



Determination of Gamma-Ray Abundance in Photopeaks with Analytic and Numeric Approaches via CERN's ROOT Analysis Framework

Ilker Can Celik^{1*}

^{1*} Harran University, Faculty of Art and Science, Department of Physics, Sanlıurfa, Turkey, (ORCID: 0000-0002-2320-6584), ilkercan0066@harran.edu.tr

(1st International Conference on Applied Engineering and Applied Natural Sciences ICAENS 2021, May 10-13, 2022)

(DOI: 10.31590/ejosat.1107756)

ATIF/REFERENCE: Celik, I. C. (2022). Determination of Gamma-Ray Abundance in Photopeaks with Analytic and Numeric Approaches via CERN's ROOT Analysis Framework. *European Journal of Science and Technology*, (36), 160-167.

Abstract

For years in gamma-ray spectrometry, there have been numerous numeric methods implemented and dedicated to integrate the photopeak areas. With the advance of new technology and software, determination of peak areas has been digitized. In addition, new statistical approaches have given diverse perspectives in analysing the experimental data. In this case study, Maestro software was used to extract the experimental gamma-ray data from radioactive ⁶⁰Co source. In general, analysts have been challenged by the calculation of background area under certain gamma-ray peaks when there are especially low counts. For instance, this problem leads them to find different net area values in various accuracies for the same peak. Therefore the precision of determination based on different statistics might affect the results. The systematic error, which is caused by the shapes and models of the background estimation, has been investigated by CERN's ROOT software. This detailed analysis of digital data from NaI(Tl) scintillation detector has been tabulated with its comparisons in various modern analytic methods and numeric approaches dated back to 1960s and 1970s.

Keywords: Gamma-ray spectrometry, photopeak, Maestro, ROOT, background estimation, net area.

Farklı Analitik ve Nümerik Yöntemlerle Fotopiklerdeki Gama Işını Miktarının CERN'ün ROOT Analiz Programıyla Hesaplanması

Öz

Yıllardır gama ışını spektroskopisinde, fotopik alanlarının integrasyonu için birçok nümerik yöntem geliştirilmiştir. Yeni gelişen teknoloji ve yazılımlarla, fotopik alanlarının hesaplanması dijital ortama aktarılmıştır. Ek olarak, yeni istatistiksel yöntemlerde deneysel verinin analiz edilmesinde farklı bakış açıları sağlanmıştır. Bu örnek çalışmada, Maestro yazılımı radyoaktif ⁶⁰Co kaynağından deneysel verinin alınması için kullanılmıştır. Genel olarak, düşük sayım sayısı olan durumlar belirli piklere ait taban sayımı hesaplamalarında analiz yapanların işini zorlaştırmaktadır. Örneğin, bu sorun onların aynı pik için farklı duyarlılıklarda net alan hesaplamalarına yol açmıştır. Bu yüzden, farklı istatistiksel yaklaşımlara bağlı bulunan doğruluk paylarının sonuçlara etkileri olacaktır. Taban sayımının hesabında kullanılan şekiller ve modellerden kaynaklanan sistematik hatalar, CERN tarafından geliştirilen ROOT yazılımı ile araştırılmıştır. NaI(Tl) sintilasyon detektörü ile yapılan bu dijital verinin detaylı analizi, 1960 ve 1970'lere uzanan nümerik metotlar ve modern analitik metotlarla karşılaştırılmalı olarak tablo halinde verilmiştir.

Anahtar Kelimeler: Gama ışını spektroskopisi, fotopik, Maestro, ROOT, tabansayımı tahmini, net alan.

* Corresponding Author: ilkercan0066@harran.edu.tr

1. Introduction

Statistics in counting gamma-rays in specific are unavoidable mathematical concept which nuclear physicists will encounter naturally. In most laboratories, nuclear physicists investigate randomly decaying radioactive substances. In addition to the counting process, this randomness also requires a statistical approach by nature. Therefore, all measurements become indeed just an estimate of a decay rate. In gamma-ray spectroscopy, counts versus channel number plots in a raw data are the first to analyse. Then, this raw data can be calibrated in terms of energies. For this study, analysis will be processed on channel numbers for the simplicity. Every measured rate in each event can be related directly and proportionally to the number of atoms in the radioactive source to define the activity. The integration of peak area measurements, namely the total counts under full gamma-ray peaks, is the main input to reach the activity of any radioactive substance. Some of the theoretical tools for this investigation have been different versions of Covell methods [1,2,4,5,8] and Total Peak Area (TPA) methods [2,3, 4, 6,7,8]. For this study, the ROOT coding via fittings by using Gaussian functions for the peak and polynomial functions for the background will be used to modernize the old procedures. For this perspective, standard deviation, coverage factor and confidence limit concepts will be introduced to make the future analyses more efficient and user friendly.

