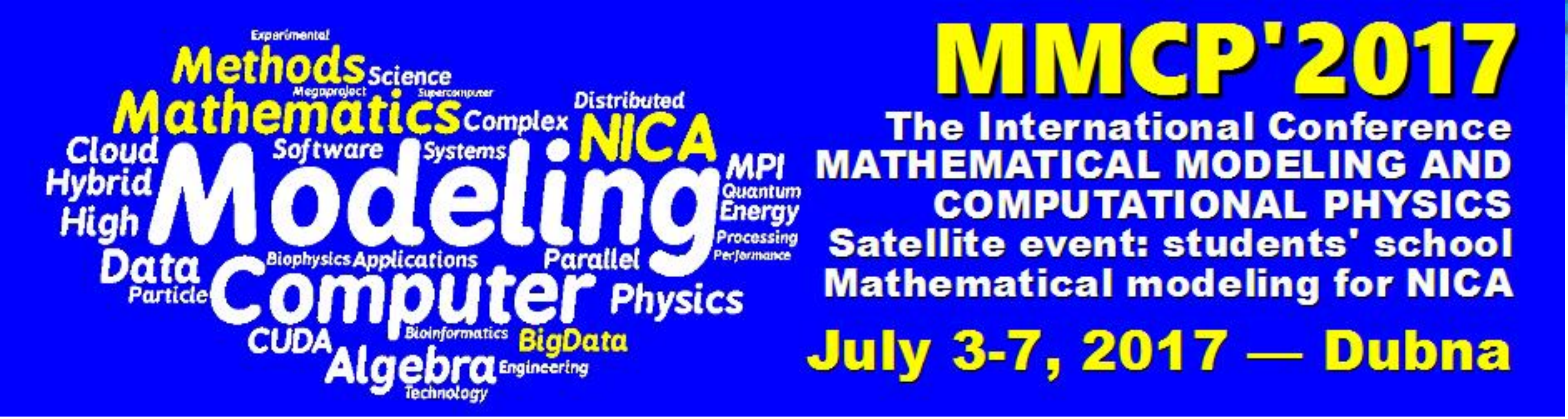


Machine-learning algorithms for classification and separation of noisy signals

O.I Streltsova [1], Y.A. Butenko [1] and A.I Streltsov [2]

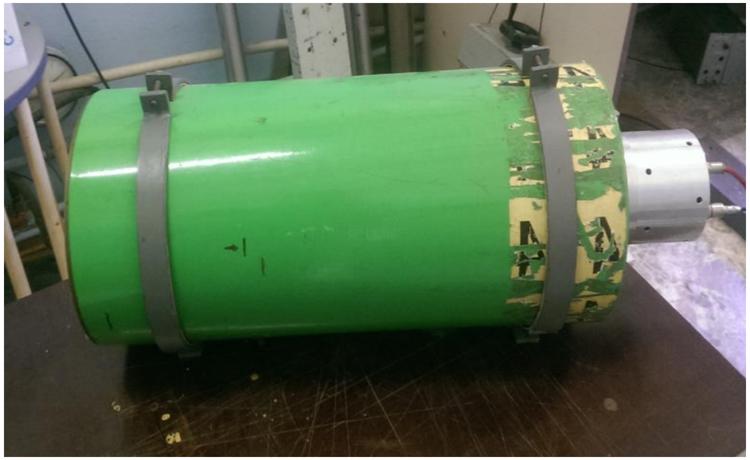
[1]*Joint Institute for Nuclear Research, LIT, Dubna*

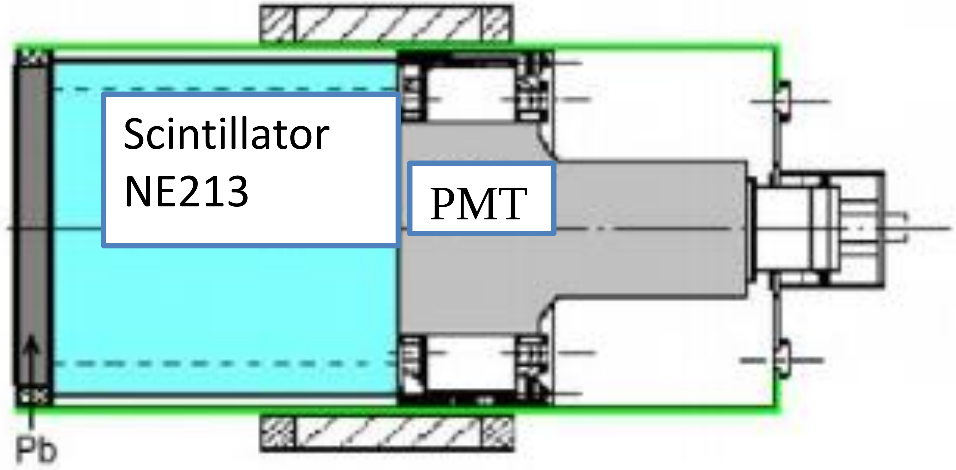
[2]Institut für Physik und CINSaT, Universität Kassel, Kassel, Germany

In collaboration with group of
E.M. Kozulin (Flerov Laboratory of Nuclear Reactions, JINR)

