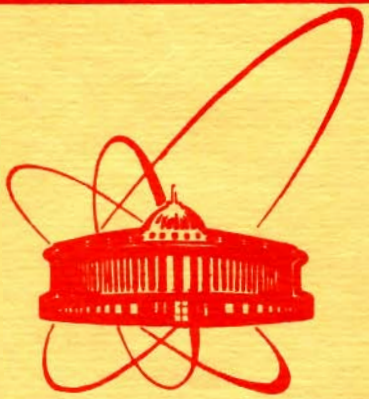


объединенный
институт
ядерных
исследований
дубна

E7 - 12219

G.Müller, D.Lehmann, G.Musiol, G.Zschornack

STATISTICAL LIMITS FOR THE UNFOLDING OF CLOSELY OVERLAPPING PEAKS IN THE X-RAY SPECTRUM ANALYSIS

1979

E7 - 12219

G.Müller, D.Lehmann, G.Musiol, G.Zschornack

STATISTICAL LIMITS FOR THE UNFOLDING OF CLOSELY OVERLAPPING PEAKS IN THE X-RAY SPECTRUM ANALYSIS

Submitted to "Nuclear Instruments and Methods"

Мюллер Г. и др.

E7 - 12219

Статистические пределы разрешимости близко расположенных пиков в спектрах характеристического рентгеновского излучения

В экспериментах по измерению характеристического рентгеновского излучения ионизированных атомов возникают суммарные пики, содержащие несколько линий. При разделении суммарного пика на отдельные пики (до 5 пиков) учитывается содержимое всех каналов в окрестности суммарного пика. Переопределенная система уравнений решается методом наименьших квадратов. С помощью числа обусловленности системы определена необходимая минимальная статистика, нужная для разделения суммарного пика. При этом отдельный пик моделировался гауссовым пиком или реальной аппаратурной функцией. Сглаживание суммарного пика не улучшает разрешимость.

Работа выполнена в Отделе новых методов ускорения ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна 1979

Müller G. et al.

E7 - 12219

Statistical Limits for the Unfolding of Closely Overlapping Peaks in the X-Ray Spectrum Analysis

The conditional number of the linear equation system for the unfolding of maximum 5 overlapping peaks with well-known positions was computed. The peaks were simulated not only by a Gaussian function but also by a more realistic function constructed with a random generator. Using the conditional number, the statistics and the area of the compound peak needed for good unfolding were evaluated. The chance of failure is increased by smoothing the compound peak.

The investigation has been performed at the Department of New Acceleration Methods, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna 1979

1. INTRODUCTION

The characteristic X-ray lines of ionized atoms get out of place as a function of the degree of ionization. One can determine shifts in the energies to a precision of the order of one thousandth of the natural line width [1]. But if a particle beam consists of ions with different ionization states, then for inner shell ionization processes a group of lines can be expected instead of one line the positions of which can be determined by the Dirac-Fock-Slater method [2]. The ionization state in electron-ion rings of the heavy ion ERA (JINR, Dubna) is a function of time [3], and the central point of an experiment for beam diagnostics [4] is to evaluate the time function of the ion charge spectrum in the electron ring by means of unfolding of measured compound X-ray peaks.

In the present paper we are interested in the error of the evaluated single peak heights after unfolding. This error is produced by the statistics of the compound peak and the single peak fit inaccuracy. At first a simple symmetric Gaussian function is taken for the single peak. This model is sufficient to investigate the error improvement in the solution of the equation system. But a better model is needed for good unfolding. For this purpose we propose to use the statistical counts of all channels in the range with a measured single peak in the center without fit, respectively. For this investigation we constructed single and compound peaks artificially with a random generator on a computer. In the case of very closely overlapping peaks the effect of the base line shape [5] and the change of the full

width at half maximum (fwhm) is sufficiently small. In all other cases the possibility for peak separation goes worsening.

