



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

The effect of energy resolution on the extraction of information content from gamma-ray spectra

K. E. Nelson, T. B. Gosnell, D. A. Knapp

December 27, 2010

Nuclear Instruments and Methods in Physics Research A

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

1 **The effect of energy resolution on the extraction of information content**
2 **from gamma-ray spectra**

3
4 Karl E. Nelson^{*}, Thomas B. Gosnell, David A. Knapp

5 Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
6

7 **Abstract**

8 Rapid unambiguous identification of illicit radioactive materials is a matter of
9 international concern that has yet to be fully attained. This is particularly true for shielded
10 fissile materials in the presence of background radiation and routinely encountered
11 benign radionuclides. We present a systematic treatment of the effect of detector
12 resolution on the identification of nuclides in simple and complex shielded γ -ray spectra.
13 Case studies of some problematic spectra with 1000 counts suggest that, at moderate
14 count rates, near unambiguous identification of potential fissile sources requires a
15 detector with an energy resolution of at least 1–2% at 662 keV. While somewhat limited
16 in scope, the study establishes a quantitative basis for the comparative evaluation of
17 detectors of comparable efficiency.

18 *Keywords:* Gamma-ray spectrometry, Radionuclide identification, Nuclear security
19

20 **1. Introduction**

21 Detecting fissile materials with a low probability of error is a challenging task for first
22 responders, airport security personnel, and U.S. port and border inspectors. International
23 concern over the transport of such materials has resulted in the development and use of a
24 variety of γ -ray detectors to scan for and identify radioactive sources in many types of
25 containers.

26 Fissile material detection is hampered by a wide variety of benign radioactive materials
27 that are routinely present in a broad range of shielding conditions. Rapid characterization
28 of these sources as benign, with a low probability of error, is required in order to
29 minimize their effect on normal activities, in particular, the flow of commerce.

30 While there has been some analysis of performance of various gamma-ray detector
31 systems [1, 2], the general problem of accurate nuclide identification measurements in the
32 presence of unknown and variable shielding has not been systematically addressed. The
33 importance of good energy resolution for high-quality spectral measurements has been
34 recognized for decades. Experience has shown that when detectors of different
35 resolutions are appropriately sized to an inspection venue, accuracy of nuclide
36 identification is more dependent on differences in energy resolution than on differences

* Corresponding author. kenelson@llnl.gov (Karl Nelson).

37 in detection efficiency. In this paper we quantify this qualitative observation. We
38 eliminated all effects of variable detector size and efficiency and focused only on the
39 effects of energy resolution on source identification.

40 We performed computer simulations of spectra from shielded sources for detectors with
41 energy resolutions in the range 0.25%–8% FWHM at 662 keV. The spectra were chosen
42 to represent radionuclides most commonly found in worldwide commerce as well as a
43 few fissionable materials. We chose the methodology of Principal Component Analysis
44 (PCA) [3, 4] as an unbiased means of assessing the information content of these spectra
45 in a global sense. While this approach does not focus solely on the information normally
46 extracted by detailed analysis with common spectral analysis tools, it does provide a near
47 quantitative measure of all available information as a function of energy resolution for
48 fixed statistical quality as well as a quantitative measure of the effects of counting
49 statistics on the information that can be extracted from a spectrum.

50 We first present a brief discussion of PCA methodology. This is followed by an
51 assessment of PCA methodology as a quantitative measure of the information content of
52 γ -ray spectra and then by its application to extract information from shielded spectra in
53 the absence and presence of statistical noise. Finally, a few case studies of sources that
54 have proven to be problematic are presented to assess the effect of resolution on the
55 accuracy of fissile source identification.

56 **2. PCA as a metric for information content**

57 The most prominent features of γ -ray spectra normally considered are full-energy peaks,
58 escape peaks, and Compton edges. In the case of shielded spectra, these features can be
59 diminished in intensity or obscured and modified by the energy-dependent attenuation of
60 an absorber. Information content can be expressed as the total number of independent
61 spectral shapes that can be represented by the response of a detector. The information
62 content cannot exceed the total number of spectral channels used to store a spectrum.

63 A high-resolution detector requires many spectral channels, while a lower resolution
64 detector requires fewer spectral channels. However, this simple view of information
65 content does not account for correlations between the channels. To gain this underlying
66 information content, we must apply statistical analysis.

67 For statistical analysis, we treat the spectral data as feature vectors. Each vector
68 represents a particular source and shielding configuration measured by a detector with a
69 particular energy resolution. We construct a representation of the universe of possible
70 spectral shapes by sampling models of different sources and shielding configurations.
71 Details of the methodology of selection of this sampling are covered in later sections.
72 Using the library of potential spectral shapes, we compute a correlation matrix from the
73 feature vectors. The coefficients of this matrix represent how much each feature varies
74 with a corresponding change in another. For example, increasing the intensity of a photo-
75 peak increases the intensity of the corresponding Compton edge as well. The correlation
76 matrix captures these relationships. Correlated channels represent a reduction of total
77 possible information content, as common information is not useful for distinguishing
78 individuals in the set. The more independent dimensions there are in the underlying
79 correlation matrix, the greater the information content.

80 PCA is a convenient means to estimate this information content for both varying
 81 resolution and varying statistical quality of a spectrum. Applying PCA to a correlation
 82 matrix produces a linear transformation (a rotation) of a data set into a basis set of
 83 eigenvectors. The number of eigenvectors, or principal components, is the same as the
 84 dimensionality of the data before the transformation. The corresponding eigenvalues
 85 represent the variance in that rotated dimension. As the information to distinguish
 86 between one spectral shape and another is based on the spectral differences, the variance
 87 represents information content in that dimension. A few eigenvectors, containing the
 88 most variance, can account for the vast bulk of the total variance. By ordering the
 89 eigenvalues by decreasing magnitude, the minimum number of eigenvectors required to
 90 represent a specified fraction of the total variance is then a direct measure of the
 91 information content of the spectral set.

92 In this study, we are interested in assessing the general effect of resolution on the
 93 information contained in a broad range of spectral distributions. Our results do not pertain
 94 to any specific γ -ray spectrum taken under a specific set of experimental conditions or
 95 under any specific spectral analysis technique or algorithm. Rather, they represent a
 96 global measure of the effect of detector resolution on the spectra from a large set of
 97 radioactive sources under a well-defined set of conditions.