Heterogeneous Computation Team, *HybriLIT*

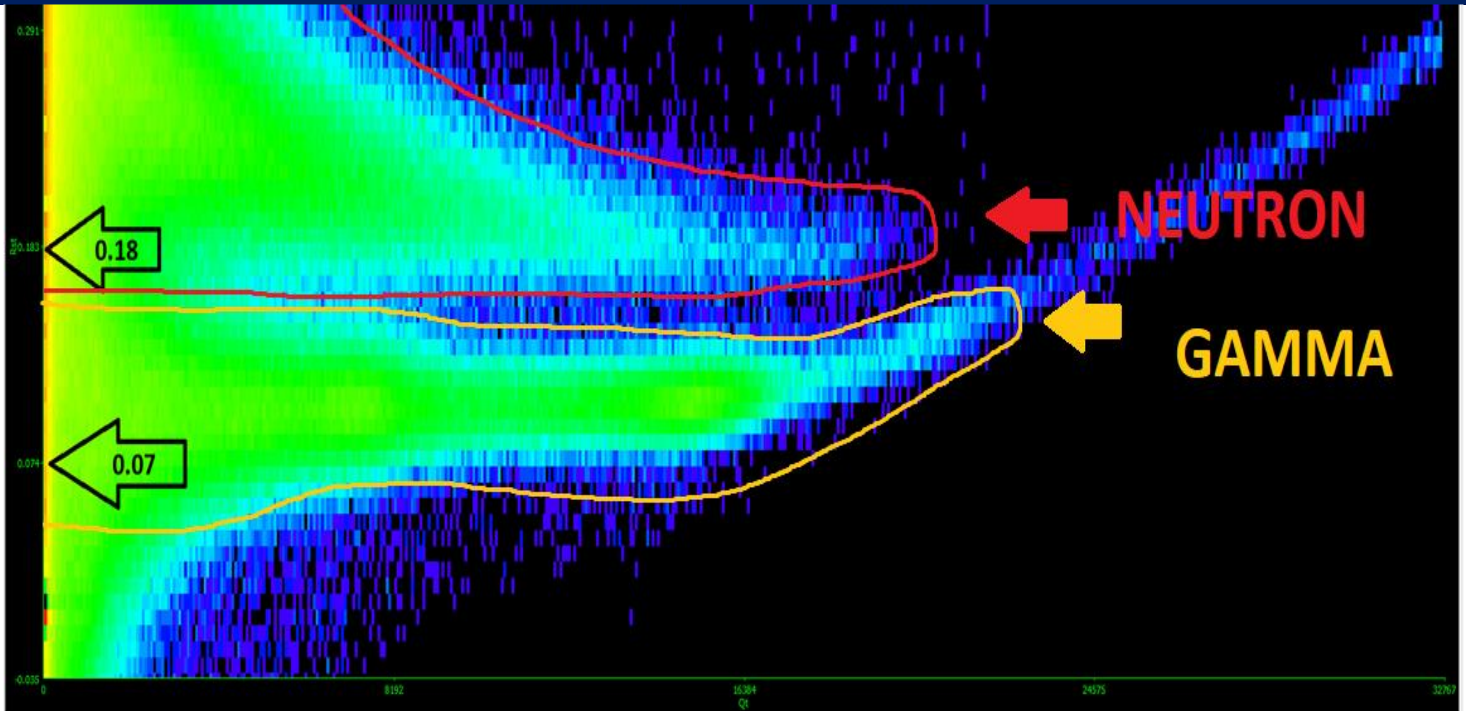
Spectrometer DEMON (DEtecteur MOdulaire de Neutrons)





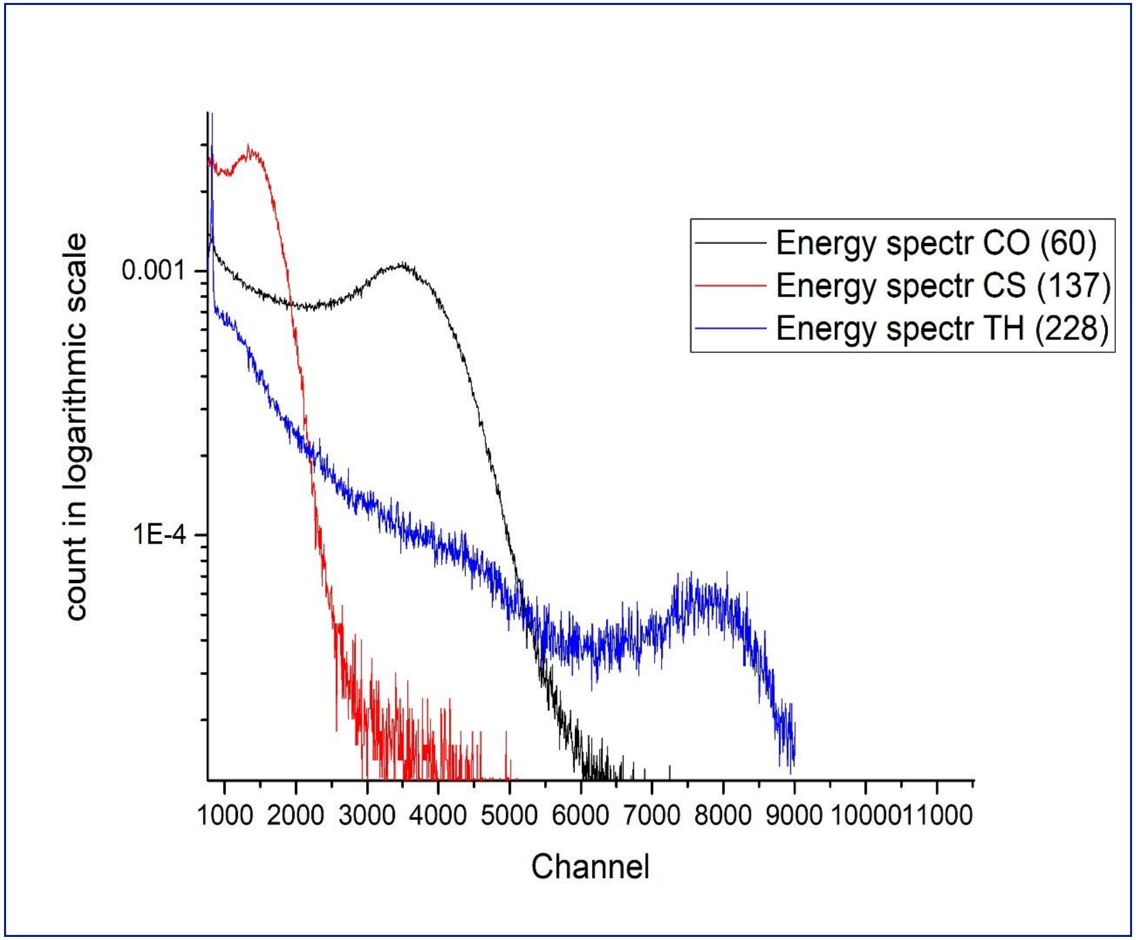
The **DEMON** neutron detectors use a ***NE213*** organic scintillator, which is coupled to a XP4512B photomultiplier tube. The effective diameter of the scintillator cell is 16 cm and the length is 20 cm. The effective diameter of the photo-cathode of the photomultiplier tube is 11 cm. A thick layer of aluminum tube (2 cm thick) is added between the cell and the outside housing of a 20 cm diameter steel tube in order to minimize the cross talk between detectors.

