Measurement Science
and Technology

PAPER • **OPEN ACCESS**

Novel machine learning and differentiable programming techniques applied to the VIP-2 underground experiment

To cite this article: Fabrizio Napolitano *et al* 2024 *Meas. Sci. Technol.* **35** 025501

OPEN ACCESS
IOP Publishing
Measurement Science and Technology
Meas. Sci. Technol. **35** (2024) 025501 (9pp)
https://doi.org/10.1088/1361-6501/ad080a

Novel machine learning and differentiable programming techniques applied to the VIP-2 underground experiment

Fabrizio Napolitano[1,*], Massimiliano Bazzi[1], Mario Bragadireanu[1,2], Michael Cargnelli[3], Alberto Clozza[1], Luca De Paolis[1], Raffaele Del Grande[1,4], Carlo Fiorini[5], Carlo Guaraldo[1], Mihail Iliescu[1], Matthias Laubenstein[6], Simone Manti[1], Johann Marton[3], Marco Miliucci[1,8], Kristian Piscicchia[1,7], Alessio Porcelli[1,7], Alessandro Scordo[1], Francesco Sgaramella[1], Diana Laura Sirghi[1,2,7], Florin Sirghi[1,2], Oton Vazquez Doce[1], Johann Zmeskal[1,3] and Catalina Curceanu[1,2]

[1] INFN, Laboratori Nazionali di Frascati, Via E. Fermi 54, Frascati I-00044, RM, Italy
[2] IFIN-HH, Institutul National pentru Fizica si Inginerie Nucleara Horia Hulubei, Str. Atomistilor No. 407, Bucharest-Magurele, Romania
[3] Stefan-Meyer-Institute for Subatomic Physics, Austrian Academy of Science, Kegelgasse 27, 1030 Vienna, Austria
[4] Physik Department E62, Technische Universität München, James-Franck-Straße 1, 85748 Garching, Germany
[5] Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria and INFN Sezione di Milano, 20133 Milano, Italy
[6] INFN, Laboratori Nazionali del Gran Sasso, Via G. Acitelli 22, 67100 Assergi, AQ, Italy
[7] Centro Ricerche Enrico Fermi–Museo Storico della Fisica e Centro Studi e Ricerche 'Enrico Fermi', Via Panisperna 89a, 00184 Roma, RM, Italy

E-mail: fabrizio.napolitano@lnf.infn.it

Received 26 May 2023, revised 24 July 2023
Accepted for publication 30 October 2023
Published 6 November 2023

Abstract

In this work, we present novel machine learning and differentiable programming enhanced calibration techniques used to improve the energy resolution of the Silicon Drift Detectors (SDDs) of the VIP-2 underground experiment at the Gran Sasso National Laboratory. We achieve for the first time a full width at half maximum in VIP-2 below 180 eV at 8 keV, improving around 10 eV on the previous state-of-the-art. SDDs energy resolution is a key parameter in the VIP-2 experiment, which is dedicated to searches for physics beyond the standard quantum theory, targeting Pauli exclusion principle violating atomic transitions. Additionally, we show that this method can correct for potential miscalibrations, requiring less fine-tuning with respect to standard methods.

[8] Current position: Italian Space Agency, Via del Politecnico, s.n.c, 00133 Roma, RM, Italy.
[*] Author to whom any correspondence should be addressed.

Keywords: VIP-2, SDD, silicon drift detector, differentiable programming

1. Introduction

The Pauli Exclusion Principle (PEP) is a key ingredient of the quantum theory, and its violation, albeit tiny, could be motivated by physics beyond the Standard Model, in scenarios such as the violation of Lorentz invariance, existence of extra dimensions, and quantum gravity scenarios, as discussed in recent studies [1–3]. In this context, the VIP-2 experiment [4] at the Gran Sasso National Laboratory (LNGS) searches for signals of PEP violation in the form of anomalous x-ray transitions in copper atoms, using several arrays of state-of-the-art Silicon Drift Detectors (SDDs). In the VIP-2 experiment, the SDD calibration is performed *in-situ* using fluorescence K_α and K_β x-rays lines from manganese and titanium, activated by a Fe-55 radioactive source, and a copper K_α line from the target material, activated by residual environmental radiation, around two orders of magnitude less intense. This approach avoids the use of a dedicated x-ray source to be switched on for calibration runs, and assures that the data can be calibrated at any given point in time.

The PEP-violating K_α transition in copper is expected to appear just a few hundred of eV below the standard copper K_α line, and the upper limit on the number of PEP-violating events in VIP-2 is determined through statistical analysis of the data. Several key factors influence this limit:

- Data taking time: increasing the data-taking time improves the available statistics and, consequently, enhances the limit. However, the improvement depends roughly on the square root of the integrated time.
- Stability and precision of the calibration: an optimal limit setting requires stable and precise calibration. If the calibration is unstable, the statistical model of the data will degrade the limit due to fluctuations in yields within the Region of Interest (ROI). The precision of the calibration directly impacts the limit setting, as a large systematic uncertainty will degrade the limit.
- Detector resolution: the resolution of the detector affects the upper limit. With the same number of signal events, a sharper line would be more significant than a broader one, which would be more compatible with background events.

In view of the experimental goal, and to ensure an accurate calibrated energy spectrum with minimal uncertainty in energy scale and resolution, precise control to determine the copper K_α transition is required. However, the difference in yield between the fluorescence lines and the copper line forces a trade-off between energy scale uncertainty and resolution. If the duration of the calibration batch is long enough, a good precision can be achieved on the copper line position. In this case, however, the calibration will be unable to capture small fluctuations on the SDDs response, worsening the energy resolution when all the spectra are added together. Conversely, if the calibration batch is short, the fluctuations could be captured, but the uncertainty on the determination of the copper position will be high, due to the low yield, again worsening the energy resolution. Various peak-finder algorithms have been tested to be used for a fit to the spectral lines, but at low yield, standard methods are sensitive to statistical fluctuations, and require extensive fine-tuning. Moreover, the fit does not produce a sufficient accuracy without well-placed initialization of the parameters. Machine Learning (ML) techniques have found great success in signal processing in the last decade, producing stable results with substantially fewer calibration and tuning requirements with respect to algorithmic methods (as an example in deep underground detectors [5]).