98 All of our computed spectra were simulated with a fixed system gain in a fixed number of
 99 channels, M . An $M \times N$ matrix \mathbf{x} of N spectra is constructed where the matrix elements
 100 are the counts or intensity in the M channels of the spectra. If x_{ji} is the number of counts
 101 in the j th channel of the i th spectrum, the vector of the channel means $\hat{\mathbf{i}}$ for the entire set
 102 is given by

$$103 \quad \hat{\mu}_j = \frac{1}{N} \sum_{i=1}^N x_{ji} \quad (1)$$

104 and the vector of the variance $\hat{\sigma}^2$ of the set about the channel means is given by

$$105 \quad \hat{\sigma}^2_j = \varepsilon + \frac{1}{N} \sum_{i=1}^N (x_{ji} - \hat{\mu}_j)^2 \quad (2)$$

106 Because γ -ray sources cannot produce any signal above the highest gamma energy
 107 detected, they can have many zero-valued channels. These higher-energy channels can
 108 cause numerical instability in the PCA decomposition. To correct for this effect, we add a
 109 minuscule but sufficient constant ($\varepsilon = 10^{-10}$) to each channel variance to stabilize the
 110 computation.

111 The $M \times M$ correlation matrix is then given by

$$112 \quad r_{ij} = \frac{1}{N} \sum_{k=1}^N \frac{(x_{ki} - \hat{\mu}_i)(x_{kj} - \hat{\mu}_j)}{\hat{\sigma}_i \hat{\sigma}_j} \quad (3)$$

113 The PCA decomposition proceeds by use of a singular value decomposition (SVD)
 114 algorithm to obtain the eigenvectors and eigenvalues.

115 We computed the number of principal components to explain a given fraction of the
 116 variance in the information content. For our analysis we wish to extract the maximum

117 number of components possible while avoiding numerical difficulties. We chose to limit
118 the maximum percent of variance extracted to 99.9% as our fundamental basis of
119 comparison. For illustrative purposes, we computed the variance fraction at both 99% and
120 99.99%. MATLAB [5] was used for all numerical calculations.

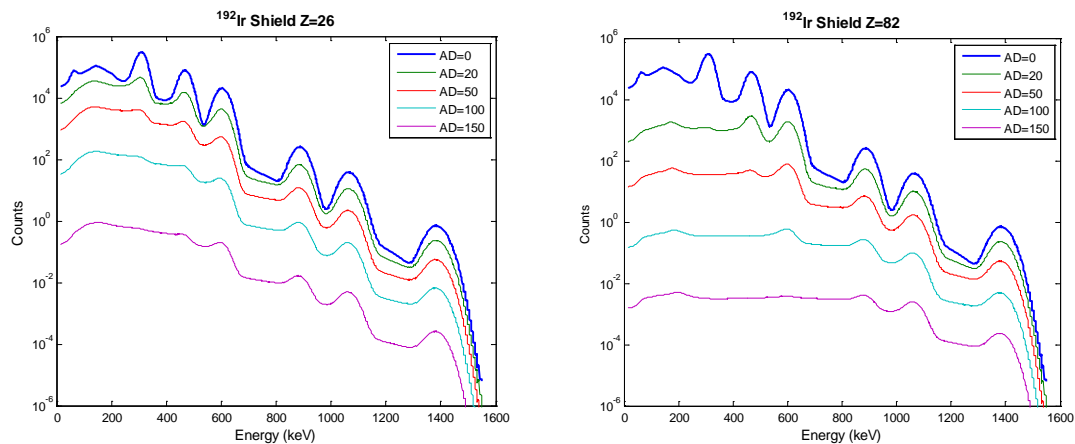
121 3. Simulation of gamma-ray spectra

122 We synthesized γ -ray spectra with a recent version of the Gamma Detector Response and
123 Analysis Software (GADRAS) suite using its one-dimensional radiation transport
124 capability [6, 7] and its detector response function model [8]. This model uses semi-
125 empirical parameterizations that can closely reproduce the response functions of a wide
126 range of detector types and configurations.

127 To eliminate all effects of variable detector size and efficiency, we first chose the full-
128 spectrum response characteristics of a generic (HPGe) detector with 140% relative
129 efficiency [9] at 1332 keV. We then broadened it numerically by adjusting only the
130 GADRAS resolution parameter to obtain responses at lower energy resolutions. Response
131 functions for resolutions of 0.25%, 0.5%, 1%, 2%, 4%, and 8% at 661.6 keV were then
132 modeled. The resolution was taken as proportional to $E^{1/2}$ and the noise from the detector
133 electronics was neglected. We simulated all of the spectra in 3000 channels at a fixed
134 system gain of 1.0 keV per channel. We generated spectra for individual radionuclides or
135 complex sources, both unshielded and shielded, using the evaluated photon energies and
136 intensities and shielding approximations given in the code. The photon data are generally
137 consistent with standard Evaluated Nuclear Structure Data File (ENSDF) evaluations,
138 with the deletion of some of the very weak transitions.

139 Examples of a series of spectra representing a ^{192}Ir source shielded by increasing
140 thicknesses of iron and lead in the range from 0 to 150 g/cm^2 obtained with the
141 procedures outlined above are shown in Figure 1.

142



143

144

145 **Fig. 1. Simulated spectra for ^{192}Ir showing changes in spectral shape for iron and lead shielding with**
146 **areal densities ranging from 0–150 g/cm^2 . Count rate is relative as time and distance can be scaled.**

147

148 **4. Test of the quantitative nature of the PCA approach**

149 In order to establish the ability of the PCA procedure to provide a quantitative measure of
150 the information content as a function of resolution alone, we created a series of detector
151 response functions for mono-energetic gamma rays equally spaced in 0.5 keV intervals
152 over the energy range 20 keV to 3 MeV for each of the resolutions given above. The
153 intensity in each channel of the simulated spectra was taken as exact with no statistical
154 uncertainty. The entire set of ~ 6000 spectra for each resolution was then subjected to
155 PCA analysis and the number of eigenvectors required to represent 99%, 99.9% and
156 99.99% of the total variance was estimated.