2. Methods

In general, main issue of integrating a peak area in gamma-ray spectroscopy was the background estimations for the peaks under overlapping circumstances. It is because statistics takes place in error calculations and every method has pros and cons in this perspective. However, gross area calculation is quite straightforward as it suggests taking all the counts under certain peak limits. For the rest of the article, the letter A will stand for either net count or net area under a photopeak. To show the count number, the letter a with subscript to denote the corresponding bin number will be used. The lower and upper limits of the peak region are shown as L and U.

Covell Method

In early years of statistical analysis digital gamma-ray spectrometry, Covell introduces us a pretty straight forward method to estimate the net counts under a photopeak. Then, this simple but elegant formula transformed itself to new forms. Let's start with Covell's original idea. According to Gilmore [6], Covell suggested to locate the highest counts in channel at first and then marking the lower and upper limits of the peak region as equal number of channels away from the centremost channel. However, this method had its own obstacles when there are overlapping peaks nearby in the usage of detectors with low resolutions. The spectrum in this article was also taken from NaI(Tl) detector. In the original Covell's method, the formulation was summarized as follows [1]:

$$a) \quad A = a_0 + \sum_{i=1}^n a_i + \sum_{i=1}^n b_i - [(2n + 1) \times \frac{(a_n + b_n)}{2}]$$

where a_0 is the highest number of accumulated counts in the centermost channel, a_i is the count in channel i in the low amplitude side of the peak whereas b_i corresponds to the count in channel i in the high amplitude side of the peak. Covell estimated the peak in approximately Gaussian shape with the total of $2n+1$ independent bins. Then, the variance of the A term can be calculated as follows:

$$Var(A) = a_0 + \sum_{i=1}^n a_i + \sum_{i=1}^n b_i + [(\frac{2n+1}{2})^2 \times (a_n + b_n)]$$

However, Gilmore's interpretation of the Covell method was slightly different than the original approach as follows [6]:

$$b) \quad A = \sum_{i=L}^U a_i - [(U - L + 1) \times \frac{(a_{L-1} + a_{U+1})}{2}]$$

Thus, the variance of the net area can be stated as below.

$$Var(A) = \sum_{i=L}^U a_i - [(\frac{U-L+1}{4})^2 \times (a_{L-1} + a_{U+1})].$$

Here, the gross area was the sum of all counts under the bin interval between L and U on the uncalibrated spectrum as shown in Figure 1 for the guidance. On the other hand, the background area was calculated by taking the mean count of one adjacent bin outside the peak region instead of the last bins on both end of the peak. Since the different interpretation and the perspective of the formula matters in statistics, other forms of the Covell's equation should be emphasized here. One of them was as follows [5]:

$$a) \quad A = \sum_{i=-n}^n a_i - [(2n + 1) \times \frac{(a_n + a_{-n})}{2}]$$

where n is the number of channels on each side of the channel zero with maximum number of counts (a_0). Thus, this equation suggests that the variance can be written as below:

$$Var(A) = \sum_{i=-n}^n a_i + [(\frac{2n+1}{2})^2 \times (a_n + a_{-n})]$$

Then, the second interpretation was shown as below [5]:

$$b) \quad A = [a_0 + \sum_{i=1}^{n-1} (a_{-i} + a_i)] - [(2n - 1) \times \frac{(a_n + a_{-n})}{2}]$$

Then, the variance may be written as:

$$Var(A) = N + [(n - \frac{1}{2}) \times (n + \frac{1}{2}) \times (a_n + a_{-n})] \text{ or}$$

$$Var(A) = [a_0 + \sum_{i=1}^{n-1} (a_{-i} + a_i)] + [(\frac{2n-1}{2})^2 \times (a_n + a_{-n})]$$

Lastly, the third variation of the Covell's formula was [5]:

$$c) \quad A = [\sum_{i=-(n-1)}^{n-1} a_i] - [(2n - 1) \times \frac{(a_n + a_{-n})}{2}].$$

which leads to the variance in a new form as shown below:

$$Var(A) = \left[\sum_{i=-(n-1)}^{n-1} a_i \right] + \left[\left(\frac{2n-1}{2} \right)^2 \times (a_n + a_{-n}) \right]$$

In all these forms of the same Covell method tries to gives diverse forms of the intensity of the radiation detected by the detection system. It's clear that all agrees on the total number of the counts under the peak, namely the gross area. In terms of background counts, all the calculations are following the same procedure except method b.

