

4. EXPERT SYSTEM BASED RADIONUCLIDE IDENTIFICATION⁷



XA9848355

4.1. SUMMARY

An expert system coupled with the gamma spectrum analysis system SAMPO has been developed for automating the qualitative identification of radionuclides as well as for determining the quantitative parameters of the spectrum components. The program is written in C-language and runs in various environments ranging from PCs to UNIX workstations. The expert system utilizes a complete gamma library with over 2600 nuclides and 80 000 lines, and a rule base of about fifty criteria including energies, relative peak intensities, genesis modes, half lives, parent-daughter relationships, etc. The rule base is furthermore extensible by the user.

This is not an original contribution but a somewhat updated version of papers and reports previously published elsewhere [1, 2, 3].

4.2. INTRODUCTION

A complete analysis of a gamma spectrum is a demanding task even for modern gamma ray spectrum analysis programs. It requires that all the peaks above a given threshold are found and their intensities and energies correctly defined. Correctly here necessarily requires also that the error estimates be in statistical control. After the *quantitative analysis* the peaks must be associated with correct radionuclides so that their intensities are all attributable to identified nuclides within the uncertainties of the statistics and measuring procedures.

We must emphasize that, in our opinion, there does not exist a computer code, which can do a complete analysis of a general gamma spectrum without strong user intervention. In some special cases, however, this may be possible. If the analyst is only interested in some specific nuclides or transitions, the analysis can be automated.

Several highly effective computer programs for the gamma spectrum analysis are also available, including programs of the SAMPO family [4]. With the latest SAMPO versions it is possible to do a complete quantitative analysis of even very complex gamma spectra. Due to automatic residual analysis of the fits even overlapping multiplets can be correctly analyzed without user intervention. The most complex cases, however, require interactive use of the code where the analyst visually verifies the fitting results.

In many cases this kind of quantitative analysis does not go quite far enough. Although it gives very accurate information about the gamma peaks present in the sample, it does not tell anything about the radionuclides that produce these peaks. This information is often crucial when doing real world measurements, whose goal is to find out about the composition and history of the sample. Many reasons for such measurements exist. One may want to find out where a specific sample originates from, i.e., is its activity of natural origin or does it come from a nuclear reactor or radioactive fallout of some kind. Procedures that try to identify the composition of the sample and to make deductions based on this information are known as the *qualitative analysis*.

To this date quantitative analysis has been the more automated discipline of the two. This can be easily understood considering the difference in nature of the two analysis phases. Quantitative methods are mathematically fairly exact and well defined, which makes it more

⁷ Prepared by P.A. Aarnio, J.J. Ala-Heikkil, T.T. Hakulinen and M.T. Nikkinen, Helsinki University of Technology, Finland.

natural to try to automate them using a computer. Qualitative analysis, on the other hand, requires many heuristic deductions in addition to exact mathematical methods. Computers have been notoriously ineffective in any problem requiring heuristics, and therefore, the qualitative analysis is even today mostly carried out by human experts who possess the required knowledge.

With the development of faster computers and suitable computing techniques for processing of heuristic knowledge, a system for the qualitative analysis has become feasible. The problem at hand can be attacked by using an *expert system*, which is a specialized program for reasoning. Expertise of a human expert has been programmed into the expert system. The system aims to reproduce the chain of reasoning of the expert in such a way that the same inputs produce similar outputs in both cases.

A considerable amount of testing is necessary for determination of the efficiency and quality of an expert system. The system must be verified against real world cases. Any deviations from the expected results indicate that the heuristics need to be reviewed. It is also essential that the tests are made as comprehensive as possible to cover a wide range of the projected applications. Our current research interest is in the integration of the quantitative and qualitative analysis phases in a user oriented environment where complete analysis of any spectra is feasible. That is, even if the user is only interested in a few transitions he will end up doing the complete analysis because it will be so simple. The final aim is to be able to do the complete analysis without user intervention.

In this paper we concentrate on the qualitative analysis of gamma ray spectra. That is, we present *SHAMAN*, an expert system for the identification of radionuclides in the measured gamma ray spectrum. *SHAMAN* is an expert system designed specifically for carrying out the qualitative radionuclide identification and quantitative activity determination on a measured spectrum. High level logic is implemented as abstract rules and mathematical algorithms as separate functions. *SHAMAN* is written in ANSI C [5] for maximum portability. It has been ported to BSD, AIX, Linux, Windows, OS/2 and DOS.

4.3. EXPERT SYSTEM CONCEPTS

An expert system is a computer program, where the flow of execution is governed by the problem specific data rather than a preprogrammed algorithm [6]. The system consists of *rules* that hold knowledge about the conditions that the data must fulfill as well as the actions taken in such a case. An *inference engine* continually monitors rules for validity as the data in the system changes and appropriately *fires* valid rules at the suitable instance thereby causing the data to be modified.

SHAMAN rules are simple if-then constructs, which are stored in the system as a list accessible to the inference engine. Rule's *antecedent* holds logical clauses that depend on the data in the system. Any C-language construct that returns zero value for false and non-zero otherwise can be used. Rule's *consequent* holds the actions to be executed when the rule is fired. These actions modify data and may cause further rules to become valid. Rules are grouped according to their application domain, which allows control over orthogonal blocks of knowledge in the system. It is thus possible to exclude rules concerning, for example, genesis modes of nuclides. Each rule has an associated numerical priority value, which is used to determine the relative precedence of rules.

