Automatic processing of linear spectra

P. M. Gopych
Khar'kov State University, Khar'kov, Ukraine

Fiz. Elem. Chastits At. Yadra **24**, 1596–1659 (November–December 1993)

The methods and problems of automatic (without human participation) computer processing of linear spectra in nuclear spectroscopy problems are reviewed. The optimization and standardization of these methods and their relation to problems in human psychology and artificial intelligence are discussed.

INTRODUCTION

Linear radiation spectra are one of the main types of primary experimental data in physics, the other natural sciences, and in applied research. The enormous amount of such data and the great complications in analyzing them quantitatively make it essential to use computers for processing them. By now there are many (dozens or even hundreds) different computer programs which have been written for solving this problem. These programs differ in their approaches to the processing, their approximation methods, their method of estimating the errors in the results, and so on. There is an extensive literature devoted to the description of the methods of obtaining linear spectra and processing them by computer; this literature includes reviews, monographs (for example, Refs. 1–5), and even textbooks (for example, Ref. 6).

Computer processing of linear spectra obtained in nuclear spectroscopy has been widely used since the early 1960s. As measurement techniques and computers have improved, the nature of the requirements imposed on spectrum-processing programs has changed. For example, the great early difficulties in the technical and programmatic realization of an active dialog between a human being and a computer have long since ceased to be a problem, and the constraints on the amount of memory available for calculations and on computing speed have nearly vanished. It has been shown that there are practically no fundamental, essentially computational, difficulties in solving the spectrum-processing problem,[4,5] and all the stages from the preliminary one to the final one in the processing can be successfully algorithmized.[7,8] The focus of research has begun to shift to the search for methods of comparing the quality of the numerous existing programs, and to optimizing and standardizing them.[9,10] The timeliness and importance of this last problem are indicated, in particular, by the fact that one of the committees belonging to the International Electrotechnical Commission is showing an interest[10] in the standardization of programs for the complete processing of linear spectra (or individual stages in this processing), since such programs are now an integral part of multichannel pulse-height analyzers; as these are used in economic activity, they require the corresponding metrological guarantee.

Among the spectrum-processing programs are automatic ones which require no human intervention, and automated ones for which a human operator is essential in the processing, mainly at the stage of deciding which hypothesis to use on the basis of visual analysis of either the spectrum as a whole or individual parts of it. Automated programs have become popular owing to the availability of computers with convenient and powerful technical and programming methods of carrying on an active dialog with the operator and clearly (on a display) presenting the spectral data in a form convenient for visual analysis by a human being. The more widespread use of automatic programs is hindered by the absence of unique, generally applicable, formalized criteria for arriving at solutions when there is a statistical uncertainty. The use of automatic programs is at present restricted mainly to preliminary spectrum processing, owing mainly to "the complexity of determining the optimal criteria according to the experimental data and the difficulties in automatic image recognition" (Ref. 11).

This situation cannot be considered satisfactory. The rate of spectrum processing by automated programs is fundamentally limited. The upper limit is determined by the throughput of the human operator as an information channel, which is only about 30–50 bits/sec (Ref. 12), which is many orders of magnitude smaller than that of modern computers. Therefore, in a system consisting of an operator plus a computer working together in dialog mode, the human is the slowest component, and it is this component which in the end determines the total (very low) rate of automated spectrum processing. It was noted long ago[1,3] that the rates at which spectral data can be obtained can markedly exceed the rates at which they can be processed by automated programs, so that the rate of spectrum processing by such programs is to a significant degree restraining the rate at which research can be done. In addition, automated spectrum-processing programs make significant use of the "imagination, experience, and intuition" (Ref. 4) of the human operator in their operation. This helps to overcome the above-mentioned difficulties in automated processing. However, this operator participation has an uncontrollable, subjective effect on the results of the spectrum analysis by automated programs, making the results obtained less reliable, as they depend on personality traits and inclinations (including mood, temperament, conscientiousness) of the operator conducting the spectrum analysis.

A significant increase in the rate of spectrum processing, elimination of the effect of uncontrollable subjective factors, and creation of favorable conditions for the optimization and standardization of methods and programs for

 1063-7796/93/060677-28$10.00

complete spectrum analysis are possible only by using automatic programs. As mentioned above, there are no essential computational difficulties in the practical design of such programs. The problem is how to apply image-recognition theory to the specific problem of linear spectrum analysis to obtain the optimal quantitative criteria for making decisions about what hypothesis to choose when there is statistical uncertainty, and using these criteria in the interpretation of the results obtained by them. In recent years this problem has been studied intensively and some progress has been made. This review is devoted to new results which have been obtained and to the description of their practical application for designing, studying, optimizing, and standardizing automatic programs for linear-spectrum processing.

1. PROBLEMS OF AUTOMATIC SPECTRUM PROCESSING

The initial data for automatic spectrum-processing programs are:[7] the measured spectrum $y(x)$ in digital form [x is the channel number and $y(x)$ is the number of counts in the channel x], the width parameter *ib* determining the characteristic width of the peaks which are being sought, the parameter l describing the peak-search sensitivity, and several other auxiliary quantities (the number of channels limiting the processed segment of the spectrum, the number of this spectrum, and so on). No other information is required. The width parameter *ib* (usually an integer[7]) is assigned a value approximately equal to the expected average (half-)width of the peaks in the spectrum (for example, 5, 10, 20,... channels). The value of this parameter determines the characteristic width of the features (peaks) being sought, and allows useful information to be distinguished from background (for example, total-absorption peaks can be distinguished from broad Compton steps or from broad inverse-scattering peaks in the γ spectra, and total-absorption peaks can be distinguished from possible "blips"—sharp dips in the spectrum in certain channels). The sensitivity parameter l is specified by imposing requirments on the processing such that a compromise, necessary for experiments of a given type, is made between the sensitivity of the search and the reliability of the results. In addition, it is implicitly assumed that the spectrum is linear, i.e., that it contains useful information in the form of sufficiently narrow peaks (lines) on a smooth background about which nothing is generally known, while the shape of the peaks (lines) is known from the results of preliminary experiments.

A program based on these initial data will automatically, without human participation, perform the following tasks (see Refs. 7 and 8 for more detail): search for peaks with characteristic channel width *ib*, divide the spectrum into informative segments which subsequently can be treated independently, determine the initial approximate values of the parameters of interest for all peaks in these information segments, and then, in the course of fitting these parameters themselves (the positions in channels and the areas in counts), estimate the errors in the results of the processing and summarize the results in tabular form. Programs to establish a correspondence between the positions and areas of the peaks and their corresponding physical quantities (calibration programs) can then operate completely independently, and we shall not consider them here (this is a separate problem).

As was shown in Refs. 7 and 8, all the stages of automatic spectrum processing listed above can be successfully algorithmized. The main obstacles to creating automatic programs are the following difficulties, which are in general rather fundamental.

Automatic search for peaks and determination of the sensitivity threshold of the search

The results of the automatic identification of peaks in spectra depend on the sensitivity threshold of the search and often are not satisfactory to the user performing the analysis: instead of doing what is desired, the program "finds too few peaks or discovers them in excess" (Ref. 1). This is related to the fact that "the problem of automatically searching for peaks is analogous to the problem of image recognition, which is a quite complex problem nowhere near being solved" (Ref. 4). Difficulties arise not only in obtaining results but also in interpreting them probabilistically, which is unavoidable owing to the statistical nature of the data being processed.

Errors in determining the spectrum parameters and optimal criterion for interrupting the fit

The problem of determining the parameters of peaks belonging to complex multiplets (often referred to as spectrum decomposition) is usually reduced[4,5] to a system of essentially nonlinear equations with terms distorted by the fundamentally unavoidable statistical fluctuations in the experimental values. Here the mathematical models used for both an individual peak and the background are known only approximately, since they are also determined by analyzing the experimental data. The solution is found by numerically minimizing the value of the weighted sum (according to the channel number x) of rms deviations of the experimental values from the values calculated for the analyzed part of the spectrum using the mathematical model chosen for it (additive background and a given number of peaks). The values of the desired parameters of the peaks (and background) which are varied during the numerical iterative minimization procedure (the fit) are approximately specified before the start of the fit from independent considerations either by a program[7,8] or by hand.[4]

The numerical problems of spectrum decomposition are often incorrectly posed, and the corresponding minimization procedures are unstable. This difficulty has successfully been surmounted mathematically[4] using the method of regularization due to A. N. Tikhonov.[13] However, the problem of optimal interruption of the iterations during the fit and the problem of estimating the errors in the parameters found by the fit, which are closely related, remain unsolved.

Many criteria have been proposed for interrupting the fit, since none guarantees that the fit will be interrupted at a true minimum of the minimizing functional. Therefore,

in practice several (at least four[4]) interruption criteria are used simultaneously, with their free parameters chosen in a computational-experimental manner in processing the experimental data in such a way that the results of the fit satisfy the requirements of the users. In estimating the errors[4] in the values found for the spectrum parameters, first the purely statistical errors are found (usually from the diagonal elements of the inverse linear-step matrix or from the diagonal elements of the regularized matrix similar to it), and then the so-called "total" errors are found, which correspond to the "statistical" errors increased in some artificial way, with a subsequent check of the reliability by a special computational-experimental or purely experimental method.[14]

However, it is well known that the spectrum-processing problem has the following very special feature. The point is that "...the reliability of the estimate is manifested in the statistics as an ensemble property in the analysis of a series of measurements of a single type. But in analyzing the data of a physics experiment one most often processes a single set of data, and either there are no other sets at all, or they are obtained under other conditions" (Ref. 5). In order for the estimates to be reliable, an additional constraint is imposed: "it is necessary that the measurement error be not simply small in the sense of the dispersion, but also that it be literally small, i.e., that it not exceed a certain level determined by the parametrization of the model" (Ref. 5). This restriction is very rigorous, and it can be satisfied only by segments of the spectrum with high statistical reliability, which are rare for real data. Therefore, the estimated values of the peak parameters obtained as a result of the fit to an actual spectrum (one set of data) are often unreliable. Accordingly, the values of the errors which are calculated are physically unsatisfactory and, as a rule, are too low.[1)] This fact is well known to experimentalists, the users of spectrum-processing programs, and some of them react to it by arbitrarily increasing the errors found by the program in the positions or areas of the peaks by fifty percent or by a factor of two or more, depending on the specific situation and previous experience.

Therefore, there are practically no criteria for optimal interruption of the fit or universal correction methods for reliably estimating the desired spectrum parameters determined by this fit. And this statement holds equally for both automatic and nonautomatic spectrum-processing programs.

Determination of the number of peaks in a multiplet

The number of overlapping peaks forming a multiplet and the initial values of the parameters of these peaks for the fit are specified for automated programs "by hand" during the demarcation, i.e., when the spectrum is scanned by the operator conducting the processing. However, the experimental spectrum is distorted by statistical fluctuations, and the mathematical model of the peak and the mathematical model of the background are known only approximately, so that equally efficient approximations of the experimental data by peaks and background can be obtained in a large number of ways. The result of the decomposition of the analyzed segment of the spectrum is therefore nonunique and depends on how many peaks, in the opinion of the operator, are contained in a multiplet and what, in the operator's opinion, is the order of the polynomial approximating the background below these peaks.

To overcome this uncertainty in the decomposition of the multiplet into constituent peaks, attempts are sometimes made to use "physical information about the spectrum, for example, the positions of the peaks in diffraction spectra can be calculated using formulas, and the peaks in γ and X-ray spectra can be calculated using the known or assumed isotopic (element) content of the sample, and so on. The number of peaks in the spectrum can be improved after a nonlinear least-squares analysis of the given spectrum...using statistical tests (for example, the χ^2 criterion). Unfortunately, this method works only in trivial cases and fails in the most complicated and interesting ones: when there are very closely spaced overlapping and weak peaks" (Ref. 5). It has been suggested that in these very complicated cases one can use the characteristic property of regularized processes of the Newton type, in which "the attempt to find by a regularized iteration process a number of regularities[2)] which is larger than the true number for each specific problem leads to a pseudosolution where either there are regularities with zero or negligible amplitude, or there are pairs of regularities with different amplitudes but whose other parameters are coincident or very close... . Coincident or very close parameters imply concident or very close positions of two lines in a given segment... . For example, for the half-width of a single line equal to 5–6 channels the calculated positions of two peaks less than 0.25–0.3 channels apart can be considered coincident" (Ref. 4). Therefore, in a "complex" case the operator must make several variants of demarcation of the analyzed segment of the spectrum assuming the existence of different numbers of peaks forming multiplets, and after the calculations using the procedure described above must select the variant for which the results of the processing contain peaks "with zero or negligible amplitude" and (or) peaks "with different amplitudes but with their other parameters coincident or very close." Furthermore, it is assumed that for this variant there are in fact no peaks "with zero or negligible amplitude," while peaks "with different amplitudes but with their other parameters coincident or very close" form a single peak whose amplitude is the sum of the amplitudes of the individual peaks. The number of peaks in the multiplet remaining after this procedure is assumed to be the true number. It is maintained that "the use of additional physical information...actually becomes necessary only when the measured data are very poor (insufficient statistics). Extraneous regularities...can be eliminated only using the regularized iteration process" (Ref. 4).

However, it is obvious from physical considerations that real spectra can also contain true peaks of small amplitude, and true peaks separated by 0.25–0.3 channels. The fact that the amplitude of a peak is small or that the

positions of two peaks are close cannot by itself imply that there is no small peak, or that at the position of two closely spaced peaks there is only one peak but of large amplitude. All variants are possible. The final decision about the structure of the analyzed segment of the spectrum is taken by a person using all the available information, including information not contained in the analyzed data. However, the main argument for making such a decision according to the results of the processing must be information on which of the possible variants of the spectrum structure is most probable. However, the method of Ref. 4 gives no recipes for calculating such probabilities. Therefore, the principal, certain advantage of the regularized iteration process is that it preserves the stability and makes it possible to obtain values of the peak parameters for various assumptions about the structure (the number of peaks) of the processed multiplet, including for cases where the peak is small or the peaks are closely spaced. The results obtained in this manner can then be used by a human to determine the number of peaks in a multiplet, but still the method does not explicitly supply the necessary quantitative recommendations for taking this solution. Therefore, now as before the acute problem arises of computing, on the basis of only the statistical reliability of the experiment, the probabilities that some (any) particular peak is a component of the analyzed multiplet or the probabilities that the multiplet consists of some set of peaks with definite positions and amplitudes. Here the shape of the peak (a mathematical model for it) is assumed to be known.

Standardization of spectrum-processing programs

The necessity of comparing the quality and of optimizing and standardizing spectrum-processing programs has long been realized and the first steps in doing this have been taken.[9,10] However, objective criteria for determining the quality of algorithms and programs are almost not used at all yet. Therefore, the actual testing of programs and the results can to some degree be subjective and random or even "a matter of misunderstanding" (Ref. 5). Comparison of program quality requires a quantitative, statistically justified criterion for such a comparison and a specific practical method for applying it. The testing will then have an objective, quantitative nature and its results can be used to formulate and justify a single standard for complete spectrum-processing programs or to standardize the individual stages of such processing. However, there is the widespread opinion that "in practice, apparently, for a long time still programs will continue to be evaluated by methods which do not give a complete and convincing picture" (Ref. 5). This will probably happen until statistical criteria for comparing programs and techniques for applying them are designed and used by the scientific community.

2. STOCHASTIC MODELS OF LINEAR SPECTRA

The problems in (automatic) spectrum processing arise from the essentially statistical nature of the spectra. For this reason, objectively ambiguous situations arise during various stages of the processing, where a solution about the choice of hypothesis can be made only in probabilistic terminology. Therefore, to solve the problems of automatic processing it is necessary to learn to calculate the probabilistic characteristics using the processed spectrum. These characteristics can be determined if only the statistical properties of the spectrum are known and specified explicitly. However, the models of linear spectra usually used (see Refs. 1–5 and many other examples) are deterministic and do not contain any information about the statistical properties of the experimental data, although it is always implicitly assumed that the counts in the spectrum channels obey a Poisson distribution (or a Gaussian distribution in the limit of large statistics) and are not correlated with each other (this hypothesis is confirmed experimentally in Ref. 15). Therefore, new progress in solving the problems of automatic processing can be achieved only by using models of linear spectra in which their statistical properties are specified explicitly. We shall refer to these as stochastic models of linear spectra.[16]

We shall write the function approximating the experimental spectrum (for example, in the sense of minimum χ^2) as the sum of two terms: one completely deterministic (not containing any random parameters), which describes the true spectrum unperturbed by statistical fluctuations (this is the usual deterministic model[1–5]), and the other random, describing the random distortions of the true spectrum related to the presence of statistical fluctuations of the experimental values.

For the background we write

$$bkgd(x)=bkgd_0(x)+u_c(x), \tag{1}$$

where x are the channels, $bkgd(x)$ is a random function approximating the experimental background, $bkgd_0(x)$ is a deterministic function describing the true background undistorted by fluctuations, and $u_c(x)$ is a random function describing the random deviations of the curve approximating the background from its true values (noise).

For the peak we write

$$u(x)=a_0u_1(x)+au_2(x), \tag{2}$$

where $u(x)$ is a random function approximating the peak, $a_0u_1(x)$ is a deterministic function describing the true peak undistorted by fluctuations (a_0 is the true amplitude of this peak), and $au_2(x)$ is a random function describing random distortions of the curve approximating the peak from its true values (a is a random amplitude). In what follows we shall assume that $u_1(x)=u_2(x)$, i.e., the function describing the shape of the peak is not distorted by random fluctuations.

For the spectrum we write

$$y(x)=bkgd(x)+u_c(x)+\sum_{j=1}^{m}(a_0+a_j)u_j(x), \tag{3}$$

where $y(x)$ is a random function approximating the measured spectrum, j is the number of the peak in the spectrum, and m is the total number of peaks.