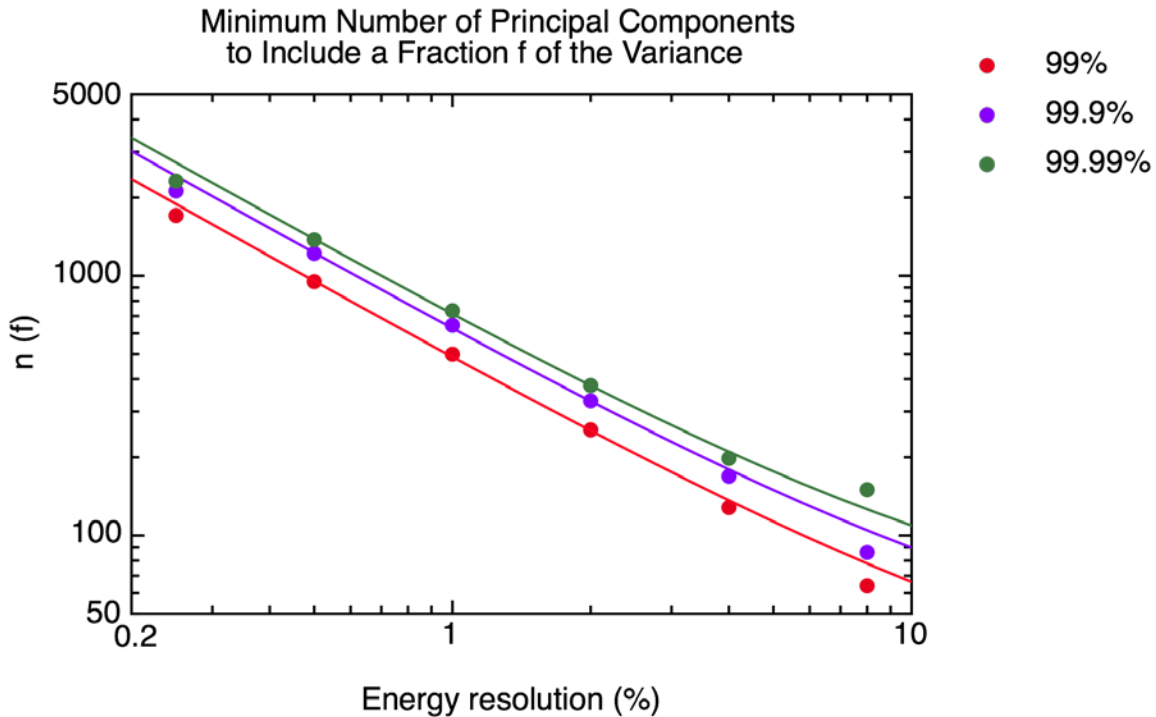
157 If each channel were statistically independent, the number of components necessary to
158 represent the spectrum to 99.9% of the total variance would be 99.9% of the total
159 channels. For a finite resolution detector, this total information content is reduced based
160 on the correlation between neighboring channels. For this oversampled set of ideal
161 spectra, our metric should estimate the maximum possible information content that can
162 be represented for a given energy resolution irrespective of what energy lines that nature
163 actually produces.

164 The results from the present calculations are shown as dots in Figure 2. The lines in the
165 graph are the “least squares” fit of the data to a function of the form

166
$$n(f) = \frac{a_1}{r} + a_2 \quad (4)$$

167 where $n(f)$ is the minimum number of eigenvectors required to represent the fraction f of
168 the total variance, and r is the resolution. We estimated the parameters a_1 and a_2 from the
169 fits when the resolutions 0.25% and 8% were excluded because of “end effects.” Apart
170 from the extremes in the resolutions, the fits demonstrate the expected inverse variation
171 of the included variance with resolution. As the fraction of variance included approaches
172 unity, the number of eigenvectors required asymptotically approaches the total number of
173 channels in the spectrum. The constants a_2 found from the fits are approximately 20, the
174 number of channels at low energy for which no peaks were generated and which, in this
175 ideal case, should not contribute to the total variance.

176



177

178 **Fig. 2. Number of principal components, $n(f)$, required to explain the specified fraction of the**
 179 **variance in the modeled spectra with evenly spaced gamma energies and in the absence of any noise.**
 180 **The values shown for 0.25% resolution are lower limits.**

181 Not counting electronic noise, 319 (~ 0.89%) of the approximately 18,000 simulated full-
 182 energy peaks have full widths at tenth maximum (FWTM) smaller than the channel width
 183 of 1 keV. They are almost all found in the spectra with resolutions of 0.25% and 0.5%. In
 184 addition, the complete widths of ~196 (~ 0.55%) of the highest-energy peaks are
 185 partially cut off at the 3000-keV spectrum limit. The majority of these are found in
 186 spectra with resolutions $\geq 2.0\%$. Because of both factors, spectral distributions near the
 187 extremes of some spectra will not follow the same regular patterns expected at
 188 intermediate resolutions.

189 Within these limitations, the simulations demonstrate that the PCA approach provides a
 190 well-behaved and near quantitative measure of the information content in a spectrum.
 191 The simulations suggest that increasing detector resolution by a factor of 2 roughly
 192 doubles the information content of an ideal spectrum. The relation between resolution
 193 and increasingly realistic γ -ray spectra is addressed in the following sections.

194 **5. Determination of the maximum possible information content of radionuclide**
 195 **spectra**

196 In order to approach a realistic measure of the effect of detector resolution on the
 197 information content of γ -ray spectra, we have chosen a set of radioactive sources that
 198 represent radionuclides typically found in worldwide commerce as well as a few
 199 fissionable materials (Table 1). The half-lives of these are mostly greater than 10 days,
 200 but a few, such as ^{99}Mo and $^{99\text{m}}\text{Tc}$, were included because of their heavy use in medical

201 imaging. Specific radionuclide mixtures of special interest, such as plutonium and highly
 202 enriched uranium (HEU), were also included.