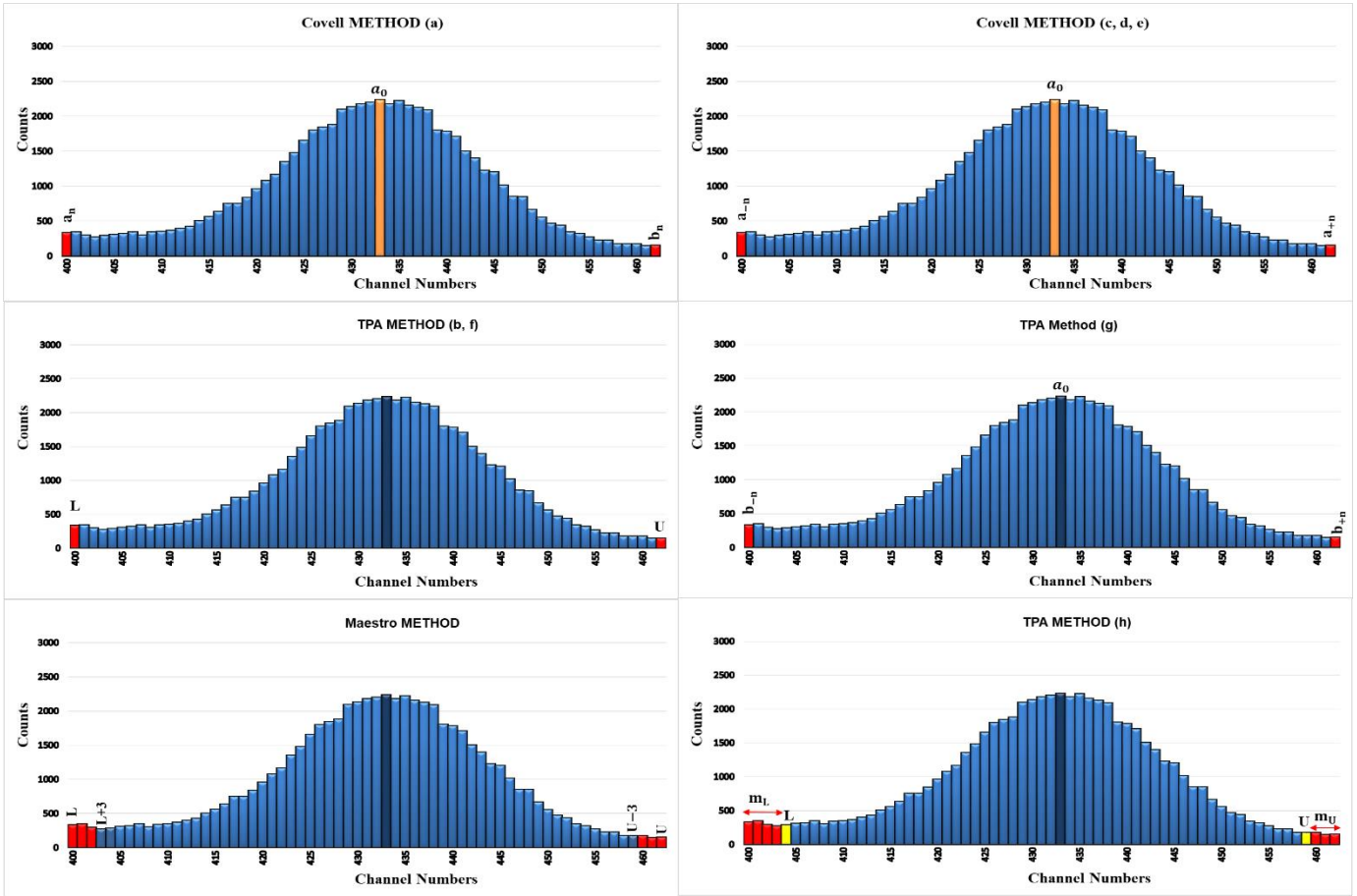


Figure 1: Visualization of different digital methods on an uncalibrated spectrum in various perspectives.

At this step, Gilmore proposes taking the first bin immediately beyond the peak limits. Ultimately, they all use the trapezium area beneath the peak. The interesting fact is the way how the variances are constructed differently as a formula. They all left the last bins on both ends of the peak separately, so the weight of the term $(a_n + a_{(-n)})$ manifests itself among the other equally weighted bin contents in evaluating the relative precision of the magnitude of the net area, i.e. $(\frac{\sqrt{Var(A)}}{A})$. For instance, the ratio of $(\frac{2n-1}{2})^2$ shows the disproportion in statistical weight. If the statistics and the ratio of the signal to background in the experiment are low, it's proven that the relative precision is better with more channel number covered by the peak. However, it's also known that the precision of the net area will not be badly affected by the accidental fluctuations of the weights in these extreme channels if the statistics are good enough [5]. In addition to Covell's original method stated here in equation 1, he also suggested the stripping method and the fraction method in his article [1], but these details will not be mentioned here. More advance calculations for finding the net area done by Sterlinski in 1968 [5] suggested to represent the radiation intensity in different size of discrete blocks of n where it represents the total channel number on both sides of the peak. The details can be found in its article. In conclusion, all these efforts are to decrease the coefficient of variation, namely the precision, in calculations.