In the field of spectroscopy, ML has been successfully applied in the context of chemical characterization, classification, noise suppression, x-ray absorption spectra prediction and more (for a non-exhaustive list of applications in infrared, Raman and x-ray spectroscopy see e.g. [6–10]), however, the calibration of detectors array in conditions of sparse data is still a challenging task. While some applications have pioneered the use of differentiable ML pipelines for detector optimization [11–13], so far none of them have addressed the issue of detector calibration.

In our experiment, we trained a deep neural network on synthetic data reproducing the spectra's features, using Convolutional Neural Networks (CNNs), to predict the position of the spectroscopic lines. Finally, the output parameters of the neural network are optimized by gradient descent within an automatic differentiation framework, where the loss function is expressed in terms of a likelihood function of the data given the spectrum's probability density function. The application of this method demonstrates gains in the detector's resolution, additionally showing the capacity of recovering miscalibrations.

The energy calibration of the SDDs is a critical experimental task in many collaborations, and its accuracy is often reflected in the precision of the final spectroscopic observables in the form of systematic uncertainty, often the most relevant. Systematic uncertainties related to the line positions are typically assessed from the linearity and stability of the energy response of the detectors. As an example, in precision spectroscopy of exotic systems at colliders (see e.g. [14] for a review), the total measurement uncertainty is largely dominated by the energy response. The method presented in this work has broad applicability to various spectroscopic measurements, offering a substantial reduction in this important source of uncertainty.

We introduce the experimental setup, detector and data taking in section 2; in section 3 we describe the novel ML and differentiable programming approach, and in section 4 the results and discussion of the comparison with the standard approach. Section 5 presents the conclusions of this work.

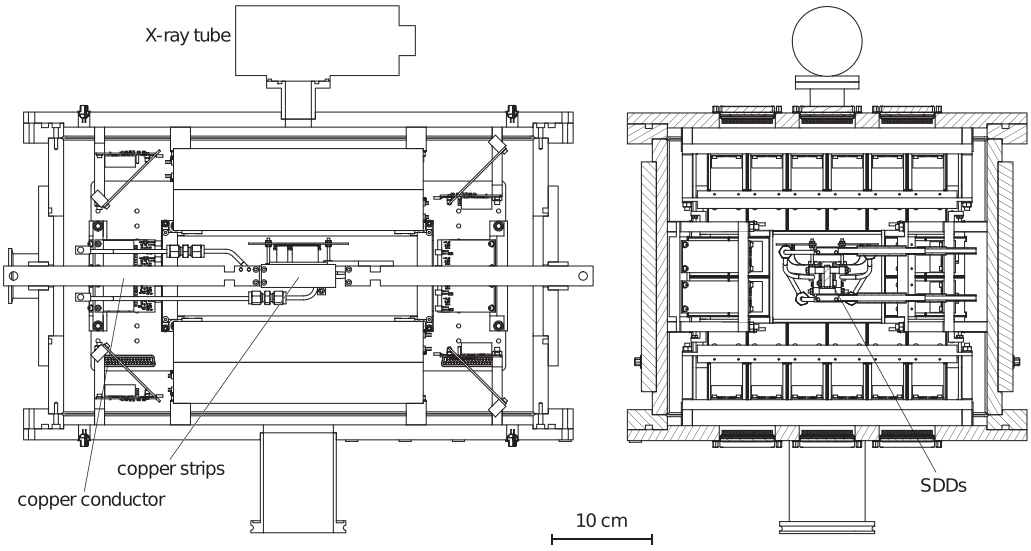

Figure 1. Lateral and transverse view of the VIP-2 apparatus. The four SDDs arrays are placed in front of the target (copper strips), two on one side, two on the opposite side [19]. A current of 180 A is circulated from the copper conductor to the target strips. The x-ray tube is not used during the VIP-2 run, and the calibration is performed *in-situ* employing Ti and Mn lines activated by a Fe-55 source, see text for more details. Reproduced from [19]. CC BY 4.0.

2. Experimental setup

The VIP-2 experiment at the LNGS is proving world-class upper limits on the PEP violation probability in the open system scenarios [15]. Protected by about 3600 m of water-equivalent shielding of the Gran Sasso d'Italia massif, and by lead and copper walls, the experimental setup employs SDDs, housed in a vacuum chamber and placed around a copper target, where a strong direct current is circulated to provide electrons to test PEP. The SDDs [16, 17] are ideal detectors for precision x-ray spectroscopy, having so far demonstrated a resolution of about 190 eV Full Width at Half Maximum (FWHM) at 8 keV [18, 19]. Those used in VIP-2 were developed in a collaboration between Stefan Meyer Institute of the Austrian Academy of Sciences, the Politecnico di Milano, INFN-LNF and Fondazione Bruno Kessler [20]. The thickness of 450 µm, an active area of 0.64 cm^2 and an efficiency of 99% at 8 keV allow for great physics reach and satisfy the scientific requirements.

The four SDD arrays of VIP-2 are placed parallel to the target surface, two on one side and two on the opposite side, each array having 2 × 4 SDD cells, and being operated at a temperature of −90° C. The vacuum chamber, where the SDDs and their front-end electronics are located, is operated at about 10^{-6} mbar. The VIP-2 apparatus is schematically shown in figure 1 with more details provided in [19].