203 **Table 1. Radioactive sources included in the resolution study of noise-free radionuclide spectra. In**
 204 **the upper left-hand corner, RG stands for reactor grade, WG stands for weapons grade [8], and**
 205 **HEU stands for highly enriched uranium.**

Pu, RG	⁵⁸ Co	⁹¹ Nb	¹¹¹ In	¹³⁴ Cs	¹⁵⁶ Eu	¹⁹⁰ Ir	²³² Th	²⁴⁰ Pu
Pu, WG	⁵⁹ Fe	⁹² Nb	¹¹³ Sn	¹³⁶ Cs	¹⁶⁰ Tb	¹⁹² Ir	²³² U	²⁴¹ Am
HEU	⁶⁰ Co	⁹⁵ Nb	^{114m} In	¹³⁷ Cs	^{166m} Ho	^{194m} Ir	²³³ U	²⁴¹ Pu
⁷ Be	⁶⁵ Zn	⁹⁵ Zr	¹²³ I	¹³⁹ Ce	¹⁷⁵ Hf	²⁰¹ Tl	²³⁴ Th	²⁴² Pu
¹⁸ F	⁶⁷ Ga	⁹⁹ Mo	^{123m} Te	¹⁴⁰ Ba	¹⁷⁶ Lu	²⁰² Tl	²³⁴ U	²⁴³ Am
²² Na	⁶⁸ Ge	⁹⁹ Rh	¹²⁴ Sb	¹⁴¹ Ce	¹⁷⁷ Lu	²⁰³ Hg	²³⁵ U	²⁴⁴ Cm
²² NaF	⁷⁵ Se	⁹⁹ Tc	¹²⁵ Sb	¹⁴⁴ Ce	^{177m} Lu	²⁰⁴ Tl	²³⁶ Np	²⁴⁹ Cf
⁴⁰ K	⁸⁵ Kr	^{99m} Tc	¹²⁶ Sb	¹⁴⁷ Nd	¹⁸² Ta	²⁰⁷ Bi	²³⁶ Pu	²⁵² Cf
⁴⁶ Sc	⁸⁵ Sr	¹⁰³ Pd	¹²⁷ Xe	¹⁵² Eu	¹⁸⁸ Pt	²¹⁰ Po	²³⁶ U	
⁵¹ Cr	⁸⁶ Rb	¹⁰³ Ru	¹³¹ I	¹⁵³ Gd	¹⁹⁰ Ir	²²⁶ Ra	²³⁷ Np	
⁵⁴ Mn	⁸⁸ Y	¹⁰⁶ Ru	^{131m} Xe	¹⁵³ Sm	¹⁹² Ir	²²⁷ Ac	²³⁸ Pu	
⁵⁶ Co	⁸⁹ Sr	¹⁰⁹ Cd	¹³³ Ba	¹⁵⁴ Eu	^{194m} Ir	²²⁸ Th	²³⁸ U	
⁵⁷ Co	⁹⁰ Sr	^{110m} Ag	¹³³ Xe	¹⁵⁵ Eu	¹⁸⁸ Pt	²³¹ Pa	²³⁹ Pu	

206

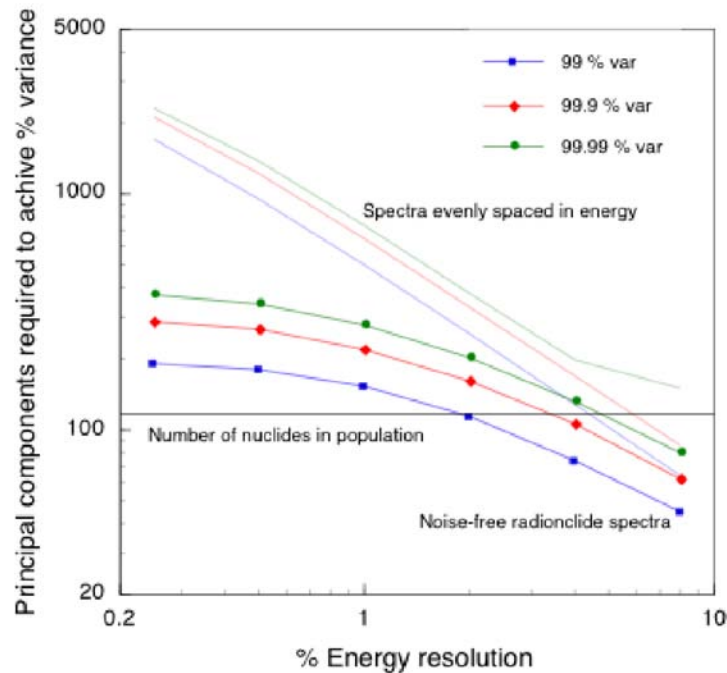
207 Essentially all of the nuclides and sources in Table 1 are normally transported within a
 208 shielded container. In order to simulate the range of shielding effects expected, we chose
 209 the elements Li ($Z = 3$), Fe ($Z = 26$) and Pb ($Z = 82$) to represent low-, medium- and
 210 high-Z shields, respectively. Approximately uniform sampling of the changes in spectral
 211 distributions resulting from energy-dependent attenuation and scattering was achieved by
 212 choosing thicknesses such that the χ^2 difference between successive spectra was equal to
 213 one standard deviation for a spectrum containing 10^4 counts at a detector resolution of
 214 1%. The number of thicknesses required varied with the complexity and maximum
 215 photon energy of the source. The synthesized spectra were convolved with detector
 216 response functions in the same manner as described above.

217 The covariance matrix produced using the method of constant χ^2 differences was largely
 218 invariance to selection of the χ^2 . We build library representations with as many as 24
 219 thousand templates and as few as six thousand. The resulting information content when
 220 evaluated in terms of the 99.9% threshold was the same. We have interpreted this
 221 invariant as indicating that we have properly covered the information space, and thus the
 222 population contained more information than could be represented in any energy
 223 resolution. The number of representatives of an individual nuclide was given by the
 224 range of spectral shapes. Tests found that moderately increasing (x2) or decreasing (x0.5)

225 the number of spectra for an individual nuclide did not affect the information content
226 results. Adding or removing nuclides would have a larger effect on information content.

