This article has been accepted for publication in IEEE Sensors Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2025.3544656

Advanced Spectroscopy Time-Domain Signal Simulator for the Development of Machine and Deep Learning Algorithms

Dima Bykhovsky[1],Zikang Chen[2] ,Yiwei Huang[2] ,Xiaoying Zheng[2] and Tom Trigano[1]

[1] *Department of Electrical and Electronics Engineering, Shamoon College of Engineering, Israel*
[2] *Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China*

Abstract—Machine learning methods, particularly deep learning, have become essential for advanced signal processing. These methods often depend on annotated datasets, which can be limited or even unavailable in many cases. One area significantly affected is nuclear spectroscopy, where the lack of annotated datasets is due to the challenges of manually labeling signals recorded in the time domain. To address this issue, it is necessary to use simulators to generate annotated signals, ensuring that the generated time signals are as realistic as possible. This paper introduces a novel simulator designed to generate time-domain signals for gamma spectroscopy. Unlike traditional energy-spectrum simulators, our approach simulates raw sensor output for training advanced deep learning (DL) models. The simulator is analytically trackable, highly customizable, and lightweight, enabling researchers to tackle challenges such as pile-up events and noise suppression. Case studies demonstrate its practical application in high-activity measurement scenarios.

Index Terms—gamma spectroscopy, simulation

I. INTRODUCTION

Machine learning (ML) and deep learning (DL) techniques are significantly used in gamma spectroscopy research. For example, DL has demonstrated its ability to identify individual radioisotopes in mixtures based on their energy spectra [1], [2]. Although most DL applications focus on spectral data, these findings indicate DL's potential to effectively process raw spectroscopic signals derived from recorded time-domain data [3].

The application of ML and DL to time-domain data remains a significant challenge, particularly for supervised learning methods. One of the main obstacles in the field is the scarcity of time-domain per-event labeled data, which is essential for algorithm development and performance evaluation. Moreover, for certain spectroscopic scenarios, such as high activity measurements, labeling is often infeasible.

In numerous applications where real annotated data is scarce (e.g. seismology or nuclear science), simulators play an important role in resolving the data bottleneck. In this specific case, synthetic time series are simulated based on realistic physical models, which allows the automatic annotation of the resulting generated data [4]. In the special case of nuclear science, most of the existing tools in gamma spectroscopy produce energy spectra rather than raw time signals. Consequently, they lack essential time-domain per-event details needed to address challenges such as pile-up rejection or correction.

This paper describes a time-domain spectroscopy simulator [5] that aims to simulate a gamma detector output. The key features of the proposed simulator are:

1) Generating realistic time-domain pulses using Poisson process modeling and tailored pulse-shape functions (double-exponential or gamma);
2) Offering adjustable event rates, user-defined spectra and other parameters;
3) Providing per-event details such as the precise pulse timing and energy;
4) Incorporating a flexible, open-source structure (in Python) that allows further collaboration and customization.

In the rest of the paper, have compiled and organized all the essential theory related to the topic, providing a solid foundation for both new learners and experienced researchers.

II. THEORY

A. Signal Definition

In this section, we outline the fundamental theory and various considerations that underpin simulations that can be used to model the output signal of a detector. For continuous time, the theoretical detector signal can be expressed by

$$y(t) = \sum_{m=0}^{\infty} E_m h_m(t - T_m) + \varepsilon(t), \tag{1}$$

where the sequence $\{T_m,\ m \geq 0\}$ represents the arrival times of particles and is a sample path of a homogeneous Poisson process with normalized intensity λ, the shape of the electrical pulse associated with the m-th impinging particle is represented by $h_m(\cdot)$, E_m represents its energy, and $\varepsilon(t)$ indicates the additive noise present.