Problems for DEMON separating method



We see that double integration method cannot successfully define gamma and neutron over the whole energy interval. However, we can use it as pre-sort for ML.

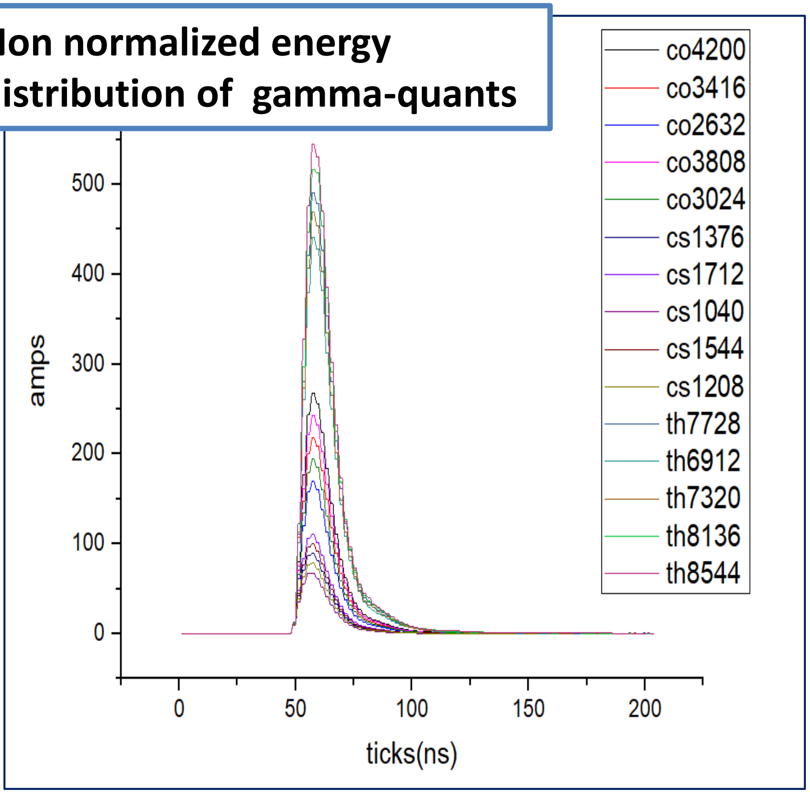
Energy distribution for gamma sources

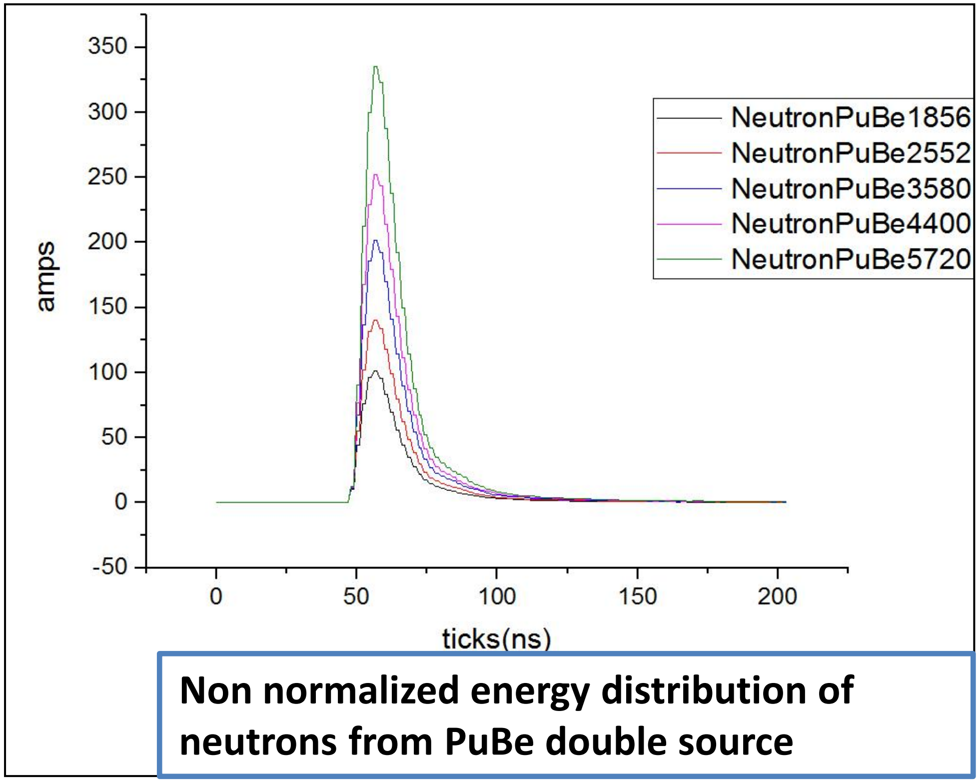


From this graph, we can see in which intervals of energies (channels) the maximum of the detected signals for three sources lay. On the basis of the presented graph, it is possible to reveal the interval boundaries for the supposed "**reliable**" gamma-quants.