227 5.1 Shielded and noiseless spectra

228 We then examined a global means of assessing the maximum information content that
229 can be extracted from common spectra under a wide range of shielding conditions. All of
230 the spectra discussed above for a fixed resolution were considered to be exact and, in the
231 absence of statistical noise and background radiation, were subjected to PCA analysis as
232 single data sets at each resolution studied. The minimum number of eigenfunctions
233 required to represent the majority of the variance in the spectra are as shown in Figure 3
234 along with those from Figure 2 for reference.



235

Fig. 3. Minimum number of principal components required to represent 99%, 99.9% and 99.99% of the variance of noise-free shielded spectra for the radionuclides in Table 1.

236 Compared to the near ideal case represented by the simulations considered in Figure 2,
237 the number of components required to extract the majority of the information content of
238 the shielded spectra, and hence the information content of these, is seen to be smaller by
239 factors of ~ 6-8 at the highest resolutions. This results primarily from the limited
240 information content of the unshielded spectra for which the typical spectrum contains
241 only ~2-4 strong γ -rays. With perfect resolution, the number of principal components
242 required to account for the variance in a spectrum should roughly be proportional to the
243 number of prominent gamma rays emitted in the decay of a source. The number of such
244 gamma rays for the 109 sources in Table 1 is ~ 200-400 and this is on the order of the
245 minimum number of eigenfunctions needed to account for the variance at the highest
246 resolution considered. At the lowest resolution, the number of eigenvectors required is
247 ~30-80, smaller by a factor of ~ 5 and clearly suggests that even under noiseless
248 conditions, the information content of relatively simple γ -ray spectra is strongly

249 compromised by loss in resolution. This arises not only from the effects produced by the
250 shielding, but also from the inability to distinguish the presence of closely spaced full-
251 energy peaks.

252 We also note that, for the worst resolution considered (8%), the number of eigenvectors
253 required to represent all of the information in the set of spectra is far less than the number
254 of nuclides used; this behavior means that for such detectors, some combinations of
255 nuclides are indistinguishable. The resolution is simply insufficient.

256 **5.2 Shielded spectra with statistical noise**

257 Estimates of the effect of statistical noise on the extraction of information from spectra
258 were obtained by assuming Poisson-distributed counts in each statistically independent
259 channel for all of the shielded spectra considered above. The noise only affects the
260 diagonal elements of the correlation matrix and was computed from the means of each of
261 the simulated spectra scaled by the total number of counts in a spectrum that varied in the
262 range 10^2 – 10^6 . Given the large number of spectra studied, the statistical bias from the
263 Poisson statistics closely approximates a normal distribution. The contribution to the
264 diagonal elements is then the mean from all spectra scaled by the number of counts in
265 each.

266 With the addition of statistical noise, the number of eigenfunctions required to include a
267 fixed fraction of the total variance is no longer a direct measure of spectral information.
268 Ordered in terms of their magnitudes, the first few eigenvalues are expected to represent
269 almost solely the variance from the average spectral fluctuations in the absence of noise.
270 Succeeding eigenvalues can be expected to include an increasing contribution from
271 statistical fluctuations. When the variance from the real spectral fluctuations is almost
272 exhausted, the eigenvalues of additional components will approach a constant value that
273 represents variance from statistical fluctuations alone. As a result, the number of
274 eigenfunctions required just before the region of constant eigenvalues is reached can
275 approximate the spectral information content.

276 As an example, the eigenvalues for the ordered set of eigenfunctions obtained from the
277 PCA analysis on the data set for a resolution of 1% for total spectrum counts in the range
278 10^2 – 10^6 are shown in Figure 4. The eigenvalue distributions initially demonstrate rough
279 exponential decreases followed by a decreasing slope to a region of approximately
280 constant values. With increasing statistical quality, the magnitude of the constant
281 eigenvalues decreases very roughly as the ratio of σ/N where σ is the standard deviation
282 of the number of counts, N , in a spectrum.

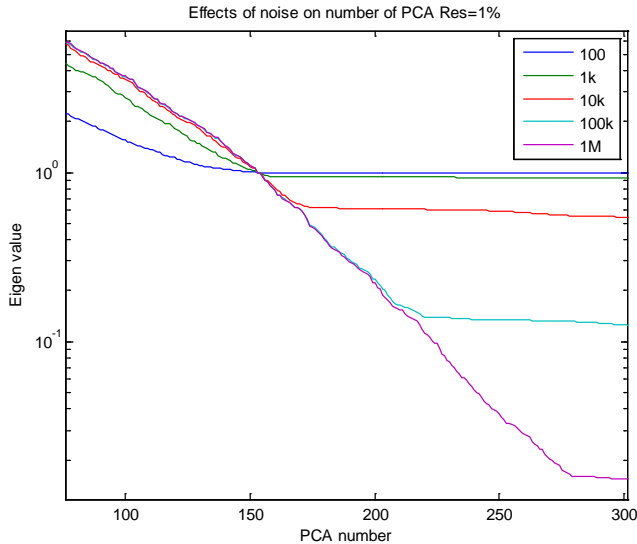


Fig. 4. Eigenvalue magnitude versus principal component number for a 1% resolution detector. The values are derived from theoretical spectra with expected Poisson statistical noise from a range of total observed counts. As the number of counts increases, the point beyond which the eigenvalues are equal (the “knee”) moves to a higher number of principal components. Beyond this point no further information can be extracted.