2.1. Data

The data used in this work, aiming to test the novel techniques, corresponds to about a month of acquisition during 2022/2023 data taking campaign of the VIP-2 experiment. During this period, no current was circulated on the target. The spectrum calibrated with the standard approach is shown in figure 2,

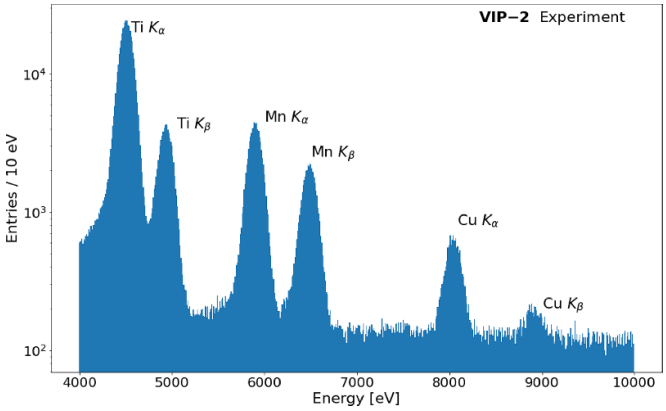

Figure 2. VIP-2 energy spectrum calibrated with the standard approach in the energy range 4000–10 000 eV. The spectroscopic lines are indicated on the plot. This spectrum corresponds to about 1 month of data taking, and it is used as baseline for this study.

where the calibration lines, Ti K_α, Ti K_β, Mn K_α, Mn K_β and the copper K_α and K_β lines are indicated.

The batch used for the optimization study is around 2 days of data taking, while the standard calibration batch previously used for the VIP-2 experiment is around a month.

The self-triggered preamplified SDDs signals are read out by the front end electronics for all the detector channels. The raw data were saved in Analogue-to-Digital Converter (ADC) channels, together with the time information. A preprocessing step was performed, removing overflow events, as well as SDD channels without spectroscopic response.

3. Method

In section 3.1, we employ a ML approach used to predict the centroids of the spectroscopic peaks, less sensible to statistical

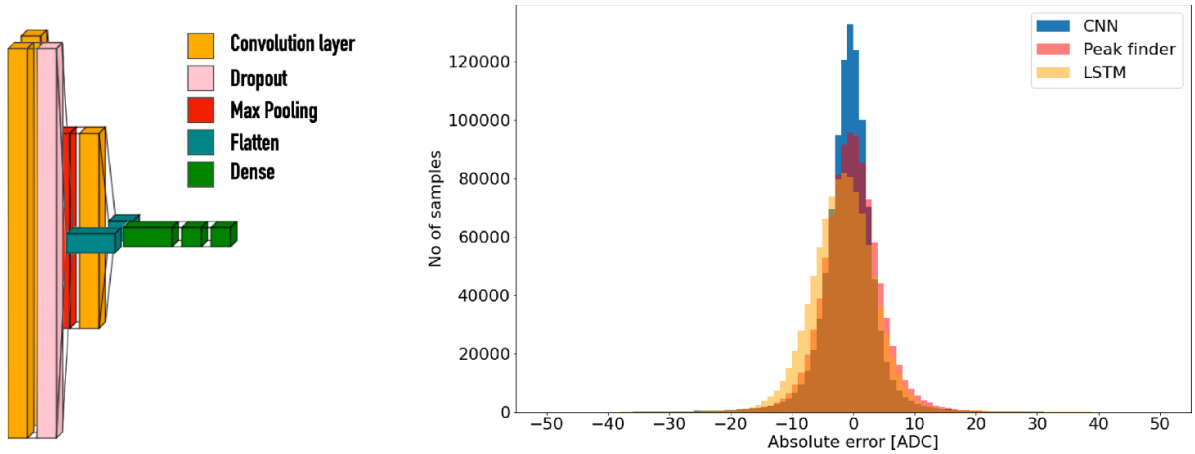

Figure 3. On the left, the schematic of the architecture employed on VIP-2 data. The input is taken from two parallel convolutional networks. The first one has kernels of bigger size, in order to learn features of higher scales, the second one has a smaller size, which instead is applied with maximum granularity. The combined output is then used as input to three layers of a fully connected network, where the output nodes are the normalized position of the calibration peaks. In the diagram, orange represents a convolutional layer, pink a dropout, red a max pooling, teal a flatten layer, and green a fully connected one. On the right, the per-spectrum average of difference between predicted and true centroid from simulated data, for the CNN, peak finder and LSTM. The result shows a good agreement within ten ADCs, and a superior performance of the CNN with respect to the peak finder and LSTM. In case of the peak finder, mis-identified peaks are conservatively not included.

fluctuation than standard peak-finder algorithms (section 3.1). In the following section 3.2 we introduce a novel approach based on differentiable programming to optimize the calibration of the SDDs, reaching below 180 eV FWHM at the copper line.

3.1. Identification of the spectroscopic centroids at low yield

We employ a neural network architecture trained on synthetic data, reproducing the features of the real data. The model architecture is described in figure 3(left), using two convolutional towers to learn features of bigger and smaller kernel size. The model architecture is inspired from the image recognition architectures [21]. The synthetic data is produced with large differences in peak positions, widths and relative amplitudes, in order to stabilize the performance even in case of important fluctuations. This conservative approach assures stability of the method in a wide range of cases. This training data was generated as follows:

- The initial peak position were set to 800, 950, 1100, 1200, 1500, corresponding to Mn, Ti and Cu peaks, representing most typical ADC values found in data. Each generated spectrum includes these peak positions, which are further randomly and independently shifted within twice the width size for the calibration lines and five times the width size for the copper line.
- The peak widths were chosen at 20 ADC, with 20% random variation.
- The relative yields of the lines were independently and randomly varied within 50% and 200% of the most typical values found in data.
- The total yield was randomly varied down to 5% to simulate sparser spectra.
- The continuum was modeled with three different components, two for the calibration lines and one of for the copper region. A shoulder was added to reproduce the transition to the copper region.

The validation data was randomly extracted from 30% of the entire synthetic dataset.