It is assumed that the probability density distribution $\mu(a_j)$ for the random parameter a_j (the random amplitude of the jth peak) has the form of a symmetric Gaussian with zero average value

$$\mu(a_j)=(2\pi\sigma_j)^{-1/2}\exp(-a_j^2/(2\sigma_j^2)), \tag{4}$$

and the function $u_c(x)$ is Gaussian white noise with zero average value and constant power density N_0. The value σ_j of the dispersion of the random deviation of the amplitude of the jth peak from its true value, which is related to the effect of statistical fluctuations of the experimental data, can as a first approximation be taken equal to the error in determining its amplitude found as a result of the spectrum processing. In order to express the noise power density N_0 in terms of the background parameters, we write down the expressions for the correlation function $R(\tau)$ of white noise:

$$R(\tau)=N_0\sigma(\tau)/2; \quad R(0)=n_0^2, \tag{5}$$

where τ is the correlation length and n_0^2 is the rms value of the noise amplitude. Setting

$$n_0=bkgd^{1/2}, \tag{6}$$

for N_0 from (5) we find

$$N_0=2bkgd_0. \tag{7}$$

Therefore, the spectrum model (3) contains additive random functions of two types describing the distortions of the true spectrum by statistical fluctuations of the experimental values: the noise and m random functions $a_ju_j(x)$ linear in the random parameters a_j. The statistical properties of these functions are clearly specified by Eqs. (4)–(7). Since the random parameters a_j enter into the model (3) linearly, we shall refer to it as the linear stochastic model. The model includes distortions of the true spectrum by both background fluctuations and fluctuations of the number of counts in a peak. But when in (2) we assumed that the function $u_1(x)$ describing the shape of the true peak coincides with the function $u_2(x)$ describing the random deviations from this curve, we thereby established the fundamental limit of this model: fluctuations in the number of experimental counts at a peak distort only its true amplitude, while the shape and position of the peak are not changed. Meanwhile, fluctuations in the number of background counts (noise) lead in this model to distortion of the amplitudes, the positions of all the peaks, and the shapes of all the peaks in the spectrum.

Among the diversity of problems leading to information of the spectral type, a common case of practical importance can be distinguished in which the model corresponds almost completely to the determination by spectrometric methods of the ratio of the components in a sample of known content. In this case the characteristic spectrum of each component is known, and therefore the peak positions in the processed spectrum can be considered to be given. Then the limitation of the model owing to neglect of random distortions of the positions due to fluctuations in the number of counts at the peaks can be considered unimportant. The linear model (3) can be generalized to include random distortions of the amplitudes, the positions, and the shape of the peaks owing to fluctuations in the number of counts at the peak. However, in this case the model becomes nonlinear in the random parameters and it becomes difficult to use it to obtain analytic results.

In Ref. 17 the linear stochastic model of one-dimensional linear spectra[16] was generalized to the case of linear spectra of arbitrary dimension k, which can be useful, for example, in analyzing the two- or three-dimensional spectra obtained in correlation experiments in nuclear spectroscopy ($\gamma\gamma$, $\gamma\gamma t$, etc. coincidences).

3. THE AUTOMATIC PEAK SEARCH AND ITS QUALITY INDICATORS. DETERMINATION OF THE SEARCH SENSITIVITY THRESHOLD

Let us assume that on the basis of some criterion it has been determined that the spectrum (3) contains $m-1$ peaks in addition to the background, the information on which is not important. We shall seek a quantitative criterion and indicators of its quality which can be used as the basis for making a statistically reliable conclusion about which of the following two alternative hypotheses is most probable. Hypothesis H_0: the mth peak is absent in the spectrum (3) and the spectrum $y(x)$ contains $m-1$ peaks. Hypothesis H_1: the mth peak is present in the spectrum (3) and the spectrum $y(x)$ contains m peaks. If such a criterion is known, m successive applications of it can be used to determine the total number of peaks in the spectrum.

We rewrite Eq. (3) as

$$y(x)=u_c(x)+\sum_{j=1}^{m}a_ju_j(x)+u_s(x)+u_i(x), \tag{8}$$

where

$$u_s(x)=bkgd_0(x)+\sum_{j=1}^{m-1}a_{0j}u_j(x);$$

$$u_i(x)=a_{0m}u_m(x). \tag{9}$$

Therefore, the problem of choosing hypothesis H_0 or H_1 is the problem of discovering the fully deterministic process $u_i(x)$ on the background of the noise $u_c(x)$ of m interfering processes with random amplitudes (a_j for the jth process) and the interfering fully deterministic process $u_s(x)$. Analogous problems also arise in the theory of finding radiolocation signals in the presence of noise.

We shall write functions of the channel number x in the form of multidimensional vectors whose coordinate components are essentially Kotel'nikov discrete steps.[18] Then Eq. (8) takes the form

$$\mathbf{y}=\mathbf{u}_c+\sum_{j=1}^{m}a_j\mathbf{u}_j+\mathbf{u}_s+\mathbf{u}_i. \tag{10}$$

The probability density $p_0(\mathbf{u})$ of the random realization of some vector $\mathbf{u}$ by Gaussian white noise with zero average value and constant power density N_0 is

$$p_0(\mathbf{u})=(\pi N_0)^{\nu/2}\exp(-\mathbf{u}^2/N_0), \tag{11}$$

where ν is the dimensionality of the vector space to which the vector $\mathbf{u}$ belongs. The probability density $p_m(\mathbf{u})$ of the random realization of the vector $\mathbf{u}$ by noise and m interfering processes with random amplitudes a_j can be found using Eq. (11) and the superposition rule for the probability densities:

$$p_m(\mathbf{u})=p_{m-1}(\mathbf{u})\left(\frac{N_0/\sigma_m^2}{2\mathbf{u}_m\mathbf{R}_m+N_0/\sigma_m^2}\right)^{1/2}$$
$$\times\exp\left[\frac{2(\mathbf{u}\mathbf{R}_m)^2}{N_0(2\mathbf{u}_m\mathbf{R}_m+N_0/\sigma_m^2)}\right], \tag{12}$$

where

$$\mathbf{R}_1=\mathbf{u}_1;\quad \mathbf{R}_m=\mathbf{u}_m-2\sum_{j=1}^{m-1}\frac{\mathbf{R}_j(\mathbf{u}_m\mathbf{R}_j)}{2\mathbf{u}_j\mathbf{R}_j+N_0/\sigma_j^2}. \tag{13}$$

It is assumed that the probability density distribution for the random parameter a_j has the form of a symmetric Gaussian (4). The probability density $p_{m+1}(\mathbf{u})$ of the random realization of the vector $\mathbf{u}$ by noise, m interfering processes with random amplitudes, and the fully deterministic process $\mathbf{u}_s$ is

$$p_{m+1}(\mathbf{u})=p_m(\mathbf{u})\exp[-(\mathbf{u}_s\mathbf{R}^s_{m+1}-2\mathbf{u}\mathbf{R}^s_{m+1})/N_0], \tag{14}$$

$$\mathbf{R}^s_{m+1}=\mathbf{u}_s-2\sum_{j=1}^{m}\frac{\mathbf{R}_j(\mathbf{u}_s\mathbf{R}_j)}{2\mathbf{u}_j\mathbf{R}_j+N_0/\sigma_j^2}. \tag{15}$$

Similarly, the probability density $p_{m+2}(\mathbf{u})$ of the random realization of the vector $\mathbf{u}$ by noise, m interfering processes with random amplitudes, and the fully deterministic processes $\mathbf{u}_s$ and $\mathbf{u}_i$ can be written as

$$p_{m+2}(\mathbf{u})=p_m(\mathbf{u})\exp[-(\mathbf{u}_{s+i}\mathbf{R}^{s+i}_{m+1}-2\mathbf{u}\mathbf{R}^{s+i}_{m+1})/N_0], \tag{16}$$

where $\mathbf{u}_{s+i}=\mathbf{u}_s+\mathbf{u}_i$ and the expression for $\mathbf{R}^{s+i}_{m+1}$ is analogous to Eq. (15).

Dividing Eq. (16) term-by-term by Eq. (14), we obtain the probability ratio

$$l(\mathbf{u})=\frac{p_{m+2}(\mathbf{u})}{p_{m+1}(\mathbf{u})}=\exp[-(\mathbf{u}_i\mathbf{R}^i_{m+1}+2\mathbf{u}_s\mathbf{R}^i_{m+1}$$
$$-2\mathbf{u}\mathbf{R}^i_{m+1})/N_0]. \tag{17}$$

The probability functional (17) shows how many times larger the probability of realizing the vector $\mathbf{u}$ by noise, m interfering processes with random amplitudes, the deterministic interfering process $\mathbf{u}_s$, and the deterministic process $\mathbf{u}_i$ subject to discovery is than the probability of realizing the vector $\mathbf{u}$ by only interfering processes. It also specifies for the model (3) a statistically optimal algorithm for discovering a peak [the process $\mathbf{u}_i$ in the spectrum (10)] and thereby a statistically reliable quantitative criterion for choosing between hypotheses H_0 and H_1. If this ratio for some function $\mathbf{u}$ (the experimental spectrum $\mathbf{y}$ or part of it) has a value above threshold

$$l(\mathbf{u})>l_0, \tag{18}$$

hypothesis H_1 should be chosen, and hypothesis H_0 should be chosen otherwise. These results were first obtained in Ref. 16.

For alternative hypotheses such as H_0 and H_1, the quality indicators of the discovery algorithm are the conditional probability F for false discovery (statistical reliability of the criterion) and the conditional probability D for true discovery (the strength of the criterion). To calculate them we find the probability density of the random effect of optimal processing of noise and interfering processes. For this we introduce the logarithmic partial probability functional

$$z=\mathbf{u}\mathbf{R}^i_{m+1}$$
$$=\mathbf{u}_c\mathbf{R}^i_{m+1}+\sum_{j=1}^{m}a_j\mathbf{u}_j\mathbf{R}^i_{m+1}+\mathbf{u}_s\mathbf{R}^i_{m+1}$$
$$=z_c+\sum_{j=1}^{m}z_j+z_s \tag{19}$$

and find the law for the probability distribution and dispersion for z. Using the fact that z_c and z_j are uncorrelated random variables obeying a Gaussian distribution with zero average value and z_s is a constant, for z we obtain the following probability distribution law:

$$\mu(z)=(2\pi\sigma_z)^{-1/2}\exp[-(z-z_s)^2/(2\sigma_z^2)] \tag{20}$$

with the dispersion

$$\sigma_z^2=N_0\mathbf{u}_i\mathbf{R}^i_{m+1}/2. \tag{21}$$

Since the conditional probabilities for false and true discovery F and D of peaks are given by

$$F=\int_{z_0}^{\infty}\mu(z)dz; \tag{22}$$

$$D=\int_{z_0}^{\infty}l(z)\mu(z)dz, \tag{23}$$

where z_0 is the threshold value of the partial probability functional z, then by using (17), (20), and (21) from (22) and (23) we find

$$F=[1-\Phi(q_0)]/2, \tag{24}$$

$$D=[1+\Phi(q-q_0)]/2, \tag{25}$$

where $\Phi(q)$ is the error function, and for q_0 and q we have

$$q_0=(z_0-z_s)/\sigma_z, \tag{26}$$

$$q=(2\sigma_z/N_0)(2\mathbf{u}_i\mathbf{R}^i_{m+1}/N_0)^{1/2}. \tag{27}$$

Equations (24)–(27) determining the conditional probabilities for false and true discovery also determine the quality indicators of the optimal peak-discovery algorithm for the linear stochastic model of the spectra (3). They were also first obtained in Ref. 16. The parameter q_0 entering into these formulas is determined from the given maximum acceptable probability for false discovery F: it is related to the threshold value z_0 of the partial probability functional (or the threshold sensitivity of the search). In Ref. 16 for

the $m=1$ (singlet) and $m=2$ (doublet) examples explicit analytic expressions are given for F and D, the quality indicators of the model, statistically optimal search algorithm, assuming that the shape of the peaks in the spectrum is described by a symmetric Gaussian curve. The values calculated using the empirical expressions (48) and (54) (Sec. 4) were used as estimates of σ_j. The parameter q (27) used to determine the conditional probability D for true discovery of a peak for a given probability F for its false discovery can also be calculated for cases $m>2$ in the form of a function of the parameters of all the peaks forming the multiplet, the background, and the quantities σ_j. It is true that explicit expressions for σ_j are at present known only for the singlet (48) and the doublet (54), but this does not exclude the possibility of practical application of the results, since as estimates of σ_j for multiplets with $m>2$ as an initial approximation it is possible to use the error in determining the amplitudes of the peaks forming the multiplet, obtained in the processing.

In deriving Eqs. (24)–(27) we nowhere made use of the information about the peak shapes, so that these expressions can be used for any particular form of the functions $u_j(x)$. In particular, they can be specified in tabular form.

Therefore, for the linear stochastic model of the processed spectrum (3) in a general form we have obtained an algorithm for the statistically optimal discovery of peaks in the presence of background and an arbitrary number of interfering peaks, and we have determined the quality indicators of this algorithm: the conditional probabilities F and D. Using the results for a multiplet of arbitrary order m (consisting of m peaks) and the values of the parameters of all the peaks and their errors (the amplitude errors) found by processing it, it is possible to make a statistically reliable choice between two alternative hypotheses: the mth peak is present in the multiplet or the mth peak is absent from it.

The linear stochastic model of k-dimensional linear spectra can be used to generalize the results given here to the case of peak identification in spectra of arbitrary dimension k (Ref. 17). It has been shown that the optimal model algorithm for finding k-dimensional peaks and its quality indicators are also given by Eqs. (17), (24), and (25), but in this case written not for vectors, but for $k\times k$ matrices.

Let us now consider in more detail the identification of isolated peaks ($m=1$), which is the case most often encountered in practice. We shall use the general expression (17) to study the model, statistically optimal algorithm for discovering such peaks. We assume that the spectrum is composed only of peaks which are separated far enough that their overlap can be neglected. The identification of peaks from doublets ($m=2$) or higher-order multiplets ($m>2$) is a separate problem, and each case has its own optimal discovery algorithm.

From the general expression (17) for identifying isolated peaks we obtain the following formula for the logarithmic probability ratio (without loss of generality we assume that the spectrum contains only a single peak):

$$L(\mathbf{y})=\alpha(a_0/bkgd_0)(\mathbf{yu}-\mathbf{bkgd\,u}-a_0\mathbf{u}^2/2), \tag{28}$$

where $L(\mathbf{y})=\ln l(\mathbf{y})$, $\mathbf{y}$ is the spectrum, $\mathbf{u}$ is the shape function of the peak, and a_0 is the true amplitude of the peak. The background **bkgd** below the peak will for simplicity be assumed to be constant with amplitude $bkgd_0$. For the coefficient α we have

$$\alpha=bkgd_0/(bkgd_0+\mathbf{u}^2\sigma^2). \tag{29}$$

The quality indicators (24) and (25) of the optimal discovery algorithm (28) are respectively

$$F=[1-\Phi((\alpha/(2k))^{1/2}L_0)]/2, \tag{30}$$

$$D=\{1+\Phi[(\alpha/(2k))^{1/2}\times(L_0-a_0k(b/bkgd)^{1/2})]\}/2, \tag{31}$$

where Φ is the error function, $L_0=\ln l_0$ is the threshold value of L [l_0 is the right-hand side of the inequality (18)], and α is determined by Eq. (29). It is assumed that the peak has the shape of a symmetric Gaussian with a half-width of b channels. Then $\mathbf{u}^2=kb$ with $k=(\pi/(8\ln 2))^{1/2}$.

We see that the optimal model algorithm for the discovery of isolated peaks (28) and its quality indicators (30) and (31) explicitly depend on the true values of the parameters of the identified peak (a_0,$\mathbf{u}$) and on the true value of the background below the peak ($bkgd_0$). Since these parameters are *a priori* unknown in the search, it is in principle impossible to fully realize the optimal discovery algorithm (28).

In order to realize at least partially the model optimal discovery algorithm, we use the partial probability functional (28) to construct the partial probability functional $z(\mathbf{y})$ which no longer depends explicitly on the true peak amplitude a_0:

$$z(\mathbf{y})=(\mathbf{yu}-\mathbf{bkgd\,u})(bkgd_0\mathbf{u}^2)^{-1/2}. \tag{32}$$

Since for $z(\mathbf{y})$ the peak shape function $\mathbf{u}$ and the background **bkgd** can be found approximately beforehand (before or during the search), the functional (32) can already be used in practice as an algorithm for seeking peaks with discovery criterion $z(\mathbf{y})>z_0$, where z_0 is some threshold value of z. The discovery algorithm based on the functional (32) will be referred to as suboptimal for the spectrum model (3).

A computer program (the ALISA program; Ref. 19) based on this algorithm has been written. Instead of the scalar products (integrals) in each channel x_i of the processed spectrum in (32), it computes the sum

$$z(x_i)=(bkgd_0b)^{-1/2}\sum_{x=x_i-b}^{x_i+b}[y(x)-bkgd_0(x)]u(x-x_i), \tag{33}$$

where $y(x)$ is the number of counts in channel x, $bkgd_0(x)$ is the estimate of the true background in the channel x, which is taken to be a straight line passing through the point with the coordinates (x_i-2b,y_-) and (x_i+2b,y_+), y_- and y_+ are the average values of the number of counts $y(x)$ according to the three channels lying closest to

(x_i-2b) and (x_i+2b), and $u(x-x_i)=\exp(-\ln 16(x-x_i)^2/b^2)$ is the shape function of the identified peak (b is its half-width). For this choice of it $u^2=kb$ [here the unimportant coefficient k in (33) has been dropped]. There is also a variant of the program with an additional criterion which eliminates from consideration peaks with small half-width [a peak is assumed to be identified if $z(x_i)>z_0$ in two or more adjacent channels or if these channels are separated by a single channel for which $z(x_i)<z_0$].

The quantitative quality characteristics (F and D) of this program realizing the model suboptimal search algorithm will be discussed later. For now we stress the fact that the conditional probability F for false discovery is a parameter of the search algorithm, so that the operator conducting the processing must specify its value on the basis of independent considerations, governed only by the requirements imposed on the processing. The conditional probabilities F and D form a system of equations

$$D+P_D=1; \quad F+P_F=1, \tag{34}$$

where D is the probability of finding a peak with the condition that it really exists in the spectrum, P_D is the probability of not finding the peak with the same condition, F is the probability of finding a peak with the condition that the peak is actually absent in the spectrum in question (i.e., F is the probability of false processing, when random statistical fluctuations of the experimental data are interpreted as a peak), and P_F is the probability of not finding a peak with the same condition. Since the probabilities F and D are determined for different conditions and obey Eqs. (24) and (25), we have

$$F+D\neq 1; \quad 0<F\leqslant D<1. \tag{35}$$

Owing to the statistical nature of the experimental data, it is never impossible to exclude completely the possibility of interpreting random fluctuations in the spectra as a true peak, because always $F>0$. For the same reason it is impossible to eliminate completely the possibility of not discovering a true peak of arbitrarily large amplitude in the spectrum, as it can be masked by random fluctuations of the experimental values, since always $D<1$. It also follows from (34) that $F=D$ in the identification of peaks of zero amplitude, when spectra without peaks are processed. Therefore, the ideal situation in which in the (automatic) processing of the spectrum all true peaks (including small ones) are always found and no false peak is identified is impossible in principle ($F>0$ and $D<1$ always).