TPA Method

Similarly to the Covell, the TPA method may give similar results and equations when the peak shape is assumed to have symmetrical Gaussian distribution. Here is the first suggested TPA formula [3,4,7,8]:

$$d) A = \sum_{i=L}^U a_i - [(U - L + 1) \times \frac{(a_L + a_U)}{2}]$$

The variance of this equation is as stated below:

$$Var(A) = \sum_{i=L}^U a_i + [(\frac{U - L + 1}{2})^2 \times (a_L + a_U)]$$

At first sight, it looks the same as the equation c for the Covell's method. The upper (U) and the lower (L) channel number at both hand of the photopeak were used to determine the total counts under the photopeak. However, it doesn't clearly state that the channel (bin) numbers have to be equal on both side of the centroid of the peak as symmetric Gaussian shape suggests. Since testing the shape of a peak is not the aim of this article, there will be no comment on a symmetry or asymmetry of a peak shape at this time. Later on, TPA method was modified by Wasson [4, 7] in private discussions as written below.

$$e) A = \sum_{i=-n}^n a_i - [(2n + 1) \times \frac{(b_n + b_{-n})}{2}]$$

where $(b_n + b_{(-n)})$ notation was used to differentiate the background counts in the last channels, denoted by $\pm n$ as subscript, on both side of the peak as determined from a straight line fit between the left and the right end channels. This was the detailed part of Wasson's modification because previous calculations took directly the content of the last bins as an example of $(a_n + a_{(-n)})$. That's one comparison aspect of it. The variance of the Wasson's modification can be stated as below:

$$Var(A) = \sum_{i=-n}^n a_i + [(\frac{2n + 1}{2})^2 \times (b_n + b_{-n})]$$

Another detail was that this method clearly indicated the symmetrical shape of a Gaussian function due to equal number (n) of channels on both side of the centremost channel of the peak. On the other hand, Gilmore came up with more flexible approach in the case of peak overlapping for low energy resolution detectors. In this method, the background count approximation was done by the user dependent channel number on both side of the peak region

beyond its limits. Therefore, it might resemble the extended version of the equation b as the following [6]:

$$f) \quad A = \sum_{i=L}^U a_i - [(U - L + 1) \times \frac{(\sum_{i=L-m_L}^{L-1} a_i + \sum_{i=U+1}^{U+m_U} a_i)}{(m_L + m_U)}]$$

The variance of this equation is shown as indicated below:

$$Var(A) = \sum_{i=L}^U a_i + [(\frac{U-L+1}{m_L+m_U})^2 \times (\sum_{i=L-m_L}^{L-1} a_i + \sum_{i=U+1}^{U+m_U} a_i)]$$

Here, Gilmore's TPA method introduces two new parameters as m_L and m_U to describe the extended bin numbers outside the peak area depending on the user's choice. This choice should be made by an investigation of neighbouring peaks. Thus, the analyst could be somehow flexible in terms of choosing different bin numbers on each side of the peak of interest. Another difference from the original Covell method was that Covell's background bins were the last bins within the peak region instead of outside the peak limits. While Figure 1 indicates how to perform the Gilmore's TPA method shown in equation h, it also emphasizes the effect of different bin numbers, namely m_L and m_U , for the estimation of the background region for a peak. It's known that the uncertainty after 3 bins starts to be stable [6, 10]. Therefore, the percentage in relative standard deviation when selecting more than 3 bins for the background calculations will not alter the results so much. For that reason, software like MAESTRO uses 3 channels symmetrically on both sides of the peak by default for its background calculation. Maestro gives also the option to set this number differently, but it's not variable as in the m_L and m_U values. They are always equal.

The Calculation behind the Maestro Software

The details of the working principle and the usage of Maestro software can be checked from its manual [9]. Here, only the calculations written by the Maestro manual will be mentioned. Note that the Maestro software results from its region of interest (ROI) report given by the program itself does not match with its manual calculations as stated below! This has to be checked by their authorities. For the net counts under the selected ROI was calculated as follow [9]:

$$g) \quad A = \sum_{L+3}^{U-3} a_i - [(U - L - 5) \times (\frac{\sum_{i=L}^{L+2} a_i + \sum_{i=U-2}^U a_i}{6})]$$

Thus, the variance of the net counts was written as below:

$$Var(A) = \sum_{L+3}^{U-3} a_i - \left[\left(\frac{(U - L - 5)}{6} \right)^2 \times \left(\sum_{i=L}^{L+2} a_i + \sum_{i=U-2}^U a_i \right) \right]$$

As said earlier, the Maestro software set the total number of channels as 3 on both sides of the peak region. For this reason, the software calculated the net area within the channel range of $U-L+1-6$. In that way, it removes the 6 channels in total from the user's peak limit selection in the case of overestimation while there is a neighbouring peak. Note that the Maestro software starts its channel number from zero. Thus, make sure the channel numbers in all other analysis match with each other.

Background Knowledge in Data Counting in Gamma-ray Spectroscopy

In the contemporary methods, the first assumption was about the shape of the curves in gamma-ray spectroscopy. Statistically in worldwide, the Gaussian shape for the signal was regarded as the most common function to implement. However, the Binomial and the Poisson distributions can be tested if needed. There are possibilities these distributions may also apply sometimes [6]: By nature, the radioactive decay possesses binomial character due to its duality about decaying or not decaying in a given time. Also, the frequency of the observation does not affect the probability of a radioactive source throughout counting. Each atom is independent from each other's decaying. As a contribution to this statement, Sterlinski [5] stated that one of the approximations was the events were independent in each channel of the detection

system. At this diversity, Gilmore also stated if the total decay number is unknown, then the Poisson distribution may statistically apply to these events. However, there is mathematically surprising fact which suggests the expected count will be non-zero in the case of no detection [6].

