{i=1}^{n}\Omega_{ij}y_i) = 0, \quad \sum_{j=1}^{n} y_j(\sum_{i=1}^{n} a_{ij}y_i - \sum_{i=1}^{n}\Omega_{ij}y_i) = 0$$

which leads to the equality

$$a_{ij} = \Omega_{ij}. \tag{B4}$$

The expansion coefficients a_{ij} can be easily found from Eq. (B2)

$$\sum_{j=1}^{n}(\Delta_j'')^2 = \ldots + y_ky_{k+2} - 4y_ky_{k+1} + 6y_k^2 - 4y_{k+1}y_k + y_{k+2}y_k \ldots \tag{B5}$$

The operator $\hat{\Omega}$ is then a quadratic symmetrical matrix of elements given by Eq. (B5)

$$\hat{\Omega} = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & -4 & 6 & -4 & 1 & 0 & 0 & \cdot & \cdot \\ \cdot & 0 & 0 & 1 & -4 & 6 & -4 & 1 & 0 & 0 & \cdot \\ \cdot & \cdot & 0 & 0 & 1 & -4 & 6 & -4 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

APPENDIX C

A priori probability distribution

The a priori probability distribution $P(\bar{y})$ should fulfill the two main conditions

$$\int (\bar{y}, \hat{\Omega}\bar{y})P(\bar{y})d\bar{y} = \omega \tag{C1}$$

and

$$\int P(\bar{y})d\bar{y} = 1. \tag{C2}$$

This is unsufficient for finding unique form of $P(\bar{y})$. In order to introduce as little arbitrariness as possible one should choose the form of $P(\bar{y})$ which minimizes the functional

$$I[P(\bar{y})] = \int P(\bar{y}) \ln P(\bar{y})dy. \tag{C3}$$

The conditional minimum of the functional can be found by means of the Lagrange multiplicators method

$$\int P(\bar{y}) \ln P(\bar{y}) d\bar{y} - \lambda \int (\bar{y}, \hat{\Omega}\bar{y}) P(\bar{y}) d\bar{y} - \mu \int P(\bar{y}) d\bar{y}, \tag{C4}$$

where λ and μ are Lagrange multiplicators. Setting the first variation of (C4) equal to zero one obtains

$$\ln P(\bar{y}) - \lambda(\bar{y}, \hat{\Omega}\bar{y}) - (\mu - 1) = 0$$

hence

$$P(\bar{y}) = c_2 \exp [\lambda(\bar{y}, \hat{\Omega}\bar{y})]. \tag{C5}$$

Eq. (C5) can be also written in the form

$$P(\bar{y}) = c_2 \exp [-\tfrac{1}{2}(\bar{y}, -2\lambda\hat{\Omega}\bar{y})]. \tag{C6}$$

The constant c_2 can be calculated according to Eq. (C2). Since $P(\bar{y})$ is a normal distribution

$$c_2 = \frac{(-2\lambda)^{n/2} |\hat{\Omega}|^{1/2}}{(2\pi)^{n/2}}. \tag{C7}$$

From Eq. (C1) one obtains

$$\frac{(-2\lambda)^{n/2} |\hat{\Omega}|^{1/2}}{(2\pi)^{n/2}} \int (\bar{y}, \hat{\Omega}\bar{y}) \exp [\lambda(\bar{y}, \hat{\Omega}\bar{y})] d\bar{y}$$

$$= \frac{(-2\lambda)^{n/2} |\hat{\Omega}|^{1/2}}{(2\pi)^{n/2}} \frac{d}{d\lambda} \int \exp [\lambda(\bar{y}, \hat{\Omega}\bar{y})] d\bar{y} = \omega.$$

Taking into account Eqs (C2) and (C7)

$$\frac{(-2\lambda)^{n/2} |\hat{\Omega}|^{1/2}}{(2\pi)^{n/2}} \frac{d}{d\lambda} \left[\frac{(2\pi)^{n/2}}{|\hat{\Omega}|^{1/2} (-2\lambda)^{n/2}} \right] = (-2\lambda)^{n/2} (-2)^{-n/2} \lambda^{-n/2-1} = -\frac{n}{2\lambda}.$$

Hence

$$\lambda = -\tfrac{1}{2} \frac{n}{\omega}. \tag{C8}$$

Finally

$$P_\alpha(\bar{y}) = c_2 \exp \left[-\frac{\alpha}{2} (\bar{y}, \hat{\Omega}\bar{y}) \right]. \tag{C9}$$

APPENDIX D

A posteriori probability distribution

Once the distributions $P_\alpha(\bar{y})$ (Eq. (16)) and $P(\bar{f}\,|\,\bar{y})$ (Eq. (13)) are calculated the a posteriori probability distribution $P(\bar{y}|\bar{f})$ can be found from Bayes theorem (Eq. (8))

$$P(\bar{y}|\bar{f}) = c_3 P_\alpha(\bar{y}) P(\bar{f}\,|\,\bar{y})$$

$$= c_3 \exp \{-\tfrac{1}{2}(\bar{y}, [\hat{B} + \alpha\hat{\Omega}]\bar{y}) + (\bar{y}, \hat{R}^+ \hat{W}\bar{f}) - \tfrac{1}{2}(\bar{f}, \hat{W}\bar{f})\}. \tag{D1}$$