Our approach enables researchers to control the signal in Eq. (1) by adjusting pile-up probabilities, noise levels, and other parameters, ensuring flexibility for various experimental setups as detailed below:

- The sampling of the continuous-time signal is discussed in Sec. II-B.
- The generation of the energy values, E_m, is described in Sec. II-C.
- Poisson process characterization that is related to times T_m and number of events $N(t)$ is provided in Sec. II-D.
- The generation of the signal shapes, $h_m(\cdot)$, is described in Sec. II-E.
- Noise $\epsilon[n]$ characteristics are detailed in Sec II-F.

B. Sampling Time

The detector output, $y(t)$, is a continuous-time signal. In practice, it is typically converted into a discrete-time signal. This conversion utilizes a sampling time, Δt, and a corresponding sampling frequency,

This article has been accepted for publication in IEEE Sensors Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2025.3544656

$F_s = 1/\Delta t$. The sampled signal, comprising L samples, is expressed as follows:

$$\begin{aligned} y[n] &= y(nT), \qquad n = 0, \ldots, L-1 \\ &= \sum_{m=0}^{M-1} E_m h_m(n\Delta t - T_m) + \epsilon[n], \end{aligned} \tag{2}$$

where M represents the number of events occurring within the observed time span. Further details on the value of M are discussed in Sec. II-D. In some scenarios, the discrete-time signal $y[n]$ may be defined directly, bypassing the continuous-time representation of $y(t)$. In such scenarios, a sampling interval of $\Delta t = 1$ sec/sample may be adopted.

C. Energies

A gamma-energy spectrogram displays the number of counts (or detections) at various energy levels. Under a probabilistic interpretation, the probability of each energy level (typically measured in keV) is directly proportional to its observed count.

Each energy level, denoted as E_i, is associated with a count, c_i. The pairs of values E_i and c_i can be obtained either through experimental measurements, such as those provided by the OpenGammaProject [6], or can be user-defined, .e.g obtained from simulators such as in [7]. Once E_i, c_i values have been acquired, the next step involves converting the counts from the gamma-energy spectrogram into a probability distribution with probabilities

$$p_i = \frac{c_i}{\sum_j c_j}, \tag{3}$$

where the denominator represents the total count across all energy levels, ensuring that the probabilities sum to one. Once this probability distribution has been established, random sampling of M energy values in accordance with these probabilities is straightforward.

D. Arrival Times

The arrival of gamma particles can be accurately modeled by a homogeneous Poisson process characterized by a rate of λ occurrences per unit time. In simulations of arrival times, the inter-arrival times follow an exponential distribution with a mean of $1/\lambda$. In the simulation context, the number of events is linked to the signal duration:

$$M = N(L\Delta t), \tag{4}$$

where $N(t)$ is the number of events that have happened up to and including time t.

It is important to note that the average number of events, as described by the estimator

$$\hat{\lambda} = \frac{N(t)}{t}, \quad \lim_{t\to\infty} \hat{\lambda} \to \lambda, \tag{5}$$

can vary across different signal realizations. For practical applications involving signals of finite duration, $\hat{\lambda}$ is an estimator of the true activity, whose variability is particularly critical when analyzing short-duration signals.

Pile-up events, which occur when more than one pulse appears in a single trace, are analyzed using the duty cycle (DC). The DC, calculated as the product of the arrival count rate and the pulse shape duration,

$$DC = \lambda T_{shape}, \tag{6}$$

represents the average number of overlapping pulse shapes during the time T. The probability of observing n events in time t is given by

$$\Pr(N(t) = n) = \frac{(\lambda t)^n}{n!} \exp\{-\lambda t\}, \tag{7}$$

and thus, the probability of pile-up is expressed as:

$$\begin{aligned} p_{pile-up} &= \Pr\big(N(T_{shape}) > 0\big) \\ &= 1 - \Pr\big(N(T_{shape}) = 0\big) \\ &= 1 - \exp\{-DC\}. \end{aligned} \tag{8}$$

DC values below 0.1 indicate a low probability of pile-up. Scenarios involving higher DC values are further explored in studies focusing on pile-up phenomena [8], [9].