Data items are contained in special tagged structures, which exhibit selected features of object oriented frame systems, e.g., *inheritance* and automatic creation of certain standard *handlers*. Instances of data structures have additional data fields for the data type tag and the last modification time. This information is necessary for the inference engine.

Rules and data objects are written in the system using a special rule language, which can be machine-translated to standard C-code using a separate rule compiler. This allows proper type checking for the objects and enforces internal consistency of the knowledge base whenever rules and objects are modified.

The inference engine is conceptually fairly straightforward implementing only *forward chaining* as its inference strategy. Inference operates in a loop. After a rule has been fired it is checked for each rule in the system whether its rule group is enabled and whether the data the rule depends on has been modified since the last check. If these criteria hold, rule's antecedent is checked. If the antecedent *matches*, i.e., evaluates true, the rule is placed in the *agenda*, which is a list holding rules scheduled for firing in the relative order of their priority. If there are yet unfired rules in the agenda, whose antecedent no longer matches, they are removed from the agenda. It is also possible to define the rule *edge-sensitive*, which means that the rule stays in the agenda even if the antecedent becomes false. When all rules have been checked, the first enabled rule in the agenda is fired and the rule entry removed from the agenda. The inference loop continues in the above manner as long as there are enabled rules in the agenda, after which the solution has hopefully been reached.

4.4. MEASURED SPECTRUM

Spectra to be identified are measured using germanium detectors and the quantitative analysis is carried out using SAMPO-90 gamma ray spectrum analysis program [7, 8]. After this the analysed data are converted to a special format required by *SHAMAN*.

The quantitative analysis proceeds roughly as follows: The measured spectrum is analysed using SAMPO-90. Calibrations for energy, efficiency, and peak shape are assumed to have been correctly determined earlier. The analysis is assumed complete, i.e., all peaks, whose statistical significance exceeds the value of peak search significance, are actually found in the spectrum and correctly analysed. This analysis yields a peak table where peak centroids, areas, intensities, and respective statistical errors are calculated. In addition, SAMPO-90 outputs the energy ranges that were used in peak fitting and area determination including information about the actual functions fitted to the peaks and the background.

These data are further processed to form a special format input file for *SHAMAN*, because some of the data provided by SAMPO-90 are unnecessary, whereas some essential information is missing. The following information is necessary for the identification:

- (a) All found peaks in the spectrum, their energies and intensities with respective errors. The identification process is considerably simplified if it can be assumed that all detectable peaks in the spectrum have been found. In *SHAMAN* detection and determination limits are calculated according to Currie [9].
- (b) The background as a function of energy. It is formed for each channel in the spectrum by summing channels around the centroid. The summing interval is the peak width multiplied by a confidence factor. Background is required for the detection limit calculations of the candidate nuclides.
- (c) Measured energy ranges. Gamma lines of the candidate nuclides, which fall outside of the measured range, are excluded from analysis. Piecewise energy ranges are possible.
- (d) Sample mass, which is required for calculating the limiting activities of candidate nuclides.

- (e) Irradiation, cooling and counting times. If the end of irradiation is unknown, a conservative estimate is the time, when the sample was received for measurement.
- (f) The origin and the physical state of the sample. This information allows one to make deductions about the possible genesis modes of the sample.
- (g) Energy, efficiency and shape calibration. Calibrations are obtained from SAMPO-90, but the actual efficiency calibration is different for candidates that are classified as background nuclides and must therefore be carried out in *SHAMAN*.
- (h) Information about the measurement environment. The efficiencies of background nuclides in lower energies are generally affected by matrix effects, which depend on the materials present in the measurement room. If the spectral background has not been subtracted in the measurement time, an approximate correction can be applied.

4.5. RADIONUCLIDE LIBRARY

The radionuclide library used by *SHAMAN* is compiled from the databases ENSDF and NUDAT retrieved from NNDC [10]. It contains 2616 nuclides with about 80 000 lines. The data base was converted into a more concise form to save space and to speed up file operations. Some additional data were taken from other publications [11, 12, 13, 14]. In particular, all parent-daughter relationships and respective branching ratios were completed, which is essential for correct activity calculations. The following data of all the nuclides are used by *SHAMAN*: mass and element numbers, half life, mass and element numbers and the respective branching ratios of immediate daughters, mass and element numbers of immediate parents, genesis modes and various discrete status information. For each gamma ray line the following data are used: energy and absolute intensity of the transition with corresponding errors and some status information.

The library is organized as a random access binary file, which is indexed according to mass and element numbers as well as gamma line energies. An index file is a sorted binary file, which can be searched sequentially to locate a pointer to the main data file. The library is basically static, which means that only casual updates are necessary based on new compilations of nuclide data in the future. This makes it possible to keep the library structure fixed, thereby simplifying library management.

SHAMAN includes a library manager for easy access to the library. It carries out all requests for candidates. Search can be performed according to mass number, element number, and the energy of a gamma transition or any combination thereof. Additional discrimination criteria can be given for individual gamma ray lines based on their status information to rule out, e.g., all X ray lines or lines.

4.6. IDENTIFICATION RULES AND METHODS

4.6.1. Selection of preliminary candidates

The identification process begins by reading in the preprocessed spectrum. Thereafter candidate nuclides are read from the library and checked against several conditions in order to be accepted into final solution.