Gamma and Neutron sources

Non normalized energy distribution of gamma-quants



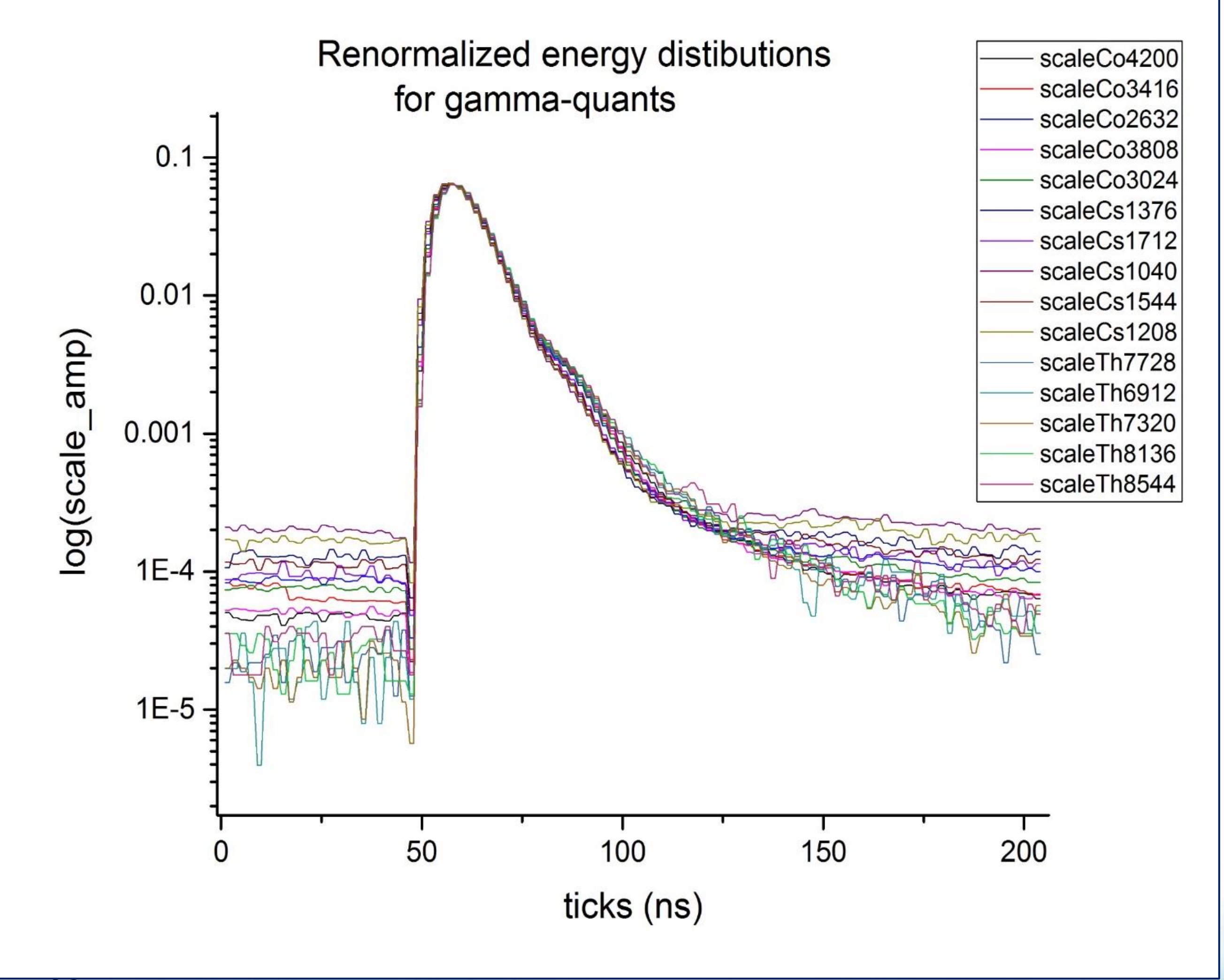


Non normalized energy distribution of neutrons from PuBe double source

***Co* (60)** – 2632, 3024, 3416, 3808, 4200,
***Cs* (137)** – 1040, 1208, 1376, 1544, 1712,
***Th* (228)** – 6912, 7320, 7728, 8136, 8544;