2. SINGLE PEAKS WITH GAUSSIAN SHAPE

In the first type of investigation we accept that the sum peak contains n single peaks with Gaussian shape.

$$z(x) = \sum_{k=1}^{n} H_k e^{-\frac{(x-x_{ok})^2}{2\sigma^2}} \qquad (1)$$

in which x is the channel number, z are counts in this channel, H_k is the height and x_{ok} is the well-known position of the k-th peak, and σ is the standardized variance. If z is known, the problem of unfolding is to determine the single peak heights H_k. We have used the channels in the interval

$$(x_0 - 2\sigma) \leq x \leq (x_0 + 2\sigma) \qquad (2)$$

which are mostly far enough from the background, if x_0 is the sum peak channel with the maximum counts. Usually the number of equations is larger than that of unknown peak heights. For example, we have 41 equations if $\sigma = 10$. Therefore the system is overdetermined and must be solved with a minimization method. For this purpose a residual vector (R) is needed

$$(Z) = (A)\,(H) + (R) \qquad (3)$$

in the matrix style of writing. The matrix (A) contains in the columns discrete values of the Gaussian function shifted in the position of the k-th single peak in accordance with eq. (1). In our case it has a poor condition and the exactness of the solution is limited. Using a favourable solution method /6/ (orthogonalization of the column vectors) or double precision numbers, we can practically exclude the errors stipulated by a computer.

Therefore the reason for the inexactness of the solution is only the inaccuracy of the vector (Z) the components of which have statistical oscillations These oscillations are amplified by the unfolding, and we want to determine the multiplier for the amplification now. With this aim we assume that an exact vector (Z) gives an exact solution vector (H) and we solve only the system

$$(\Delta Z) = (A)\,(\Delta H) + (R) \tag{4}$$

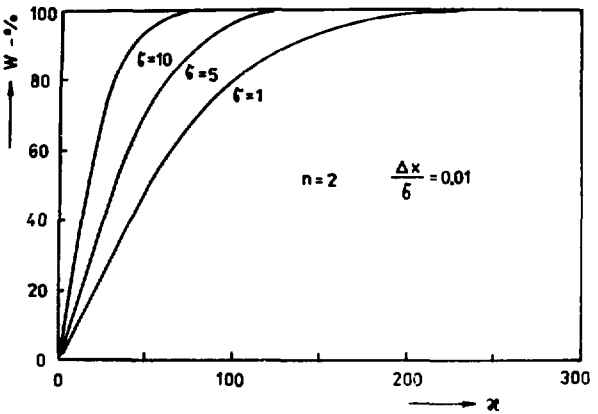
for some (we use 1 000) different vectors (ΔZ) the components of which are normally distributed. In accordance with the peak shape, we take

$$\rho = \sqrt{e^{-\frac{(x-x_0)^2}{2\sigma^2}}}\,\Delta z_{max} \tag{5}$$

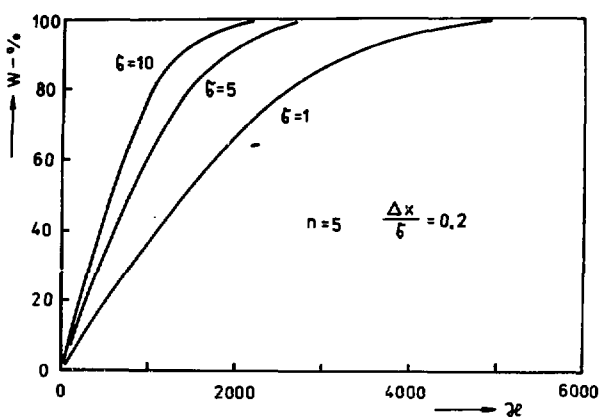
for the standardized variance in the distribution function. Taking only the amount of the largest component in the solution vector (ΔH), the multiplier is defined by

$$\kappa = \frac{|\Delta H_{max}|}{\Delta z_{max}} \tag{6}$$

and is also a rate for the condition /8/ of the matrix (A) as in quadratic matrices. Therefore we want to call κ the condition number of (A). The distribution of κ is approximately normal and the integral gives the probability W for a given upper boundary of κ. This function is shown in fig. 1 with two examples of the parameters n and $\Delta x/\sigma$ in which Δx is the distance between the single peaks. In the following we use the upper boundary for such κ with a probability of 68.3% and designate it as κ_σ. In fig. 2 κ_σ is shown as a function of the distance between the single peaks. For $n>2$ we assume that all distances are equal. One can see that in the double logarithmic graph all characteristics are straight lines. This means that we have the proportion



a)



b)