Since the last term in the figure bracket is constant Eq. (D1) can be written in the form

$$P(\bar{y}|\bar{f}) = c_3 \exp\{(\bar{y}, \hat{D}\bar{y}) - 2(\bar{y}, \hat{R}^+\hat{W}\bar{f})\}$$

$$= c_3 \exp\{([\bar{y} - \hat{D}^{-1}\hat{R}^+\hat{W}\bar{f}], \hat{D}[\bar{y} - \hat{D}^{-1}\hat{R}^+\hat{W}\bar{f}]) - (\hat{D}^{-1}\hat{R}^+\hat{W}\bar{f}, \hat{B}\hat{W}\bar{f})\}$$

where $\hat{D} = \hat{B} + \alpha\hat{\Omega}$. Since $(\hat{D}^{-1}\hat{R}^+\hat{W}\bar{f}, \hat{B}\hat{W}\bar{f}) = \text{const}$

$$P(\bar{y}|\bar{f}) = c_3 \exp\{([\bar{y} - \hat{D}^{-1}\hat{R}^+\hat{W}\bar{f}], \hat{B}[\bar{y} - \hat{D}^{-1}\hat{R}^+\hat{W}\bar{f}])\}. \tag{D2}$$

This is a multidimensional gaussian distribution (cf. [9]) with the covariance matrix $\hat{D}^{-1} = (\hat{B} + \alpha\hat{\Omega})^{-1}$.

APPENDIX E

Determination of $\alpha_{\max}$

According to Eqs (20) and (21) the a posteriori probability is equal to

$$P(\alpha|\bar{f}) = c_4 \alpha^{\frac{n-r}{2}} \int P(\bar{f}|\bar{y}) P_\alpha(\bar{y}) d\bar{y}, \tag{E1}$$

r is the degree of degeneration of matrix $\hat{\Omega}$. The extremum of this distribution can be found by means of the logarithmic derivative

$$\ln P(\alpha|\bar{f}) = \ln c_4 + \frac{n-r}{2} \ln \alpha + \ln \int P(\bar{f}|\bar{y}) \exp\left\{-\frac{\alpha}{2}(\bar{y}, \hat{\Omega}\bar{y})\right\} d\bar{y}$$

hence

$$\frac{\partial}{\partial\alpha} \ln P(\alpha|\bar{f}) = \frac{n-r}{2\alpha} - \frac{1}{2} \frac{\int P(\bar{f}|\bar{y}) \exp\left\{-\frac{\alpha}{2}(\bar{y}, \hat{\Omega}\bar{y})\right\} (\bar{y}, \hat{\Omega}\bar{y}) d\bar{y}}{\int P(\bar{f}|\bar{y}) \exp\left\{-\frac{\alpha}{2}(\bar{y}, \hat{\Omega}\bar{y})\right\} d\bar{y}}$$

$$= \frac{n-r}{2\alpha} - \frac{1}{2} \langle (\bar{y}, \hat{\Omega}\bar{y}) \rangle_\alpha. \tag{E2}$$

The interpretation of the latter expression is as follows: logarithmic derivative of the a posteriori probability of α-parameter equals to half the difference between the value $\frac{n-r}{\alpha}$ and the mean value of the functional $\hat{\Omega}$ over the distribution $P(\bar{y}|\bar{f})$, $\frac{n-r}{\alpha}$ being the mean value of $\hat{\Omega}$ over the a priori distribution. Finally setting $\frac{\partial}{\partial\alpha}(\ln P(\alpha|\bar{f})) = 0$, $\alpha_{\max}$ can be found from the following equation

$$\frac{n-r}{2\alpha_{\max}} - \frac{1}{2} \operatorname{Tr}\{\hat{\Omega}\hat{D}^{-1}\} - \frac{1}{2}(\langle\bar{y}\rangle_{\alpha_{\max}}, \hat{\Omega}\langle\bar{y}\rangle_{\alpha_{\max}}) = 0. \tag{E3}$$

Eq. (E3) is to be solved by the successive approaches method.

REFERENCES

[1] D. L. Phillips, *J. Associat. Comput. Machin.* **9**, 84 (1963).
[2] V. F. Turchin, V. P. Kozlov, M. S. Malkevich, *Sov. Phys. Usp.* **13**, 681 (1971).
[3] L. S. Turovtseva, V. F. Turchin, *Dokl. Akad. Nauk SSSR* **212**, 561 (1973).
[4] E. A. Wolicki, in *New Uses of Ion Accelerators*, J. F. Ziegler, ed., Plenum Press 1975.
[5] A. Barcz, A. Turos, L. Wieluński, *Int. J. Appl. Radiat. Isot.* **26**, 227 (1975).
[6] A. Turos, L. Wieluński, A. Barcz, *Nucl. Instrum. Methods* **111**, 605 (1973).
[7] S. G. Rautian, *Usp. Fiz. Nauk* **66**, 475 (1958).
[8] J. Z. Ziegler, J. E. E. Baglin, *J. Appl. Phys.* **42**, 2031 (1971).
[9] W. T. Eadie, D. Drijard, F. E. James, M. Roos, B. Sadoulet, *Statistical Methods in Experimental Physics*, North Holland 1971.
[10] L. Brillouin, *Science and Information Theory*, Academic Press 1962.