When specifying the probability F of false discovery, it is necessary to find a compromise between two extreme possibilities which is acceptable for experiments of a given type. The first possibility is to choose F to be large. This ensures high probability of a true discovery, even for small peaks. However, the probability for false discovery is also large, i.e., among the identified peaks there are many false ones. In this case the search sensitivity is high, but the reliability of the result is low. The second possibility is to choose F to be small. Here the probability for a true discovery is smaller, because some of the peaks, especially small ones, can be missed in the search. However, the probability of false discovery is also small, i.e., there will be few false peaks among the identified ones. In this case the search sensitivity is low, but the reliability of the result is high.

To find the relation between the numerical values of the quality indicators of the peak-search programs (algorithms) (F and D), the parameters of the processed spectra, and the parameters of these programs (algorithms) themselves, it is convenient to use[20] the results of a systematic model experiment done by computer. This involved the numerical modeling (construction) of a large number of short [5–10 (half-)widths of an identified peak] spectra with quasiexperimental statistics and constant true background which did or did not contain peaks whose shape was known exactly (usually a symmetric Gaussian). Here the peak amplitude a, the half-width b, and the amplitude of the background below the peak *bkgd* were known exactly, since they were specified in the construction. The processing of a large number ($\sim 10^4$–10^5) of such spectra with quasiexperimental statistics makes it possible, for any specific peak-identification algorithm, to determine the value of the peak sensitivity threshold corresponding to any *a priori* specified value of the conditional probability F of false discovery, and to determine the dependence of the conditional probability D of true discovery on the values of the peak and background parameters (on a, b, and *bkgd*) for this value of F.

The numerical experiment of Ref. 20 was carried out in two stages. First, the dependence of F on the search sensitivity threshold was determined. For this a large number of quasiexperimental spectra with constant background *bkgd* not containing peaks ($a=0$) were constructed. The program being studied, running with a fixed value of the sensitivity threshold, was used to search for peaks in all these spectra. Then the value of F corresponding to a given value of the search sensitivity threshold is

$$F=p/n, \tag{36}$$

where p is the number of peaks found and n is the number of processed spectra (it is assumed that there can be only a single peak in each spectrum). Second, the dependence $D(a)$ of the conditional probability D of true discovery on the amplitude of the identified peak a was determined for a fixed value of the conditional probability F of false discovery (the peak half-width b and the amplitude of the background below the peak *bkgd* were also fixed in this case). For this a large number of spectra with peaks of different amplitude a (and constants b, *bkgd*) were constructed, and the search program with sensitivity threshold corresponding to the chosen value of F (for example, $F=0.01$) was used to seek peaks in all these spectra. Then

$$D=p(a)/n(a), \tag{37}$$

where $p(a)$ is the number of identified peaks with amplitude a and $n(a)$ is the number of processed spectra containing such peaks. The errors ΔF and ΔD in the values of

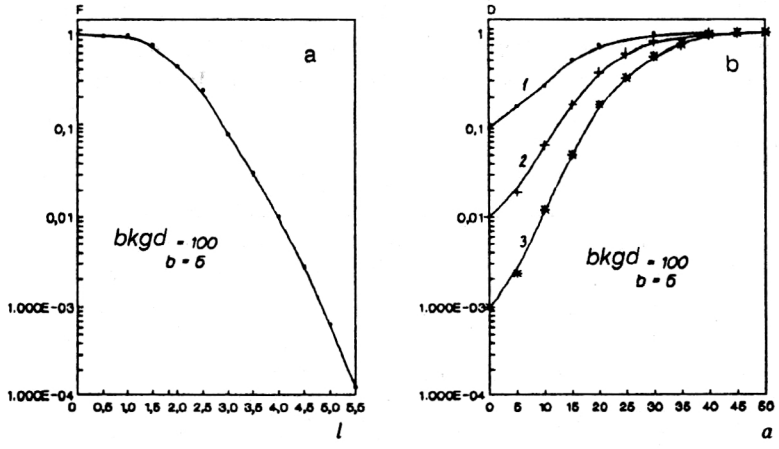

FIG. 1. (a) Conditional probability F of false discovery as a function of the search sensitivity threshold l and (b) dependence of the conditional probability D of true discovery for $F=0.1$, 0.01, and 0.001 (curves *1*, *2*, and *3*, respectively) on the amplitude a of the identified peak for the program of Ref. 19 realizing the model suboptimal search algorithm (32). The points are the result of a model experiment.

F and D that are found are calculated using the expression for the errors in quantities obeying a binomial distribution law. For example,

$$\Delta F=(F(1-F)/n)^{1/2}. \tag{38}$$

In Figs. 1–3 we show as an example several results of studies, using the method of Ref. 20, of the properties of the program (the ALISA program[19]) realizing the suboptimal model search algorithm (32), (33).

The dependence of the conditional probability F of false discovery on the search sensitivity threshold l is shown in Fig. 1a, and the dependence of the conditional probability D of true discovery of a peak on its amplitude a for $F=0.1$, 0.01, and 0.001 is shown in Fig. 1b. An example of the dependence of F on the background amplitude *bkgd* found for fixed search sensitivity threshold l is shown in Fig. 2. It follows, in particular, from this figure that in spectra with variable background, when peaks are identified with constant search sensitivity threshold, the value of the conditional probability F for false discovery varies over the spectrum, although not very strongly. Therefore, in order to obtain $F=$const in a search for such spectra using the program of Ref. 19, it is neccessary to carry out the search for a variable value of the sensitivity threshold l.

In Fig. 3 for $F=0.1$ we compare the dependences of the probability $D(a)$ for true discovery of peaks on their amplitude a for two variants of the ALISA program (curves *2* and *3*), for the ALISA program when the true background and half-widths of the Gaussian peaks in the spectrum are known exactly *a priori* during the running of the program (curve *4*), and with the dependences $D(a)$ for the optimal model search algorithm calculated with different assumptions (curves *1* and *5*). Curves *2*, *3*, and *4* were constructed from the values obtained in the model experiment performed by computer, and curves *1* and *5* were calculated using Eqs. (30) and (31). Curve *5* was obtained for $\alpha=1$, and curve *1* was obtained for the value of α determined by (29), when the potential error in de-

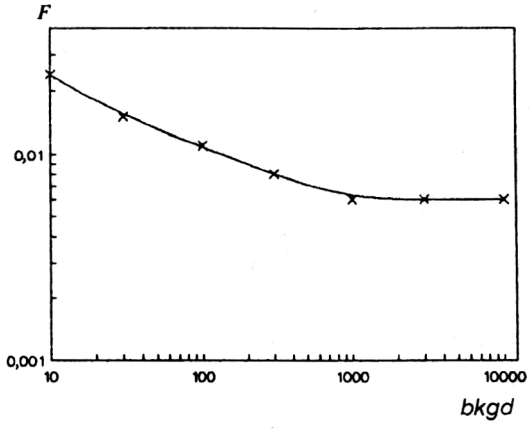

FIG. 2. Dependence of F on the true background amplitude *bkgd* in the processed spectrum for fixed ($l=4.0$) search sensitivity threshold for the peak-identification program of Ref. 19. The points ($\times$) are the result of a model experiment.

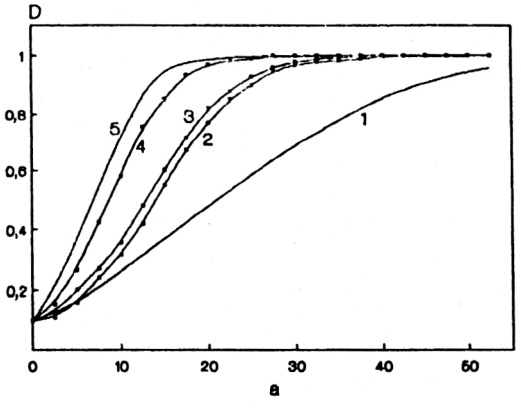

FIG. 3. Dependence of the conditional probability D of true discovery of a peak of symmetric Gaussian shape with half-width $b=5$ channels for constant background with a true amplitude of $bkgd=100$ counts on the peak amplitude a with the conditional probability of false discovery $F=0.1$: (*1*) calculation (31) with σ from (48); (*2*) ALISA (Ref. 19) without an additional criterion; (*3*) ALISA (Ref. 19) with an additional criterion; (*4*) ALISA (Ref. 19) with *a priori* exactly known true background and half-width of the identified peak; (*5*) calculation of (31) with $\alpha=1$. The points on curves *2*, *3*, and *4* are the result of the model experiment. The calculation of $D(a)$ using Eq. (31) with σ from (39) leads to complete coincidence of curves *1* and *3*.

termining the amplitude of an isolated peak using Eq. (48) is taken as the estimate of σ. The value $\alpha=1$ is obtained from (29) for $\sigma=0$, i.e., the case corresponds to a stochastic model simpler than (3) in which random distortions of the true spectrum of statistical fluctuations of the experimental values of only the background are taken into account, and the increase in the number of counts in the region of the peaks is neglected. We see from this figure that curves *1* and *5* bound the experimental dependences *2*, *3*, and *4* above and below. The at first glance paradoxical fact that the theoretical curve *1* for the optimal search algorithm lies below curves *2*, *3*, and *4* for different variants of the program realization of the suboptimal search algorithm is explained by the incomplete correspondence between the spectrum model (3) and the estimate of the dispersion σ (48) used (for lack of anything better) in constructing curve *1*. The point is that the estimate (48) was found for the case where the error σ is determined both by the background counts and by the counts at the peak, and an estimate determined only by the counts at the peak is needed for substitution into (29). Therefore, the use of the value of σ from Eq. (48) in calculating α, F, and D leads to double counting of the background and, consequently, to calculated values of $D(a)$ which are too low. The calculated curve *1* is completely compatible with the empirical curve *3* if σ is calculated using the expression

$$\sigma=3.4[1+2(a/bkgd)^{-1/2}]b^{-1/2}, \qquad (39)$$

which was also found empirically in fitting curve *3* by Eqs. (29)–(31) with σ as a variable parameter (function). This correspondence between curve *3* found from the results of the model experiment and the calculated model curve implies that the stochastic model (3) with σ defined according to Eq. (39) allows the quality of the program realizing the suboptimal search algorithm to be estimated completely quantitatively. The calculated curve *5* lies much higher than curves *2*, *3*, and *4*. This happens because in calculating $D(a)$ using Eq. (31) with $\alpha=1$ the increase of the fluctuations near the peak was neglected, and the amplitude a, the peak half-width b, and the background under the peak *bkgd* were assumed to be known exactly *a priori*. The calculated curve *5* can approach the empirical curve *3* if the effective background $bkgd_{\text{eff}}\sim 4bkgd$ is used in the calculations, but then the empirical curve is not reproduced at all.

It also follows from Fig. 3 that the dependence $D(a)$ for the ALISA program with an additional test (curve *3*) is everywhere higher than for the same algorithm without this test (curve *2*). This suggests the possibility of improving the quality of the search by introducing into the program additional criteria containing *a priori* information about the spectrum (in this case the content of this *a priori* information is that there can be no peak in the spectrum with a half-width of less than two channels). However, the quality of the search is affected much more by the use of *a priori* information about the true background *bkgd* in the processed spectra (curve *4* in Fig. 3 lies much higher than curves *2* and *3*). Therefore, the quality of the search programs is better, the larger the amount of *a priori* information about the processed spectrum that is used. It is possible, for example, to obtain this additional information about the spectra by iterative processing of them as a whole.

The effect of additional criteria on the quality of peak-search programs was studied specially in Ref. 21. For the four programs from Refs. 22–24 the technique of Ref. 20 was used to check the suitability of the use of the additional criteria in them for improving the quality of the search programs.

In comparing the original authors' texts in Fortran for the programs PEAK FIND (Ref. 22), PEAKS (Ref. 23), PEAK, and PEAKI (Ref. 24) with the desciptions of their algorithms from the corresponding studies, in all the programs groups of several operators were isolated who check the results of the calculations using the algorithms described for correspondence to some additional criteria. Two versions of each of these programs were studied: the original author's version and the version from whose original text the groups of operators realizing the author's additional criteria were removed. These versions of each program were then treated as independent programs and were studied independently.

A large number of spectra 40 channels long containing symmetric Gaussian peaks of amplitude a or spectra without peaks were constructed by computer. In each of these spectra each of the programs (the two versions) performed a search for peaks with a definite value of the parameter determining the search sensitivity threshold. In processing the spectra without peaks ($a=0$) for each of the programs the dependence of the conditional probability F for false discovery on the search sensitivity threshold and the values of these thresholds corresponding to $F=0.1$ were found. Then the dependences $D(a)$ for $F=0.1$ were found for all the programs by processing the model spectra containing peaks of different amplitude a. The results are shown in Fig. 4, where we give the dependences $D(a)$ for $F=0.1$ for all four programs in both versions. The results for the programs with additional criteria are shown by the open circles, and those for the programs without them are shown by the dark circles. We see from this figure that the dependences $D(a)$ are practically identical for the programs of Refs. 22 and 23 with and without additional criteria (although the quality of PEAK FIND of Ref. 22 with an additional criterion is somewhat better than without it), so that the quality of the peak identification by both versions of these programs is also practically the same. The use of additional criteria in the programs of Ref. 24 noticeably worsens their quality in almost the entire range of amplitudes of the identified peaks. This suggests that these programs are easily improved by removing from them the criteria which their authors introduced to improve the quality of the peak identification. Then the characteristics of these programs will be on the whole the same as those of the search programs of Refs. 22 and 23. These results indicate that care must be taken in making additions to algorithms of peak-search programs, because the effect opposite to that desired can be obtained. They give a good illustration of the well known rule that no transformations

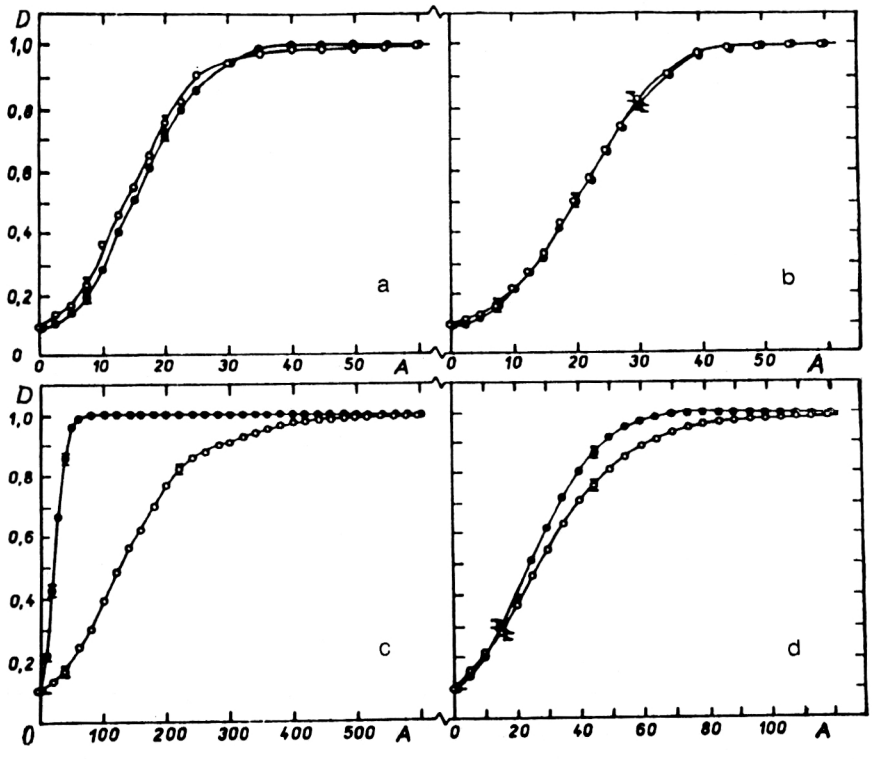

FIG. 4. Effect of additional identification criteria on the quality of peak-search programs. The points are the values (from the results of a model experiment) of the conditional probability D for true discovery of symmetric Gaussian peaks with an amplitude of a counts and a half-width of $b=5$ channels on a background $bkgd=100$ counts with value of the search sensitivity threshold corresponding to the conditional probability of false discovery $F=0.1$ for the peak-search programs (a) PEAK FIND, (b) PEAKS, (c) PEAK, and (d) PEAKI with additional criteria (○) and without them (●).

of the initial spectrum can increase the amount of information extracted from it if no additional *a priori* assumptions about the properties of this spectrum are made.

Algorithms for searching for isolated peaks are procedures for taking decisions for simple alternative hypotheses, the quality indicators of which are the conditional probability F of false discovery (significance of the criterion) and the conditional probability D of true discovery (power of the criterion). To estimate the quality of the algorithm or of a specific peak-search program it is necessary to find the dependence of the conditional probability D for true discovery (for $F=$const) on the parameters of the identified peaks and the background in the spectrum. To compare the quality of the different algorithms or peak-search programs with each other, it is necessary to compare the dependences of D found for them on the parameters of the identified peaks and the background in the spectrum for the same value of F, identical for all the programs compared. Then the quality of peak identification will be best for the program which has the largest value of D (for $F=$const and identical values of the peak and background parameters). This method of comparing automatic peak-search programs using the results of the model experiment with the technique of Ref. 20 for calculating the values of F and D characterizing them was suggested in Ref. 25. The comparison will be correct if the probability distribution function $\mu(z)$ in Eqs. (22) and (23) for different search programs is actually described by a symmetric Gaussian curve. In Ref. 26 for the example of several real programs it was shown that this assumption holds. The quality of the automatic peak-search programs described in Refs. 7, 19, 22–24, 27, and 28 was compared quantitatively in Ref. 29.

