Information retrieval methods for high resolution γ-ray spectra

WU Hexi[1,2,*] ZHANG Huaiqiang[2] LIU Qingcheng[1,2,*] YANG Bo[2]
WEI Qianglin[2] YUAN Xinyu[2]

[1]*Engineering Research Center of Nuclear Technology Application (East China Institute of Technology, Fuzhou) Ministry of Education, Nanchang 330013, China*

[2]*School of Nuclear Engineering and Geophysics, East China Institute of Technology, Fuzhou 344000, China*

Abstract A program based on MATLAB 7.0 platform was developed to locate characteristic peak position and calculate net area of characteristic peak. The formula for the calculation of relative standard deviation of net peak area by Sterlinski's method was found excellent in searching single peaks and resolving overlapping peaks in high resolution gamma-ray spectrum. Gaussian function fitting method using Levenberg-Marquardt algorithm was applied to calculate net area of peaks. A standard test spectrum supplied by the IAEA in 1995 was analyzed by the program and another two widely used commercial software. The analysis results show the program was superior to the latter two in searching single peaks and resolving overlapping peaks. The optimized fitting indexes are found between 0.962 and 0.996, which shows that the program adopted is feasible and accurate for extracting the net peak area in high resolution gamma-ray spectra.

Key words High resolution, Peak searching, Gauss fitting, Levenberg-Marquardt algorithm

1 Introduction

Gamma decay is usually accompanied by beta or alpha decay, and the induced gamma-ray is found to be generated by the interaction between neutrons and the matter. Therefore, the analysis on γ-ray energy spectrum can be regarded as an important means for identifying the kinds of radioactive nuclide and its content in most cases. The energy resolution of γ-ray spectrometer is a primary factor in accuracy determination of gamma radionuclide analysis. Meanwhile, high precise energy spectrum analysis software is also required. Qualitative analysis of γ-ray spectrometer is based on the accurate localization of characteristic peak position of gamma radionuclide, while quantitative analysis is determined by accurate calculation of net area of characteristic peak especially in the case of multiple. However, many of weak peaks are lost and a lot of false peaks are found in analysis reports of Genie2000 and GammaVision 32[1], which are commercially available and have been widely applied in γ-ray energy spectrum analysis[2], and the net area of single peaks of an overlapping peak can not be calculated. The above commercial packages are often set up some 'black boxes' to users, by which users are unable to get more useful information. In this paper, a program is introduced to locate characteristic peak position and calculate net area of characteristic peak especially for overlapping peak, which can provide an accurate basis for analysis and help to determinate radioactive nuclide emitting γ-ray.

2 Peak Searching Scheme

When net peak area was calculated by Covell's method (COL_m) and Sterlinski's method (STE_m), which are described as relative standard deviation (RSD) of net peak area respectively[3]. And m is the offset of the peak position to its boundary.

Supported by National Natural Science Foundation of China (No. 41174089 and 41164003) and Open-ended Foundation (No. HJSJYB2010-07) from the Chinese Engineering Research Center.

* Corresponding author. *E-mail address:* qchliu@ecit.cn

Received date: 2011-12-05

COL_m and STE_m were only used to determine width of peak, which had been discussed in Ref.[3]. In these processes, the m corresponding the minimum COL_m and STE_m was picked out. The width of peak was $2m+1$. On the contrary, suppose the m was fixed, if COL_m and STE_m were calculated through the whole spectrum, the abnormity may appear in peak. It is very likely that COL_m and STE_m can be used to locate the peak position in high resolution gamma-ray spectra. In this section, COL_m and STE_m were studied in peak position identification. The detailed analysis is as follows.

Ideal characteristic peak of γ-ray spectrum is described as Eq.(1)

$$y_i = \sum_{j=1}^{k} A_j \cdot e^{-\left(i-m_{p,j}\right)^2 / 2\sigma_j^2} + B_i \qquad (1)$$

where the parameters of j-th Gaussian function are A_j, m_{pj} and σ_j. k is the total number of ideal Gaussian peak. B_i is the background counting in i-th channel. An ideal characteristic peak of γ-ray spectrum can be obtained by equation 1 with k=1, A_1=200, m_{p1}=56, σ_1=2 and $B_i \equiv$ 40. It has been transformed by COL_m and STE_m under m=2, and the results are shown in Fig.1.