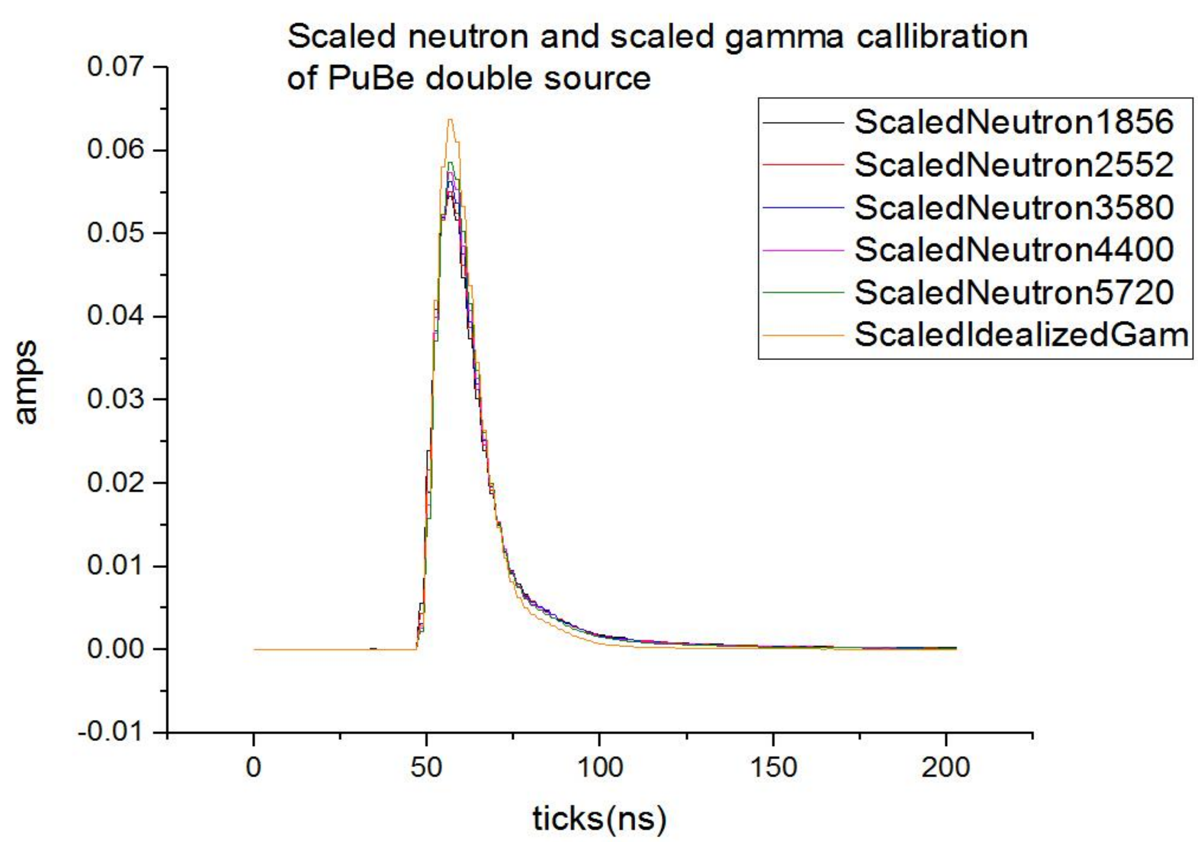
PuBe – 1856, 2552, 3580, 5720

Renormalized energy distribution for gamma-quants



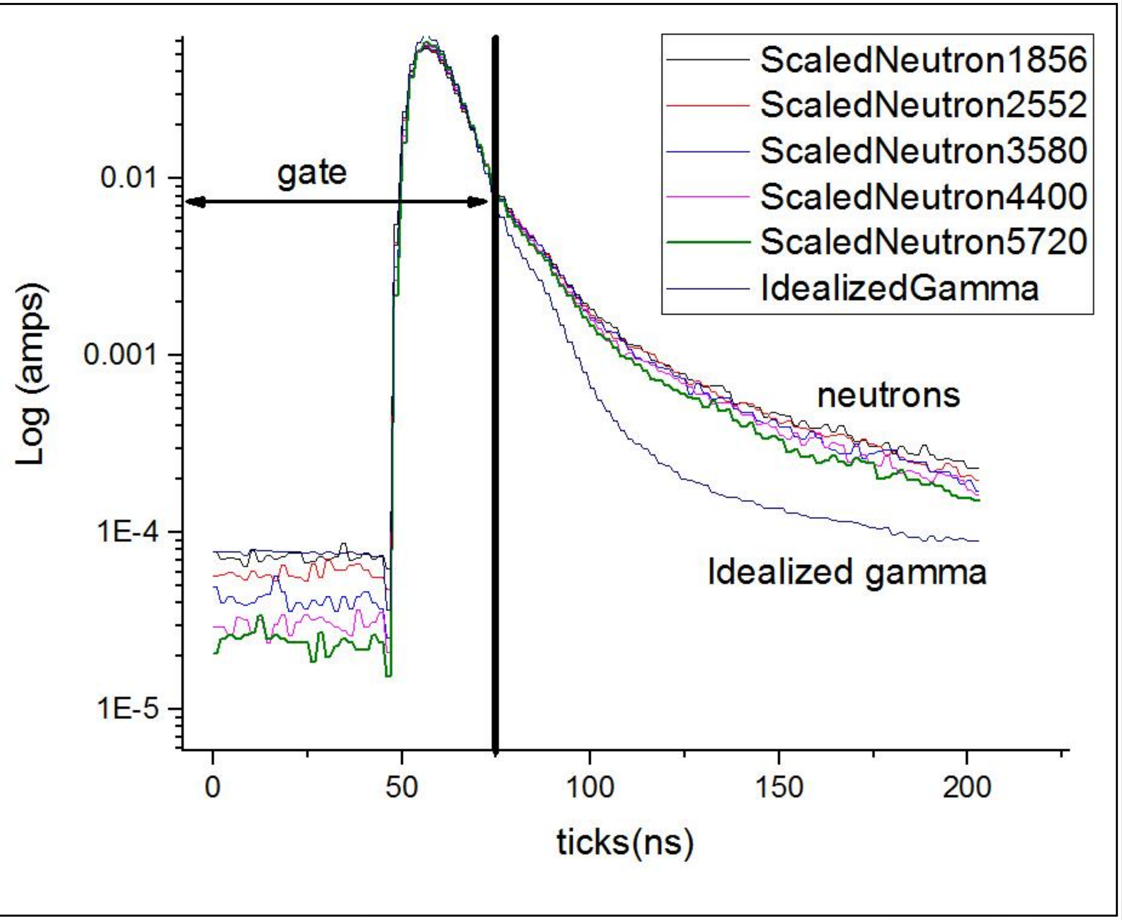
The main conclusion is that the shots (experimental detection) of all gamma-quants have the same shape/profiles. This fact allows us to introduce a concept of an idealized gamma-quant -- by renormalizing all the gamma-quants by their energy. We rescale the obtained profiles by the corresponding energy (number of the channel). The above figure confirms our observations.