Each search program must contain at least two parameters which can be used to tune it: the sensitivity threshold and the (half-)width parameter (we recall that it determines the characteristic width of the maxima in the spectrum, which should be interpreted as peaks). The search sensitivity threshold is actually among the formal parameters of each of the programs compared, but unfortunately the width parameter does not enter as a formal parameter in all these programs (they are often crudely tuned by the authors to search for peaks with half-width of ~ 5 channels, characteristic of the γ spectra from semiconductor detectors). The width parameter is a formal parameter only for 4 of the 8 programs compared from Refs. 7, 19, 22–24, 27, and 28, and the programs in Ref. 29 could not be tuned in this parameter. Therefore, the width parameter was not varied, but instead the half-width of the peaks in the quasiexperimental spectra, the results of whose processing were used to compare the quality of the programs, was chosen to be equal to 5 channels, the most typical value of the half-width of the peaks in γ spectra from semiconductor detectors. By using each of the compared programs to analyze $\sim 10^5$ short quasiexperimental spectra without peaks, the dependences of F on the search sensitivity threshold (as in Fig. 1a) were found for all the programs studied, and for each program the value of the threshold corresponding to the conditional probability of false discovery $F=0.1$ was determined. The dependences of the conditional probability $D(a)$ of true discovery on the amplitude a of the identified peak were found for these values of the search sensitivity parameter corresponding to $F=0.1$ by using each of these programs to analyze another $\sim 10^5$ short quasiexperimental spectra with symmetric Gaussian peaks and different amplitudes. In the determi-

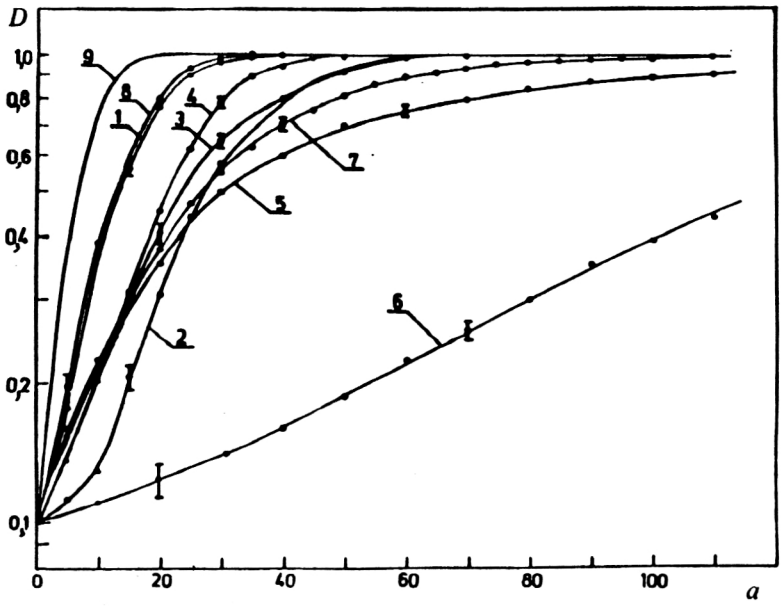

FIG. 5. Comparison of the quality of automatic peak-search programs for fixed value of the conditional probability of false discovery $F=0.1$: (*1*) PEAK FIND (Ref. 23); (*2*) SUCH (Ref. 29); (*3*) SUPIK (Ref. 7); (*4*) PEAKS (Ref. 24); (*5*) FIND (Ref. 30); (*6*) PEAK (Ref. 25); (*7*) PEAKI (Ref. 25); (*8*) ALISA (Ref. 19); (*9*) calculation of $D(a)$ for the model optimal search algorithm (31) with $\alpha=1$.

nation of F and D the constant component in the quasiexperimental spectra was 100 counts. The results obtained are shown in Fig. 5.

Comparing the dependences $D(a)$ in Fig. 5, we see that most of the programs are close in quality (curves *2–5*, *7*, and *8*). The low value of $D(a)$ for the program PEAKI (curve *6*) is related to the worsening of the quality of this program by the additional search criterion used in it (see Fig. 4c of Ref. 21). The largest value of D for all amplitudes a of the identified peaks is that of the ALISA program,[19] and consequently it is the best among the programs compared. Curve *9* gives the dependence $D_0(a)$ for the optimal model search algorithm, assuming that the spectrum is distorted only by statistical fluctuations of the background and that the parameters of the background and the identified peak are known exactly [Eq. (31) with $\alpha=1$]. Curves *8* and *9* in Fig. 5 are the same as curves *3* and *5* in Fig. 3, for which it was concluded that the quality of the ALISA program could be improved by using *a priori* information about the spectrum. Since ALISA is the best of the programs studied, to this conclusion we can add the statement that it is most likely that the use of additional *a priori* information about the spectrum is still the only way to improve the quality of peak identification using the model suboptimal search algorithm (32), (33).

Unfortunately, it is at present impossible to compare the quality of real programs for automatically searching for isolated peaks in multidimensional linear spectra, simply owing to the absence of such sufficiently complete programs even for processing two-dimensional spectra. Therefore, as an illustration we shall give only the quality indicators for the identification of isolated, symmetric Gaussian peaks on a constant background in two-dimensional spectra, calculated for the optimal model peak-search algorithm found in Ref. 17. The dependences given in Fig. 6 of the conditional probability $D(a,b)$ for true discovery of two-dimensional peaks on their amplitude a and half-width b were calculated using Eq. (74) for σ, the dispersion of the random deviation of the amplitude of a two-dimensional peak from its true value related to statistical fluctuations of the experimental data. Earlier we saw from Ref. 3 that the use of Eq. (48), analogous to (74), for σ in the case of one-dimensional spectra leads to values of $D(a)$ which are too low (the corresponding curve *1* is the lowest curve in this figure), so that the values of $D(a,b)$ in Fig. 6 should be interpreted only as lower limits on the values of the conditional probability for true discovery of two-dimensional peaks by the corresponding optimal model search algorithm.

The quality indicators for peak identification, the probabilities F and D, are conditional relative to the initial hypothesis about the presence of a peak in the spectrum, i.e., F is the probability of discovering, if known, that there is no peak in the spectrum (hypothesis H_0), and D is the probability of discovering, if known, that there is in fact a peak (hypothesis H_1). The conditional probabilities F and D are determined under different conditions, so that, as noted above, $F+D\neq 1$, which is unusual and often inconvenient. The users of these programs are in fact interested in probabilities, conditionless relative to the initial hypothesis of the presence of a peak in the spectrum, that the peak found in a search is a true peak P_t or a false one P_f. In other words, in analyzing the results of measurements it is necessary to determine the probability of the existence of some line if *a priori* its 100% presence (or absence) in the spectrum is not stipulated. This problem is easily solved[30] if the conditional probabilities F and D are known for each peak discovered and if the result of the experiment is taken to be not the directly measured spectrum, but the result of its processing.

Let the result of a measurement be event A, the discovery of a peak. Then $P_f=P(H_0|A)$ is the probability that event A is a consequence of the realization of hypothesis H_0 (no peak in the spectrum); $P_t=P(H_1|A)$ is the probability that event A is a consequence of the realization of hypothesis H_1 (a peak in the spectrum); $F=P(A|H_0)$ is

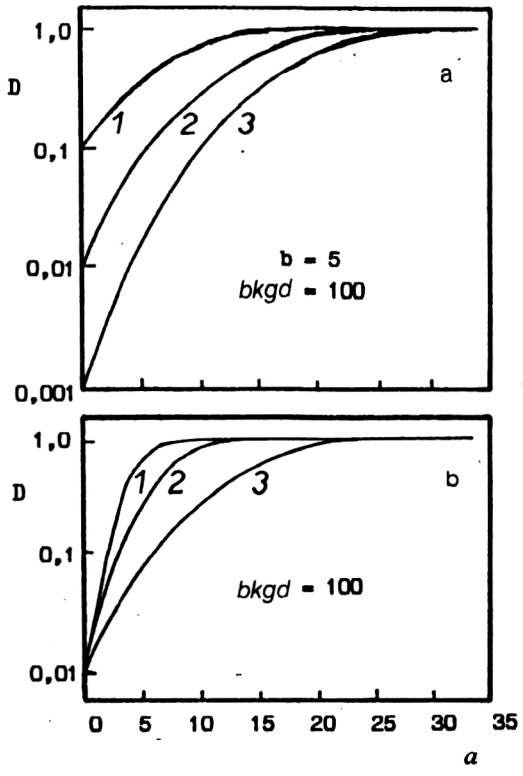

FIG. 6. Quality indicators of the model optimal algorithm for searching for isolated peaks in two-dimensional spectra (Ref. 17): dependence of the conditional probability $D(a,b)$ of true discovery of symmetric Gaussian peaks on their amplitude a and half-width b on the constant background $bkgd=100$ counts. (a) Curves *1*, *2*, and *3* are the calculation for $b=5$ and $F=0.1$, 0.01, and 0.001, respectively; (b) curves *1*, *2*, and *3* are the calculation for $F=0.01$ for $b=15$, 10, and 5, respectively. All the calculations were carried out using the value of σ from Eq. (72).

the probability for event A with the condition that hypothesis H_0 is true; $D=P(A|H_1)$ is the probability of event A with the condition that hypothesis H_1 is true, and from the well known Bayes formula we have

$$P(H_i|A)=\frac{P(H_i)P(A|H_i)}{P(H_0)P(A|H_0)+P(H_1)P(A|H_1)}. \quad (40)$$

From this we find

$$P_f=\left(1+\kappa\frac{D}{F}\right)^{-1}; \quad P_t=\left(1+\kappa^{-1}\frac{F}{D}\right)^{-1};$$

$$P_t+P_f=1, \quad (41)$$

where $\kappa=P(H_1)/P(H_0)$; $P(H_0)$ and $P(H_1)$ are the *a priori* probabilities of hypotheses H_0 and H_1; P_t and P_f are the conditionless *a posteriori* probabilities of true and false discovery.

The coefficient κ in the equations for P_t and P_f takes into account the additional *a priori* information about the spectrum. The presence of this coefficient in the equations for the conditionless probabilities quantitatively reflects the fundamentally important fact that the final decision about the presence or absence of a peak is made using all the available *a priori* information about the spectrum, including subjective information like the experience of the experimentalist. The question of choosing a specific numerical value for κ is complicated; it has hardly been studied at all and depends on many factors, including the type and goals of the experiment being carried out. However, in most cases it can be assumed as an initial approximation that for a researcher processing only a single spectrum and not possessing additional information about the experiment, hypotheses H_0 and H_1 are equally probable and $\kappa=1$. Since F and D always satisfy $0<F\leqslant D<1$, for $\kappa=1$ for the conditionless probabilities P_t and P_f we have

$$0<P_f\leqslant 0.5; \quad 0.5\leqslant P_t<1, \quad (42)$$

i.e., from the results of a single measurement the fact that a peak exists can be proved with nearly 100% probability, but the fact that a peak is absent is very difficult to prove without using additional information. The reliability of peak identification can be raised by repeated measurements, so that it is necessary to be able to calculate the *a posteriori* probabilities of true and false discovery using the results of an arbitrary series of experiments.

Suppose that a series of n identical measurements of the same spectrum, for example, a γ spectrum, has been carried out, and a peak, an event A^{in}, has been discovered in i measurements during the processing. Then by analogy with the derivation of Eq. (40) we have the following: $P_f^{in}=P(H_0|A^{in})$ and $P_t^{in}=P(H_1|A^{in})$ are the probabilities for hypothesis H_0 (no peak) and hypothesis H_1 (a peak) with the condition that the event A^{in} was observed; $F_{in}=P(A^{in}|H_0)$ and $D_{in}=P(A^{in}|H_1)$ are the probabilities of realizing event A_{in} assuming that hypotheses H_0 and H_1 are true. Now, using the Bayes formula with this notation, we have

$$P_f^{in}=\left(1+\kappa\frac{D_{in}}{F_{in}}\right)^{-1}; \quad P_t^{in}=\left(1+\kappa^{-1}\frac{F_{in}}{D_{in}}\right)^{-1}, \quad (43)$$

where $\kappa=P(H_1)/P(H_0)$ is the ratio of the *a priori* probabilities of hypotheses H_0 and H_1.

The values of F_{in} and D_{in} can be found[30] if the peak-identification quality indicators, the conditional probabilities F and D, are known for each of the processed spectra:

$$F_{in}=\frac{n!}{i!(n-i)!}\prod_{k=1}^{i}F_k\prod_{l=i+1}^{n}(1-F_l), \quad (44)$$

$$D_{in}=\frac{n!}{i!(n-i)!}\prod_{k=1}^{i}D_k\prod_{l=i+1}^{n}(1-D_l), \quad (45)$$

where the subscript k on F and D labels the spectra in which a peak was found and l labels spectra in which a peak was not found.

In the classical theory of testing statistical hypotheses the conditional probabilities F_{in} and D_{in} have the following meaning. The value of F_{in} is the probability of taking a false peak for a true one if the peak is considered to have been discovered when it is found in i out of n measurements. The value of D_{in} is the probability of repeating the statistical experiment, i.e., if the peak exists, the probability of finding it i times in carrying out n measurements is D_{in}. The probabilities F_{in} and D_{in} can also be used to make a decision about the presence of a peak if n spectra are pro-

cessed together, or if a single spectrum is processed but using several (n) independent peak-identification algorithms. Then F_{in} and D_{in} will be quantitative characteristics of the quality of the processing methods used (a first attempt to test the suitability of using several search programs simultaneously in processing the same spectrum is described in Ref. 31).

Finally, let us discuss the choice of the specific numerical value of the search sensitivity threshold l for which it is appropriate to search for peaks using automatic programs. However, since it is impossible to meaningfully interpret the value of l (it is related to the specific features of the programming of the search algorithm), we shall discuss not l itself, but the corresponding value of the conditional probability F of false discovery, since F and l have a one-to-one correspondence (see Fig. 1a, for example). This change also ensures uniformity in the description of the properties of different search programs, since the same value of F corresponds for each program to a particular numerical value of l depending on the features of the algorithm, the approximations used in it, details of the programming, etc. We shall assume that it is appropriate to search for peaks at the value of F (and the corresponding sensitivity threshold l) typical of a human operator who identifies peaks by hand by visual analysis of the processed spectrum (as occurs in automated processing programs). Actually, in this case the user, i.e., the operator himself, must certainly be satisfied with the results of the search, so that here there is the optimal relation between the search sensitivity and the reliability of the results corresponding to the requirements on the spectrum processing in experiments of a given type. In order to find the desired value of F it is necessary to study the quantitative characteristics of the human operator in dealing with the problem of peak identification. These can be described by the same conditional probabilities F and D as for automatic programs.

Such a study has been carried out in Ref. 32 using a repeated model experiment by a technique completely analogous to that of Ref. 20, which was used earlier to compare the quality of automatic programs.[21,29] A large number of short quasiexperimental spectra with or without peaks were modeled (the same spectra as in Refs. 20, 21, and 29), but the peaks in these spectra were identified not by a search program, but by a human operator who visually analyzed these spectra on a computer screen. The program package ATOS (automatic testing of spectrum processing) was created to solve this problem. It models spectra with a quasiexperimental statistical length of 50 channels, each containing either one symmetric Gaussian peak with a half-width of 5 channels on a constant background, or only background. The sequence of modeled spectra with or without a peak, the amplitude, and the position of the peak maximum in the spectra with a peak are formed randomly, so that the human operator being tested does not know either the ordering of the spectra with and without a peak, or the position and amplitude a of a possible peak in these spectra. The background in all cases was taken to be constant and equal to 100 counts. Since in searching for peaks using computer programs the

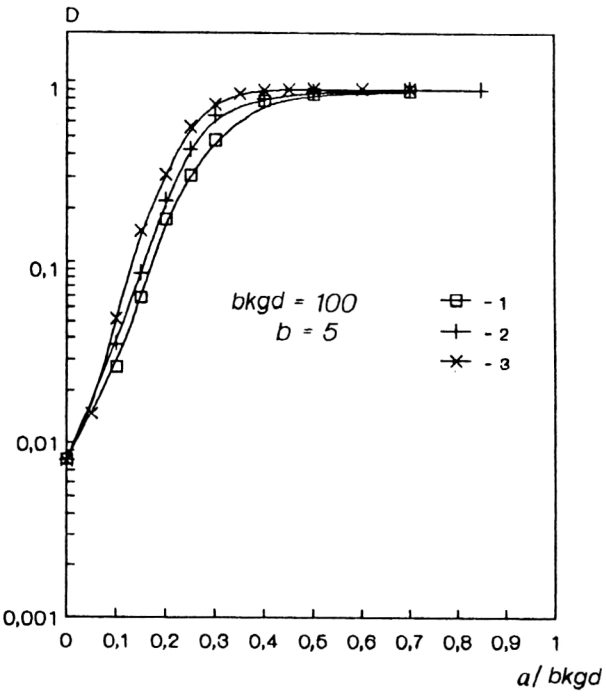

FIG. 7. Dependences of the conditional probability $D(a)$ of true discovery of a symmetric Gaussian peak with half-width $b=5$ channels and amplitude a on a constant background ($bkgd=100$ counts) for professional operators [$D_{op}(a)$, curves *2* and *3*] and for an automatic program (Ref. 19) realizing the model suboptimal search algorithm (32) [$D_{pr}(a)$, curve *1*].

shape of the peak is assumed to be known, the operator being tested is also given information about the line shape undistorted by statistical fluctuations, displayed as a graph on the computer screen during the processing. The calculation of the values of F and $D(a)$ for each operator tested is done after the operator finishes searching for peaks in all the quasiexperimental spectra assigned to him. Observation of the test while it was being carried out was not allowed.

The quality of the peak search by visual analysis of the spectra by a human significantly depends on his previous experience, and therefore two types of operator were tested: professionals with a great deal of experience in spectrum processing, and nonprofessionals (schoolchildren, liberal-arts students, first-year physics students) who first became acquainted with the problem of peak identification before the start of the tests from a specially prepared information package.