According to the first rule all nuclides having a gamma transition close enough to a peak in the spectrum are read in. Peaks are processed sequentially starting from the highest energy peak, because the density of gamma transitions per unit energy decreases roughly exponentially as a function of energy. Since the number of candidates is low and the

resolution good at high energies, less candidates remain to be handled for low energy peaks. The energy tolerance used is the peak width multiplied by an adjustable confidence factor.

No candidates are accepted on the basis of the annihilation peak only. Annihilation peak is present for positron emitters and high-energy gamma ray emitters due to pair production. In both cases other lines must be visible too, because otherwise it would be impossible to discriminate between several candidates. A special dummy nuclide is associated to the annihilation peak to account for the possible unexplained portion of intensity due to external annihilation radiation. Thus, the annihilation peak will always be fully explained and the peak shares of other nuclides will not be distorted in the activity calculation.

Upon reading candidates into the system, several calculations are done:

- (a) The distance to the nearest peak and the significance compared to the background are calculated for all lines. The significance value is corrected for matrix effects.
- (b) The *primary line* is determined. It is the most significant line in the measured energy range.
- (c) All parent nuclides of the candidate are read in by tracking the decay chain upwards. In general, if decay chains can be assumed to be in equilibrium, the *effective half life* of the nuclide is defined as the half life of the longest lived precursor in the chain. This is especially true for background nuclides that belong to natural decay chains. For newly irradiated samples decay chains are often not in equilibrium, which renders this kind of calculation invalid.
- (d) The *pure nuclide activity* is calculated. It is defined as the activity that the candidate nuclide would have immediately after the irradiation, if the sample consisted entirely of that isotope. The *effective pure nuclide activity* is calculated using effective half life instead. Pure nuclide activity is meaningless for background nuclides.
- (e) If the candidate has gamma lines that are not found in the spectrum, the *threshold line* is determined. It is the most significant unfound line in the measured energy range.
- (f) The *threshold activity* is calculated. It is defined as the minimum detectable activity of the candidate based on the threshold line, i.e., it is the maximum activity that the candidate nuclide in the sample can have for the threshold line to remain unseen in the spectrum.

After all the above quantities have been calculated, pruning rules in the knowledge base become effective. Rules are assigned priorities so that the most efficient pruning rules are handled first. Five possible states exist for a candidate. The candidate is *possible*, when it is considered a part of the solution, *impossible*, when there are strong reasons for discarding the candidate, *improbable*, when the candidate is believed to be absent but there are doubts, *insignificant*, when the candidate strictly speaking is present in the sample but its effect for some reason is negligible, and *unfitting*, when the candidate is part of an interfering set of nuclides, and activity calculation indicates that this candidate is extraneous. Impossible candidates are never looked into again but improbable, insignificant or unfitting nuclides may be reconsidered if a satisfactory solution cannot otherwise be reached.

4.6.2. Early pruning

Early pruning rules operate on candidate nuclides before the activity calculation has taken place. They are used to strip the amount of candidates to the minimum to yield as small initial interference matrices as possible, which considerably eases activity calculation.

- If the primary line of the candidate is not found in the spectrum, the candidate is judged impossible. This is the most efficient rule being typically responsible for 80% of all discarded candidates.
- If neither the candidate nuclide nor any of its parent nuclides have suitable genesis modes, the candidate is judged impossible. The knowledge about valid genesis modes comes from the user in some form and is based on the origin of the sample and the particular measurement setup.
- If the physical state of the sample is other than gaseous and the sample is in an open space, all noble gas candidate nuclides are judged improbable. Since noble gases are inert and form very few compounds, they will soon fly away from a porous sample.
- If there is a candidate, whose gamma line branching ratios match closely (within a given tolerance) to the intensities of the respective spectrum peaks, all other candidates, whose most significant lines are associated to these peaks are judged improbable. The more found gamma lines the nuclide has, the more reliable this judgment is. In particular, this rule cannot be used if the candidate has only one found gamma ray line.
- If the candidate nuclide is an isomer, whose longer-lived parent has not been found, the candidate is judged improbable. It is assumed that an isomeric nuclide cannot be present in detectable quantities except as a decay product.
- If no candidates are associated to a peak, it is checked, whether it is an escape peak. If there is a stronger peak at 511 keV or 1022 keV higher energy, the unassociated peak is marked as escape peak and not included in the activity calculation. Strictly speaking, estimates for the relative escape peak efficiencies should be calculated to be sure [15, 16], but this would require knowledge of some additional detector characteristics.
- If no candidates have been associated to a peak, and it does not seem to be an escape peak, discarded candidates that have earlier been associated to this peak are reconsidered. Reconsidered candidates cannot be discarded again on the same grounds. Candidates marked as impossible are never reconsidered.
- If the user has especially told *SHAMAN* not to accept the candidate for any reason, it is discarded. The judgments as well as the reasoning are given by the user. This will offer the user a way to affect the inference process.

4.6.3. Activity calculation

When initial candidates have been read in for all spectrum peaks and there is no more pruning to be done, activity calculation is performed. In general, most candidate nuclides are associated to several spectrum peaks as well as most peaks to many candidates. This means that a best estimate approximation is required. The standard method is to fit candidate nuclides to the spectrum in the least squares sense.