Renormalized energy distribution for neutrons and idealized-gamma



The graph shows that the scaled neutron signals differ from the ideal gamma. Neutrons have a smaller maximum, but they have a predominance in the tail.

Previous analysis (with DEMON separating method)



Renormalized neutron and gamma-shots in logarithmic scale. Gate marks the region used in previous analysis to separate gamma from neutron-signals according to the square under the tail rule:

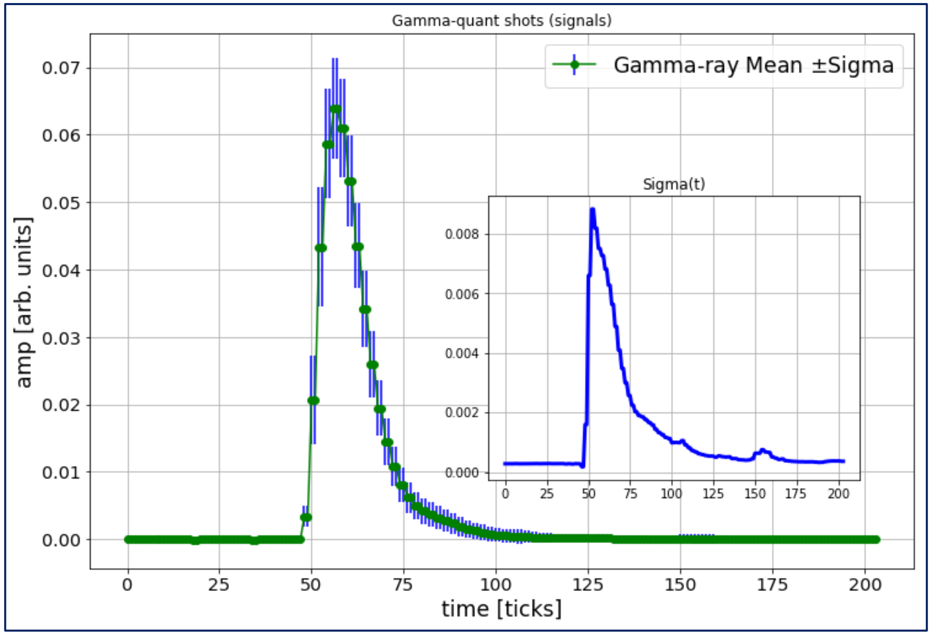
Gamma if

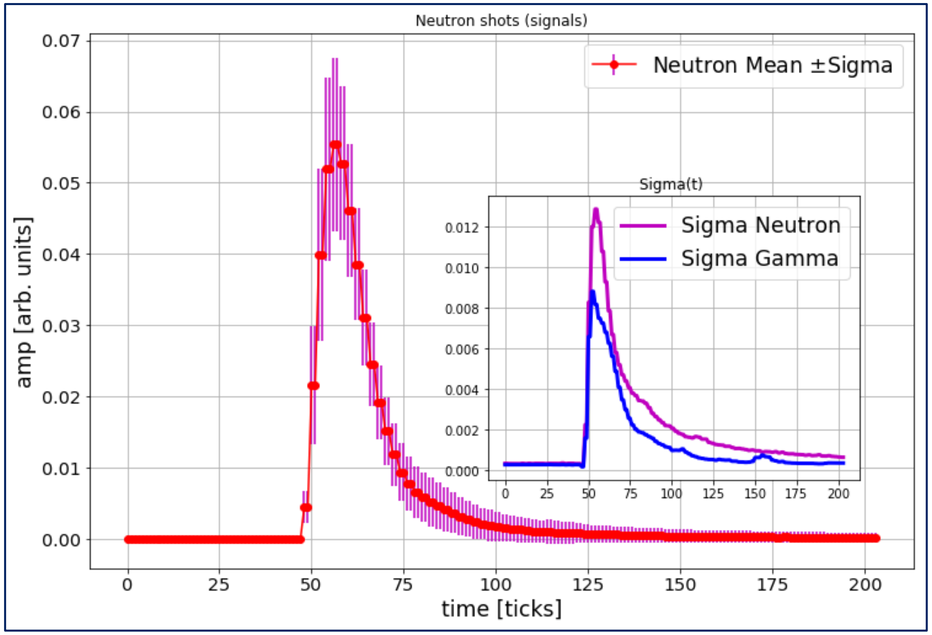
$$\frac{S_{total} - S_{gate}}{S_{total}} < 0.1$$