An example of the results of the tests for a group of professionals is shown in Fig. 7. We see that the quality of the peak search by different professional operators is nearly identical [the corresponding dependences $D_{op}(a)$ are close]. In addition, the value of the conditional probability F for false discovery was nearly identical for all professionals: $F=0.012\pm0.004$. This can be interpreted as a quantitative manifestation of the fact that the requirements imposed on peak identification by professional operators and their possibility of such identification are objectively close, and $F\sim0.01$ corresponds to "good" quality of the peak

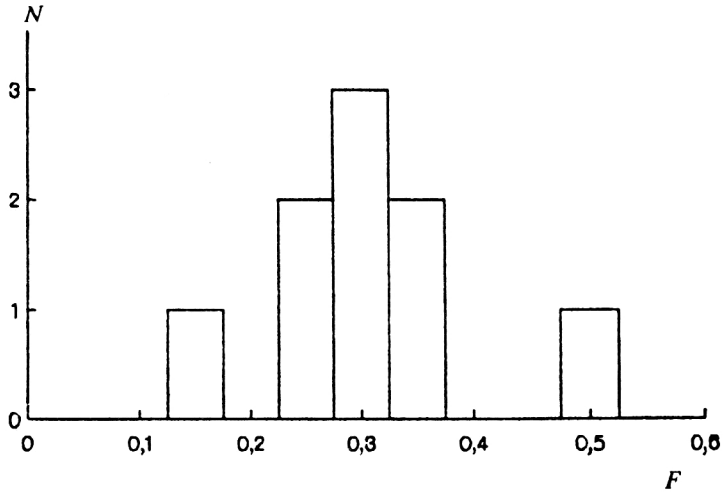

FIG. 8. Histogram of the distribution of values of the conditional probability F of false discovery constructed from the results of tests by nonprofessional operators. The testing was carried out[32] using a repeated model experiment.

search. This leads to the conclusion that an automatic peak search in linear spectra by means of computer programs is most appropriately carried out at the value of the search sensitivity threshold corresponding to the conditional probability for false discovery $F \sim 0.1$. In Fig. 7 we also show the dependence $D_{pr}(a)$ for $F \sim 0.01$ for the program of Ref. 19 realizing the suboptimal model search algorithm (32). We see that for all amplitudes $a > 0$ of the identified peaks $D_{pr}(a) > D_{op}(a)$, while in the amplitude range $0.1 \lesssim a/bkgd \lesssim 0.2$ (*bkgd* is the amplitude of the background below the peak) $D_{pr}(a) \cong 2D_{op}(a)$, i.e., the quality of the peak search by an automatic program can be better than that of professional operators, in some cases by about a factor of two. This implies that automatic programs may not be inferior to humans in searching for isolated peaks in linear spectra if the human and the program are provided with identical *a priori* information.[33] Moreover, the fact that the values of $D(a) = D_{pr}(a)$ for an automatic program exceed those of $D(a) = D_{op}(a)$ for professional operators when $F = \text{const}$ implies that the machine program uses the *a priori* available information on the spectrum (the shape and width of identified peaks) more efficiently than a human does.

The results of the tests for nonprofessionals differ strongly from those for professionals. In Fig. 8 we show the values of F for tested nonprofessionals in the form of a histogram, from which it is clear that there is a large spread in the values found during the tests, and that the results are grouped around $F \cong 0.3$. The large spread in these values may be related to the difficulties nonprofessionals experience in first realizing exactly what a peak is. And the fact that nonprofessionals nonuniformly classify spectra which *a priori* are without peaks as spectra with and without peaks requires additional study. This might be explained on the basis of human reflexes. According to Ref. 34, in the absence of a preferred choice between two alternative hypotheses with unclear (undefined) criteria, a human will make a choice on the basis of personal preferences, but not with equal probabilities. The probability of choosing one of the variants is 0.617, and that of choosing the other is $1 - 0.617 = 0.383$. On the basis of the closeness of the quantitative value $F \sim 0.3$ and the predicted[34] value of the probability of a choice in the case of undefined criteria (0.383), it can be concluded that nonprofessional operators, when making a decision about the presence (absence) of a peak in spectra which definitely do not contain a peak, also obey the laws of human reflexes and make a choice on the basis of personal preferences which do not follow directly from the experimental data. In this case the question of the interpretation of the numerical value of F which is obtained lies outside the scope of physics or mathematics proper and belongs to human psychology. On the other hand, this implies that a set of model spectra with or without peaks and perturbed by statistical fluctuations can be viewed as a uniquely defined and easily reproduceable simple mathematical model of an object convenient for quantitative studies of the characteristics of visual image recognition by humans when random noise is present. A set of computer programs (like ATOS, Ref. 32) for generating such spectra, displaying them on a computer screen, enabling a dialog, and presenting the results of the testing can serve as the experimental basis for research into human psychology. Such research can be of both fundamental and applied value from the viewpoint of studying the effect of the human factor on the reliability of particularly complicated and accident-prone technical systems. The study of this problem is extremely important at present.[35]

4. POTENTIAL ERRORS IN DETERMINING THE PEAK PARAMETERS AND OPTIMAL CRITERIA FOR INTERRUPTING THE FIT

In order to overcome the difficulties in the determination of the peak parameters from the results of fitting the processed spectrum associated with unsoundness of the estimates, the authors of Ref. 36 proposed that the results of a repeated model experiment carried out by computer be used. This would make it possible to establish the relation between the parameters of the (deterministic) mathematical model of the spectrum and the errors in determining them, which arise only from the statistical nature of the

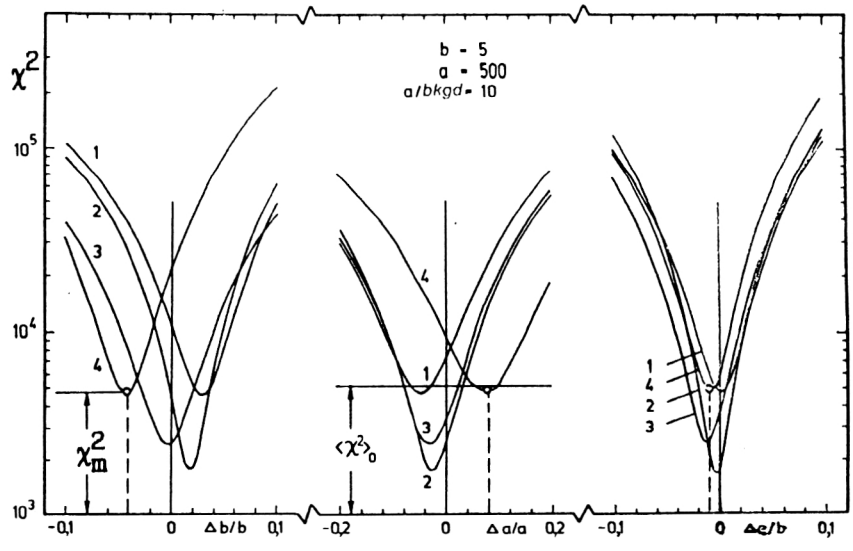

FIG. 9. Shape of the χ^2 minima for four quasiexperimental spectra with a single symmetric Gaussian peak on a constant background, constructed for identical values of the parameters of the mathematical model of the spectrum but with different quasiexperimental statistics. Here *a*, *b*, and *bkgd* are the values of the parameters specified in the construction (the peak amplitude and half-width and the background amplitude); Δa, Δb, and Δe are the deviations in the amplitude, the half-width, and the peak position from their specified values. The χ^2 functional is defined in (46).

experimental data, i.e., to find the potential errors in determining the peak parameters. Empirical formulas for calculating the potential errors in determining the parameters of isolated peaks and peaks from a doublet as functions of the parameters of these peaks and the parameters of the spectrum background are given in Ref. 37 (such functions are referred to as physical reliability functions, PRFs, in Ref. 36).

When we speak of potential errors we have in mind the errors found as a result of processing a single measured spectrum (a single statistical realization of the experiment), but including the spread of the results also over all the other possible statistical realizations of these data (as though it were possible to repeat the experiment an arbitrary number of times under identical conditions and then find the errors in the values of the peak parameters by processing all such spectra). This statement is illustrated in Fig. 9, where we show the form of the χ^2 minimum for four (curves *1–4*) model quasiexperimental spectra containing an isolated peak on a constant background, which were constructed for identical parameters of the mathematical model of the spectra (the peak amplitudes and half-width and the ratio of the peak amplitude to the background amplitude), but each time with different sequences of statistical fluctuations in each channel. For the fourth spectrum (curve *4*) in this figure the circle shows the position of the χ^2 minimum. Along the horizontal axis in relative units are the deviations of the half-width, the amplitude, and the peak position from their true values specified in the construction. The spread in the positions of the minima for all the fragments in Fig. 9 is due only to the effect of the specific distribution of the statistical fluctuations of the experimental values, i.e., it is related only to the different statistical realization of the experiment carried out under identical conditions (in the model experiment "identical experimental conditions" are ensured by using identical spectrum parameters specified in the construction, and different statistical realizations of the experiment correspond to a different set of random fluctuations of the number of counts in each channel). The solid vertical lines show the values of the peak parameters specified in the construction of the model spectra, and the dashed lines are the values of these parameters found in the fit for one of the spectra (curve *4*). The error in determining the parameters related to the fitting technique is smaller than the size of the circles. The distance between the dashed and solid lines is equal to the value of the shift of each of the parameters from its true value, which is a consequence of only the specific statistical realization of the experiment. The rms values of these shifts found from the repeated model experiment done by computer are taken[37] to be equal to the values of the corresponding potential errors (the PRFs) for a given multiplet with given parameters of its mathematical model, i.e., the errors arising only from the statistical reliability of the experimental data. When the errors are estimated using the PRFs, their values for the peak parameters in all spectra with the same statistical reliability will be identical. This approach to estimating the errors in the approximation includes not only a given distribution of statistical fluctuations obtained in a particular experiment (as occurs in estimating the errors using the diagonal elements of the inverse linear iteration matrix), but also other possible statistical realizations, and in this sense it contains integrated information averaged over the statistical ensemble of spectra with the same statistical reliability but with different sequences of fluctuations of the experimental values. The potential errors are also reliable estimates of the errors, since "the reliability of the estimate is manifested in the statistics as an ensemble property in analyzing series of measurements of the same type" (Ref. 5).

By studying the curves in Fig. 9 we can formulate a rule for the optimal interruption of the iteration process of the fit during the spectrum processing. In fact, in estimating the errors in the results of the fit by calculating the potential errors using the PRFs, the fit itself need not at all be performed in such a way that values of the parameters corresponding to the χ^2 minimum (for example, for curve *2*) are found, since the information obtained here can turn out to be redundant. It is sufficient to carry out the fit until the value $\chi^2 = \langle\chi^2\rangle_0$ is reached, which is the average of the

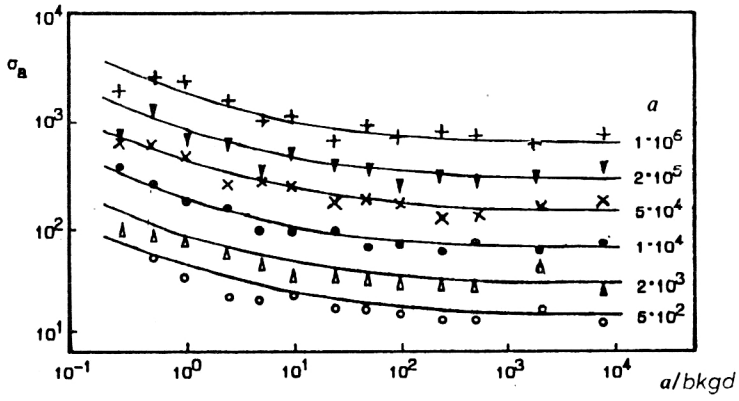

FIG. 10. Potential error σ_a in determining the amplitude a of an isolated symmetric Gaussian peak on a background of amplitude *bkgd* as a function of the ratio *a/bkgd*. The points are from a model experiment, and the curves were calculated using Eq. (48).

minimum values $\chi^2=\chi^2_m$ over the ensemble of spectra measured under identical conditions and having the same statistical reliability. Here the limiting value $\chi^2=\langle\chi^2\rangle_0$, which must be reached in the course of the fit, must be known in the form of a function depending on the parameters of the peaks belonging to the processed multiplet and the parameters of the background. Such functions have been found in Ref. 36 for isolated peaks and doublets, but assuming that the quadratic functional maximized in the fit is defined in a somewhat nonstandard manner:

$$\chi^2=\frac{10^6}{k}\sum_{i=1}^{k}\frac{(y(i)-z(i))^2}{z(i)+1}, \tag{46}$$

where $y(i)$ is the experimental value in channel i (with the background subtracted), and $z(i)$ is the corresponding calculated value for the mathematical model used for the processed spectrum. This restricts the region of applicability of the results. However, the use of Eq. (46) for the minimizing functional may be preferable in some cases.[8]

Below, we give the expressions for the potential errors in determining the parameters of isolated peaks and doublets (the PRFs) found by using the repeated model experiment.[37]

For isolated peaks:

$$\sigma_x=0.5(1+2(a/bkgd)^{-1/2})(b/a)^{1/2}, \tag{47}$$

$$\sigma_a=1.4(1+2(a/bkgd)^{-1/2})(a/b)^{1/2}, \tag{48}$$

$$\sigma_b=0.8(1+2(a/bkgd)^{-1/2})(b/a)^{1/2}, \tag{49}$$

$$\sigma_s=1.4(s(1+2.83(a/bkgd)^{-1}))^{1/2}, \tag{50}$$

$$\langle\chi^2\rangle_0=2(1+12(a/bkgd)^{-1})(1+10^6(1+a)^{-1}), \tag{51}$$

where σ_x, σ_a, σ_b, and σ_s are the potential errors in determining the positions x, the amplitudes a, the half-widths b, and the areas s of isolated peaks of symmetric Gaussian shape on a constant background *bkgd*. Equation (51) is used to calculate the limiting value of χ^2 [determined by (46)] which must be reached in the fit (the optimal rule for terminating the iterations). For the examples of Fig. 9 the value of $\langle\chi^2\rangle_0$ is shown by a horizontal line. The relative accuracy of calculating the errors according to Eqs. (47)–(50) is ~20%. They are obtained in the parameter ranges

$$10^{-1}\leqslant a/bkgd\leqslant 10^4;\quad 10^2\leqslant a\leqslant 10^6;\quad 2\leqslant b\leqslant 20.$$

For doublets:

$$\sigma_{x1}=1.34\left(\frac{b}{az}\right)^{1/2}\exp\left[1.3\left(\frac{a}{bkgd}\right)^{-0.4}\right]d^{0.4(a/bkgd)^{-1/4}-3.2}, \tag{52}$$

$$\sigma_{x2}=2.24\left(\frac{b}{a}\right)^{1/2}\exp[5\exp(-4.5d)]\left[1+\left(\frac{a}{bkgd}\right)^{-1}+\frac{1}{4}\ln(0.05+z)\right], \tag{53}$$

$$\sigma_{a1}=\sigma_{a2}=3.36\left(\frac{a}{b}\right)^{1/2}\left(\frac{a}{bkgd}\right)^{-1/3}d^{-3}, \tag{54}$$

$$\langle\chi^2\rangle_0=\frac{5\cdot 10^4}{a}\exp\left(\frac{37.3}{6.2+\ln(a/bkgd)}+\frac{2.4\cdot 10^3}{a^2/bkgd}\right), \tag{55}$$

where $a=a1+a2$, $d=|x2-x1|/b$, $z=a1/a2$; $a1$, $a2$, $x1$, and $x2$ are the amplitudes and positions of the smallest (1) and largest (2) peak in the doublet; b is the half-width of these peaks, and *bkgd* is the background amplitude. The empirical formulas (52)–(56) were obtained for the following parameter ranges:

$$0.125\leqslant a1/a2\leqslant 1.0;\quad 0.4\leqslant a/bkgd\leqslant 50;$$

$$0.2\leqslant d\leqslant 1.0;\quad 2500\leqslant a\leqslant 100\,000.$$

We note that Eq. (55) for the limiting value $\langle\chi^2\rangle_0$ which must be obtained during the fit in determining the parameters of the peaks in the doublet (the optimal rule for terminating the iterations) depends only on the total amplitude a and on its ratio to the background amplitude, *a/bkgd*. The value of $\langle\chi^2\rangle_0$ does not depend on either the amplitudes $a1$ and $a2$ or on the distance $|x2-x1|$ between the peaks making up the doublet. The degree to which the data of the model experiment and the calculations using the empirical formulas given above agree is shown by the examples in Figs. 10 and 11 [σ_a for isolated peaks and $\langle\chi^2\rangle_0$ for doublets].

It is interesting to compare the potential errors calculated using the empirical formulas with the results of an actual experiment. For this, repeated (11 times) measurements were made[37] of the spectrum of γ rays from a stan-

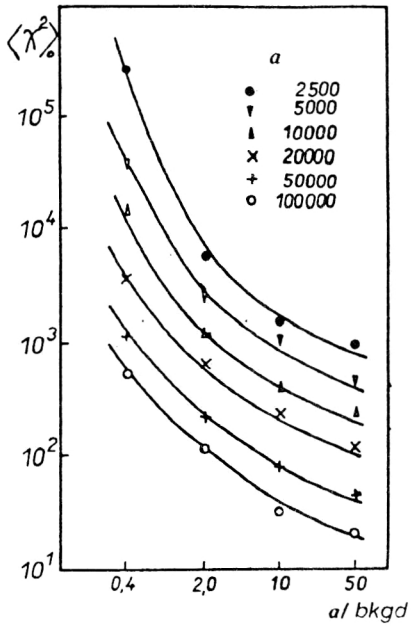

FIG. 11. Limiting value $\langle\chi^2\rangle_0$ which needs to be reached in fitting doublets. Here $a=a1+a2$, $a1$ and $a2$ are the amplitudes of the peaks forming the doublet, and *bkgd* is the amplitude of the true background. The points are from a model experiment, and the curves were calculated using Eq. (55).