283

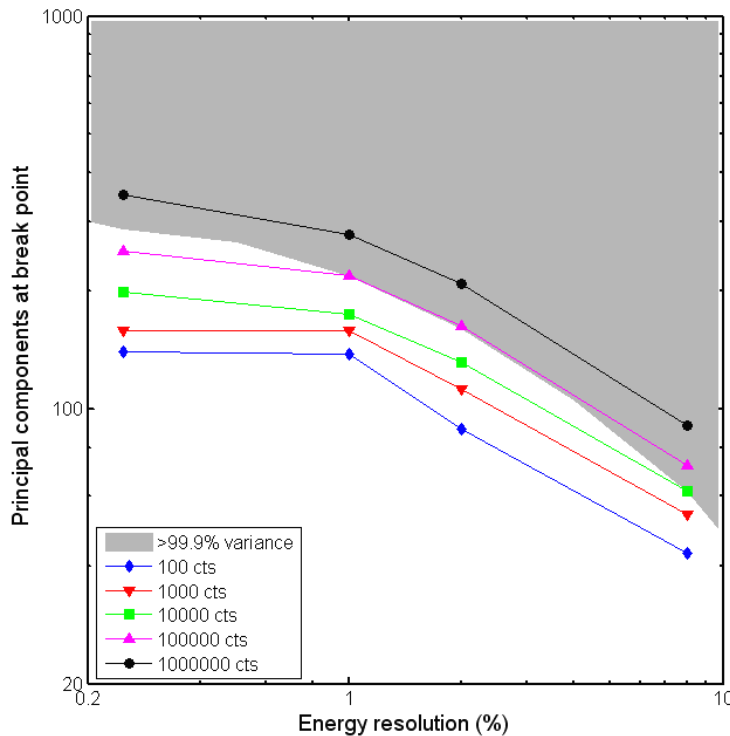
284 Using the estimated number of eigenfunctions just before the region of constant
 285 eigenvalues, the number of eigenfunctions required to extract the spectral information
 286 content from the entire data set for all shielded sources of Table 1 is shown in Fig. 5 as a
 287 function of detector resolution and statistical quality. For reference, the number of
 288 eigenfunctions required to include 99.9% of the total variance for noiseless spectra
 289 (Figure 3) is included in the figure.

290 Over the entire range in resolution, a change in the number of counts in the spectra by a
 291 factor of 10^2 changes the information content by a factor of ~ 3 -3.5 on average. Of
 292 particular interest is the fact that for spectra with 10^6 total counts, the number of
 293 components required to extract the information content is significantly larger than that for
 294 the equivalent noiseless data set. This reflects the fact that the eigenvectors do not cleanly
 295 separate the real spectral variance from the statistical noise. The analyses demonstrate the
 296 limited benefits of improving detector resolution when treating spectra of low statistical
 297 quality. For the present set of spectra, improving energy resolution from 8% to 1%
 298 increases the extractable information content by a factor of ~ 3 but no significant increase
 299 in extractable information is observed with continued resolution improvement for spectra
 300 containing $\leq 10^3$ counts. Even with 10^4 total counts, the extracted information from a
 301 detector resolution of 0.25% only increases by about 16% compared to that extracted
 302 from a spectrum taken at a resolution of 1%.

303 In comparison with methods of spectral analysis that rely on full-energy peak analysis,
 304 the results discussed above differ because the present approach is based on the
 305 information content included in the continua of the spectra. This bias can underestimate
 306 the ability of analysis procedures that search for individual pre-defined full-energy γ -ray
 307 peaks to extract useful information from spectra of low statistical quality with increasing
 308 energy resolution.

309 The value of our global analysis is limited, as our measure of information content does
 310 not illustrate the ability to distinguish between particular nuclides. It is possible that the
 311 distinguishing features between a benign source and a fissile source may only be

312 available in that final 16% mentioned above. We address this aspect of the problem in the
 313 following section.



314 **Fig. 5. Information content of the same radionuclide spectra used to produce Fig. 3, but with Poisson**
 315 **statistical noise added, for range of energy resolutions. The plotted points are the number of**
 316 **principal components at the “knees” in the curves, as illustrated in Fig. 4, for detectors of varying**
 317 **resolution. The “knee” values from Fig. 4 for 1% detectors are shown in the oval. The heavy black**
 318 **line indicates the number of principal components required to explain 99.9% of the available**
 319 **population variance in noise-free spectra (see Fig. 3).**

320 6. The effect of energy resolution on the identification of problematic radionuclides

321 In order to estimate the effect of detector resolution on specific source identification, we
 322 examined sources that have proven difficult for existing algorithms to properly classify
 323 [10] using a NaI(Tl) scintillator system. In particular we found that some benign sources
 324 were often misclassified as fissile sources (false positives) and that some fissile sources
 325 were often misclassified as benign sources (false negatives).

326 We parceled our collection of 103 sources into three schematic classes: fissile, benign,
 327 and background. For the specific fissile materials used in our case study, we chose
 328 weapons-grade Pu (WGPu) [11] and ^{237}Np , a fissile nuclide of increasing international
 329 concern [12, 13], from the fissile class. From the benign class, we chose ^{67}Ga , ^{131}I , ^{133}Ba ,
 330 and ^{192}Ir . For background, we used the entire class: ^{232}Th and ^{226}Ra and their daughters
 331 and ^{40}K . We then calculated the probability of misclassification of a spectrum by a
 332 regression-based identification algorithm as defined in the Appendix.

333 A straightforward application of the data sets to produce the principal components used
 334 for the analyses described in the previous sections is not appropriate to the case studies
 335 because, in calculating the variances about the mean value, the entire data set of all

336 nuclides in Table 1 was used. Any individual γ -ray spectrum will have its own unique
337 mean. It is this mean for each of the nuclides examined that must be explicitly
338 considered. To resolve this situation without having to perform a completely new set of
339 spectrum simulations, the magnitude of the mean for the initial data set was artificially
340 reduced to 1% of its original value. The PCA procedure was then applied and only the
341 eigenfunctions that contained the principal portion of the variance (99.9% of the total) for
342 each nuclide were retained. This limited set was shown to reproduce the spectral
343 distributions of the original nuclides with negligible error when measured in terms of the
344 χ^2 statistic for 10,000 count spectra and did not appreciably change the results of any of
345 the previous analyses. This procedure was used to generate approximate spectra for each
346 of the sources considered and for the entire set of shielding conditions discussed in
347 Sec. 5.

348 We selected a sample (without noise) from either the fissile or benign classes and then
349 solved by brute-force multiple linear regression to find the best possible linear
350 combination of all the spectra in the fissile and benign classes with the addition of
351 background subject to positive constraints. That is, for a specific sample, we would
352 produce the best linear combination of background alone, background plus sample, and
353 background plus benign spectra.

354 A non-negative least squares procedure, based on the complementary block pivot
355 algorithm [14], used for this purpose considerably reduced the number of conditionally
356 acceptable combinations. The choice of a non-negative algorithm is required to maintain
357 the physical requirement that spectra can only be described by the sum of different source
358 combinations and not their differences. A linear combination of templates can be
359 compactly represented as a linear combination of the magnitudes of the principal
360 components for those spectra; thus, the fit can be performed in lower-dimensional
361 principal-component space rather than as a fit to the spectra themselves.