E. Pulse Shapes

Analytical shape modeling is the predominant method employed for pulse shape analysis, striking a balance between the simplicity of mathematical and computational expressions, and considerations of physical and electronic circuit dynamics. The two most commonly applied models for this purpose are the double exponential model and the gamma model, although alternative models may also be utilized [8], [10].

1) Double-Exponential Model: The double exponential model is described as follows [11]:

$$h(t) = a\left(\exp\left(-\frac{t}{\tau_2}\right) - \exp\left(-\frac{t}{\tau_1}\right)\right), \tag{9}$$

where a is the normalization factor, and τ_1 and τ_2 are characteristic decay times with $\tau_1 \ll \tau_2$.

An important characteristic of this model is the rise time, t_r, which complements the fall time $t_f = T_{shape} - t_p$, where T_{shape} is the assumed finite duration of the pulse shape. For example, the rise time for the double-exponential model is given by

$$t_p = \frac{\tau_1 \tau_2}{\tau_2 - \tau_1} \ln \frac{\tau_2}{\tau_1}. \tag{10}$$

2) Gamma-Shapes: Another commonly used model is the gamma function model:

$$h(t) = at^{\alpha} e^{-\beta t}, \tag{11}$$

where a is the normalizing factor, α represents the shape parameter, and β is a scale parameter. The rise time for the gamma model is defined by

$$t_p = \frac{\alpha}{\beta}. \tag{12}$$

Both of the models are considered to effectively capture the dynamics of pulse shapes.

F. Noise

All practical electrical measurements are influenced by additive noise. This noise can originate from multiple sources, such as the thermal motion of electrons within electrical conductors (thermal noise), electromagnetic interference from the environment, quantization noise, and more. Typically, when aggregated, these noise contributions are modeled as white Gaussian noise (WGN) according to the Central Limit Theorem.

Users can parameterize the noise level via variance σ_n^2 or equivalent signal-to-noise ratio (SNR), which evaluates the relationship between the signal's power and the noise's power. It is important to note that the power of the noise is directly proportional to the noise variance.

This article has been accepted for publication in IEEE Sensors Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2025.3544656

III. APPLICABILITY CONSIDERATION

A. Evaluation Modes for Signal Boundaries

In the process of evaluating algorithms using simulated signals, two distinct operational modes are considered, each relating to the placement of pulse shapes within the signal boundaries:

1) Boundary-Constrained Pulse Placement: In this mode, the start and end points of any pulse shape must lie within the confines of the signal, ensuring that no shape exceeds the signal boundaries. This constraint often simplifies evaluations of certain spectroscopic algorithms. However, it tends to result in an underestimation of the expected rate $\hat{\lambda}$ as defined in Eq. (5).

2) Boundary-Independent Pulse Placement: This mode allows pulse shape locations to be independent of the signal's start and end points, providing a more realistic representation of signal behavior. While this setting enhances accuracy, it may complicate the use of the signal for debugging and prototyping algorithms due to potential edge effects and partial pulse shapes.

Each mode offers distinct advantages and challenges, necessitating careful consideration based on the specific requirements of the algorithm testing and development process.

B. Combination of Multiple Sources

When the composite spectra is of interest, the spectral lines from each source overlap and are combined into the same gamma-energy spectral histogram. The conversion of the histogram to the probability distribution has been discussed in Sec. II-C.

C. Implementation Workflow

The implementation of our spectroscopy activity simulator is structured in accordance with the theory outlined in Sec. II:

1) Sampling: All time-related parameters are normalized according to the sampling frequency.

2) Data Acquisition: The energy spectrum data is downloaded and converted to a probability distribution.

3) Time-Stamp Generation: Time stamps T_n for the arrival of each pulse are generated.

4) Pulse-Shape Dictionary Generation: A dictionary of pulse shapes is generated. This dictionary serves to simulate various pulse responses based on predefined or dynamically generated parameters.