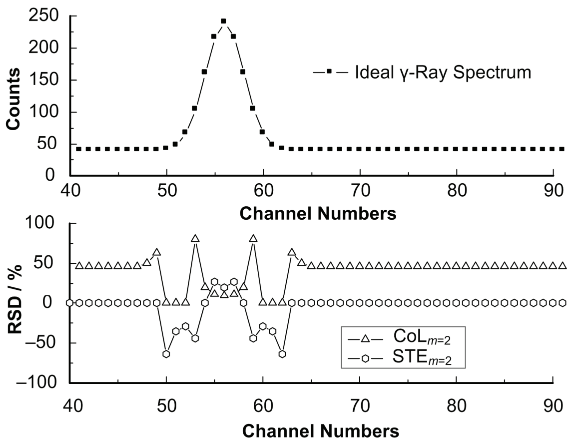

Fig.1 Transformed results of ideal γ-ray spectrum (A=200, m_{p1}=56, σ=2, B_i≡40) by COL_m and STE_m

From Fig.(1), the results transformed by STE_m appear positive anomaly around the peak position; and by COL_m around the peak position and at the boundaries. It was considered preliminarily that STE_m can be regarded as peak-seek algorithm. But its resolving ability to overlapping peaks remains undiscovered. It is studied as follows.

An ideal two-overlapping peaks of gamma-ray spectrum is calculated by equation 1with k=2, A_1=300, m_{p1}=50, σ_1=2, A_2=100, m_{p2}=56, σ_2=2 and B_i≡30. The results transformed by STE_m are shown in Fig.2. (1) Positive anomaly discontinuities occur twice in the width of above double peak while m=2, which indicates that this method has the ability to resolve the overlapping peak. (2) Only one continuous positive discontinuity is found in the transformed results while m=6. In the follow-up peak-seek processing, we should firstly determine characteristic peak in the energy spectrum for following analysis under $m=3\sigma$, then find out the number of overlapping peaks under $m=\sigma$. (3) Because the transformed result has no meaning for searching peaks when it's less than 0, it will be set zero by program in subsequent calculations. In fact, the program just needs the maximal value within a successive positive value, the others will be removed.

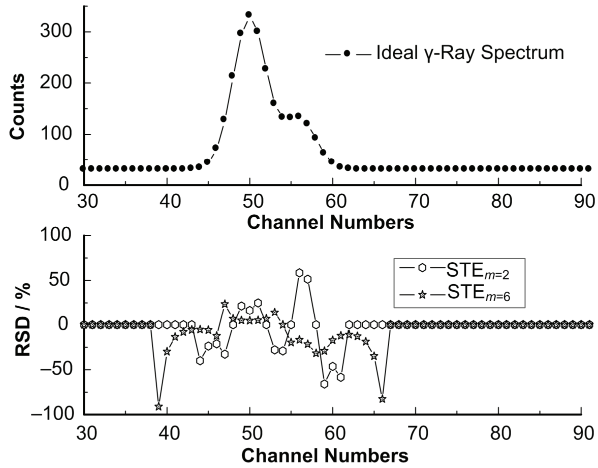

Fig.2 Transformed results of overlap of peaks in ideal γ-ray spectrum by STE_m

In the above discussion, scattering background functions are constants. When the scattering background line has statistical fluctuation, it is necessary to verify the validity of the transformation results. So we use a formula [3] instead of a constant to describe the scattering background spectrum (Eq.2).

$$B_i = b + [-2b\ln(x_1)]^{0.5}\sin(2\pi x_2) \qquad (2)$$

where B_i is the background counting in i-th channel. b is the average counts of scattering background. x_1, x_2 are random numbers between 0 and 1, which were acquired by Mersenne Twister algorithm[5] in the follow-up program.