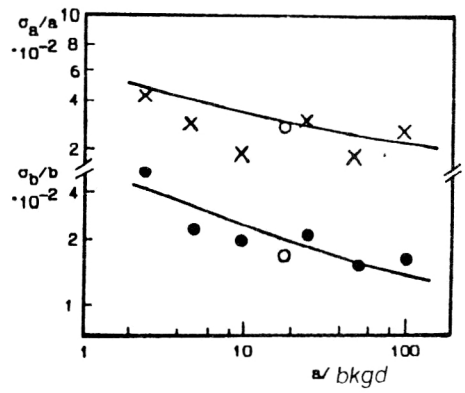

FIG. 12. Comparison of the calculation using Eqs. (48) and (49) (curves) of the potential errors in determining the amplitude $a(\sigma_a)$ and half-width $b(\sigma_b)$ of an isolated peak with the results of a model experiment (×,●) and an actual experiment (○). In the actual experiment $a=1018$ counts, $b=5.8$ channels, $t_1=3$ min, and $t_2=2$ min. The calculation was done for $a=1000$ counts and $b=5$ channels.

dard ^{22}Na source in a typical spectrometer with a Ge(Li) detector and a TRIDAC-C multichannel pulse-height analyzer. Each spectrum was measured for fixed experimental geometry in two stages: without digital shift of the channels during time t_1 and with digital shift of the channels during time t_2. The peak from the isolated total γ absorption line at 1275 keV and the Compton plateau from the same line were summed instrumentally in this way. In the case of the doublet the measurements were made in three stages: the background (measured without digital shift of the channels in time t_1), the first peak of the doublet (digital shift of k channels, time t_2), and the second peak of the doublet (digital shift of $k+\Delta k$ channels, time t_3). The experimental data obtained were processed like those for the earlier model quasiexperimental spectra. The calculations based on the empirical formulas are compared with the results of processing these measurements in Figs. 12 and 13. We see from these figures that the expressions from Ref. 37 can be used successfully to analyze real spectra. In Fig. 14 the experimental (from the results of 11 successive identical measurements) positions of the same isolated peak (the points in the figure) are compared with each other and with their potential errors (error bars). We see that the amplitude of the instrumental dynamical instabilities at the location of the peak is more than twice as large as the errors arising from the statistics, and therefore for σ_x, σ_{x1}, and σ_{x2} they are not compared with the results from analyzing the experimental data. It also follows from this that the instrumental instabilities manifested in the measurement with high energy resolution of the γ spectra do not always allow the complete realization of the statistical accuracy of the experiment attainable in such cases.

Above, we gave the expressions found empirically for the potential errors in determining the parameters of isolated peaks and peaks belonging to a doublet by analyzing the results of a repeated model experiment.[36,37] This is quite complicated work, and for multiplets with $m>2$ it is extremely difficult to realize. Therefore, in Ref. 38 the empirical results of Ref. 37 were used together with the maximum-likelihood method with the linear stochastic spectrum model[16] to find expressions for calculating the potential errors in determining the parameters of peaks belonging to a multiplet of arbitrary complexity.

We write the likelihood function for the part of the spectrum containing a multiplet of m peaks as[38] (the notation is the same as in Sec. 3)

$$l=p_m(\mathbf{y}-\mathbf{u}_s)\left(\frac{N_0/\sigma_m^2}{2\mathbf{u}_m\mathbf{R}_m+N_0/\sigma_m^2}\right)^{1/2}$$
$$\times\exp\left[\frac{2[(\mathbf{y}-\mathbf{u}_s)\mathbf{R}_m]^2}{N_0(2\mathbf{u}_m\mathbf{R}_m+N_0/\sigma_m^2)}\right]. \qquad (56)$$

In order that l depend only on the true parameters of the spectrum, we assume that the squared dispersion σ_i^2 of the

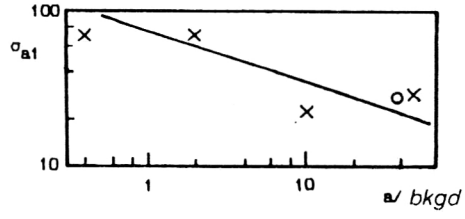

FIG. 13. Comparison of the calculation using (54) (curve) of the potential error σ_{a1} ($=\sigma_{a2}$) in determining the amplitude $a1$ of the smallest peak in a doublet with the results of a model experiment (×) and an actual experiment (○). In the actual experiment the average values are $a=a1+a2=2499$ counts, $a1/a2=0.667$, $d=1.03$, and $b=5.8$ channels. The calculation is for $a=a1+a2=2500$ counts, $a1/a2=0.5$, $d=1.0$, and $b=5$ channels; *bkgd* is the amplitude of the true background.

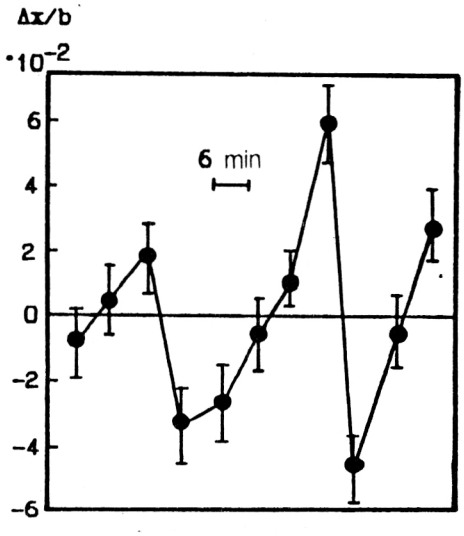

FIG. 14. Positions of the same isolated peak found in a series of 11 successive identical measurements of the γ-ray spectrum in a spectrometer with Ge(Li) detector. The horizontal line is the peak position averaged over the results of all the measurements, and $\Delta x/b$ is the deviation from the average value in units of the half-width. The error bars were calculated using (47). The average parameters of the peak in the experimental spectrum are: amplitude $a=1018$ counts, half-width $b=5.8$ channels, and ratio of the peak amplitude to the background amplitude $a/bkgd=18.5$.

random parameter a_i is proportional to the true amplitude a_{0i} of the ith peak (ν is the coefficient of proportionality):

$$\sigma_i^2=\nu a_{0i}. \tag{57}$$

For definiteness we assume that the function $u_i(x)$ describing the shape of an isolated peak is a symmetric Gaussian with a half-width of b channels:

$$u_i(x)=\exp[-\beta(x-\lambda_i)^2/b^2], \tag{58}$$

and that $\beta=4\ln 2=2.772\,588\,7\ldots$. Now the likelihood function l is a function of $2m+1$ parameters: $a_{01},\ldots,a_{0m}$, $\lambda_1,\ldots,\lambda_m$, and N_0. For the logarithmic likelihood function L, from (56)–(58) we have

$$L(a_{01},\ldots,a_{0m},\lambda_1,\ldots,\lambda_m)=L_1+L_2+L_3, \tag{59}$$

where

$$L_1=-\frac{1}{N_0}\left(\mathbf{y}-\sum_{i=1}^{m}a_{0i}\mathbf{u}_i\right)^2, \tag{60}$$

$$L_2=-\frac{1}{2}\sum_{i=1}^{m}\ln[1+fa_{0i}(\mathbf{u}_i\mathbf{R}_i)], \tag{61}$$

$$L_3=\frac{f}{N_0}\sum_{i=1}^{m}a_{0i}\left[\left(\mathbf{y}-\sum_{j=1}^{m}a_{0j}\mathbf{u}_j\right)\mathbf{R}_i\right]^2 \times[1+fa_{0i}(\mathbf{u}_i\mathbf{R}_i)]^{-1}, \tag{62}$$

$$f=2\nu/N_0. \tag{63}$$

In the expressions for L we have dropped the terms independent of a_{0i} and λ_i.

Estimates of the mathematical expectation value of the errors in determining the multiplet parameters are sought following the standard procedure.[39] We define the matrix

$$\mathbf{A}=\begin{vmatrix}\left\langle-\frac{\partial^2 L}{\partial a_{0i}\partial a_{0j}}\right\rangle & \left\langle-\frac{\partial^2 L}{\partial a_{0i}\partial\lambda_j}\right\rangle\\ \left\langle-\frac{\partial^2 L}{\partial\lambda_j\partial a_{0i}}\right\rangle & \left\langle-\frac{\partial^2 L}{\partial\lambda_i\partial\lambda_j}\right\rangle\end{vmatrix}=\begin{vmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{vmatrix}, \tag{64}$$

where A_{ij} are $m\times m$ block matrices, and $\langle\ldots\rangle$ denotes averaging of the derivatives over the noise and random amplitudes a_i of the linear stochastic spectrum model. The matrix $\mathbf{B}$, the inverse of the matrix $\mathbf{A}$, gives the correlation moments of the effective estimates of the positions and amplitudes of all the peaks ($\mathbf{B}_{11}$ for the amplitudes and $\mathbf{B}_{22}$ for the positions). Using the inversion rule for block matrices

$$\mathbf{B}_{11}=[\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}],\quad \mathbf{B}_{22}=[\mathbf{A}_{22}-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}], \tag{65}$$

all the calculations involved in obtaining the estimates can be reduced to operations on $m\times m$ matrices. In practice the number of peaks in a multiplet is small ($m<5$), so that there are no computational difficulties in the problem, especially when a computer is used. There is some awkwardness in the calculations associated with the need to refer back to the recursion relations (13), which are used to specify the resolving functions $\mathbf{R}_m$.

The elements of the matrix $\mathbf{A}$ are second partial derivatives averaged over the noise and ensemble of random functions $a_i\mathbf{u}_i$. The derivatives can be calculated by direct differentiation of L_1, L_2, and L_3 from (60)–(62). The results are averaged using the fact that for any smooth vector function $\mathbf{s}$ the average of the following scalar product is equal to zero:

$$\left\langle\left(\mathbf{y}-\mathbf{bkgd}-\sum_{i=1}^{m}a_{0i}\mathbf{u}_i\right)\mathbf{s}\right\rangle=0. \tag{66}$$

Taking into account (66), fairly simple but cumbersome expressions can be obtained for the second partial derivatives of L. These expressions specify a complete set of rules for calculating the potential errors in determining the parameters of the peaks in a multiplet of any complexity using the linear stochastic spectrum model.[16] However, the coefficient ν introduced in (57) is still unknown. To find ν, we write out explicitly the expressions for the potential errors in the simplest case of determining the parameters of isolated peaks and compare them with the results of the model experiment and with the empirical formulas from Ref. 37. Then the coefficient ν is found as a fitted parameter in this comparison.

For the case of a singlet ($m=1$) the matrix $\mathbf{A}$ has the form

$$\mathbf{A}=\begin{vmatrix}\langle-\partial^2 L/\partial a^2\rangle & 0\\ 0 & \langle-\partial^2 L/\partial\lambda^2\rangle\end{vmatrix}. \tag{67}$$

Therefore, the errors σ_a and σ_λ in determining the peak amplitude a and position λ are given by

$$\sigma_a^2=-(\partial^2 L/\partial a^2)^{-1};\quad \sigma_\lambda^2=-(\partial^2 L/\partial\lambda^2)^{-1}. \tag{68}$$

Direct calculations give

$$\sigma_a^2=N_0[1+fa(\mathbf{uR})]\{2\mathbf{u}^2[1+(\nu/N_0)\mathbf{u}^2 \times(1+fa(\mathbf{uR}))]\}. \tag{69}$$

Using the fact that for a symmetric Gaussian $\mathbf{u}^2=b(\pi/2\beta)^{1/2}$, and setting

$$N_0=2r\,bkgd,\quad \nu=q[b(\pi/2\beta)]^{-1}, \tag{70}$$

from (69) we find (r and q are new parameters)

$$\sigma_a=a\left\{\frac{2^{1/2}}{s(a,b)}\left(q+\frac{r}{a/bkgd}\right)\left[1-\frac{q^2 2^{1/2}}{s(a,b)} \times\left(q+\frac{r}{a/bkgd}\right)^{-1}\right]\right\}^{1/2}, \tag{71}$$

where $s(a,b)$ is the area of a Gaussian peak with amplitude a and half-width b. Always $s(a,b)\gg 1$, so that instead of (71) we can write the approximate expression [for $s(a,b)>50$ it differs from (71) by less than 1%]

$$\sigma_a\cong a\left[\left(q+\frac{r}{s(a,b)}\right)\frac{2^{1/2}}{s(a,b)}\right]^{1/2}. \tag{72}$$

Comparing the calculation using (72) with the results of the model experiment of Ref. 37 (these are shown in Fig. 10 for σ_a), for the parameters q and r we find

$$q=1.4\pm 0.3;\quad r=9\pm 2.$$

The calculations of the values of σ_a using the empirical formula (48) and Eq. (72) just found with these values of q and r describe the results of the model experiment almost equally well and in this sense are equivalent.

Therefore, all the parameters in the general expressions for the errors are completely determined and now they are suitable for calculating the potential errors in the parameters of multiplets of any order.

For the error σ_λ in determining the position λ of an isolated peak we find by analogy[38]

$$\sigma_\lambda=\left(\frac{(2/\pi\beta)^{1/2}br}{a(a/bkgd)}\right)^{1/2}. \tag{73}$$

This expression differs considerably from the corresponding Eq. (47), but in the limit $a/bkgd\ll 1$ it gives a good description of the results of the model experiment. Therefore, the expressions given here correctly reproduce the errors in the amplitudes, and the correct results for the errors in the peak positions are obtained only in the limit of small (compared with the background) amplitudes. This is related to the fact that the linear stochastic spectrum model[16] presupposes the presence of distortions in the peak positions only owing to background fluctuations (see Sec. 2).

Let us also give the expression for calculating the potential error in the determination of the amplitude of an isolated symmetric Gaussian peak in two-dimensional spectra:[17]

$$\sigma_a=1.4(1+2(a/bkgd)^{1/2})(a/b)^{1/2}f(b), \tag{74}$$

where, as usual, a and b are the amplitude and half-width, and

$$f(b)=0.4[\exp(-b/3)+\exp(-b/17)]. \tag{75}$$

We see that Eq. (74) is the product of the empirical formula (48) for calculating the potential error σ_a in the determination of the amplitude of an isolated peak in the one-dimensional spectrum and the empirical function $f(b)$ which was found in Ref. 17 in the reduction of the study of the two-dimensional spectrum to a set of one-dimensional ones (the intersections of the two-dimensional peak by different planes passing through its vertex). At present there are no other results pertaining to the potential errors in determining the peak parameters in multidimensional spectra.

5. DETERMINATION OF THE NUMBER OF PEAKS IN A MULTIPLET

Owing to the statistical nature of experimental data, the question of the number of peaks contained in a multiplet can be answered only in probabilistic terminology. Therefore, the probabilities of finding peaks in a multiplet calculated on the basis of data on the statistical reliability of the experiment are, in general, the only objective quantitative information which can be extracted and which is needed to solve the problem of determining the number of peaks contained in a multiplet. A definitive solution (whether or not there is a peak) on the basis of these data is taken only by a human using all the available information, including information not contained in the processed spectrum.

The general problem of calculating the probabilities of finding peaks in a multiplet of arbitrary order m within the linear stochastic spectrum model was solved in Ref. 16. However, in order to use these general results it is necessary to know the errors σ_i in the determination of the amplitude of each peak i in the analyzed multiplet ($i\leqslant m$), which for arbitrary m are unfortunately determined only implicitly[38] via Eqs. (59)–(66). Such a determination of σ_i for the case of arbitrary m makes the results difficult to interpret, and specific calculations of the probabilities of finding peaks belonging to a multiplet have so far been published only for doublets ($m=2$; Ref. 37). The general expressions from Ref. 16 have been used to calculate (Fig. 15) the conditional probabilities D for true discovery of peaks belonging to a doublet for fixed probability F of false discovery using the estimates of the errors σ_a in determining the amplitudes of the peaks in the doublet by the empirical formula (54). We see from this figure that for the smallest of the peaks in the doublet the conditional probability D_{a1} for true discovery, as expected, depends strongly on the distance d between the peaks. For fixed d the conditions for finding peaks of equal amplitude $a1=a2$ are the most favorable. For fixed ratio $a1/a2$ as the distance d between the peaks of the doublet increases the peaks affect each other less and the probability of true discovery of them approaches the value characteristic for each of the peaks when it is isolated. For the peak of large amplitude ($a2$) dependence of the conditional probability

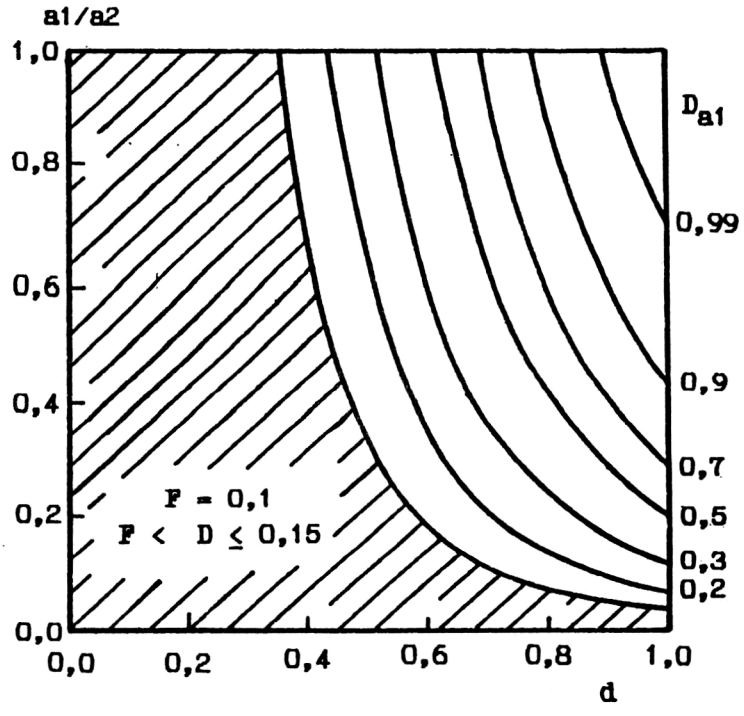

FIG. 15. Conditional probability D_{a1} of true discovery of the smaller of the peaks in the doublet for the value of the conditional probability of false discovery $F=0.1$. The curves are the calculation of D_{a1} as a function of $a1/a2$ (ratio of the amplitude of the smallest peak to that of the largest) and d (distance between the peaks in units of the half-width). Example for a peak half-width of $b=5$ channels, amplitude of the largest peak $a2=100$ counts, and background amplitude $bkgd=100$ counts.

D_{a2} of its true discovery on the ratio $a1/a2$ is practically absent (the probability decreases weakly for $a1/a2 \to 1$).