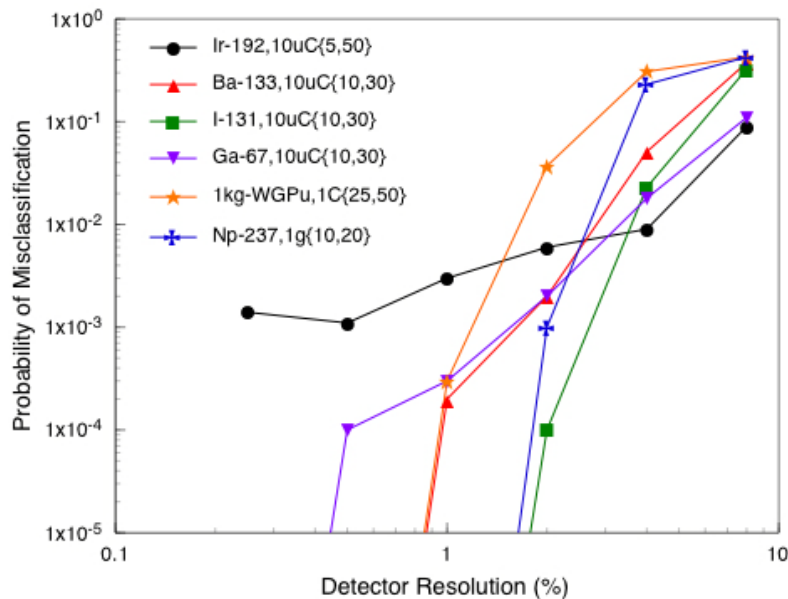
362 We then assumed a desired signal level of 1000 total counts from the source from which
363 we can estimate the expected noise. Samples which are well represented (in terms of χ^2
364 for the desired noise level for the best possible resolution) by the background set alone
365 are rejected as they do not have distinguishing information from which we could assess
366 the value of energy resolution.

367 For each source-plus-background spectrum, it was verified that the best fit for the correct
368 source choice was significantly better statistically than the best fit to background only.
369 The covariance matrix was then used to estimate a multivariate Gaussian uncertainty and
370 then to estimate the probability of misidentification of the source actually in the
371 spectrum.

372 We then estimated the rate of misclassification by computing the variance of the sample
373 assuming Gaussian statistics, as the integral of a Gaussian distribution assuming that the
374 decision is between the best representative sample in the fissile and benign sets. We
375 deemed this estimate as the ambiguity. An ambiguity of 0.50 would mean that calling the
376 sample the wrong class is a coin toss. An ambiguity of 0.0001 would indicate that an
377 error is estimated to occur only once in 10,000 times. We have summarized this in the
378 appendix.

379 The probability for misidentification of a source is shown in Figure 6 as a function of the
 380 assumed detector resolution for its selected conditions of shielding. The choice of
 381 shielding materials and thicknesses was made to roughly represent the type of shielding
 382 that might be expected in normal shipments found in commerce.

383 The γ -ray spectra from decay of ^{192}Ir and from decay of weapons-grade plutonium both
 384 have their principal intensities located below ~ 600 keV. When subject to significant
 385 attenuation from shielding, they are strongly modified from their unshielded spectral
 386 distributions. The distributions shown in the figure indicate that the probabilities for
 387 misidentification of these sources in the presence of background are quite small for
 388 resolutions better than $\sim 2\%$. ^{131}I , whose spectrum is dominated by a single transition at
 389 364 keV, is identified correctly so long as the resolution is better than $\sim 4\%$. So is that for
 390 ^{67}Ga , whose main intensity is concentrated in the same energy region. On the other hand,
 391 ^{237}Np , whose spectrum contains a large number of relatively weak transitions in the
 392 energy range 62–280 keV, is subject to a relatively high misidentification in the presence
 393 of shielding for detector resolutions worse than $\sim 2\%$.



394

Fig. 6. Misclassification probabilities for various scenarios and detector resolutions. The descriptors for each graph are the radionuclide, the source strength or mass and, in parentheses, the atomic number and areal density (g/cm^2) of the shielding material. The values in braces correspond to the effective Z and areal density (g/cm^2) of the shielding material. All scenarios assume a total of 1000 observed gamma rays in the detector.

395 For this set of tests and this set of radionuclides with the shielding chosen, the results
 396 shown in Figure 6 suggest that so long as detector resolutions are better than ~ 1 – 2% , the
 397 information content of spectra is sufficient to permit identification of sources with very
 398 low rates of misidentification. However, the probability for misidentification increases
 399 rapidly with worsening resolution

400 Our examples were done as a bounding analysis that only reveals the minimum required
 401 resolution for the sources we selected. We expect that there will be other source
 402 selections that are even more challenging. Real instrumental field spectra will also be
 403 more challenging to correctly interpret owing to such factors as added spectral distortion

404 from energy calibration imprecision, gain drifts, greater scattering complexity, and
405 changing background conditions.

406 **7. Conclusion**

407 To our knowledge, the effect of detector resolution on the information content of γ -ray
408 spectra has never been explored in a systematic and quantitative manner. This is of
409 relatively little concern in scientific studies where the resolution of typical HPGe
410 detectors approximates or approaches that required for the majority of detailed spectral
411 studies. But the ability to properly identify sources in the contents of containers, under
412 the rather stringent time restrictions determined by potential emergencies or the need to
413 minimize interruption of the normal flow of commerce can be very sensitive to the
414 statistical quality of spectra and their energy resolution. The global studies reported here
415 have demonstrated that the methodology of PCA provides an essentially quantitative
416 means of assessing the information content that can be extracted from γ -ray spectra as a
417 function of these parameters as well as shielding. Because of the global approach taken,
418 the analyses do not provide detailed information on any single source spectrum and its
419 shielding and statistical quality. Nevertheless, the global approach strongly suggests that
420 detectors designed for nuclide identification under generic screening conditions should
421 have resolutions at 661.6 keV no worse than 1–2% to achieve low misidentification rates.