The following steps were used to compile program in MATLAB 7.0 platform.

(1) 10 001 pieces of scattering background spectrum were obtained by Eq.(2) when $b=i\times100$ and i varied from 1 to 10 000 at the total channels of 16 384.

(2) All of the above scattering background spectrum were disposed by the optimal filtering algorithm[1] to reduce the influence of statistical noise.

(3) Each scattering background spectrum was transformed by STE_m with m varying from 1 to 15, the maximum $MAX_{m,b}$ of the transformed results can be acquired with fixed b and m.

(4) The maximum and minimum of $MAX_{m,b}$ were acquired from different scattering background levels with the same m, described as MAX_m and MIN_m respectively, the results are shown in Fig.3.

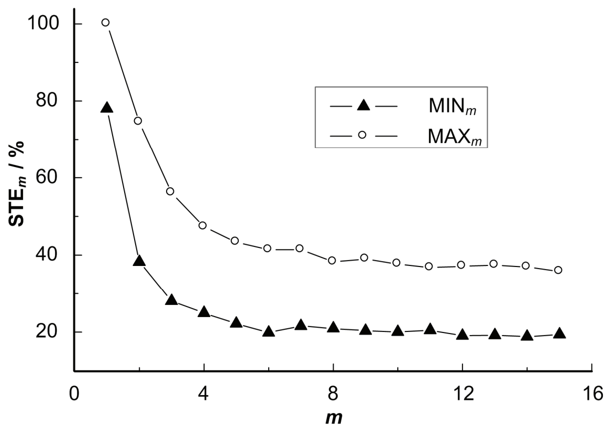

Fig.3 Calculated results of background spectrum from Spravochnik's function by STE_m

From Fig.3, the MAX_m and MIN_m decrease with m. For a certain m, if the maximum value MD in the continuous positive abnormal points of the transformed results is greater than MAX_m, there is a peak. If MD is less than MIN_m, no peak can exist at the channel. Otherwise, there may be a suspected peak.

The capacity of removing false peaks and, at the same time, retaining real peaks are primary factors in judging the peak searching algorithm. Detection limit is used to distinguish suspected peak in this paper. The specific steps are as follows.

(1) The scattering background of full spectrum is deducted by SNIP method[1,6,7].

(2) If the peak width is less than the FWHM, the suspected peak will be removed.

(3) If the peak width is not less than the FWHM, net area of discovered peak is calculated by Gaussian fitting using Levenberg-Marquardt Method.

(4) Input confident probability $(1–\alpha)$ into the program to calculate the confident interval k_α; and the detection limit (L_D) by Eq.(3). If net area of discovered peak is lager than L_D, the peak will remain. Otherwise the peak will be removed.

$$L_D=k_\alpha^2 +2k_\alpha(2I_B)^{0.5} \quad (3)$$

where I_B is the total counts of scattering background in the range of width of discovered peak.

3 Gaussian Function Fitting Method

There are lots of methods in calculating characteristic peak area at present. Total peak area method is widely used in most commercial software, in which fitting function of scattering background is step function. With the development of computer technology, Gaussian fitting method is increasingly used to calculate net area of single-peak or overlapping peak[8–10]. However, different authors used different fitting algorithms, such as three point Gaussian fitting, Newton method, and so on. The full spectrum scattering background was deducted by SNIP method in this article. Then, the remaining counts were fit by Levenberg-Marquardt algorithm[11] in the range of width of peak. The fitting quality was evaluated by the fitting optimization index (R^2)[12].

$$R^2=1-[\Sigma(y'_i-y_i)^2]/[\Sigma(y'_i-y_p)^2] \quad (4)$$

where R^2 is the correlation coefficient, y_i' is the fitting data, y_i is the original data, and y_p is the average of original data.

The initial parameters of Gaussian function must be inputted before fitting by Levenberg-Marquardt algorithm. Here, we got the information of the peak position and peak height by the peak searching method described above. The initial value of FWHM at the same channel was calculated by calibration curve of FWHM which was obtained by interpolation method[13]. Before the fitting to overlapping peaks, the initial FWHM should be obtained by the FWHM of the calibration curve at the maximum counting.