3.1.1. Network architecture. The network is a CNN with two branches. The input is a 1D sequence of length 300, which represent the uncalibrated energy spectrum. The convolutional branch consists of two 1D convolutional layers with Rectified Linear Unit (ReLU) activation. The first layer has 5 filters and a kernel size of 50; the second layer has 5 filters and a kernel size of 10. Max pooling is applied with a pool size of 2. The size of the kernels and pooling were optimized upon consideration of the feature sizes and peaks structure. The dense branch starts with a 1D convolutional layer with 5 filters and a kernel size of 1, followed by a dropout layer. The outputs of both branches are flattened and concatenated, and three dense layers with ReLU activation are applied, with 500 units, 50 units, and 5 units respectively. This network architecture captures local and global patterns in the input data through convolutional operations, pooling, and dense layers. The final layer produces the network's output with the five node representing the position of the five spectroscopic centroids.

3.1.2. Benchmark. We use another ML approach, and a classical peak finder algorithm [22, 23] as benchmarks to evaluate this network. Most of the common ML technique can be adapted to treat 1D sequences. However, their training time, performance metrics, interpretability and robustness can vary depending on the architecture design, the challenge to solve and the available hardware.

Apart from CNN, Recurrent Neural Network (RNN) and transformers are common architecture that can handle sequential data. Long Short-Term Memory (LSTM) [24] is a type of RNN that employs memory cells and gates to control the flow of information, and allowing the network to remember information for longer periods. Although LSTM is commonly used in natural language processing, where the precise order of the input sequence is important, in this work, retaining information about the entire structure of the spectrum, as done by RNNs, is not strictly a requirement. Additionally, LSTM networks are known for much longer training times with respect to CNN (about an order of magnitude [25] due to the more complex architecture layout). However, since these networks have been employed as robust peak detection in health applications (e.g. [26]), we construct a LSTM with two bidirectional layers, each having 64 LSTM units (deeper than the architecture used in [26]), followed by the same fully connected layers of 500, 50 and 5 units as mentioned above. The LSTM network was trained on the same synthetic data used for the training of the CNN.

3.1.3. Performance. The trained model is subsequently fine-tuned on synthetic data, but with much lower statistics, for the CNN and benchmark LSTM. The CNN model, could be trained successfully on commercially common hardware (Intel(R) Core(TM) i5-9600 CPU @ 3.10GHz), while the LSTM required dedicated GPU (NVIDIA Tesla T4). On figure 3 (right), the absolute error of the network prediction with respect to the true values used to generate the data is shown, for the CNN (blue) and LSTM and peak finder benchmarks (orange and red). The standard deviation of the distributions are about 4.6 (CNN), 5.8 (LSTM) and 4.9 (peak finder) ADC channels, or around 19, 25 and 21 eV (the precise value depends on each SDD channel), respectively. The CNN is found to outperform the LSTM approach, and even the peak finder by almost 20% in precision of the peak identification. Additionally, while the CNN and LSTM identified 100% of the peaks, the peak finder did not identify 25% of them. This method demonstrates stable performances even for low count numbers, corresponding to small integrated times, mitigating the need for extensive fine-tuning of the algorithms, and offering high precision of the reconstruction. In figure 4, the ADC spectrum of a single SDD acquired in 2022 by the VIP-2 experiment, together with the corresponding neural network prediction, are shown.

The predicted centroids were then used as seeds for a subsequent shape fit. Finally, the calibration constants were derived assuming a polynomial energy response:

$$\mathrm{E} = \mathcal{C}\left(\mathrm{ADC}, p_0, p_1, p_2\right) = p_0 + p_1 \times \mathrm{ADC} + p_2 \times \mathrm{ADC}^2. \tag{1}$$

The SDDs energy response is known to have a linear dependency on the ADC [27], however a quadratic term was also introduced to correct for possible distortions at the level of the front-end electronics. Additionally, the relatively simple form enables a differentiable programming approach, as shown in the next subsection.

3.2. A differentiable programming approach to calibration optimization

Differentiable programming is a computational paradigm that enables the automatic differentiation of mathematical functions, making it a valuable tool for scientific computing and ML, already pioneered in High Energy Physics [11–13]. Within this paradigm, it is possible to employ high-level syntax to form fully differentiable complex models that can be optimized using gradient-based methods. The key feature of differentiable programming is its ability to automatically compute function gradients with respect to their inputs. This is achieved through a process called reverse-mode automatic differentiation, which propagates gradients backward through the function graph using the chain rule of calculus. This allows for efficient computation of gradients, even for functions with numerous inputs and outputs.

We used an approach based on differentiable programming to optimize the p_0, p_1 and p_2 constants for each one of the 32 SDD cells and for each calibration batch. For this purpose, JAX [28], an open-source numerical computing library designed for high-performance ML research, was used.

Since we are dealing with three calibration constants for each $i \in 1 \ldots N$ calibration batch, we define the calibration matrix as:

$$\mathbf{P} = \begin{bmatrix} p_{0,1} & p_{1,1} & p_{2,1} \\ \vdots & \vdots & \vdots \\ p_{0,i} & p_{1,i} & p_{2,i} \\ \vdots & \vdots & \vdots \\ p_{0,N} & p_{1,N} & p_{2,N} \end{bmatrix}. \tag{2}$$