Although the problem is mathematically straightforward, situations exist that are difficult for many algorithms to handle. The first of these cases appears, when the coefficient

matrix is singular, i.e., the matrix contains some degenerate vectors. The other case occurs, when the matrix is underdetermined, i.e., there are less linearly independent equations than free variables. In activity calculations of gamma spectra both of these cases appear frequently.

To reach a mathematically optimal solution in all cases, a sophisticated algorithm is required. Standard methods are Gauss-Jordan elimination, QR-factorisation and singular value decomposition SVD [17, 18]. Although all these methods in principle work in cases, where clearly singular matrices have been eliminated, rounding errors typical for numerical calculations may cause the solution to be unstable or even push a regular matrix to singular. To prevent this, one has to analyse the matrix in advance taking appropriate actions whenever ill-conditioned cases are encountered. This is precisely what SVD-algorithm allows one to do. The algorithm produces a set of *singular values* of the initial coefficient matrix, which are used to take into account nearly singular cases. The only drawback of SVD is its longer execution time compared to the other methods.

The SVD-algorithm produces also a covariance matrix, whose diagonal elements are the variances of calculated activities, as well as a vector containing peak residuals characterizing the unexplained portions of the peaks. Negative residuals indicate over-explained peaks.

In addition to the solution provided by SVD, *SHAMAN* uses also information about the particular interference setup: which candidate nuclides belong to which interference group, how the submatrices are determined and whether they are columns degenerate.

4.7. ACTIVITY BASED RULES

After the activity calculation is complete, several new pruning rules become active. Most of these rules are only effective on the particular interference setup, which is subject to change after each recalculation of activities.

- If the calculated activity is greater than threshold activity, the candidate is judged unfitting. In other words this means that the calculated activity is so great that the threshold line should be seen in the spectrum.
- If the decay corrected pure nuclide activity of the sample is smaller than the smallest detectable activity, the candidate is judged impossible. This rule is quite effective in eliminating candidates with short half-lives, and given the correct values of sample mass, decay time and effective half-life, it is also extremely reliable. However, for background nuclides the mere concept of pure nuclide activity is meaningless and this rule therefore useless.
- If the decay corrected value for the candidate nuclide mass in the sample based on the calculated activity exceeds the sample mass, the candidate is judged impossible. The relative concentration of individual nuclides is typically small, which makes this rule very reliable.
- If the calculated activity of the nuclide is negative, the candidate is judged unfitting. A negative activity value indicates that other interfering candidates are able to (more than adequately) explain all the peaks in question. It is possible that by discarding some other nuclide the results would be better, but as a first approximation this is most likely the false candidate.
- If the error of the calculated activity is large, the candidate is judged unfitting. The error value is produced by the fitting algorithm and depends on the interference setup. By default *SHAMAN* uses a 50% threshold.

- If the same peaks are associated to both a parent nuclide and its daughter, which is a shorter lived isomeric state of the parent; the daughter nuclide is judged insignificant. This rule actually discards a nuclide that is present in the sample. This is necessary, however, so that the activity of the parent could be correctly determined.
- If the same peaks are associated to both a parent nuclide and its non-isomeric longer lived daughter, the parent nuclide is judged insignificant. This rule is based on the assumption, that all short lived parents have decayed into their respective daughters, which is true if the decay chain is in equilibrium.
- If a peak is poorly associated, i.e., its relative intensity residual is great, and it does not seem to be an escape peak, discarded candidates that have earlier been associated to this peak are reconsidered. By default SHAMAN uses an 80% threshold.

4.8. USER INTERFACE AND REPORT GENERATION

The user interface of *SHAMAN* is command line oriented. When the system is started, the user is placed to command interpreter prompt, where it is possible to view and set system parameters and run inference. Even as identification proceeds, it is possible to suspend inference to view system status etc. *SHAMAN* has also a simple macro facility to help automate repetitive tasks.

SHAMAN outputs currently two distinct reports about the inference results. These reports are quite detailed and aimed for an expert, who wishes to easily evaluate the validity of the result. Most of the numerical data and reasoning that *SHAMAN* produces are presented in some form in these reports.

Spectrum Identification Report presents information about the spectrum peaks: nuclides associated to each peak, explanation percentages, various status information and lists of nuclides that have once been associated to each peak.

Nuclide Identification Report presents detailed information about each accepted candidate: half lives, genesis modes, parent-daughter relationships and calculated activity values. For each gamma line the report contains its energy, intensity, distance from the nearest spectrum peak, peak share if the line is associated to a peak, the statistical significance and various status information. Information is also presented about the inference process: the statistical parameters used, lists of all discarded nuclides, background nuclides and reconsidered nuclides with appropriate reasoning.

For complete information about the identification process it is also possible to have *SHAMAN* log all its actions to file. This enables the user to follow the course of inference in great detail.

4.9. TEST SPECTRA

In order to evaluate the performance of *SHAMAN* a set of test spectra was created. The spectra were chosen with the general-purpose nature of *SHAMAN* in mind, i.e., they include spectra from several fields of applications of gamma spectroscopy. From a practical point of view, it is well motivated to keep the contents of the library constant during the development of the system. In this way the effects of modifications can easily be detected by comparing the new results to the previous ones.