Fig. 1. Two examples for the probability W as a function of the upper boundary of κ. a) n = 2, b) n = 5 (all distances are equal).

$$\kappa_\sigma \sim \left(\frac{\sigma}{\Delta x}\right)^{n-1} . \qquad (7)$$

Using quadratic matrices, we have also proved the proportionality for n=2 and n=3 analytically.

In the case of a given inaccuracy for the single peak heights it is possible to determine the area of the sum peak needed for the unfolding. For this purpose we introduce the well-known equation for the statistical oscillations of the counts

$$\Delta z_{max} = \sqrt{z_{max}} \qquad (8)$$

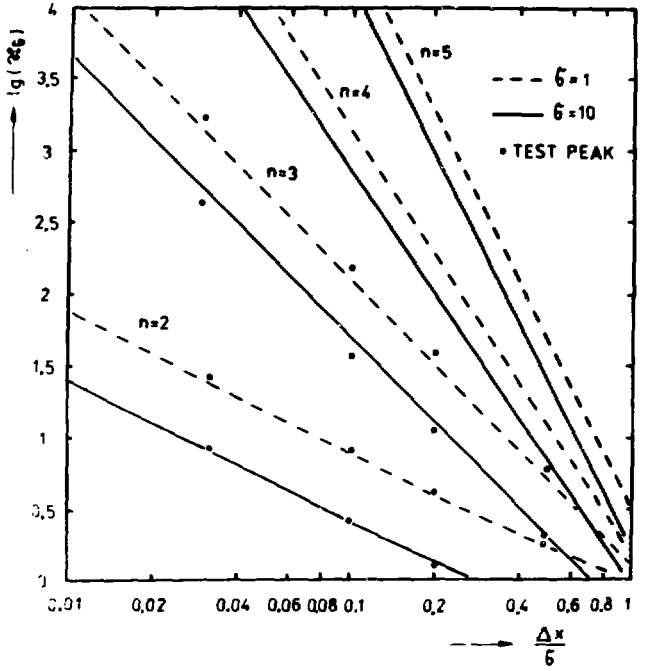


Fig. 2. Conditional number as a function of the single peak distance.

into eq. (6) and obtain

$$\frac{|\Delta H_{max}|}{z_{max}} = \frac{\kappa_\sigma}{\sqrt{z_{max}}} \cdot \tag{9}$$

Finally

$$z_{max} = \left(\frac{\kappa_\sigma}{\Delta H_{max}/z_{max}}\right)^2 . \tag{10}$$

Multiplying by 2.5σ we obtain an approach to the peak area

$$F \approx 2.5\sigma \left(\frac{\kappa_\sigma}{\Delta H_{max}/z_{max}}\right)^2 . \tag{11}$$

which is sufficient for an error less than $\frac{\Delta H_{max}}{z_{max}}$ with a 68.3% probability. In many cases we have studied the function $\sigma \cdot \kappa_\sigma^2$ and found that the dependence on σ is very small. Figure 3 demonstrates this for two examples.