Some information for deciding whether or not a peak is present in a multiplet can also be obtained[36] from the expressions for calculating the potential errors in determining the parameters of these peaks, for example, Eqs. (52)–(54) for the potential errors in the determination of the parameters of peaks belonging to a doublet. The point is that for some values of the multiplet parameters the spread in their values associated only with statistics admits values which have no real physical meaning, for example, negative amplitudes. Therefore, it is possible to introduce criteria determining the region of values of the multiplet parameters (the physical region) which correspond to sufficiently accurate results of interest to the user. The boundary of the physical region of parameter values for a multiplet can be determined, for example, for a doublet using the inequality $\sigma_{a1}/a1 < 1$ ($a1$ is the amplitude of the smallest of the peaks making up the doublet, and σ_{a1} is the potential error in its determination). If this inequality holds for the parameters of a real multiplet, it can be assumed that its decomposition into component peaks is admissible. These arguments have also served as the basis for referring to the expressions for calculating the potential errors in the determination of the peak parameters as "physical reliability functions" (PRFs; Ref. 36). The PRFs clearly can also be used for planning experiments.

6. A PROGRAM FOR AUTOMATED SPECTRUM PROCESSING UNDER STATISTICALLY DETERMINED CONDITIONS: ETAP

The theoretical-methodological results presented above and obtained using the linear stochastic spectrum model[16] have been partially realized[40] in the form of a computer program for the automated processing of linear spectra under statistically determined conditions, ETAP (first reported in Ref. 41). In the creation of this program the problem initially posed (in 1977) was that of the stage-by-stage solution of the problem of automatic spectrum processing with gradual algorithmization and realization in the form of computer programs of only those problems in automatic spectrum processing for the solution of which there was already a thoroughly worked out and reliable theoretical base. Therefore, the ETAP program interprets all peaks in the processed spectra as isolated peaks, and the problem of multiplet decomposition is treated in the next stage of automatic spectrum processing, the theoretical study of which is continuing. In return, ETAP processes spectra with isolated peaks under fully statistically determined conditions, and the results of the processing are, as will be shown below, close to statistically optimal.

The ETAP program[40] is designed for the rapid automatic processing of linear spectra, for example, γ spectra from semiconductor detectors. It automatically searches for peaks for a fixed value of the conditional probability F of false discovery, and calculates the positions, amplitudes, half-widths, and areas of the peaks it finds with the corresponding errors, along with the conditional probability D of true discovery of all the peaks found. The peaks are assumed to be symmetric Gaussians, and no fit of their parameters or decomposition of complex multiplets (when no minima lie between the maxima of the component peaks) is performed.

The technique of automatic processing of spectra containing only isolated peaks is fairly well developed,[7] and therefore in designing ETAP the attention was focused on optimization and correct interpretation of the results of the processing. Among the parameters controlling the operation of the program is the value of F, the conditional probability for false discovery of peaks at which the spectrum

must be processed. Iterative processing of spectra was also provided for. It is used to include the effect of varying the half-width of the peaks along the spectrum for a given background, for splitting of the spectrum into informative segments which can be treated independently, and for the quality of the peak identification. The optimization of the structural scheme of ETAP and its individual subprograms has made it possible to create a compact and fast program.

ETAP searches for peaks for a constant value of the conditional probability F of false discovery fixed by the operator at the beginning. Since most peak-search programs have[29] similar values of the conditional probability D of true discovery of peaks for $F=\text{const}$ (i.e., close in quality), the main factor determining the choice of search algorithm is that F be independent of the amplitude of the background. Therefore, in Ref. 40 preference was given to the search algorithm of Ref. 27, for which independence of F from the amplitude of the background *bkgd* in the spectrum (for $bkgd > 100$ counts) arises from the fact that the peak search is carried out in a transformed spectrum, where the square root of the experimental value is taken in each channel. The independence of F from the amplitude of the background was checked for the search program of Ref. 27 by means of a repeated model experiment using the technique of Ref. 20. An empirical formula for $F(\mathbf{1})$ relating the value of the conditional probability F and the sensitivity threshold l was found in this way in Ref. 40. This expression is used in the ETAP program to determine the required sensitivity threshold l from the value of F specified in the initial data.

The area s of each peak is calculated in the ETAP program from the expression $s=1.064ab$, where a and b are the amplitude and half-width of a symmetric Gaussian peak (inclusion of various corrections and calculation of s by channel-by-channel summation of the counts below the peak is also possible). The errors in the position, amplitude, half-width, and area are calculated using the empirical formulas (47)–(50). We recall that the values of the errors determined by these formulas are the potential errors in the determination of the parameters of an isolated symmetric Gaussian peak for a given statistical reliability of the data. The conditional probability D of true discovery of each of the identified peaks is also calculated using an empirical formula[40] found by interpolation of the results of the model experiment, in the course of which the dependences $D(x)$ with $x=a/bkgd^{0.477}$ were obtained for the values of the conditional probability of false discovery $F=0.001$, 0.01, and 0.1.

Complex testing of the ETAP program was carried out by using it to process a control quasiexperimental spectrum (the parameters of which were specified in its construction, so that they were known exactly) and the experimental γ-ray spectrum of radionuclides from a set of sample standard gamma sources (SSGS) [measurements made using a Ge(Li) detector and a standard spectrometric apparatus]. The results of the ETAP program were also compared with the results of the Katok-F program (Ref. 42) and the Moment Method program (Ref. 43). Comparison showed that the values of the parameters found by all these programs for all isolated peaks in the experimental spectrum practically coincide within the error bars. Since ETAP calculates the potential errors, the results for isolated peaks obtained using it can qualify as results obtained with the maximum accuracy for a given statistical reliability of the experiment, i.e., they are close to statistically optimal.

Therefore, the main feature of ETAP is the fact that peaks are sought automatically in the processed spectrum for a value of F, the conditional probability for false discovery, which is fixed and the same for the entire spectrum and with the conditional probability D for true discovery calculated for all the identified peaks. Consequently, the spectrum is processed under statistically determined conditions. Therefore, the summary table of the results of the ETAP program (a fragment of it from Ref. 40 is given in Table I as an example) differs significantly from the summary tables of all the other spectrum-processing programs. The main difference is that ETAP gives not only the values found for the peak parameters with their errors, but also the conditional probability D of true discovery of each peak for a conditional probability F of false discovery which is fixed and identical over the entire analyzed spectrum ($F=0.001$ for the table given here).

This difference is of fundamental importance. "The current conditions for publishing an article, as a rule, do not allow a description of the experimental conditions or records of the main and control measurements. This is bad, because the absence of such information makes critical analysis of the results extremely difficult" (Ref. 44). To partially compensate for this deficiency, articles containing experimental results are often illustrated with figures of typical experimental spectra or the most important fragments of them. Then it becomes possible for the reader to visually estimate the statistical reliability of the experimental data (the number of counts in the channels, the relations between the peak and background amplitudes, the degree of separation of peaks in multiplets, and so on), and, consequently, to qualitatively estimate the contribution of statistical effects to the total experimental error and the reliability of the primary experimental data. But this material is insufficient for a quantitative comparison of the results of various experiments. A statistically correct quantitative comparison of the quality of the primary experimental data using only the results of their processing is possible if the conditional probabilities D of their true discovery for (identical) known values of the conditional probability F of their false discovery are given for all the peaks in the spectra being compared. The initial data needed for this are provided by the ETAP program. Then such a comparison does not require analysis of the complete measurement protocols. It is sufficient to use the results of the processing of the compared spectra by the ETAP program (in particular, the values of D calculated with it for $F=\text{const}$) and the expressions for F_{in} and D_{in} (44), (45), found in Ref. 30.

Therefore, in different laboratories with different researchers the spectral data need to be analyzed by computers using programs which ensure spectrum processing under statistically determined conditions (with the

TABLE I. Fragment of the summary table of results of the automatic spectrum-processing program ETAP (example from Ref. 40); $F=0.001$.

Peak number	p	Δp	s	Δs	b	Δb	D
1	39,94	0,53	1265	281	4,78	1,56	0,55
2	101,41	0,14	4940	278	4,62	0,39	1,00
3	109,80	0,05	22661	356	5,65	0,13	1,00
4	121,00	0,24	2416	261	4,36	0,71	1,00
5	139,22	0,36	2193	281	5,41	1,05	0,98
6	149,79	0,44	1070	239	3,89	1,30	0,68
7	164,43	0,45	1743	282	4,49	1,32	0,86

Note. p, Δp, s, Δs, b, and Δb are the values of the positions (channels), areas (counts), and half-widths (channels) found by the program for the identified peaks with the corresponding errors. F and D are the conditional probabilities of false and true discovery for all the peaks.

calculation of D for fixed F), and the values of F and D must be given in published tables of results along with, for example, the energy and intensity of each line (in any case, for lines which correspond to peaks with values of D considerably less than unity). From 1977 until now the only program satisfying these requirements has been the ETAP program.[40]

During its existence the ETAP program has developed and absorbed new results of research on the problem of automatic spectrum processing. Compared with the published version,[40] the present version now incorporates the following additions: calculations, which can be requested by the user, of not only the values of D but also the conditionless *a posteriori* probabilities of true and false discovery of peaks P_t and P_f (according to the results of Ref. 30 with $\kappa=1$), and a peak search using the program of Ref. 19, which, as shown in Ref. 29, is the best among existing search programs. In addition, the choice of the specific numerical value of the conditional probability F of false discovery at which it is appropriate in most cases to perform automatic spectrum processing, $F\cong 0.01$ (Ref. 32), is justified, which completely eliminates the earlier indeterminacy which for many was a psychological barrier in the practical use of the ETAP program.

7. STANDARDIZATION OF SPECTRUM-PROCESSING PROGRAMS

The early technique and experience of comparing spectrum-processing programs described in Ref. 9 are important in the positive sense, since this was the first time that a large collaborative effort was made to try to solve the problem of the quantitative determination of the quality of computer programs for processing linear spectra for the example of γ spectra of high energy resolution. This was definitely a step forward compared with the (until recently widespread) practice of estimating the quality of such programs on the basis of experience in using them.

The technique of Ref. 9 is based on the following principle. Several reference γ spectra measured (or synthesized from measured ones) using a standard apparatus are determined. These spectra contain some number of characteristic features: isolated peaks of large and small amplitude on large and small, linear and nonlinear backgrounds, and doublets with different ratios of the areas of their component peaks and different distances between these peaks. Some quantitative criteria of the quality of the programs are also determined; they are calculated from the results of using these programs to process all the reference spectra. Thus, the main idea of the method is comparison of the results of using different programs to process a small number of identical "reference" spectra. The main defect of the approach is the use of a small statistical sample (the reference spectra), which cannot encompass all the possible diversity of statistical situations, so that it is impossible to make full use of statistically reliable quantitative comparison criteria.

Let us consider the possibility of comparing computer programs for spectrum processing on the basis of a different approach, based on the use of the results of a repeated model experiment.

Spectrum-processing programs usually form a part of measurement information processing systems (MIPSs) which, in turn, are one of the elements of information-measurement systems (IMSs), the principal way of obtaining information in, for example, nuclear physics. Both IMSs and MIPSs can be standardized. MIPSs, as a rule, deal only with the digital equivalent of some physical quantity (a set of data), performing calculations stipulated by the program. Therefore, the attestation of a MIPS and its elements possesses a number of features[45] related to the need to estimate the methodological errors due to the specific features of the program algorithm, the word size of the computer, and so on. The metrological attestation of the algorithm (program) within a MIPS can be accomplished in two ways: either by computational methods or by using a model experiment.[45] Therefore, the technique proposed in Ref. 20 for studying the quality of peak-search programs on the basis of using the results of a repeated model exper

iment completely corresponds to one of the approaches to the attestation of algorithms and programs used in metrology. The possibility of using it for the standardization of peak-search programs was first suggested in Ref. 46. It is discussed in more detail in Ref. 47, where the need to use the conditional probabilities F and D introduced in Ref. 16 for metrologically characterizing the quality of the identification algorithms was stressed.

In the formal error model, the MIPS is treated as a black box whose input and output are the only things accessible to observation. To the input is fed a signal (a digital set) with known informative parameters, constructed in accordance with a mathematical model which adequately reflects the properties of the studied phenomenon, and the correspondence of the informative parameters of the output signal (a digital set) with the expected values is analyzed. The noninformative parameters of the signal can be random, but must lie in the range of values allowed by the attested system. By varying the values of the parameters of the input signal it is possible to obtain influence functions, dependences of the changes in the metrological characteristics on changes of the influencing factors or noninformative parameters (for the conditions under which the system is used). In the case of attestation of peak-search programs, the input signal is a set of a large number of model quasiexperimental spectra, the known informative parameters of which (the parameters of the identified peaks and the noninformative parameters are specified in the construction) are the search sensitivity threshold and the amplitude of the background below the peak. The output signal is the fact that a peak is (or is not) discovered. The metrological characterists (influence functions) which are sought are the dependences of the conditional probability D of true discovery of peaks on the values of the conditional probability F of false discovery (the search sensitivity threshold), the amplitude a of the identified peak, its half-width b, and the amplitude of the background below the peak $bkgd$ on the parameters of the program itself. Therefore, in Refs. 19–21, 29, and 40 the metrological attestation of a number of automatic peak-search programs has actually already been carried out (for different programs with different amounts of data), since the influence functions $D(a,b,bkgd)$ determining the metrological properties of these programs for different values of F have been found for them. These influence functions have been found according to a standard metrological scheme, so that the technique of Ref. 20 and the results of the attestation using it of search programs have a natural place in the existing scheme of metrological justification of MIPSs. The work performed on the attestation of search programs can serve as the foundation for the development of a regional, national, or local standard. Only one more step needs to be taken in this direction: determination of the admissible variations of the program characteristics which can be considered to satisfy the requirements of the proposed standard. This is not difficult to do using the results and experience already gained.

The other stage of the complete processing of linear spectra which can now be considered to have the necessary theoretical-methodological and metrological foundation for standardization and for which sufficient experience in using it has been accumulated is the stage of estimating the errors in determining the peak parameters (so far only for isolated peaks and peaks forming doublets). In the attestation of the methods of estimating the errors in the MIPS input signal there is also a large set of model quasiexperimental spectra, the known informative parameters of which have been specified in the construction of the peak and background parameters. The noninformative parameters are the parameters controlling the fitting program and the parameters of the mathematical model of an isolated peak used in the fit. The output signal is the spread of the values of the peak parameters found in the processing. The desired influence functions are the dependences of the rms spread of the values found for the peak parameters on the statistical reliability of the data (on the true values of the parameters of the constructed peaks and background). In the case of the singlet and doublet the influence functions (expressions for calculating the potential errors or the physical reliability functions) have already been found in Ref. 37 using such a model experiment; these are Eqs. (47)–(50) and (52)–(54). For multiplets with more than two peaks, using the results of Ref. 38 it is at present possible to calculate only the potential errors in the determination of their amplitudes. Therefore, the equations from Ref. 37 for the potential errors in the determination of the singlet and doublet parameters are, owing to the specific features of the method by which they were obtained, influence functions by definition, i.e., they are the metrological characteristics of the error-estimation procedure, and they can be used to check the correctness of estimating the errors by means of other methods.

We note that in the case of multiplets there is a significant effect on the errors in determining their peak parameters from the error with which the shape of an isolated peak is known in the fit (when the peak is Gaussian, this is the error with which its half-width is known). This has been demonstrated in Ref. 8 for the case of doublets, but so far there are no systematic quantitative results on this problem. In spite of this, in the fit of the proposed standard the allowed range of error in the parameters of the mathematical model of an isolated peak used in the multiplet decomposition still needs to be specified.

Thus, the approach of Ref. 47 to the metrological attestation of complete spectrum-processing programs or individual stages of such programs based on the use of the results of a repeated model experiment performed by computer fully corresponds to the accepted system of metrological guarantees of MIPS elements and can therefore serve as the basis for a fit of the corresponding standards. The preliminary studies needed for standardization of the peak-search methods (programs) and estimation of the errors in determining their parameters have already been carried out. The use of a large number of model quasiex

perimental spectra in such a metrological attestation of the programs does not present any technical difficulties at present.

8. AUTOMATIC PROCESSING OF LINEAR SPECTRA AND THE PROBLEM OF ARTIFICIAL INTELLIGENCE

The determination of the limits of the possibilities offered by computers in solving intellectual problems (for example, automatic image recognition in the processing of images obtained in physical experiments) has been under active discussion for many years in connection with the problem of creating artificial intelligence, and this is an important topic today. Many results have been obtained by studying this problem for the example of image recognition and the geometrical reconstruction of photographed events from track devices in high-energy physics like spark and bubble chambers. The experience which has been gained from the design of computer-based systems for processing track information has significantly affected the current ideas about the possibilities offered by computational technology and its use in solving various problems. Attempts to completely replace the human intellect by automatic, computer-based image processing systems operating without human intervention have proven unsuccessful, in spite of the enormous efforts made to design powerful specialized programs based on a diversity of mathematical methods of image recognition. The end result can be summarized as follows. "As a rule, it is possible to process 80–90% of the information in photographs automatically. By increasing the size and complexity of the program realization,... it is possible to raise this to 90–95%. However, further progress in the processing of relatively complex events always proves to be impossible... . This stability at the 90% level for the automatic processing of information in images obtained in track-chamber experiments may be related to a well known natural tendency..., which is grasped by humans at the intuitive level and transferred to the exploitation of human-designed research involving the creation of experimental setups... . Here we may be encountering a manifestation of a heretofore unknown law of nature, from which...definite conclusions can be made. Namely: the use of computers in image-processing system ... is inherently limited... . About 10% of the image information which can be processed must be processed by a human being... . It is also not impossible that the very fact of the establishment of this rule may be used to solve problems in constructing genuine artificial intelligence" (Ref. 48).

The ETAP program[40] for the automatic processing of linear spectra has been used for many years at various physics centers to process spectrometric information obtained in fundamental and applied research by nuclear-spectroscopy methods. Experience in using it indicates that about 90% of the peaks in experimental spectra are isolated and can be efficiently processed automatically. The other 10% of the peaks form a part of multiplets of various complexity, and ETAP does not automatically decompose these into their component peaks. In such cases multiplets are processed, when necessary, by other programs whose operation involves participation of a human operator. Therefore, in the case of the images obtained in spectrometric experiments, i.e., in linear spectra, there again appears the "stability at the 90% level for the automatic processing of information," first established in Ref. 48 in analyzing image-processing methods for track-chamber photographs. Therefore, again in the processing of linear spectra "we may be encountering a manifestation of a heretofore unknown law of nature," as mentioned above, and this fact "may be used to solve problems in constructing genuine artificial intelligence."