The varying complexity of the test spectra required consideration of several aspects:

- The complexity of the nuclide identification depends primarily on the number of peaks in a spectrum. The number of peaks found in a gamma spectrum is typically 10 ...100, but in a very complicated spectrum even a few hundred peaks may be found.
- The complexity of the nuclide identification depends also on the quality of the peaks. Small peaks usually have bigger relative intensity errors than larger peaks. More candidate nuclides can be assigned to low energy peaks than to high-energy peaks. The X ray, escape and sum peaks are often difficult to distinguish from normal photopeaks during the analysis. Only if a peak can not be explained with the photopeaks of the candidate nuclides can it be recognized to be an X ray, escape or sum peak.
- The complexity of the nuclide identification depends also on the time parameters of the measurements. The shorter the cooling time the more short-lived nuclides must be regarded as possible candidates. Some conclusions can also be drawn from the length of the irradiation time, if the irradiation conditions are known accurately enough.

The selected 51 spectra are of varying complexity. Half of the spectra are real measured ones, whereas half of them have been synthesized in order to achieve a more versatile test set. The number of peaks in the spectra varies from 10 to 130 (except for the very complicated Chernobyl fallout spectrum with 275 peaks) and the number of present nuclides varies from 1 to 52. The measured spectra include some X ray, escape and sum peaks, but the synthesized spectra do not. The irradiation and cooling times vary from zero to one year.

The measured spectra in the library can be categorized as follows:

- Ten spectra of different combinations of 8 standard sources (^{22}Na , ^{54}Mn , ^{57}Co , ^{60}Co , ^{137}Cs , ^{133}Ba , ^{152}Eu , ^{241}Am).
- Two spectra with different measuring times of a weak natural uranium source.
- Two spectra with different measuring times of an irradiated uranium oxide sample.
- One spectrum of an irradiated indium sample.
- Four background spectra measured in the same laboratory with different measuring times.
- Six spectra of environmental, NAA and reactor samples measured with different detectors and in different geometries.
- One spectrum of a wipe sample of the Chernobyl fallout.

The synthesized spectra have been done with a non-commercial subroutine package of SAMPO-90. The synthesis model includes a tailed Gaussian photopeak and a Compton continuum, but no X ray, escape, sum or backscatter peaks. The Compton continuum has been modeled using the peak-to-Compton ratio as the only parameter. The spectra of the different gamma energies are summed together and statistics are added to the summed spectrum.

The versatility of the spectrum library can easily be enhanced by including synthesized spectra. Another advantage is that the nuclides present and their activities are known in advance, whereas measured spectra must be carefully analysed manually. However, there are some disadvantages with synthesized spectra. First, the spectra are only as good as the model. *If the model is crude, so are the synthesized spectra. Second, the spectra are correct only as long as the input data are correct.* In our case, the same gamma library was used in the synthesis and analysis and thus if the spectra are not exactly correct, at least they are

consistently biased. Third, the modelling of decay chains requires a complete decay database and a carefully implemented decay calculation. In our spectrum synthesis model, the decay chains have been totally neglected for simplicity, but this is not a serious defect as the activities of the daughter nuclides have been calculated by hand and given as input to the synthesis program.

The synthesized spectra in the library can be categorized as follows:

- Four spectra of totally artificial nuclide combinations.
- Four spectra with typical radionuclide inventories of the coolant and the gaseous waste of a nuclear power plant.
- Four spectra with some radionuclides generated by charged particle reactions (in an accelerator).
- Seven NAA-spectra with totally artificial element combinations.
- Six NAA-spectra with realistic element combinations found in literature.

The test spectrum library is comprehensive enough for an evaluation of the expert system *SHAMAN* or any other nuclide identification system. Eventual new interesting test cases can still be added to the library, but there should be no need to discard any of the current spectra.

4.10. PERFORMANCE

The evaluation of the expert system *SHAMAN* has been done on an IBM RS/6000 workstation, but the results do not essentially depend on the computer environment. The eventual differences can be explained with rounding errors associated to different floating-point representations in different computers.

The general requirements for an acceptable identification can be listed as follows:

- (a) *Good qualitative and quantitative identification of the nuclides.* Every nuclide present should be identified and no extra nuclides should be included. In fact, it is not very serious to have a few spurious nuclides in the solution as long as none of the nuclides really present are missing. It frequently happens that the decision between some nuclides can not be made with the rule base available to the expert system. Good quantitative identification means correct determination of the nuclide activities and other numerical parameters.
- (b) *Good qualitative and quantitative explanation of the peaks.* Every peak should be explained as a photopeak, X ray peak, escape peak or sum peak of the present nuclides. Every photopeak should also be quantitatively explained within the intensity error limits.
- (c) *Working rule set.* The rules should form a set with which an acceptable solution can be achieved and they should not be oversensitive to variations in the identification parameters. A working rule set is actually presupposed by the two previous requirements.
- (d) *Speed.* The speed requirements vary from one application to another, but it is clear that given the same task, the expert system should be faster than a skilled human analyst. This is also a strong function of available computing resources, a factor that is constantly improved by advancing technology.

The test runs were made with an older version of *SHAMAN* dated from January 1992, after which several corrections and new features have been implemented. Therefore, the results given herein do not necessarily apply directly to the current development version, but rather serve to give a general understanding of the performance of the system. A new evaluation of the performance will be done in the near future when the next frozen version of *SHAMAN* is available.