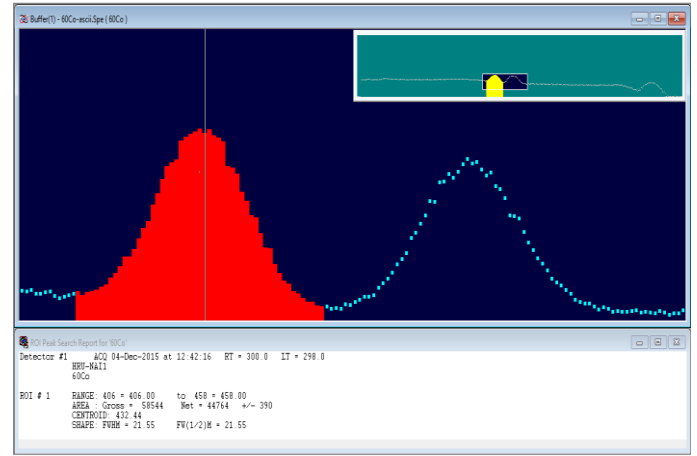


Figure 2. Visualization of a peak selection and its report in Maestro software on an uncalibrated spectrum.

Finally after some approximations and the details in counting time, total counts, signal to background ratio and detection efficiency, the agreement in representing the signal function can be settled down mostly on the Gaussian form. According to some papers mentioned in Kokta's article [7], the analyst should trust more on the real measurements by the analysis device and the detectors in the spectrum. It's because there might be undistinguishable theoretical impracticability due to time instability of the measuring device, Compton scatterings structure in counts contributed from higher energetic areas on to the low energy region, possible energy calibration errors and many others. Therefore, some modifications on Gaussian function were even suggested and experimented on the data in that article.

Suggested Contemporary Method in This Study

One important thing was that the peaks under investigation should normally be tested for their statistics in count numbers, so that the desired functions can be fitted. For instance, if the background to signal ratio is high, then we may have no sufficient statistics to fit the data as Maestro software is designed to give warning at this kind of occasions. Here in Table 1, the data does not have problem about this issue. For the analysis, some scientists do the analytical analysis through the data via the histogram, and others model the data distribution for a given channel limits via functions. All the methods explained above are regarded as the analytical bin to bin analysis. In gamma-ray spectroscopy, many methods and perspectives were devoted to find either the integration of the net peak area or the approximation of the accurate and precise background area. While the real data calculations in discrete values as counts of each individual bins at certain channels in a histogram of a spectrum stay on the more trustworthy side, the approximations in numerical calculations by fitting some well modelled functions give much broader flexibility and predictability to some very complex calculations. In this sense, both methods have their own advantages. Particularly, suggested method in this article is based on to model the signal shape as a normalized Gaussian function (gausn) and background shape as a polynomial function (polN) at different degrees depending on the user's preference. This is because different shapes of background other than the straight line can significantly alter the accuracy of the integrated net area values [5]. In conclusion, a computer program in C++ is written to run by the ROOT as a data visualization tool created by CERN. In this open-source data

analysis framework, everything can be coded as a user friendly and interchangeable way by the coder. The ROOT has predefined and user defined functions already given out for the convenience. For fitting, it has also various techniques and classes to test different results for the best of the analyst's interest. Before explaining the procedure, I'd like to emphasize that there might be two common cases to encounter in the gamma-ray spectrum. In the first scenario, there might be one solely standing peak. For the rest of the situation, there will be more peaks in a close range with overlaps. In this later case, it's hard to eliminate the extra contributions from each other while there are already other technical details, but only the digital integration issues and ideas will be dealt with here. One single peak case will be skipped here due to its simplicity. Firstly, the mathematical model which the code was implemented on has one total function constituted by two normalized Gaussian functions and one polynomial function at adjustable degree. This code is tested on two well-known peaks of ⁶⁰Co around 1173keV and 1332keV gamma-ray energies. All the analysis is processed on uncalibrated data, so the abscissa throughout the analysis represents the channel number, namely the bin number of the histogram in the spectrum. For the background estimation, the broken line method with the first order polynomials was used. Thus, there will be two separate straight lines individually, which will be seen in Figure 3, for each peak's background. In this method, three coordinates were recorded as x and y data points. First x₁ and y₁ coordinates represent where the left limit of the first peak started. Then, the second x₂ and y₂ coordinates are where the two peaks intersect together. The last x₃ and y₃ coordinates show the point where the second peak in higher energy fades out before stabilizing with the background level. This is namely the right edge of the second peak. This coordinates can be captured even by eye after zooming the corresponding area in the spectrum. This is also called an initialization step for the parameters, so the fit will go smoothly.