4 Experimental Analyzing

A standard spectrum namely STRAIGHT was applied to test peak searching and net area fitting method in this paper. This spectrum is widely used to test peak searching algorithm supplied by the IAEA in 1995[14]. The IAEA also gives both exact peak position and net area of all peaks in the STRAIGHT spectrum. Accordingly, the peaks in STRAIGHT spectrum were classified by net area of peak. If the net peak area is larger than 400, the peak is classified as super-peak. The peak is defined as strong peak with the net peak area varying from 100 to 400. Likewise, the weak peak has net peak area varying from 10 to 100. All other peaks are classified as ultra-low peak.

Table 1 Analysis results of the program and several γ-ray spectrum analysis software for test spectra STRAIGHT.

Information of STRAIGHT test spectra (Range characteristic peak by net area S)			GammaVision 32	Genie2000			this paper
				ST^*=2	ST=1.5	ST=1	
Super-high peaks	S>400	41	40	25	33	39	41
High peaks	400>S>100	46	22	6	8	19	28
Weak peaks	100>S>10	71	2	1	4	8	10
Ultra-low peaks	10>S>0	15	0	0	0	0	0
Total		173	64	32	45	66	79
Misses			109	141	128	107	94
False peaks			1	1	18	37	1

*ST is significance threshold for peak searching.

Two kinds of commercial gamma-ray spectrum analysis software, Genie2000 and GammaVision 32, were used on STRAIGHT spectrum and the analysis results were shown in Table 1. Before searching peaks by Genie2000, users must set the value of significance threshold to remove the false peaks in terms of past experiences. This is a great challenge for new users. The larger threshold, the fewer false peak, but more real peaks will be removed at the same time. In the capability of both removing false peaks and retaining real peaks, GammaVision 32 is better than Genie2000 by comparing analysis results in table 1. Only one super-peak is left out by GammaVision 32, which is an overlapping peak at the positions of 683.3 and 685.7. All other overlapping peaks are distinguished by GammaVision 32, such as peak positions of 871.8 and 879.4, 967.1 and 972.4. The above analysis shows that the resolution capability of overlapping peak should be strengthened when two peaks are close to each other.

The program was proposed in this article by following steps: (1) The FWHM at mid-channel of gamma-ray spectrum was obtained after FWHM calibration, defined as $FWHM_m$. (2) Filtering instrument spectrum. (3) Transforming instrument spectrum by STE_m under m=round ($3\times FWHM_m/2.355$). Both real peaks and suspected peaks in above instrument spectrum were marked by comparing transformation results with MAX_m and MIN_m. (4) The above step was repeated under m=round ($FWHM_m$/ 2.355). (5) The channel number that corresponds to the maximal value within the range from m_p-round ($1.5\times FWHM_m$) to m_p+round ($1.5\times FWHM_m$) was found to replace peak position m_p with step 3 and 4, and the peak was marked as single peak. If the same channel was obtained at neighboring peak position, the channel number at sub-maximum was used to replace the later peak position. Then the two peaks were marked as an overlapping peak. Like this, all overlapping peaks can be verified and marked. (6) If a peak was verified under above different m conditions and was marked as single peak, the peak is judged as a real peak. Net-area of all other peaks was acquired by the above Gaussian function fitting method. Finally, detection limit was calculated to distinguish all other peaks.

The above steps were used to develop program in MATLAB 7.0 platform and analysis results were shown in the last column of Table 1. In Table 1, the developed program can distinguish all overlapping peaks, and the capability of both removing false peaks and retaining real peaks are superior to the above two widely used commercial software.

Gaussian function fitting method was applied in full spectrum and the optimal fitting indexes were between 0.962 and 0.996, which shows that the instrument spectrum is extracted accurately by Gaussian-function fitting method. A segment of fitting results were shown in Fig.4.