Both the measured and synthesized test spectra were analysed with analysis program SAMPO-90. The peak finding and fitting was performed interactively to ensure that all multiplets and deformed peaks were included in the analysis results. The analysis results were converted using a special utility program into a format suitable for *SHAMAN*.

In the evaluation, each spectrum was identified using the full nuclide library. Also, the identification parameters of *SHAMAN* were kept constant, except for three spectra where drift in the gain required a larger energy tolerance. The use of some rules was also controlled by the user. These included the genesis modes with charged particle and photon reactions, which were defined unacceptable in NAA spectra. This kind of data is usually available to the analyst, but since it cannot be deduced from the spectrum itself, its manual introduction to the expert system is well justified. All the modifications above can be achieved using the scripting facility of the system, so an automated testing procedure can be created. With the exception of the items above *SHAMAN* identified the spectra with minimum of user intervention.

The results of the evaluation show that *SHAMAN* performs generally very well even for relatively difficult spectra. Over 90% of the nuclides present in the test spectra are identified by *SHAMAN*, and most of the unidentified nuclides have been discarded either because the activity calculation has turned against them or because some of their prominent gamma peaks are missing from the spectrum, i.e., they are near their detection limit. Also, almost all peaks are correctly explained, with the exception of some X ray, escape or sum peaks, whose identification is currently not on a very sophisticated level.

There are, however, extra nuclides in the identification results, which shows that *SHAMAN* is overconservative when discarding nuclides. An average spectrum in the library has 12 nuclides and 54 peaks. Of them *SHAMAN* correctly identifies 11 nuclides, but there will also be 6 spurious nuclides included in the result. The problem with spurious accepts is especially difficult for strongly interfering nuclides, since it is generally quite difficult to decide which nuclide, if any, should be dropped to improve the solution. This problem can be partially solved by developing the processing of the interference groups. As a whole, however, the current rule base operates well.

Recently a coincidence correction scheme has been implemented in *SHAMAN* which corrects for true cascade decay coincidences at close measuring geometries [3]. It calculates both the loss and increase of the apparent intensities of the cascade lines. It also calculates the intensities of the coincidence sum peaks, which are not part of the decay scheme of the nuclide in question.

The effect of coincidence corrections can be best illustrated with some representative examples by comparing the explanation of peaks with and without coincidence correction. Three example nuclides with strong coincidence summing effects from a spectrum measured at a close geometry are shown in Table III. The explained peak shares should ideally be 100%. With coincidence correction they seem to be well in agreement with each other, with the exception of three clearly over-explained peaks: 1.5424 MeV of ^{110m}Ag , 1.32551 MeV of ^{124}Sb

and 1.4005 MeV of ^{134}Cs . These are peaks with strong summing-in effects, as can be seen by comparing the intensities in columns 3 and 5 in Table III. These peaks have not been explained at all or they have been clearly under-explained without coincidence correction, so that even their explanation has improved with coincidence correction. Thus, the introduction of coincidence correction in *SHAMAN* will lead to better results when identifying spectra measured in close geometries.

The quantitative results have received less attention in these tests, because the activity calculation has been tested earlier and has been found to be quite reliable. The execution speed was not a major concern either, especially because it depends heavily on the hardware, the operating system and the particular spectrum. The running time of *SHAMAN* varies from 1 to 20 minutes in a typical Unix workstation.

TABLE 1. IDENTIFICATION RESULTS OF THE MEASURED TEST SPECTRA

Spectrum label	Present nuclides	Unidentified nuclides	Spurious nuclides	Number of peaks	Description
Meas0	10	0	5	45	8 standard sources
Meas1	6	1	5	56	3 standard sources
Meas2	6	0	0	16	3 standard sources
Meas3	5	2	5	33	4 standard sources
Meas4	7	1	1	17	4 standard sources
Meas5	9	0	2	20	4 standard sources
Meas6	6	1	5	37	4 standard sources
Meas7	6	0	1	14	5 standard sources
Meas8	7	1	1	10	5 standard sources
Meas9	7	0	1	17	6 standard sources
Meas10	14	0	3	49	Natural uranium
Meas11	18	0	8	114	Natural uranium
Meas12	21	2	2	56	Irradiated uranium
Meas13	23	4	3	94	Irradiated uranium
Meas14	2	0	3	22	Irradiated uranium
Backgd0	19	2	14	126	Room background
Backgd1	14	1	10	74	Room background
Backgd2	9	1	3	30	Room background
Backgd3	8	0	1	15	Room background
Stuk1	10	0	2	46	Environmental sample
Stuk2	16	1	10	50	Environmental sample
Stuk3	23	1	14	112	Environmental sample
Stuk4	33	2	17	87	NAA sample
Stuk5	25	3	11	74	NAA sample
Stuk6	6	2	7	46	Gaseous reactor waste
Chernobyl	52	8	63	275	Chernobyl fallout
Average	13.9	1.3	7.6	59	