5) Pulse Assignment: A corresponding pulse shape and energy value are assigned for each generated time stamp. This step integrates the temporal and spectral data, forming the basis of the simulated signal.

6) Signal Synthesis: The simulated signal is synthesized according to the model defined in Eq. (1). This synthesis involves the superposition of all assigned pulses within the signal framework, adjusted for noise and other factors.

IV. CASE STUDY EXAMPLE

In a case study, we have used a raw HPGe measurement signal of a source including a mixture of Cesium ^{137}Cs and Americium ^{241}Am to provide gamma-shape examples. The signal was sampled at 10 MHz frequency, using the ADONIS system detailed in [12]. The one-second excerpt of the recording was cleaned with pile-up rejection and the removal of dead-time events. The resulting signal has about 70×10^3 shapes [13]. An example of the resulting signal is presented in Fig. 1. The corresponding energy histogram of this signal is presented in Fig. 2. The shapes in the signal were fitted with gamma shapes, and the resulting joint distribution of the α and β parameters is presented in Fig. 3

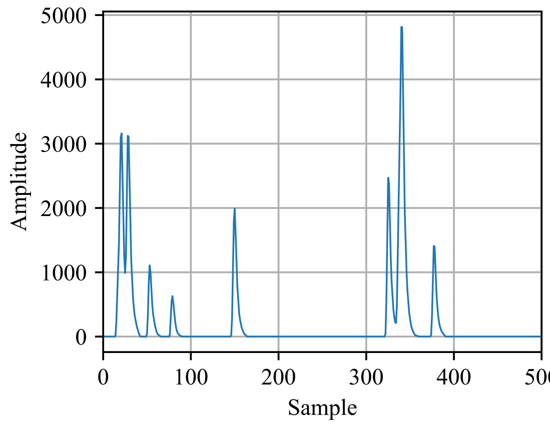

Fig. 1. The experimental measurement example (short-time excerpt).

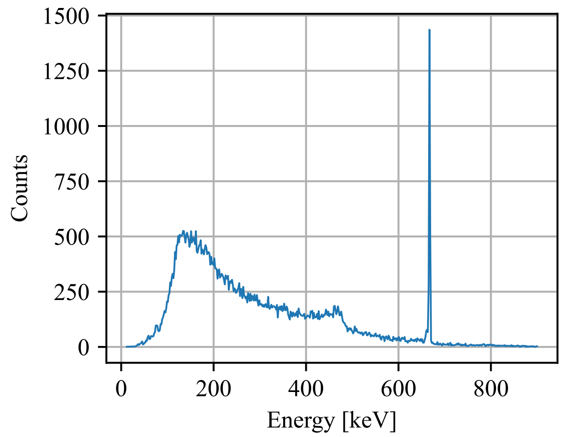

Fig. 2. The energy histogram of the signal in Fig. 1. The maximal energy peak is around 667 keV.

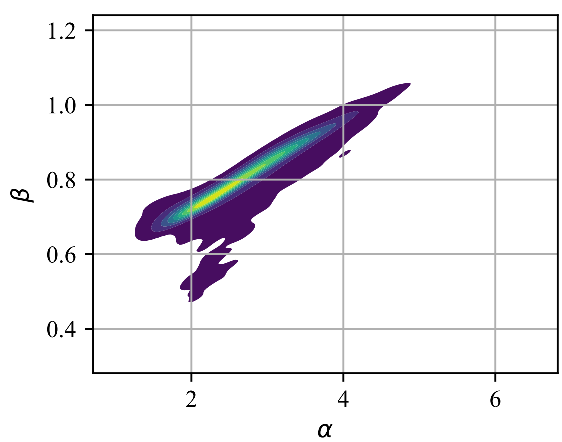

Fig. 3. Joint distribution of α and β parameters of the signal in Fig. 1.

This article has been accepted for publication in IEEE Sensors Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2025.3544656

Finally, the signal with the same energy histogram and the same shape parameters was recreated by simulation with two distinct activity values that differ from the original one, demonstrating flexibility in replicating real-world scenarios. Figure 4 compares the simulated signals, showing accurate replication of key features.