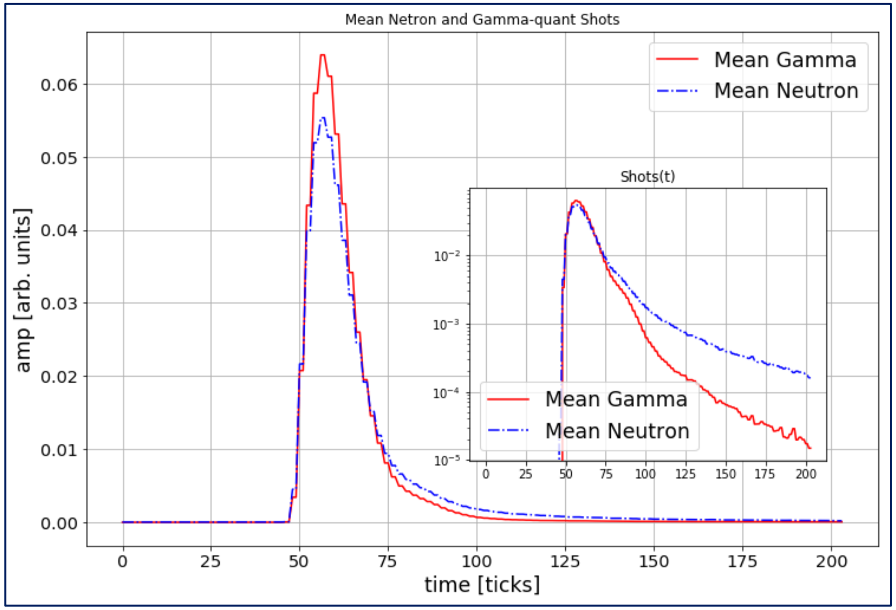
and **Neutron** if

$$\frac{S_{total} - S_{gate}}{S_{total}} > 0.2$$

Normalized profile for gamma and neutron







SOFTWARE for ML

ANACONDA

Anaconda is a free open source distribution of the **Python** and R programming languages for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment

jupyter

The **Jupyter Notebook** is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.

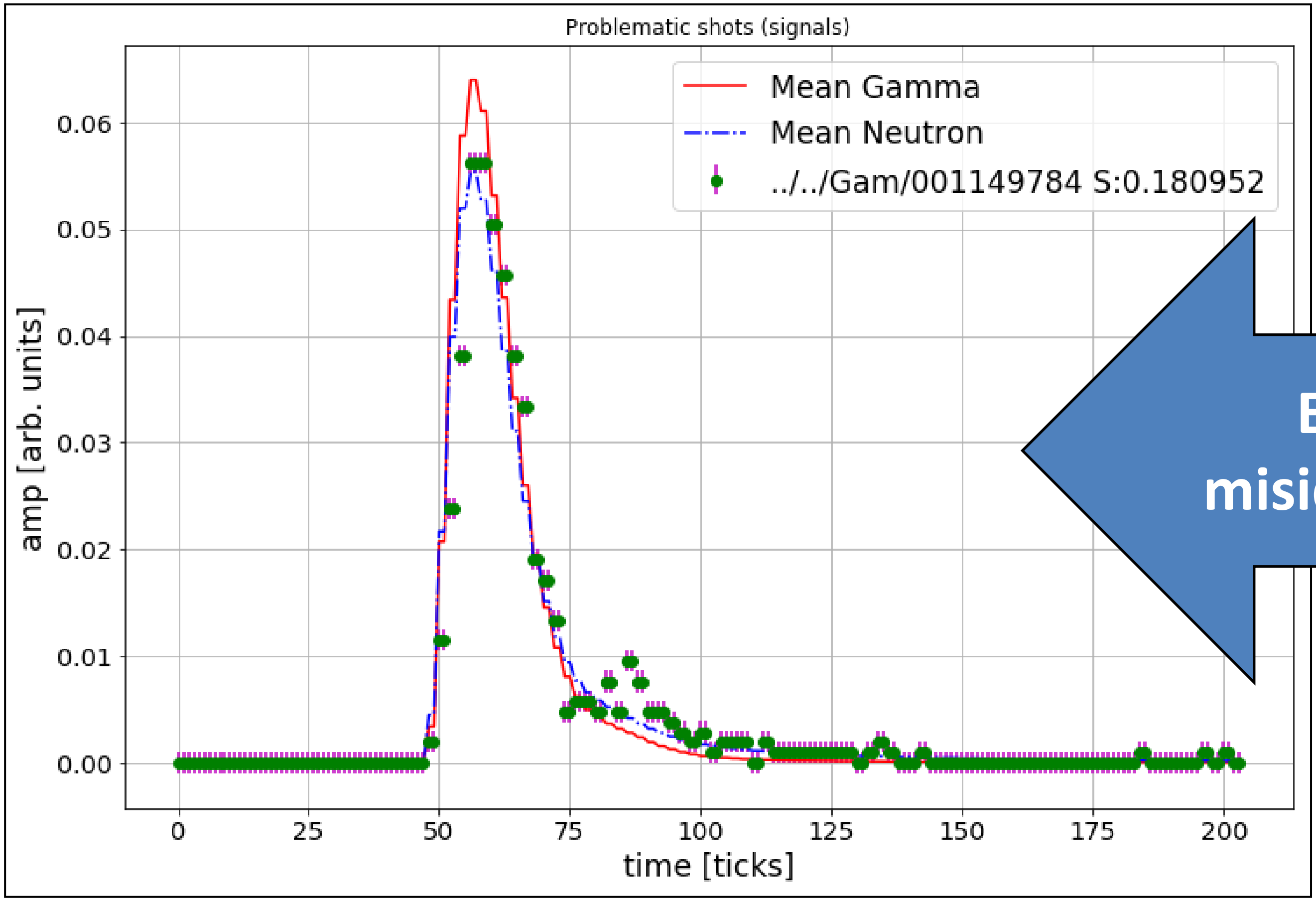
scikits learn machine learning in Python

Scikit Learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries **NumPy and SciPy**