The goal is to optimize this matrix to get a better spectroscopic response of the SDDs. In order to do that, leveraging differentiable programming, we need to define a loss-function. We define a likelihood-based loss-function, taking into account the shape of the copper K_α line, since it is the closest to the PEP violating ROI (7000–8500 eV):

$$\begin{aligned} \mathcal{L}(\mathbf{ADC}, \mathbf{P}) = \prod_i \prod_{j \in i} \big(& \mathrm{Gauss}(\mathcal{C}(\mathrm{ADC}_{i,j}, P_i) - \mu_{\mathrm{Cu}_{K_{\alpha 1}}}, \sigma) \\ & + T_1(\mathrm{Gauss}(\mathcal{C}(\mathrm{ADC}_{i,j}, P_i))) \\ & + \mathrm{Gauss}(\mathcal{C}(\mathrm{ADC}_{i,j}, P_i) - \mu_{\mathrm{Cu}_{K_{\alpha 2}}}, \sigma) \\ & + T_2(\mathrm{Gauss}(\mathcal{C}(\mathrm{ADC}_{i,j}, P_i)))\big) \end{aligned} \tag{3}$$

where $\mathrm{ADC}_{i,j}$ is the jth value of the ADC in the batch i, Gauss is the Gaussian distribution, $\mu_{\mathrm{Cu}_{K_{\alpha 1}}} = 8047.8$ eV, $\mu_{\mathrm{Cu}_{K_{\alpha 2}}} = 8027.8$ eV [29, 30], and σ the width. P_i is the set of calibration constants $\{p_{0,i}, p_{1,i}, p_{2,i}\}$ in the ith batch. Finally, the two contribution T_1 and T_2 are the tail functions, which reproduce incomplete charge collection [31], and are written as:

$$T_i(x) = \frac{A_i}{2\beta\sigma} \times e^{\frac{x-\nu}{\beta\sigma}\frac{1}{2\beta^2}} \times \mathrm{erfc}\left(\frac{x-\nu}{\sqrt{2}\sigma} + \frac{1}{\sqrt{2}\beta}\right) \tag{4}$$

with A_i the normalized amplitude of the tail, β the tail's slope, erfc the complementary error function. ν is $\mu_{\mathrm{Cu}_{K_{\alpha 1}}}$ for $i = 1$ and $\mu_{\mathrm{Cu}_{K_{\alpha 2}}}$ for $i = 2$.

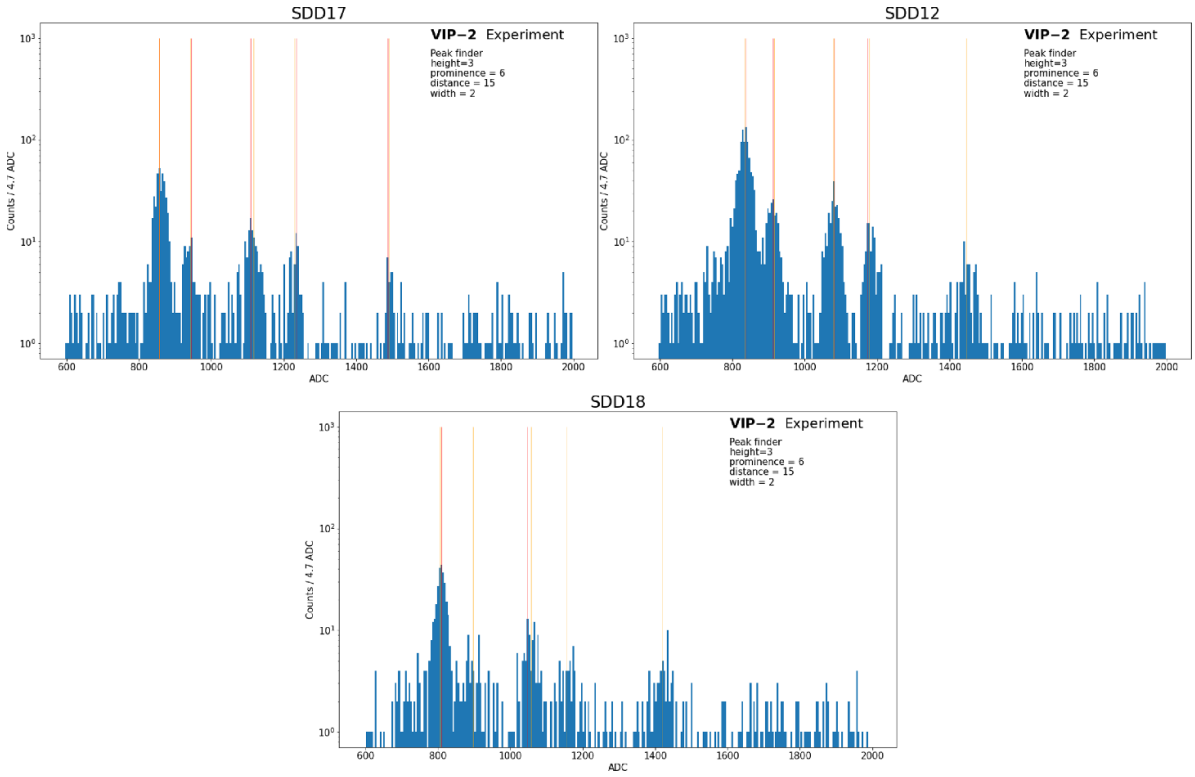

Figure 4. The ADC spectrum of selected SDDs (17, top left; 12 top right; 18 bottom) taken by the VIP-2 experiment in 2022, which exhibit a similar, yet slightly different relative yields. The orange lines are obtained by the neural network prediction, and are found to correctly identify the spectroscopic lines of interest (titanium, manganese and copper), even with low counts and statistical fluctuations in all the considered cases. The red lines represent `SciPy` peak finder algorithm, with the algorithm parameters printed on the top right of the plot. Due to the different levels of statistical fluctuations, the algorithm does not identify copper K_α (SDD12) and Ti K_β, Mn K_β (SDD18) with the same set of parameters, requiring extensive calibration and fine-tune.