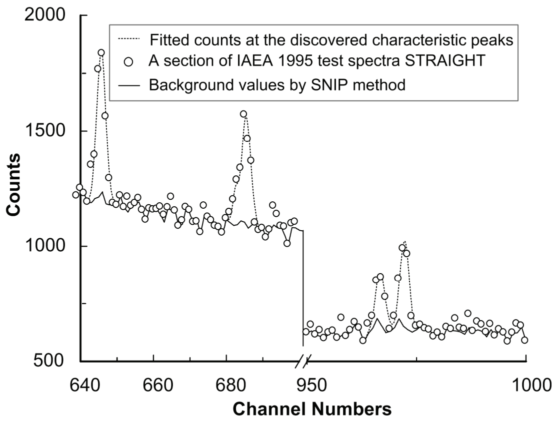

Fig.4 Gamma-ray characteristic peaks by Gaussian- function fit method after deducting the background by SNIP method based on Levenberg-Marquardt method.

5 Conclusions

By analyzing RSD of net area of ideal Gaussian peak, positive anomaly was found around the peak position. The transformed results of an overlapping peak show that positive anomaly discontinuity occurs twice in the width of the overlapping peak under $m=\sigma$. These primary analyses indicate that STE_m can not only search single peak but also has the ability to resolve an overlapped peak.

To validate the feasibility of the algorithm when scattering background has strong statistical fluctuations, we simulated 10 001 pieces of scattering background by Spravochnik formula with b varying from 100 to 1 000 000. By analyzing the scattering background, MAX_m and MIN_m were obtained. At a certain m, if the maximum value MD in the continuous positive anomaly points of the transformed results is greater than MAX_m, there is one peak at least. If MD is less than MIN_m, no peak exists at the channel. If the value of MD is between MAX_m and MIN_m, there may be a suspected peak, which can be detected by detection limit. So this method can overcome the effect of scattering background.

A program based on MATLAB 7.0 platform was developed and the validity was tested by STRAIGHT spectrum (provided by IAEA in 1995). The test showed that the program proposed in this paper can distinguish all overlapped peaks, and the capability of both removing false peaks and retaining real peaks is better than Genie2000 and GammaVision 32. At the same time, this program has an algorithm to calculate net area of any peaks. The optimal fitting indexes were from 0.962 to 0.996 by test, which proved the accuracy of the net peak area calculation algorithm. But the validity in quantitative analysis needs further study.

References

1 Wu H X, Yuan X Y, Liu Q C, *et al.* Atom Energy Sci Technol, 2012, **46:** 1142–1146.

2 Xie F, He X B, Jiang W G, *et al.* Nucl Sci Tech, 2011, **22:** 304–306.

3 Pang J F. gamma-ray energy spectrum data processing. Xian, Shanxi scientific and technical Publishers, 1990, 365–375. (in Chinese)

4 Abramovitz M, Stigan I. Spravochnik po special'nym funkciyam (Handbook of Special Functions). Moscow, Nauka, 1979.

5 Matsumoto M, Nishimura T. ACM transactions on modeling and computer simulation, 1998, **8:** 3–30.

6 Wu H X, Liu Q C, Yang B, *et al.* Nucl Tech, 2010, **33:** 513–516.

7 Morháč M, Matoušek V, Turzo I, *et al.* Nucl Instrum Meth Phys A, 2006, **559:** 76–80.

8 Josef U, Gerg R, James T. Nucl Instrum Meth Phys Res A. 2010, **619:** 457–459.

9 Fu C, Wang N P. Nucl Sci Tech, 2010, **21:** 214–217.

10 Li J Q, John D V, Asad E R. Nucl Instrum Meth Phys Res A. 1999, **422:** 438–443.

11 Madsen K, Nielsen H B, Tingleff O. Technology report: Informatics and mathematical modeling, Technical University of Denmark. 2004.

12 Ni Z C. Medical statistics. Beijing, People's Medical Publishing House, 1990, 143–148.

13 Markevich N, Gertner N I. Nucl Instrum Meth Phys Res A. 1989, **283:** 72–77.

14 Blauuw M. Nucl Instrum Meth Phys Res A. 1999, **432:** 74–76.