Result of ML classifiers for pre-sorted dataset

Classifier	Seconds	Score
Linear-SVM	44.58	0.49341
RBF-SVM	29.56	0.79002
Naive-Bayes	0.13	0.78691
Decision-Tree	0.32	0.97336
Random-Forest	46.31	0.98991
AdaBoost (50 parameters)	2.76	0.94466
AdaBoost1 (200 parameters)	10.96	0.99922

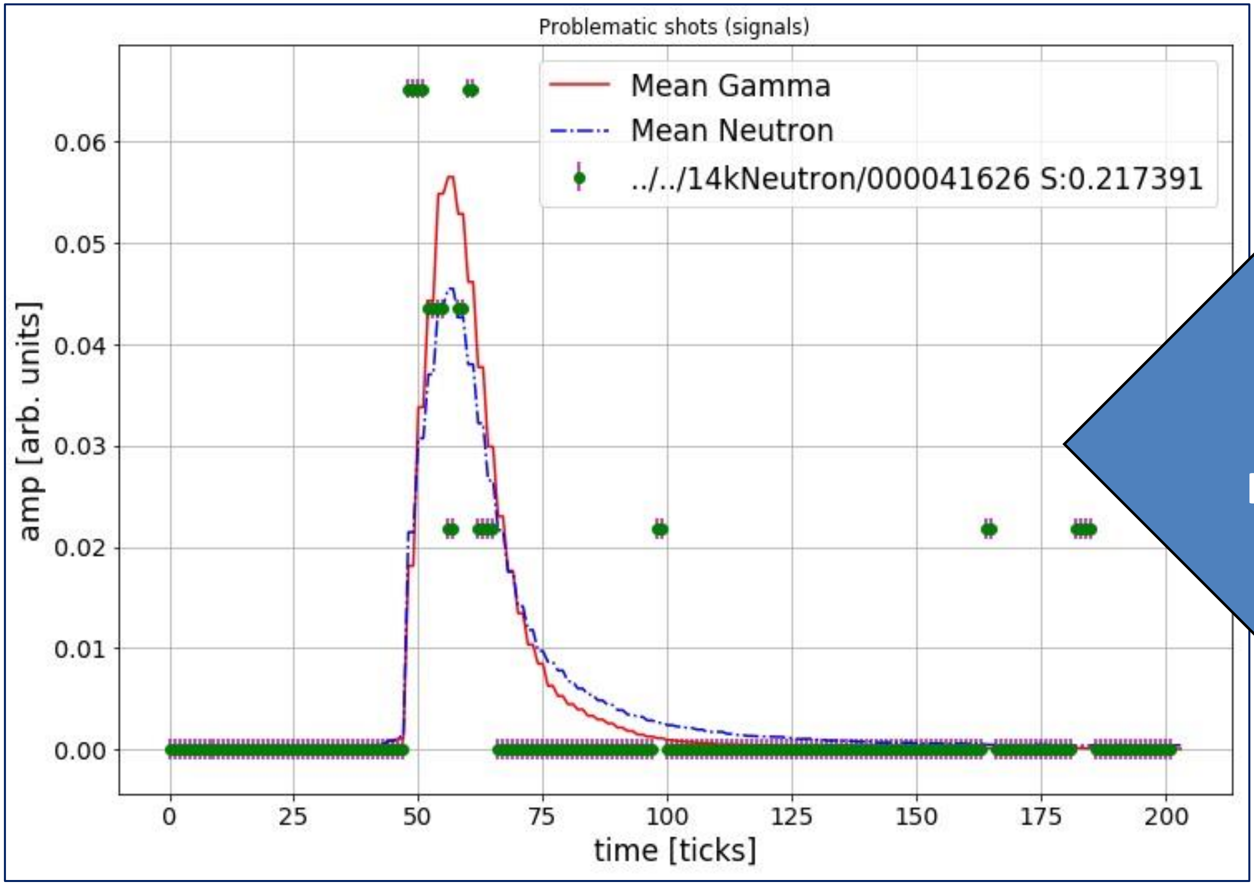
Misidentified signals (with AdaBoost1 classifier)



Example 1 of misidentified signal

We need «purification» of the data to exclude misinterpreted data from consideration by means of AdaBoost1 classifier.

Misidentified signals (with AdaBoost1 classifier)



Example 2 of misidentified signal

We need «purification» of the data to exclude misinterpreted data from consideration by means of AdaBoost1 classifier.

Results after «purification»

Total excluded after “purification”:

Gamma = **80**

Neutron = **87**

Train length = **7836** signals.

~ 2.1% of signals was excluded from the dataset.

Results after «purification»

Classifier	Seconds before	Seconds after	Score before	Score after
Linear-SVM	44.58	44.48	0.49341	0.49384
RBF-SVM	29.56	28.52	0.79002	0.79456
Naive-Bayes	0.13	0.12	0.78691	0.79378
Decision-Tree	0.32	0.38	0.97336	0.97202
Random-Forest	46.31	46.29	0.98991	0.99378
AdaBoost (50 parameters)	2.76	2.70	0.94466	0.94845
AdaBoost1 (200 parameters)	**10.96**	**10.72**	**0.99922**	**1.00000**

After «purification» Adaboostm1 have maximum score. And in our opinion is the best machine learning algorithm for separation and classification

Conclusion

- We apply several machine-learning (ML) algorithms for identification and separation of the neutron and gamma-ray signals coming from the DEMON detector.
- The ML-predictions have been contrasted with the results obtained within a standard method based on an integral-area scheme. In the situations where the standard method fails a properly trained ML-algorithm provides more adequate predictions and, therefore, performs much better.

Thank you for attention!