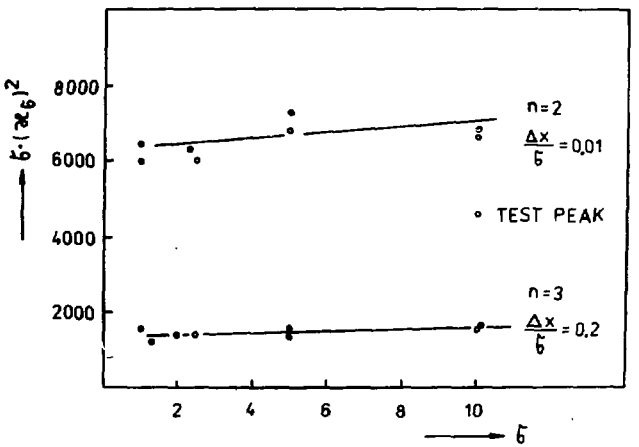


Fig. 3. The main point of the peak area as a function of σ. Hence it follows that the effect of σ or the number of channels is small in the interval (2) for the unfolding.

3. NONGAUSSIAN PEAKS

Usually in spectral analysis practice X-ray peaks do not have an exact Gaussian shape and in modern programs fits are used with two half-Gaussian curves with different widths connected at their tops. But for good unfolding a better approximation is needed because small differences between the real peak and its approximation are multiplied by the algorithm and the results are unsatisfactory. For a mathematical description we introduce a difference matrix (ΔA) into eq. (3) and obtain

$$(Z) = [(A) + (\Delta A)](H) + (R). \tag{12}$$

One can see that $(\Delta A)(H)$ is an additional error vector which will be also multiplied by κ. Therefore an exact approximation of the single peak shape is the main condition for good unfolding and we propose to construct the matrix (A) itself by a single test peak measured with a suitable source in the area of our interest. In order to construct the k-th column of (A) this test peak must be shifted to the position x_{ok} computed by the Dirac-Fock-Slater method. For shifts smaller than one channel an interpolation is necessary. The interpolation is a problem because we do not have an even function but discrete counts with statistical oscillations only. Therefore it seems advantageous to construct a polynomial function of corresponding degree through some points, and it is easy to determine the function at any positions. There is one more reason because we have found that the matrix (A) is often singular if $p<n$ in a p-point interpolation. Therefore we used a 5-point interpolation and saw that the peak shape continues invariably. We examined this method on a computer generating the test peak by the formula

$$x = \left(\sum_{k=1}^{12} \mathrm{RANF}(k)/\sigma\right)\sigma + x_o + 0.5. \tag{13}$$

where x is an integer and RANF is a random procedure for numbers between 0 and 1. The difference

between the test peak and an exact Gaussian peak is small but if thus constructed spectra consist of some independent single peaks, it is usually impossible to unfold them on the basis of eq. (1). On the other hand, the results were better using eq. (13) for construction of the matrix (A) as well. Based on this, we further evaluated the conditional number κ_σ and the function $\sigma \cdot \kappa_\sigma^2$. The results designated by "Testpeak" are shown in figs.2 and 3, respectively, and are in good agreement with the exact Gaussian peak investigation. The differences were greater only for $n>3$. Of course, in spite of better approximation the statistics of the test peak has an influence on the total error. In this case one must use for κ the multiplier

$$1 + \sqrt{\frac{z_{max}\ (\text{Sum peak})}{z_{max}\ (\text{Test peak})}} \ . \qquad (14)$$

We present now an example which demonstrates the efficiency of the method. An artificially composed compound peak consists of three single peaks with $\sigma = 5$ and the specification in the table. The area of the test peak for construction of the matrix (A) was 2 000 000.

Table

Specification of the compound peak and computed single peak heights, $n = 3$, $\Delta x/\sigma = 0.2$

k	k_{ok}	F_k (area)	H_k (exact)	H_k (unfolded)
1	99	100 000	7 871	7 664
2	100	200 000	15 742	14 970
3	101	400 000	31 483	32 453

One can see that the maximum error (in the third component) with respect to z_{max} is only 2%. Using eq. (9), we can expect an error less than 8% with a 68.3% probability. This means that we