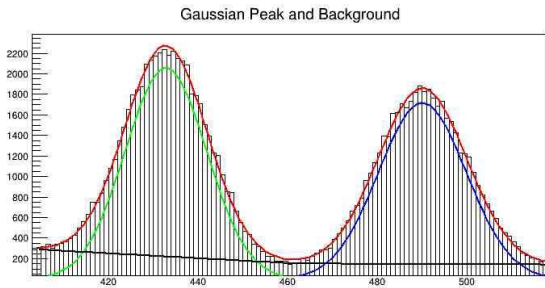


Figure 3. Broken line method with first order polynomials in the ROOT.

As a second step, the initialization values of the parameters for the two normalized Gaussian functions (gausn) should be set. The values in brackets [] in the equations below indicate certain parameters. The gausn function can be written as:

$$gausn(x) = [area] \times \frac{e^{-0.5 \times \left(\frac{x-[mean]}{[sigma]}\right)^2}}{[sigma] \times \sqrt{2\pi}}$$

It's known that the integration of a Gaussian function can be given as follows:

$$\int_{-\infty}^{+\infty} [height] \times e^{-0.5 \times \left(\frac{x-[mean]}{[sigma]}\right)^2} dx = [height] \times [sigma] \times \sqrt{2\pi}$$

When this integration is normalized to the value of 1, then the [area] in gausn function can be regarded as a new parameter to extract from the fit. What fitting gausn instead of gaus on a real data gives is the convenient way of retrieving the area and its error directly from the new parameter in hand. In addition to that, it will correspond to the integration from +∞ to -∞. In the case of defining Gaussian functions instead, an integral function has to be called in the ROOT for the numerical analysis where we can

retrieve both the area and its error in a more complicated way. This is because the covariance and correlation matrices will be in place to get the errors in the calculations. In the programming perspective, this only means more lines in the code more than a problem. The integral function in the ROOT can be used in the case of both Gaussian and gausn function while representing the signal. However, the detail of the difference in the usage might matter for the efficiency and practicability for the user. In the usage of Gaussian, the integral within certain peak limits gives the fraction of the area. This part has similarities with the reduction method by getting the fraction of the peak mentioned by Covell [1] and another method called partial peak area method by Kennedy [10]. However, they both used the summation values of discrete values in terms of counts in a specific channel range. In the usage of gausn, the net area does not need a correction for the full coverage. Since the usage of integral symbolise the continuity, any x axis range at any two random points can be calculated easily. This x values do not need to be integer values unlike channel numbers. To correct the fraction of the area while using integral with Gaussian function, z score transformation in statistics can be used. The z score when finding the proportion of the selected area is equal to $\frac{x-mean}{sigma}$ [11]. Instead of looking up this value from z score tables, some predefined statistical functions in the ROOT can be benefited as follows:

- Cumulative distribution function of the normal distribution for the lower tail:

$$normal_cdf(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \times e^{-\frac{t^2}{2}} dt$$

- Complement of the cumulative distribution function of the normal distribution for the upper tail:

$$normal_cdf_c(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \times e^{-\frac{t^2}{2}} dt$$

- Error function: $erf(x) = \frac{2}{\sqrt{\pi}} \times \int_0^x e^{-t^2} dt$

- Complementary of error function: $erfc(x) = \frac{2}{\sqrt{\pi}} \times \int_x^{+\infty} e^{-t^2} dt$.

The combination of z scores for the x coordinates of the peak limit and the special statistical functions stated above provides the correction ratio for the retrieval of the whole area. Usually, this alters the outcome a little if there is no adjacent peak nearby in the range of interest. However, it matters in the case where the user restricts the limits in the integral due to overlaps. This correction factor is also mentioned in probability intervals and confidence limits section as a table by Gilmore [6].

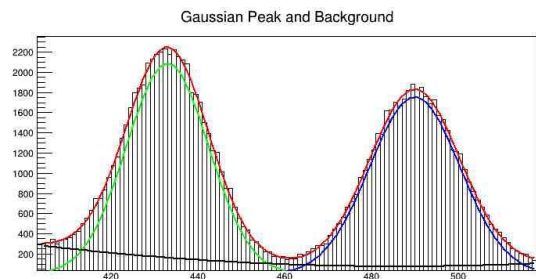


Figure 4. Second order polynomial fits for the background in the ROOT.

In the second part of the analysis, second order polynomial function was used to represent the background as shown in Figure 4. Note that the background line in gamma-ray spectra follows the declining trend giving the impression of trapezoid area along the x axis. This is due to the different contributions from higher energetic peaks to the lower part of the spectrum. To estimate the

proportion of all these contributions may not be feasible because of unequally distributed case in each peak along the way. With regard to the broken line method, the usage of a straight line as a background might be sometimes either an underestimation or an overestimation depending on the shape of a higher degree polynomial background. At that point, Sterlinski suggests that when the peak limit range is short, the background tends to be the straight line [5]. However, it deviates from this structure for the increased interval. That also means the curvature from the secant between the limits of a photopeak becomes apparent in higher energies due to the broadening. Thus, it will be wise to compare different polynomial degrees for presenting the background. In this article, only first and second degrees (pol1 and pol2) will be mentioned. For the ROOT fits, the Migrad algorithm was used by default in Minuit2 library. For details, the ROOT website can be visited. The total function consisting of 2 gauss and polN function was modelled to fit on a real data to extract all information about the peaks. For the broken line method, some parameters, such as breaking point coordinates between the two peaks and the slopes of the background lines under each peak were set as fixed parameters for ROOT to handle the fitting procedure correctly. Thus, errors of the background parameters were zero, so the background integral error was as seen in Table 1.