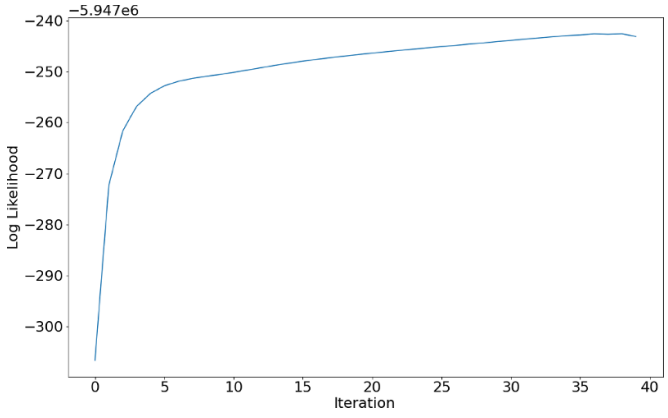

Figure 5. The value of the Log Likelihood defined in equation (3) as a function of the number of iterations in the gradient-descent.

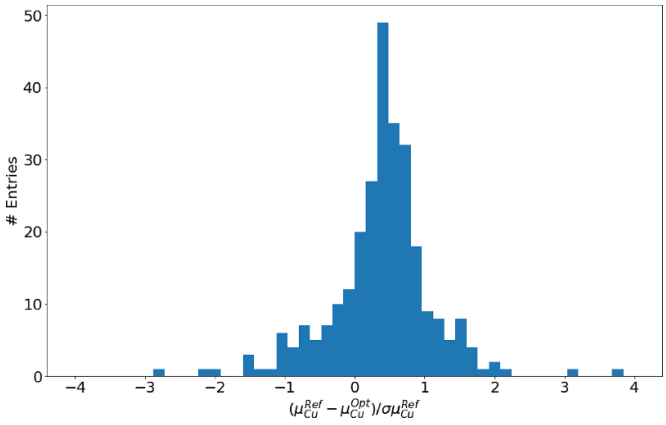

Figure 6. The relative difference of the Cu K_α positions before and after gradient descent, normalized with respect to its statistical error.

The gradient of this function with respect to the calibration parameters $\mathbf{P}$ is computed with JAX, allowing the optimization. For enhanced numerical and convergence stability, the logarithm of the likelihood is used.

We find that already after 40 iterations, the optimization converges to a maximum, as shown in figure 5.

Additionally, as a cross- and sanity-check, we quantified how much the optimized calibration parameters differ from the reference, by studying $(\mu_{\mathrm{Cu}}^{\mathrm{Ref}} - \mu_{\mathrm{Cu}}^{\mathrm{Opt}})/\sigma\mu_{\mathrm{Cu}}^{\mathrm{Ref}}$, where $\mu_{\mathrm{Cu}}^{\mathrm{Ref}}$ is the reference Cu K_α centroid position with statistical error $\sigma\mu_{\mathrm{Cu}}^{\mathrm{Ref}}$, and $\mu_{\mathrm{Cu}}^{\mathrm{Opt}}$ is the one corresponding to the optimized parameters. The statistical error is obtained from the fit in MINUIT [23].

The histogram of these relative differences is shown in figure 6, where it can be seen that the vast majority of the centroid displacements following the optimization procedure are well within the 1σ statistical error of the fit, ensuring that the optimization is not significantly changing the calibration parameters, as expected.

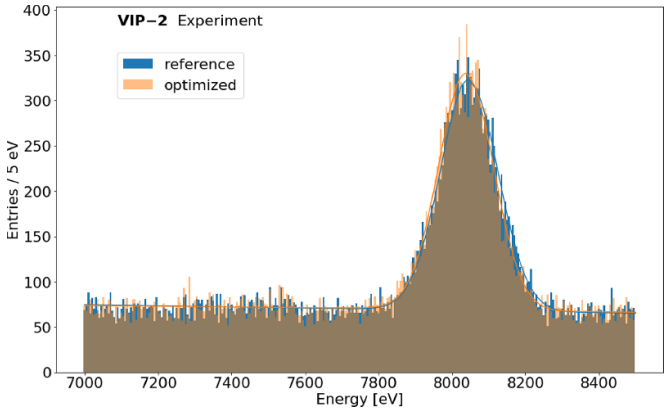

Figure 7. The copper line before (blue) and after (orange) the optimization, corresponding to the entire dataset. The solid lines represent the best fits to the spectra, in the energy range 7000-8500 eV, which is the experimental ROI of the VIP-2 experiment. The quantitative comparison of the two spectra is presented in table 1 in terms of peak position, FWHM and reduced χ^2.

Table 1. Comparison of reference and optimized positions, FWHM, and χ^2/ndf.

	Position (eV)	FWHM (eV)	χ^2/ndf
Reference	8050 ± 1	185 ± 2	1.64
Optimized	8048 ± 1	175 ± 2	1.27

4. Results and discussion

In order to assess the overall gains of the method, we compare the two FWHMs of the copper K_α line after and before the optimization. In this work, we focus on the K_α since we expect the PEP violating transition in copper; however, this method can be applied to any other spectroscopic property. In figure 7 the blue spectrum represents the reference data and the orange spectrum the data obtained with the enhanced calibration procedure, for the entire dataset. The energy range in the figure is the ROI of the VIP-2 search for PEP violation in copper. As it can be seen, the copper peak results to be more prominent, and with a smaller width. We quantify the enhancement of these spectroscopic qualities with a fit to the line shape. The best fit is shown as a blue and an orange line on the plot, and the best fit values are reported in table 1, where the values of the peak position, FWHM and reduced χ^2 are reported.

The fit function used to describe the shape is:

$$f(x,A,\mu,\sigma) = A \times \frac{51}{100} \times \text{Gauss}\,(x-\mu-20,\sigma) + T_2\,(x) + A \times \text{Gauss}\,(x-\mu,\sigma) + T_1\,(x) + m \times x + C \quad (5)$$

where the first term describes the $K_{\alpha 2}$ with relative intensity 51/100 and the second term describes the $K_{\alpha 1}$. About 20 eV is the energy difference between the two, and the continuum background is found to be best described with a linear function. The two contributions $T_1(x)$ and $T_2(x)$ are the tail functions included to reproduce incomplete charge collection as defined above.