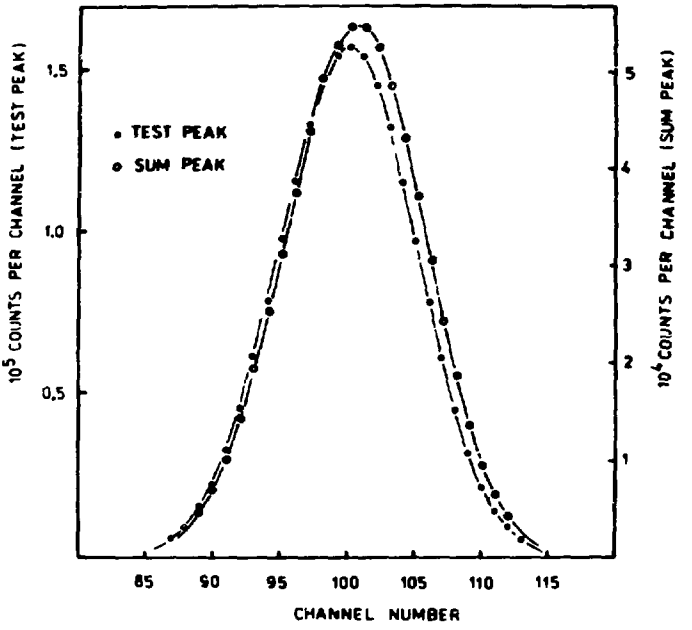


Fig. 4. Sum and test peaks.

have had a favourable statistics. The sum and test peaks for this example are shown in fig. 4. By the way, all peaks were constructed by various sequences. Therefore they are statistically independent.

In the last type of this investigation we have studied the effect of a smoothing procedure based on a five-point-mesh on the unfolding. Many computed examples have shown that the accuracy mostly goes worsening using the method of smoothing. Only in some cases we have observed an independence of the smoothing. Figure 5 shows for the above example that the separated single peak heights very strongly vary with smoothing cycles and one obtaines unsuitable results.

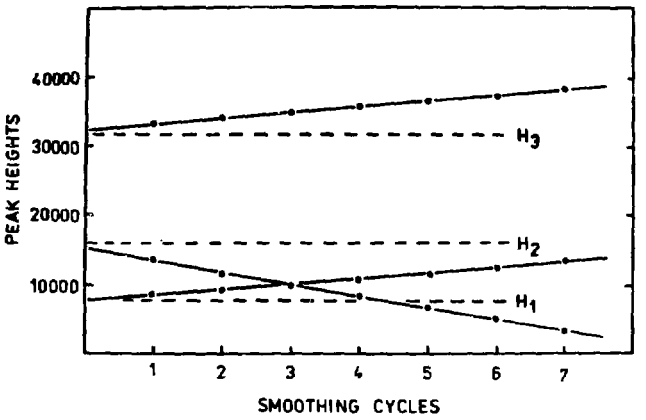


Fig. 5. Effect of smoothing on the unfolding accuracy.

4. CONCLUSIONS

In the case of well-known single peak positions, we have determined the area of the sum peak which is necessary for good unfolding. Using a high resolution spectrometer with a high intrinsic Ge detector, the normalized X-ray shifts $\Delta x/\sigma$ for multiple-ionized atoms are of the order of 0.01-0.2. In a real time one can obtain a peak area of the order of 10^7 counts which corresponds to a κ_σ of about 10^2 for a 10% inaccuracy of the single peak heights. Than it is impossible to unfold a sum peak in more than three single peaks at the present time. We have a chance to expand this limit only for higher counting rates and a better resolution of the spectrometer.

REFERENCES

1. Hansen P.G. Nucl. Instr. and Meth., 1978, 154, p.321-326.
2. Siebert H.U. et al.JINR, P9-9657, Dubna, 1976.

2. Lehmann D. et al. JINR, P9-10744, Dubna, 1977.
4. Siebert H.U. et al. JINR, P9-9366, Dubna, 1975.
5. Baba H. et al. Nucl. Instr. and Meth., 1977, 145, p. 517-523.
6. Schwarz H.G. et al. Numerik symmetrischer Matrizen (BSG B.G. Teubner Verl. Leipzig, 1969).

Received by Publishing Department
on January 31 1979.