422 By its very nature, the PCA methodology is biased by features of a spectrum not always
423 considered in the analysis of spectra with field instruments. As such, it may tend to
424 somewhat overestimate the resolution required to achieve nuclide identification with very
425 small probability of misidentification but this is not expected to invalidate the general
426 conclusions drawn here. Because of the widely-varying nature of photon spectra from
427 different radionuclides and the varying effects from attenuation by shielding materials,
428 the sensitivity for misidentification of any particular source may differ considerably from
429 the schematic results we have obtained. Nevertheless, the methodology of PCA seems
430 robust for such detailed analysis and the effect of resolution on it.

431 **Acknowledgments**

432 The authors wish to thank Stanley Prussin for contributions to the manuscript. This work
433 performed under the auspices of the U.S. Department of Energy by Lawrence Livermore
434 National Laboratory under Contract DE-AC52-07NA27344.

435 **Appendix**

436 In order to compute the ambiguity between assigning a sample to the class of benign,
437 background or fissile sources, we define matrices for the background library (A_B), the
438 fissile library (A_T), and the benign library ($A_{\bar{T}}$). For each resolution we compute a matrix
439 containing the subset of eigenvectors (U^T) that represent 99.9% of the total variance.
440 We use these components to compute for a test sample (Y) the best representative
441 sample in each set (B , $T + B$, $\bar{T} + B$). After processing we are left with the best
442 representative spectral shape in each of the potential classes ($\tilde{Y}_B, \tilde{Y}_T, \tilde{Y}_{\bar{T}}$).

$$UY = UA_B x_B \mid x_B \geq 0 \Rightarrow \tilde{Y}_B = A_B x_B$$

$$443 \quad UY = U(A_T x_T + A_B x_{BT}) \mid x_T, x_{BT} \geq 0 \Rightarrow \tilde{Y}_T = A_T x_T + A_B x_{BT} \quad (\text{A.1})$$

$$UY = U(A_{\bar{T}} x_{\bar{T}} + A_B x_{B\bar{T}}) \mid x_{\bar{T}}, x_{B\bar{T}} \geq 0 \Rightarrow \tilde{Y}_{\bar{T}} = A_{\bar{T}} x_{\bar{T}} + A_B x_{B\bar{T}}$$

444 [If Y was drawn from T then Y_T will equal Y , etc.]

445 We then compute the variance matrix (Σ_Y) of the sample assuming 1000 sample counts.

446 From this variance matrix, we can compute the variance of the decision metric (σ^2). For

447 the background set, we compute the distance metric (χ^2) from background set in the

448 space and reject the sample as being indistinguishable from background if this distance is

449 insufficient. None of the samples used in this study failed to meet this criterion.

$$450 \quad \chi^2 = (Y - \tilde{Y}_B)^T U^T (U \Sigma_Y U^T)^{-1} U (Y - \tilde{Y}_B) \quad (\text{A.2})$$

451 To get the decision variance, we compute the scalar uncertainty projected between the

452 best representative sample in the fissile and benign sets.

$$453 \quad \sigma^2 = (Y_T - Y_{\bar{T}})^T U^T U \Sigma_Y U^T U (Y_T - Y_{\bar{T}}) \quad (\text{A.3})$$

454 The decision rate error is then estimated as the integral of sample being misclassified. As

455 the sample under test was taken from one of the sets (benign or fissile), the error rate will

456 be the integral of the probability distribution in the other set, starting from the point

457 where the probability of assignment was equal.

458 References

- 459 1. C.J. Sullivan, S.E. Garner, M. Lombardi, K.B. Butterfield, M.A. Smith-Nelson, IEEE
460 Nucl. Sci. Symp. Conf. Rec. N24-64 (2007) p.1181.
- 461 2. L. Pibida, M. Unterweger, L.R. Karam, J. Res. Natl. Inst. Stand. Technol. **109** (2004)
462 451.
- 463 3. I. T. Jolliffe, *Principal Component Analysis*, 2nd Ed., Springer, New York (2002).
- 464 4. D. F. Morrison, *Multivariate Statistical Methods*, 4th ed., McGraw-Hill, New York,
465 (2005) 264–316.
- 466 5. <http://www.mathworks.com/products/matlab/>
- 467 6. D. J. Mitchell, J. Mattingly, *Rapid computation of gamma-ray spectra for one-*
468 *dimensional source models*, Trans. Am. Nucl. Soc. **98** (2008) 565–566.
- 469 7. J. Mattingly and D. J. Mitchell, *A framework for the solution of inverse radiation*
470 *transport problems*, Nuclear Science Symposium Conference Record, NSS '08, IEEE
471 (2008).
- 472 8. D. J. Mitchell, H. M. Sanger, K. W. Marlow, *Gamma-ray response functions for*
473 *scintillation and semiconductor detectors*, Nucl. Instr. and Meth. **A276** (1989) 547.
- 474 9. K. Debertin, R. G. Helmer, *Gamma- and X-ray spectrometry with semiconductor*
475 *detectors*, North-Holland, Amsterdam (1988) 135–136.
- 476 10. Labov, Simon, et al., *Foundations for Improvement to Passive Detection Systems—Final*

- 477 *Report*, Lawrence Livermore National Laboratory report, UCRL-TR-207129 (October
478 2004) pp. 62–63.
- 479 11. United States Department of Energy, *Nonproliferation and arms control assessment of*
480 *weapons-usable fissile storage and excess plutonium disposition alternatives*,
481 DOE/NN-0007 (1997) pp. 37–38.
482 http://www.osti.gov/bridge/product.biblio.jsp?osti_id=425259
- 483 12. U. S. Department of Energy, *Restricted declassification decisions, 1946 to the present*,
484 (2001).
485 <http://www.fas.org/sgp/othergov/doe/rdd-7.html#I20>
- 486 13. T. Burr, W. D. Stanbro, W. Charlton, *An evaluation of safeguards approaches for*
487 *Neptunium*, J. Nucl. Sci. and Tech, **38, 3** (2001) 209.
- 488 14. L. F. Portugal, J. J. J´udice, L. N. Vincente, *A comparison of block pivoting and interior-*
489 *point algorithms for linear least squares problems with nonnegative variables*, Math.
490 Comp. **63** (1994) 208, 625–643.