The results of the fit demonstrate that the ML and differentiable programming approach have yielded improvements in each of the examined parameters. In particular, the peak position in the optimized procedure has shown to correct for the residual miscalibration, and the centroid value is compatible with tabulated 8047.8 eV, well within the statistical uncertainty of the fit, substantially reducing the impact of the energy scale uncertainty. The FWHM of the line has reached below 180 eV for the first time in VIP-2, showing a substantial improvement with respect to traditional methods used in the past. Finally, the reduced chi-squared, used to quantify the agreement of the data against the model, also shows improvement, indicating greater compatibility between the data and the model.

4.1. Discussion

In this work, we proposed a twofold approach to enhance the detector calibration, particularly when dealing with challenging features in spectra, such as those found in VIP-2 data. Firstly, we utilized CNNs, which demonstrated performance gains in precisely identifying the positions of spectroscopic peaks. CNNs are well-suited for CPU training, fast prediction, and easy interpretation. RNNs, such as LSTM, and attention-based ML architectures are powerful tools extensively used in natural language processing, thanks to their capability to retain information across the entire sequence. However, in the context of spectroscopic data, this characteristic may not be as beneficial. Unlike natural language, where the sequence's structure is crucial, spectroscopic lines' properties are primarily influenced by their immediate surroundings, rather than by the entire spectrum. The CNN architecture we have presented in this work captures features with different size through the use of two parallel convolutional towers with different kernel sizes, enabling its use also on data with low yield.

In the second part, we have introduced a completely novel approach to detector calibration optimization based on differentiable programming. In general, the calibration process is performed leveraging known processes or quantities to characterize the response of a detector system, which also could change as a function of time and other relevant physical quantities such as temperature, pressure or electronic conditions. Several steps can be taken to increase quality and the precision of the calibration, such as assuring the stability of the experimental setup, acquiring calibration data with the sufficient statistics and so on. Often this poses a challenge to the measurement design, which must carefully consider this factor. Additionally, the degree of precision achievable in energy calibration typically represents the most relevant systematic uncertainties. With this new technique, as we have shown, a regression is performed on the initial calibration parameters, with respect to the precision of the final reference observable. Arbitrarily complex calibration chains, even with higher dimensionality, could employ this technique, making it applicable downstream of each calibration step.

Following this study applied to the most recent VIP-2 data, we outline the advantages this method brings with respect to the physics reach of the VIP-2 experiment. First, in the analysis of the 2020 VIP-2 data [19], the energy scale uncertainty was included to reflect the differences in scale between calibrated spectrum and standard lines, originating from the calibration procedure. With the new approach, and its capability to correct for potential miscalibration, we show that this effect will be substantially reduced. Second, the reduced chi-squared shows that the data calibrated with this method have a higher degree of compatibility with the model, ensuring a more coherent description of the background and of the spectroscopic lines. The PEP signal description at 7.7 keV relies on the accurate knowledge of the continuum background, and the energy scale. Finally, the better energy resolution in the form of a smaller FWHM assures a higher discovery significance of the signal, since the PEP-violating line would appear more prominently above the background.

5. Conclusions

The VIP-2 experiment at LNGS is searching for signals beyond the standard quantum theory, namely PEP forbidden transitions in copper, possibly connecting Lorentz invariance, extra dimensions and quantum gravity to atomic x-ray phenomena. The SDDs are optimal tools to inspect this energy range. However, so far their energy resolution in VIP-2 has typically never gone substantially below 190 eV. We have applied in this work an approach to leverage the calibration procedure in order to push the FWHM of the copper line at the hardware limit. We have trained a neural network with synthetic data as peak finder to exploit the peculiar spectrum and *in-situ* calibration lines even at low yields, enabling the use of much smaller calibration batches. The network output was then used as input in a fully differentiable framework, where the loss function was described in terms of the agreement with the copper line as a likelihood. The gradient descent has shown substantial improvements of the spectroscopic properties of the detectors, both in terms of the compatibility of the line with its two components, and by bringing for the first time in VIP-2 the SDD's FWHM at 8 keV below 180 eV. Last, but not least, the method has shown to correct for small miscalibrations. This novel approach has the potential to improve the physics capabilities with enhanced calibration. For VIP-2, this translates to several advantages: a smaller energy scale uncertainty, better control over the background in the ROI, and consequently a higher discovery significance for PEP violating signals. We plan to employ this method on the entire VIP-2 dataset. Furthermore, we plan to explore the advantages it yields in experiments which strongly rely on the precise determination of spectroscopic lines in the x-ray domain, such as the SIDDHARTA-2 experiment [32] at the DAΦNE collider.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the corresponding author.

Acknowledgments

We thank: the INFN Institute, for supporting the research presented in this article and, in particular, the Gran Sasso underground laboratory of INFN, INFN-LNGS, and its Director, Ezio Previtali, and the LNGS staff. We thank C Capoccia from LNF and H Schneider, L Stohwasser, and D Stückler from Stefan-Meyer-Institute for their fundamental contribution in designing and building the VIP-2 setup.

Funding

This research was funded by the National Institute for Nuclear Physics (INFN, Italy). This research was funded in whole, or in part, by the Austrian Science Fund (FWF) Grants P25529-N20, Projects P 30635-N36 and W1252-N27 (doctoral college particles and interactions). This publication was also made possible through the support of Grant 62099 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. We also thank the support of the H2020 FET project TEQ with Grant 766900. We gratefully acknowledge support from Centro Ricerche Enrico Fermi ('Problemi aperti nella meccanica quantistica' project), and from Foundational Questions Institute (Grants Nos. FQXi-RFP-CPW-2008 and FQXi-MGA-2102).