3. Conclusions and Comments

Overall result gives the impression that there might be more additional implementations in the future in this field. Especially, the combination of C++ and the ROOT has list of other possible predefined algorithms for least-square and likelihood fits. They also gives the flexibility to a pro-user to build its own algorithms if needed. If remembered, the main goal is to reach the estimated net area under a desired peak. Despite that, the core issue was unpredictable nature of the backgrounds and all the other unavoidable contributions to them. In terms of gross counts calculated directly from a histogram in methods from a to i, the results gave similar values apart from method d and e in the Table 1. That's because the coverage range for those was set from bin number $-(n-1)$ to $+(n-1)$. Therefore, they rejected the counts in the very last bins on each side of the peak by resulting in lower values. When these analytical results were compared with the numerical results of the ROOT fittings with pol1 and pol2 backgrounds, gross area values gave almost equal outcomes. Maestro software, on the other hand, deducted 3 bins on both sides of the peak limit. This was indeed the adjusted gross area value in its manual rather than the same gross area value found in method from a to c and from f to h. In terms of errors in gross counts or gross areas, all results were consistent on giving the same variances. The same pattern was seen on the outcomes of the second peak for the gross area. The ROOT fits gave still the similar answers no matter the pol1 or pol2 was used in the total function for the fitting procedure. On the background side, Gilmore's method in equation b gave lower counts because of the two bins it selected to average the background count per bin. Then, it was multiplied by the same total bin number as the other methods. However, it gave just 4% less result for the background. Method d and e represents the same background values. In comparison to method a and c, they consist extra counts coming from the last bin numbers on each side of the peak. This makes the difference in their background equations. While the background coefficient is $\frac{(2n+1)}{2}$ in method a and c, this factor is $\frac{(2n-1)}{2}$ in method d and e. However, these will not affect the result for the net counts in method a, c, d and e. This is

because they are just different forms of the same equation overall. For the TPA methods in f, g and two different cases in method h, the methods f and g had different interpretations in a way how they were constructed. In method f, the symmetric Gaussian shape of the data does not appear clearly in the equation. Only the lower and upper limits were known. However, the method g gave the impression of a symmetrical nature by noting the equal number of channels as n on both sides of the mean value. As to background counts, method a, f and g have the same values due to the usage of the same procedures for the background. One drastic change comes about in the calculation of Gilmore's background equation. According to Gilmore's method, analyst can decide the total number of bins on both side of the peak to find the average count per channel. They were stated as m_L and m_U . This decision was based on how close other peaks were to the range of interest. If there was no adjacent peak nearby, more bins could be averaged on both sides. Otherwise, the limitation on the bin number depends on the right or left side with overlapping peak. Thus, the left and the right side might not always have the same number of bins in total for the background calculation. That's why these two different variable as m_L and m_U were introduced in the equation h. Notice that this average value for the background was still multiplied by the same total number $(U-L+1)$ similar to the method f and g. In Table 1, method h had two options for the m_L and m_U as an example. Even this little change affected the answer in background counts. Therefore, it's safe to say that it will be crucial to test some different values until the results give closely related comparison with the previous methods from a to g. While comparing the results of peak 1 and peak 2, peak 2 gives interestingly much closer values to the other methods. Thus, this feature might not be generalized to all peaks. It can give better approximations in different occasions from peak to peak. For the ROOT fit values, background results were pretty good with broken line method, namely in the usage of the pol1 function. However, the pol2 function in the model gave underestimated values due to the downward direction slope instead of declining trapezoid line. That means the deviation was too much. Gilmore's approach in the background calculation in method h was closer as the bin number for m_L and m_U increases in total. When that happened, results indicated similarity with Maestro method in equation i. This is because Maestro method uses 6 bins to estimate the average counts per bin for the background. After interpreting gross and background counts and their estimations, the net area results were reached easily. Except the method h, they all agreed on pretty much the same values for the net total. Despite this, method h gave around 6% lower estimations in the example. Thus, the best solution for this method can only be found by comparing with other methods with different m_L and m_U values. Otherwise, it might easily lead to a wrong approximation. In the ROOT fit side, the net area results were the same between the one taken from gauss parameter and the one calculated directly from corrected integral values in a certain range. These results in fitting proved that the corrections as stated in the Table 1 were crucial when integrating the area with Gaussian function instead of gauss. The net area results when pol2 background function was selected approved the overestimation in comparison to other methods. It was clear to see the fact that pol2 shape lowered the total counts under the peaks. Thus, the net area came up as a higher value in the estimations.