The problem of the automatic processing of linear spectra is technically much simpler than that of the automatic processing of images on track-chamber photographs. This gives us a chance to study this problem in more detail and thereby obtain more detailed information both about the possibility of imitating the human intellect by computer in image recognition, and on the human image-recognition process itself.

Our experience in developing automatic spectrum-processing programs indicates that it is important to study automatic image recognition by computer in conjunction with the ability of the human intellect to recognize the same images. Such studies lead to results which convincingly demonstrate both the limits of computer possibilities, and the relation between the capabilities of humans and computers. For example, we have shown[32] that an experienced (professional) human operator identifies isolated peaks in spectra with conditional probability of false discovery $F=0.012\pm0.004$, while a nonprofessional operator does this for $F\sim0.3$. This is easily understood, since "much of what we possess as knowledge, experience, and information is the result of developed intuition. Long-term specialization in some form of activity makes it possible to broaden our intuitive capabilities in the corresponding subject area" (Ref. 49). This example clearly demonstrates the role of such long-term specialization in broadening the intuitive abilities of an operator in solving the particular problem of peak identification. After all, the quality of the search for peaks by nonprofessional and professional operators are limiting cases characterizing the intuitive abilities in recognizing certain images (peaks) by a fully unprepared operator and an operator who has long specialized in this form of activity. The results of such studies for professionals can be viewed as a quantitative description of the limits of the possibilities offered by the human intellect in solving particular image-recognition problems. Comparison of the dependences $D(a)$ for $F=0.01$ for professionals and for identification programs[19] realizing the suboptimal search algorithm (Fig. 7) shows that here machines are not only not inferior to humans, but beat them. Computers can be used to successfully process spectral data for any *a priori* specified value of F, and in this sense they offer broader possibilities than does the human intellect. In addition, it is only by studying humans that we can answer the question of why professionals searching for peaks have the typical value of the conditional probability for false discovery $F\simeq0.01$, while nonprofessionals have $F\sim0.3$.

These specific numerical values of F for professionals and dilettantes do not follow from the features of the experimental data or computer programs.

The experience in designing automatic programs also indicates the importance of the objective, quantitative inclusion of the available *a priori* information about the processed data (the spectrum) which is not directly contained in them. The possibility of the correct, quantitative inclusion of, in general, arbitrary additional information about the spectrum when making a decision about the presence or absence of a peak in it is reflected in the coefficient κ in Eqs. (41) and (43), which were first written down for peak identification in Ref. 30. The question of choosing the numerical value of κ for different specific cases remains unstudied, but it is clear that there is some hope in determining it by using both computer programs and the human intellect itself. In this regard let us again draw attention to some very important facts. **1**. The quality of peak searches by automatic programs is better, the larger the amount of *a priori* information about the spectrum that is used (Fig. 3). **2**. When a human and a computer program are placed under identical conditions regarding the use of *a priori* information about the spectrum, the quality of the automatic peak search is not only not worse, but in some cases is even better, than the quality of the search for the same peaks by a qualified operator (Fig. 7). Comparing these facts, we conclude that the better quality of peak identification by a human compared with that by automatic programs which actually is seen in practice is a consequence of the use in the search (often unconsciously) of additional *a priori* information about the spectra which is not contained in the analyzed data, and which is accessible to a human but not to a computer. The *a priori* information plays the largest role in the processing of so-called "complex" images, which suggests that those images whose correct recognition requires a large amount of additional information should be considered complex. It is appropriate to use a human operator to process such complex images, because a human carries the additional information needed for the recognition. The amount of such data for images which (owing to a deficiency of the *a priori* information) are poorly processed in the automatic mode and which usually require use of a human operator is about 10%. This is the percentage of data which are not amenable to automatic analysis in both track-chamber photographs[48] and linear spectra. According to the idea expressed in Ref. 48, "here we may be encountering a manifestation of a heretofore unknown law of nature." Therefore, fully automatic image processing by computers may be successful when not only mathematical recognition methods, but also (in formalized form) some *a priori* information, carried usually by a human operator and acquired by him earlier in the course of training by analyzing other similar images, are used. Therefore, the complete mimicking of the human intellect in the problem of image recognition by computer requires first an explanation of exactly what additional information it is that a human uses to process images of a certain type (for example, linear spectra or track-chamber photographs), formalization of this information, and putting it at the disposal of an automatic program, which solves the same problem as the human does. Programs for the automatic processing of linear spectra deal with relatively simple objects (spectra), so that it can be hoped that in this case the amount of additional information that a human uses is not very large, that it is fairly simple, and that it can be revealed and formalized. This can ensure the effective operation of automatic programs even in complicated cases where human operators now perform the processing. This leads to another conclusion: programs for the automatic processing of linear spectra (like other automatic programs for computers which imitate the human intellect in image-recognition problems) must be trainable or self-trainable. During the training (self-training) the automatic program must acquire the additional *a priori* information about the specific data being processed which is not contained in these data, but without which it is impossible to correctly solve the recognition problem, especially in so-called complex cases. Usually (for example, in working with automated programs) the carrier of this additional information is a human, so that to completely solve the problem of automatic image recognition by computer it is important to study the properties of the human intellect itself, in particular, in regard to how it solves recognition problems. Therefore, efforts to design automatic, without human participation, programs for the complete processing of linear spectra must lead not only to a solution of this problem, which is so important by itself, but must also help to define the limits of the possible use of computers for solving intellectual problems. These efforts will also aid the study of the human image-recognition process itself, i.e., the study of the human intellect.

CONCLUSION

A great deal of progress has been made in recent years in understanding and solving a number of problems in the automatic processing of linear spectra. This progress has led to the creation of the computer program ETAP for fully automatic spectrum processing under statistically determined conditions, the results of which are close to statistically optimal in the processing of isolated peaks.

A stochastic model of linear spectra (both one-dimensional and of arbitrary dimension k) has been proposed, and its effectiveness and productivity have been demonstrated for solving problems in automatic spectrum processing. The problem of automatically searching for isolated peaks has been solved almost completely, and the theoretical-methodological and practical results which have been obtained pose the real, current problem of creating a standard for programs (algorithms) to carry out such a search. The importance of the practical use in automatic processing programs of quantities characterizing the quality of the search algorithms—the conditional probabilities F and D for false and true discovery of peaks—has been demonstrated, and recipes have been given for calculating F and D, including those for peaks which are components of complex multiplets. Expressions have been obtained for calculating the conditionless *a posteriori* probabilities of true and false discovery of peaks, P_t and

P_f. A parameter has been found which can be used for the statistically correct inclusion of the effect on the results of additional *a priori* information on the spectrum. The influence of subjective factors on the results has been studied, and used as the basis for specifying a choice of numerical value of the search sensitivity threshold for peak-identification programs. This parameter is one of the most important ones controlling the spectrum processing (it has been found that the sensitivity threshold must correspond to the value $F=0.012\pm0.004$). The effectiveness of such studies as a tool for investigating problems in psychology, in particular, human reflexes, has been demonstrated. The concept of potential errors in determining the peak parameters, which depend only on the statistical reliability of the experiment, has been introduced, and a technique has been given for calculating them from the results of a repeated model experiment and (partially) by the maximum-likelihood method using the linear stochastic spectrum model. The model experiment has been used to find expressions for calculating the potential errors in determining the parameters of isolated peaks and doublets, and the maximum-likelihood method has been used to find expressions for calculating the potential errors in the determination of the amplitudes and positions of peaks in a multiplet of arbitrary complexity (errors in the positions only for peaks of small amplitude). The possibility of standardizing the methods of estimating the errors in the determination of the peak parameters has been demonstrated. A method of determining the optimal criteria for terminating the fit in the iterative search for the values of the parameters of the mathematical model describing the spectrum has been given, along with examples of such criteria for the case of processing isolated peaks and doublets.

In summary, we can conclude that the problem of automatic processing of spectra containing isolated peaks can be considered practically solved. Great progress has been made in understanding the problems of automatic decomposition of multiplets into their component peaks, and the theoretical-methodological results which have been obtained admit the practical realization of the most important components of a computer program which will solve in a first approximation the problem of the automatic processing of linear spectra, including the decomposition of complex multiplets. These studies will also help to obtain new results in problems of human psychology and artificial intelligence. It may be possible to completely solve the problem of automatic, without human participation, processing of linear spectra by designing trainable (self-trainable) computer programs which mimic the human intellect in solving image-recognition problems.

The author would like to thank Prof. I. I. Zalyubovskiĭ for wide-ranging help and support, M. H. S. Bakr, G. V. Vinel', L. A. Vylova, V. A. Deryuga, A. A. Marchuk, V. I. Sorokin, V. V. Sotnikov, W. Habenicht, V. M. Tsupko-Sitnikov, E. K. Chernysheva, and H. Oehler for their collaboration in obtaining a significant fraction of the results presented here, and also R. I. Gopych for helping with the text.

[1)] Of course, the actual errors can be larger than the calculated ones for reasons unconnected with statistics. For example, this can happen owing to unstable operation of the measuring apparatus (see Fig. 14).

[2)] In the terminology of Ref. 4, the "number of regularities" is the number of peaks in a multiplet.

[1] V. M. Tsupko-Sitnikov, in *Computers in Experimental Physics*, D10-5255, JINR, Dubna (1970), p. 125 [in Russian]; Report D10,11-8450, JINR, Dubna (1974), p. 300 [in Russian].

[2] P. Quittner, *Gamma-Ray Spectroscopy with Particular Reference to Detector and Computer Evaluation Techniques* (Academiai Kiado, Budapest, 1972).

[3] C. M. Lederer, in *Radioactivity in Nuclear Spectroscopy*, edited by J. M. Hamilton and J. C. Manthuruthie (Gordon and Breach, London, 1972), Vol. 1, p. 73.

[4] V. Gadzhokov, Fiz. Elem. Chastits At. Yadra **11**, 1474 (1980) [Sov. J. Part. Nucl. **11**, 584 (1980)].

[5] V. B. Zlokazov, Fiz. Elem. Chastits At. Yadra **16**, 1126 (1985) [Sov. J. Part. Nucl. **16**, 501 (1985)].

[6] N. G. Volkov *et al.*, *Methods in Nuclear Spectroscopy* [in Russian] (Énergoatomizdat, Moscow, 1990).

[7] G. Éler *et al.*, Preprints R10-6817, R10-6818, R10-6819, JINR, Dubna (1972) [in Russian]; Prib. Tekh. Eksp. No. 5, 253 (1973) [Instrum. Exp. Tech.]; All-Union Institute of Scientific and Technical Information Deposit No. 6652-73, Moscow, 1973.

[8] G. Éler *et al.*, Preprints R10-7364, R10-7365, R10-7366, R10-7367, JINR, Dubna (1973) [in Russian].

[9] R. M. Parr *et al.*, in *Proc. of the American Nuclear Society Conf. on Computers in Activation Analysis and Gamma-Ray Spectroscopy*, Mayaguez, Puerto Rico, May 1978, p. 544.

[10] B. V. Polenov and K. N. Stas', At. Energ. **60**, No. 2, 157 (1986).

[11] O. F. Nemetz and R. G. Ofengenden, Vestn. Akad. Nauk Ukr. SSR **12**, 81 (1974).

[12] Z. B. Golembo and G. V. Venikov, Itogi Nauki Tekh., Ser. Tekh. Kibernetika (All-Union Institute of Scientific and Technical Information, Moscow, 1970), p. 5.

[13] A. N. Tikhonov and V. Ya. Arsenin, *Solutions of Ill-Posed Problems* (Halsted, New York, 1977) [Russian original, Nauka, Moscow, 1974].

[14] V. S. Aleksandrov *et al.*, in *Proc. of the Twenty-Fourth Conf. on Nuclear Spectroscopy and Structure of the Atomic Nucleus. Abstracts* [in Russian] (Nauka, Leningrad, 1974), p. 484.

[15] N. D. Eckhoff, Nucl. Instrum. Methods **68**, 93 (1969).

[16] P. M. Gopych *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **39**, No. 3, 564 (1975) [Bull. Acad. Sci. USSR, Phys. Ser.].

[17] P. M. Gopych and A. A. Marchuk, in *Applied Nuclear Spectroscopy* (Énergoatomizdat, Moscow, 1983), No. 12, p. 129 [in Russian].

[18] V. A. Kotel'nikov, *Theory of Potential Antijamming* [in Russian] (Gosénergoizdat, Moscow, 1956).

[19] P. M. Gopych *et al.*, in *Proc. of the Thirty-Second Conf. on Nuclear Spectroscopy and Structure of the Atomic Nucleus. Abstracts* [in Russian] (Nauka, Leningrad, 1983), p. 582.

[20] P. M. Gopych *et al.*, Prib. Tekh. Eksp. No. 4, 90 (1977) [Instrum. Exp. Tech.].

[21] P. M. Gopych and V. A. Deryuga, Prib. Tekh. Eksp. 23 (1983) [Instrum. Exp. Tech.]; in *Proc. of the Twenty-Second Conf. on Nuclear Spectroscopy and Structure of the Atomic Nucleus. Abstracts* [in Russian] (Nauka, Leningrad, 1979), p. 475.

[22] J. Routti and S. G. Prussin, Nucl. Instrum. Methods **72**, 125 (1969).

[23] S. R. Avramov *et al.*, Report R10-9741, JINR, Dubna (1976) [in Russian].

[24] G. V. Vinel' *et al.*, Report R10-10844, JINR, Dubna (1977) [in Russian].

[25] P. M. Gopych *et al.*, in *Proc. of the First All-Union Conf. on Automatization of Scientific Research in Nuclear Physics. Abstracts* (Institute for Nuclear Research, Ukrainian Academy of Sciences, Kiev, 1976), p. 150 [in Russian].

[26] P. M. Gopych and V. A. Deryuga, Report D10,11-11264, JINR, Dubna (1978), p. 322 [in Russian].

[27] V.-D. Fromm, Report 10-9094, JINR, Dubna (1975) [in Russian].

[28] N. G. Volkov *et al.*, in *Proc. of the Twenty-Fourth Conf. on Nuclear Spectroscopy and Structure of the Atomic Nucleus. Abstracts* [in Russian] (Nauka, Leningrad, 1974), p. 483; N. G. Volkov, *ibid.*, p. 454.

[29] P. M. Gopych and I. I. Zalyubovskiĭ, in *Questions About the Accuracy*

of Nuclear Spectroscopy [in Russian] (Physics Institute, Lithuanian Academy of Sciences, Vilnius, 1988), p. 136.
[30] P. M. Gopych and V. V. Sotnikov, in *Applied Nuclear Spectroscopy* [in Russian] (Énergoatomizdat, Moscow, 1983), No. 12, p. 271.
[31] P. M. Gopych *et al.*, Report D10,11-11264, JINR, Dubna (1978), p. 326 [in Russian].
[32] P. M. Gopych *et al.*, Prib. Tekh. Eksp. No. 3, 95 (1992) [Instrum. Exp. Tech.]; in *Proc. of the Forty-First Conf. on Nuclear Spectroscopy and Structure of the Atomic Nucleus. Abstracts* [in Russian] (Nauka, Leningrad, 1991), p. 483.
[33] P. M. Gopych and V. V. Sotnikov, in *Proc. of the Int. Conf. on Nuclear Spectroscopy and Structure of the Atomic Nucleus. Abstracts* [in Russian] (Nauka, St. Petersburg, 1993), p. 361.
[34] V. A. Lefebvre, *The Fundamental Structure of Human Reflexion* (Peter Lang Publ., New York, 1990); V. A. Lefebvre, *Algebra of Conscience* (Reidel, Dordrecht, 1982).
[35] S. D. Malkin *et al.*, Atomic Energy Abroad, No. 5, 9 (1990) [in Russian].
[36] P. M. Gopych *et al.*, in *Proc. of the Twenty-Fourth Conf. on Nuclear Spectroscopy and Structure of the Atomic Nucleus. Abstracts* [in Russian] (Nauka, Leningrad, 1974), p. 479; P. M. Gopych *et al.*, *ibid.*, pp. 480, 481.
[37] P. M. Gopych, in *Questions About the Accuracy of Nuclear Spectroscopy* [in Russian] (Physics Institute, Lithuanian Academy of Sciences, 1986), p. 138; *A Study of the Structure of Transitional Nuclei in the Region* $50<Z$, $N<82$, Candidate's Dissertation, Khar'kov, Khar'kov State University, 1974 [in Russian].
[38] P. M. Gopych, in *Questions About the Accuracy of Nuclear Spectroscopy* [in Russian] (Physics Institute, Lithuanian Academy of Sciences, Vilnius, 1988), p. 139.
[39] S. Brandt, *Statistical and Computational Methods in Data Analysis* (North-Holland, Amsterdam, 1970) [Russian transl., Mir, Moscow, 1975].
[40] P. M. Gopych *et al.*, in *Applied Nuclear Spectroscopy* [in Russian] (Énergoatomizdat, Moscow, 1982), No. 11, p. 86.
[41] P. M. Gopych *et al.*, Report D10,11-11264, JINR, Dubna (1978) [in Russian], p. 330.
[42] V. Gadjokov, Report E10-12353, JINR, Dubna (1979).
[43] Ts. Vylov *et al.*, Report R6-9072, JINR, Dubna (1975) [in Russian].
[44] B. S. Dzhelepov, *Methods of Developing Complex Decay Schemes* [in Russian] (Nauka, Leningrad, 1974).
[45] I. S. Kaburov *et al.*, in *Problems of the Metrology of Information-Measurement Systems in Physico-Technical Measurements* [in Russian] (VNIIFTRI, Moscow, 1982), p. 4.
[46] P. M. Gopych, in *Accurate Measurements in Nuclear Spectroscopy* [in Russian] (Mokslas, Vilnius, 1984), p. 94.
[47] P. M. Gopych and V. A. Deryuga, in *Questions About the Accuracy of Nuclear Spectroscopy* [in Russian] (Physics Institute, Lithuanian Academy of Sciences, Vilnius, 1984), p. 21.
[48] V. P. Borodyuk and V. N. Shkundenkov, Report R10-85-744, JINR, Dubna (1985) [in Russian].
[49] E. P. Velikhov, *Fifth-Generation Computers* [in Russian] (Financy i Statistika, Moscow, 1984), Preface.

Translated by Patricia A. Millard