ORCID iDs

Fabrizio Napolitano https://orcid.org/0000-0002-8686-5923
Raffaele Del Grande https://orcid.org/0000-0002-7599-2716
Marco Miliucci https://orcid.org/0000-0002-2315-2379
Francesco Sgaramella https://orcid.org/0000-0002-0011-8864

References

[1] Brahma S, Ronco M, Amelino-Camelia G and Marciano A 2017 Linking loop quantum gravity quantization ambiguities with phenomenology *Phys. Rev.* D **95** 044005
[2] Arzano M and Kowalski-Glikman J 2016 Deformed discrete symmetries *Phys. Lett.* B **760** 69–73
[3] Piscicchia K *et al* 2023 Experimental test of noncommutative quantum gravity by VIP-2 lead *Phys. Rev.* D **107** 026002
[4] Shi H *et al* 2018 Experimental search for the violation of Pauli exclusion principle: VIP-2 collaboration *Eur. Phys. J.* C **78** 1–14
[5] Holl P, Hauertmann L, Majorovits B, Schulz O, Schuster M and Zsigmond A J 2019 Deep learning based pulse shape discrimination for germanium detectors *Eur. Phys. J.* C **79** 450
[6] Luo R, Popp J and Bocklitz T 2022 Deep learning for Raman spectroscopy: a review *Analytica* **3** 287–301

[7] Drera G, Kropf C M and Sangaletti L 2020 Deep neural network for x-ray photoelectron spectroscopy data analysis *Mach. Learn.: Sci. Technol.* **1** 015008
[8] Zhang W, Kasun L C, Wang Q J, Zheng Y and Lin Z 2022 A review of machine learning for near-infrared spectroscopy *Sensors* **22** 9764
[9] Chatzidakis M and Botton G A 2019 Towards calibration-invariant spectroscopy using deep learning *Sci. Rep.* **9** 2126
[10] Rankine C D and Penfold T J 2021 Progress in the theory of x-ray spectroscopy: from quantum chemistry to machine learning and ultrafast dynamics *J. Phys. Chem.* A **125** 4276–93
[11] De Castro P and Dorigo T 2019 Inferno: inference-aware neural optimisation *Comput. Phys. Commun.* **244** 170–9
[12] Simpson N and Heinrich L 2023 neos: end-to-end-optimised summary statistics for high energy physics *J. Phys.: Conf. Ser.* **2438** 012105
[13] Dorigo T *et al* 2022 Toward the end-to-end optimization of particle physics instruments with differentiable programming: a white paper (arXiv:2203.13818)
[14] Curceanu C *et al* 2019 The modern era of light kaonic atom experiments *Rev. Mod. Phys.* **91** 025006
[15] Piscicchia K *et al* 2020 VIP-2—high-sensitivity tests on the Pauli exclusion principle for electrons *Entropy* **22** 1195
[16] Lechner P *et al* 1996 Silicon drift detectors for high resolution room temperature x-ray spectroscopy *Nucl. Instrum. Methods Phys. Res.* A **377** 346–51
[17] Lechner P *et al* 2001 Silicon drift detectors for high count rate x-ray spectroscopy at room temperature *Nucl. Instrum. Methods Phys. Res.* A **458** 281–7
[18] De Paolis L *et al* 2022 The Pauli exclusion principle for electrons tested by VIP-2 at the LNGS laboratories *Il Nuovo Cimento* C **45** 1–4
[19] Napolitano F *et al* 2022 Testing the Pauli exclusion principle with the VIP-2 experiment *Symmetry* **14** 893
[20] Quaglia R, Luca Bombelli P B, Fiorini C, Occhipinti M, Giacomini G, Ficorella F, Picciotto A and Piemonte C 2015 Silicon drift detectors and cube preamplifiers for high-resolution x-ray spectroscopy *IEEE Trans. Nucl. Sci.* **62** 221–7
[21] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2015 Going deeper with convolutions *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 1–9
[22] Pan D, Kibbe W A and Lin S M 2006 Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching *Bioinformatics* **22** 2059–65
[23] Antcheva I *et al* 2011 Root-a c++ framework for petabyte data storage, statistical analysis and visualization *Comput. Phys. Commun.* **182** 1384–5
[24] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
[25] Weytjens H and De Weerdt J 2020 Process outcome prediction: CNN vs. LSTM (with attention) *Business Process Management Workshops: BPM 2020 Int. Workshops (Revised Selected Papers) (Seville, Spain, 13–18 September 2020)* vol 18 (Springer) pp 321–33
[26] Laitala J, Jiang M, Syrjälä E, Kasaeyan Naeini E, Airola A, Rahmani A M, Dutt N D and Liljeberg P 2020 Robust ecg r-peak detection using lstm *Proc. 35th Annual ACM Symp. on Applied Computing (SAC '20)* (Association for Computing Machinery) pp 1104–11
[27] Miliucci M, Iliescu M, Amirkhani A, Bazzi M, Curceanu C, Fiorini C, Scordo A, Sirghi F and Zmeskal J 2019 Energy response of silicon drift detectors for kaonic atom precision measurements *Condens. Matter* **4** 31
[28] Bradbury J, Frostig R, Hawkins P, James Johnson M, Leary C, Maclaurin D, Necula G, Paszke A, VanderPlas J, Wanderman-Milne S and Zhang Q 2018 JAX: composable transformations of Python+NumPy programs
[29] Otto Krause M and Oliver J H 1979 Natural widths of atomic K and L levels, K α x-ray lines and several KLL Auger lines *J. Phys. Chem. Ref. Data* **8** 329–38
[30] Alvin Bearden J 1967 X-ray wavelengths *Rev. Mod. Phys.* **39** 78
[31] Van Gysel M, Lemberge P and Van Espen P 2003 Implementation of a spectrum fitting procedure using a robust peak model *X-Ray Spectrom.* **32** 434–41
[32] Miliucci M *et al* 2021 Silicon drift detectors' spectroscopic response during the SIDDHARTA-2 kaonic helium run at the DAϕNE collider *Condens. Matter* **6** 47