TABLE 2. IDENTIFICATION RESULTS OF THE SYNTHESIZED TEST SPECTRA

Spectrum label	Present nuclides	Unidentified nuclides	Spurious nuclides	Number of peaks	Description
Synt40	1	0	5	66	¹⁷⁶ Lu source
Synt41	4	0	0	17	Artificial sample
Synt42	4	0	0	18	Artificial sample
Synt43	2	0	0	16	Artificial sample
Synt44	9	0	1	39	Gaseous reactor waste
Synt45	7	2	3	30	Gaseous reactor waste
Synt46	19	1	9	108	Reactor coolant
Synt47	13	1	1	41	Reactor coolant
Synt48	7	0	4	48	Accelerator nuclides
Synt49	7	1	11	114	Accelerator nuclides
Synt50	7	0	12	43	Accelerator nuclides
Synt51	6	0	15	99	Accelerator nuclides
Synt60	12	6	11	38	Artificial NAA sample
Synt61	9	1	6	28	Artificial NAA sample
Synt62	12	2	8	46	Artificial NAA sample
Synt63	9	0	9	16	Artificial NAA sample
Synt64	11	4	14	34	Artificial NAA sample
Synt65	21	0	7	98	Artificial NAA sample
Synt66	11	0	4	89	Artificial NAA sample
Synt67	7	0	0	17	Realistic NAA sample
Synt68	10	0	2	24	Realistic NAA sample
Synt69	8	0	1	37	Realistic NAA sample
Synt70	15	0	3	43	Realistic NAA sample
Synt71	10	1	2	40	Realistic NAA sample
Synt72	14	1	4	52	Realistic NAA sample
Average	9.4	0.8	5.3	48	

TABLE 3. EXPLANATION OF PEAKS OF THREE SAMPLE NUCLIDES IN SPECTRUM STUKI WITHOUT AND WITH COINCIDENCE CORRECTION. IDEALLY THE EXPLAINED PEAK SHARES WOULD BE 100%. THE DATA HAVE BEEN TAKEN DIRECTLY FROM THE SHAMAN REPORT FILES

Nuclide	E (MeV)	I (%)		I (%)	
		(without coinc. corr.)	Peak share (%)	(with coinc. corr.)	Peak share (%)
^{116m} Ag	0.446811	3.750	105.131	1.9510	86.795
	0.62036	2.806	124.780	1.4429	101.821
	0.657762	94.600	96.846	58.8866	95.664
	0.677623	10.350	113.368	5.2983	92.093
	0.687015	6.440	100.870	3.6439	90.569
	0.706682	16.440	101.794	8.8146	86.609
	0.744277	4.730	99.130	2.6643	88.607
	0.763944	22.290	103.266	13.6509	100.357
	0.818031	7.340	116.410	3.7198	93.617
	0.884685	72.700	101.492	44.6764	98.973
	0.937493	34.360	95.914	21.5004	95.239
	1.3843	24.280	88.385	16.4682	95.130
	1.47579	3.995	85.983	2.9009	99.077
	1.50504	13.040	85.793	8.6291	90.090
	1.5424	0.000	0.000	4.1143	156.303
1.5623	1.029	39.509	1.2635	76.981	
¹²⁴ Sb	0.60273	97.800	96.325	83.5675	106.629
	0.645855	7.380	109.249	5.2122	99.021
	0.70932	1.350	110.996	0.8465	89.321
	0.713781	2.270	110.444	1.4904	93.060
	0.722786	10.760	103.556	7.6064	93.947
	0.968201	1.888	102.941	1.2396	86.737
	1.04513	1.840	99.899	1.1538	80.391
	1.32551	1.620	74.161	2.2358	131.352
	1.36816	2.620	98.008	1.7202	82.581
	1.43656	1.230	91.333	1.0903	103.897
1.69098	47.300	91.465	37.5700	93.234	
¹³⁴ Cs	0.47535	1.460	95.614	0.9591	64.928
	0.563227	8.380	111.323	5.2990	72.764
	0.569315	15.430	113.179	9.7957	74.271
	0.604699	97.600	104.372	74.4136	82.257
	0.795845	85.400	101.923	64.9025	80.069
	0.801932	8.730	113.106	5.7351	76.806
	1.03857	1.000	98.165	0.8863	89.935
	1.16794	1.800	84.186	2.0485	99.034
	1.36515	3.040	70.668	3.8606	92.767
1.4005	0.000	0.000	6.4794	157.596	

4.11. DISCUSSION

The aim of expert system *SHAMAN* is to solve the problem of radionuclide identification in gamma spectrum analysis, which is rather difficult when using a full radionuclide library with 50 000...80 000 gamma lines. The system has been verified using a well-defined set of test spectra, and the performance has been found to be generally very good, but naturally with some room for improvement. It must be kept in mind that when evaluating performance for a radionuclide identification system, there are several factors that need consideration. The actual requirements for the accuracy and speed of identification depend strongly on the application at hand. For an early warning monitoring system high speed is preferred over pinpoint accuracy whereas the best identification results might be required in another application. An identification system such as *SHAMAN* should preferably be able to satisfy both of these somewhat contradicting requirements.

The test suite described in this paper is an excellent metric by which the performance of an identification system can be measured. After the results from the test suite have become available, the development effort of *SHAMAN* has concentrated on making the identification results more consistent and in a better agreement with the expected results. This development still continues.