Last but not least, last column in the Table 1 shows the proportion of background counts to the signal.

Table 1. The Results of all the Methods

	Methods	Gross Counts or Area	$\sqrt{\text{Variance}}$	Background Counts or Area	$\sqrt{\text{Variance}}$	Net Area or Counts	$\sqrt{\text{Variance}}$	Correction Factor %	Background to Signal Ratio %
Peak 1	Covell (a)	59985	245	12934	618	47051	665	0.99900	27.5
	Covell (b)	59985	245	12402	605	47583	653	0.99900	26.1
	Covell (c)	59985	245	12934	618	47051	665	0.99900	27.5
	Covell (d)	59547	244	12496	597	47051	645	0.99900	26.6
	Covell (e)	59547	244	12496	597	47051	645	0.99853	26.6
Peak 1	TPA (f)	59985	245	12934	618	47051	665	0.99900	27.5
	TPA (g)	59985	245	12934	618	47051	665	0.99900	27.5
	TPA (h) $m_L=4, m_U=1$	59985	245	16478	441	43508	505	0.99900	37.9
	TPA (h) $m_L=4, m_U=2$	59985	245	15690	393	44295	463	0.99900	35.4
	Maestro (i)	58603	242	12211	329	46392	408	0.99900	26.3
Peak 1	ROOT Pol1 Fit from Parameter					46977	242	0.998742	
	ROOT Pol1 Fit from integral	59968	241	12868	0	46977	241	0.998742	27.4
	ROOT Pol2 Fit from Parameter					49618	372	0.998085	
	ROOT Pol2 Fit from integral	59867	243	10033	314	49618	368	0.998085	20.2
Peak 2	Covell (a)	50502	225	8291	487	42211	536	0.99689	19.6
	Covell (b)	50502	225	8863	503	41640	551	0.99689	21.3
	Covell (c)	50502	225	8291	487	42211	536	0.99689	19.6
	Covell (d)	50211	224	8000	470	42211	521	0.99689	19.0
	Covell (e)	50211	224	8000	470	42211	521	0.99689	19.0
Peak 2	TPA (f)	50502	225	8291	487	42211	536	0.99689	19.6
	TPA (g)	50502	225	8291	487	42211	536	0.99689	19.6
	TPA (h) $m_L=4, m_U=1$	50502	225	7719	297	42783	373	0.99689	18.0
	TPA (h) $m_L=4, m_U=2$	50502	225	8100	278	42402	358	0.99689	19.1
	Maestro (i)	49488	223	8620	271	40868	351	0.99689	21.1
Peak 2	ROOT Pol1 Fit from Parameter					42210	225	0.996285	
	ROOT Pol1 Fit from integral	50533	224	8291	0	42210	224	0.996285	19.6
	ROOT Pol2 Fit from Parameter					45712	400	0.993846	
	ROOT Pol2 Fit from integral	50574	224	4807	332	45712	387	0.993846	10.5

This ratio is around 27% which gives enough statistics to fit any function to a peak. All in all, the Table 1 indicates straight forward outcomes, self-explanatory, well modelled and trustworthy perspective to the issues explained above.

For the future work and development, the detail background estimations with different bin sizes will be tested and presented.

4. Acknowledgement

Thanks to the ROOT team in the CERN and other academia in their forum who actively involved responding the questions in a very polite and encouraging way.

References

1. LOSKA, L., 1988. A modification of the "total peak area" (TPA) method for gamma ray spectra. *Appl. Radiat. Isot.*, 39(6):475-477.
2. BAEDECKER, P.A., 1971. Digital Methods of Photopeak Integration in Activation Analysis. *Analytical Chemistry*, 43(3):405-410.
3. STERLINSKI, S., 1968. Analysis of digital data from a multichannel pulse height analyzer on gamma ray total absorption peaks. *Analytical Chemistry*, 40(13):1995-1998.
4. GILMORE, G.R., 2008. *Practical Gamma Ray Spectrometry*. John Wiley & Sons, Ltd, England, 387s.
5. KOKTA, L. 1973. Determination of peak area. *Nuclear Instruments and Methods*, 112:245-251.
6. BAEDECKER, P.A., GROSSMAN, J.N. 1989. The Computer Analysis of High Resolution Gamma-Ray Spectra from Instrumental Activation Analysis Experiments. Open-File Report:89-454. U.S. Geological Survey. DOI:10.3133/ofr89454.
7. A65-B32 Software's User's Manual. Ortec Part No. 777800. Available: <https://www.ortec-online.com/-/media/ametektotec/manuals/a65-mnl.pdf> [Access on: 1th September 2020].
8. KENNEDY, G., 1990. Comparison of photopeak integration methods. *Nuclear Instruments and Methods in Physics Research A* 299: 349-353.
9. Spiegel, Murray R.; Stephens, Larry J (2008), *Schaum's Outlines Statistics* (Fourth ed.), McGraw Hill, ISBN 978-0-07-148584-5