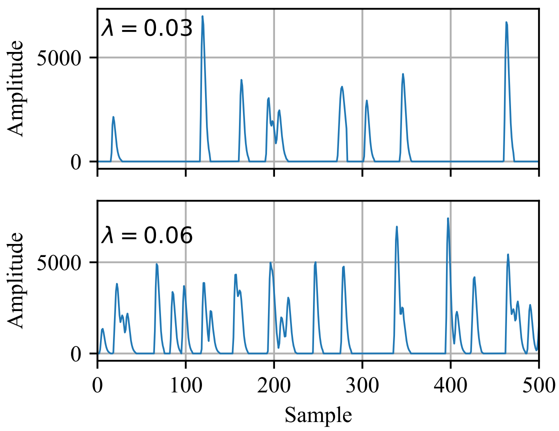

Fig. 4. Generated time-domain signals for the activities $\lambda = 0.03$ (upper) and $\lambda = 0.06$ (lower).

V. FUTURE WORK

Our open-source Python simulator provides a robust foundation for generating synthetic time-domain spectroscopic signals, but there is significant potential for future improvements. For example, it may be beneficial to implement custom non-parametric shapes derived directly from experimental data and more complex and detector-specific noise models, such as non-linear distortions and detector-specific artifacts.

REFERENCES

[1] M. Kamuda, J. Stinnett, and C. J. Sullivan, "Automated isotope identification algorithm using artificial neural networks," *IEEE Transactions on Nuclear Science*, vol. 64, no. 7, pp. 1858–1864, 2017.

[2] B. Ayhan and C. Kwan, "New results on radioactive mixture identification and relative count contribution estimation," *Sensors*, vol. 21, no. 12, 2021.

[3] T. Trigano and D. Bykhovsky, "Deep learning based method for activity estimation from short-duration gamma spectroscopy recordings," *IEEE Transactions on Instrumentation and Measurement*, 2024. accepted for publication.

[4] A. Othman, N. Iqbal, S. M. Hanafy, and U. B. Waheed, "Automated event detection and denoising method for passive seismic data using residual deep convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[5] "Gamma simulator." https://github.com/bykhov/gamma-simulator.

[6] "Gamma-spectrum-database." https://github.com/OpenGammaProject/Gamma-Spectrum-Database.

[7] F. M. de Oliveira, G. Daniel, and O. Limousin, "Artificial gamma ray spectra simulation using generative adversarial networks (GANs) and supervised generative networks (SGNs)," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1047, p. 167795, 2023.

[8] M. Lee, D. Lee, E. Ko, K. Park, J. Kim, K. Ko, M. Sharma, and G. Cho, "Pulse pileup correction method for gamma-ray spectroscopy in high radiation fields," *Nuclear Engineering and Technology*, vol. 52, no. 5, pp. 1029–1035, 2020.

[9] M.-R. Mohammadian-Behbahani and S. Saramad, "A comparison study of the pile-up correction algorithms," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 951, p. 163013, 2020.

[10] X. Wang, Q. Xie, Y. Chen, M. Niu, and P. Xiao, "Advantages of digitally sampling scintillation pulses in pileup processing in pet," *IEEE Transactions on Nuclear Science*, vol. 59, no. 3, pp. 498–506, 2012.

[11] G. F. Knoll, *Radiation detection and measurement*. John Wiley & Sons, 4th ed., 2010.

[12] E. Barat, T. Dautremer, L. Laribiere, J. Lefevre, T. Montagu, and J.-C. Trama, "ADONIS : a new system for high count rate HPGe Gamma spectrometry," in *2006 IEEE Nuclear Science Symposium Conference Record*, 2006.

[13] "Supplementary code for signal analysis." https://github.com/bykhov/gamma-simulator/tree/main/signal_analysis_example.