The existing set of rules in the expert system is a good basis for a number of site-specific rule sets, which can be used to make more advanced deductions about the particular problem. This requires a facility by which the user can easily define his own rules in the system. The current procedure of recompiling the rule base is efficient but somewhat awkward in this respect. Specialized rule sets can also be created for different application domains, like environmental measurements or nuclear reactor monitoring systems. It should be noted that all the above results have been obtained by using *SHAMAN* as a 'black box'. We are currently developing a graphical user interface that facilitates an easy user intervention and intuitive display of the results obtained. This will allow the user to easily verify the correctness of the results and guide *SHAMAN* to recalculate some results, if necessary. Currently the user intervention is mostly needed in discarding spurious nuclides. These extra nuclides are mainly due to the conservative rule base and reconsidering mechanism of *SHAMAN*.

When integrating the quantitative and qualitative analysis phases a direct feedback from identification results and activities to the peak energy and area determination becomes possible. Currently *SHAMAN* assumes that all the peaks above a given threshold are correctly analysed. If the results, however, suggest that it is not the case, the user has to do a separate analysis update and then run *SHAMAN* again. When the integration has been completed we wish to have a seamless system from gamma spectrum measurement and analysis to full identification and activity determination.

REFERENCES TO SECTION 4

- [1] AARNIO, P.A., HAKULINEN, T.T., , ROUTTI, J.T., Expert system for nuclide identification and interpretation of gamma spectrum analysis, *Radioanal. Nuclear Chem.* **160** 1 (1992) 245.
- [2] AARNIO, P.A., ALA-HEIKKIL, J.J., HAKULINEN, T.T., , ROUTTI, J.T., "Expert system for nuclide identification in gamma spectrum analysis", Paper presented at ANS Topical Conference on Methods and Applications of Radioanalytical Chemistry-III, Kailua--Kona, Hawaii, April 10–16, 1994.
- [3] ALA-HEIKKILA, J.J., Coincidence Correction in *SHAMAN* for Environmental Measurements, Rep. TKK-F-C166 (1994).



- [4] AARNIO, P.A., NIKKINEN, M.T., ROUTTI, J.T., "Gamma spectrum analysis including naa with sampo for Windows", Paper presented at ANS Topical Conference on Methods and Applications of Radioanalytical Chemistry-III, Kailua-Kona, Hawaii, April 10–16, 1994.
- [5] American National Standard for Information Systems – Programming Language C (X3J11/88-159). Washington, DC (1989).
- [6] HARMON, P., KING, D., Expert Systems. John Wiley & Sons, Inc. New York (1985).
- [7] AARNIO, P.A., ROUTTI, J.T., J.V. SANDBERG, MicroSAMPO – personal computer based advanced gamma spectrum analysis system, Radioanal. Nucl. Chem. **124** 2 (1988) 457.
- [8] AARNIO, P.A., NIKKINEN, M.T., ROUTTI, J.T., "SAMPO 90 — High resolution interactive gamma spectrum analysis including automation with macros", Paper presented at ANS Topical Conference on Methods and Applications of Radioanalytical Chemistry-II, Kona, Hawaii, April 21–27, 1991.
- [9] CURRIE, L.A., Limits for qualitative detection and quantitative determination, Anal. Chem. **40** (1968) 586.
- [10] National Nuclear Data Center Online Nuclear Data Service. NNDC/ONL-92/8 (periodically revised).
- [11] LEDERER, C.M., SHIRLEY, V.S., (Eds), Table of Isotopes, Seventh Edition. John Wiley & Sons, Inc. New York (1978).
- [12] BROWNE, E., FIRESTONE, R.B., SHIRLEY, V.S., (Eds), Table of Radioactive Isotopes, John Wiley & Sons, Inc. New York (1986).
- [13] ERDTMANN, D., SOYKA, W., The Gamma Rays of the Radionuclides. Verlag Chemie. New York (1979).
- [14] CROUCH, E.A.C., Fission-product yields from neutron-induced fission, Atomic Data and Nuclear Data Tables **19** 5 (1977) 417.
- [15] DEBERTIN, K., HELMER, R.G., Gamma- and X-ray Spectrometry with Semiconductor Detectors, North Holland. Amsterdam (1988).
- [16] HEYDORN, K., RHEE, S.K., Advance Prediction of Single- and Double-Escape Peak Areas in Gamma-Ray Spectrometry, 1989 Winter Meeting, San Francisco, California, November 26–30, 1989. American Nuclear Society, Transactions **60** (1989) 3.
- [17] GOLUB, G., VAN LOAN, C.F., Matrix Computations. North Oxford Academic Publishing Co. Ltd. Oxford (1983).
- [18] PRESS, W.H., FLANNERY, B.P., TEUKOLSKY, S.A., VETTERLING, W.T., Numerical Recipes in C, The Art of Scientific Computing. Cambridge University Press, Cambridge (1988).

5. QUALITY ASSESSMENT OF SOFTWARE FOR GAMMA RAY SPECTRUM ANALYSIS¹⁰

5.1. SUMMARY

Computer programs play an important role in the analysis of gamma ray spectra. However, computer results should be in statistical control. In this line, tests carried at Risø National Laboratory and some guidelines on QA/QC are hereafter presented.

5.2. INTRODUCTION

Although the aim of nuclear activation analysis is to determine the amount of a stable isotope in a sample, rather than the amount of a radioactive isotope, there is no substantial difference between the two types of measurement with respect to the processing of gamma ray

¹⁰ Prepared by K. Heydorn, Risø National Laboratory, Denmark.