EPJ Web of Conferences **295**, 02003 (2024) https://doi.org/10.1051/epjconf/202429502003
CHEP 2023

AI Driven Experiment Calibration and Control

Thomas Britton[1,*], *Cullan* Bedwell[3], *Abhijeet* Chawhan[3], *Julie* Crowe[3], *Naomi* Jarvis[2], *Torri* Jeske[1], *Nikhil* Kalra[1], *David* Lawrence[1], and *Diana* McSpadden[1,3]

[1]Thomas Jefferson National Accelerator Facility
[2]Carnegie Mellon University
[3]University of Virginia

Abstract. One critical step on the path from data taking to physics analysis is calibration. For many experiments this step is both time consuming and computationally expensive. The AI Experimental Calibration and Control project seeks to address these issues, starting first with the GlueX Central Drift Chamber (CDC). We demonstrate the ability of a Gaussian Process to estimate the gain correction factor (GCF) of the GlueX CDC accurately, and also the uncertainty of this estimate. Using the estimated GCF, the developed system infers a new high voltage (HV) setting that stabilizes the GCF in the face of changing environmental conditions. This happens in near real time during data taking and produces data which are already approximately gain-calibrated, eliminating the cost of performing those calibrations which vary ± 15% with fixed HV. We also demonstrate an implementation of an uncertainty aware system which exploits a key feature of a Gaussian process.

1 Offline Calibrations

Traditionally, detector calibrations are performed after data taking and rely on slow and computationally expensive track reconstruction. The offline calibration time for a typical GlueX experiment takes on the order of months, with multiple rounds of inter-dependent calibrations requiring significant attention from experts. This inevitably produces delays in the publication of experimental results. Machine learning can be used to make calibrations more efficient by reducing the amount of time (CPU and expert) spent on calibrations by reducing the number of iterations needed to reach the point of "fine-tuning" the calibrations. For this work specifically, we aim to have machine learning models predict the calibration constants in near real time and, where applicable, make control decisions to stabilize the detectors throughout data taking. Additionally, we require an estimate on the uncertainty of the prediction to be utilized for control decisions in the event the model is uncertain about its prediction. In order to make predictions quickly, we tend to favor the use of input features that do not rely on computationally expensive processes such as reconstruction.

2 The GlueX Central Drift Chamber

The Gluonic Excitations Experiment (GlueX) is located in Hall-D at the Thomas Jefferson National Accelerator Facility in Newport News, Virginia. It is designed to search for and

*e-mail: tbritton@jlab.org

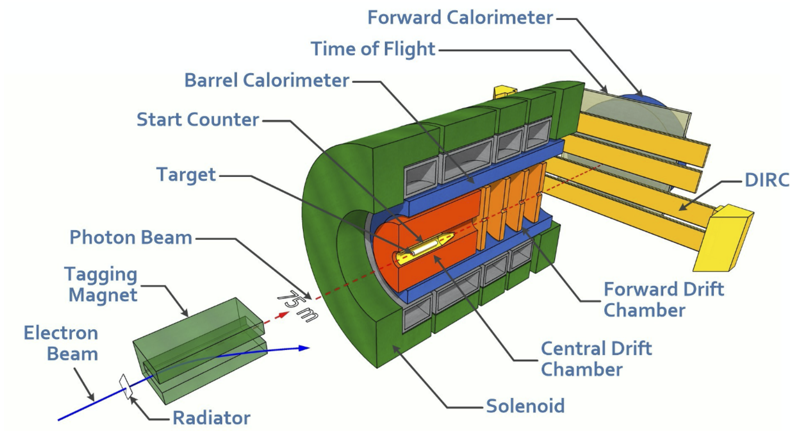

Figure 1. The GlueX spectrometer located in Hall-D at Thomas Jefferson National Accelerator Facility.

measure exotic hybrid mesons using photoproduction [1]. A schematic of the detector is shown in Fig. 1.

The GlueX Central Drift Chamber (CDC) [2] is used to detect and track charged particles with momenta greater than 0.25 GeV/c. It is a 1.5 m long by 1.2 m diameter cylindrical chamber with 3522 anode wires held at a nominal 2125 V. Each wire is surrounded by a 1.6 cm diameter straw. A gas mixture, 50:50 Ar:CO_2, flows through the chamber.

As charged particles traverse the detector, the gas is ionized, and the corresponding avalanche of electrons drift towards the wires forming a pulse hit. The pattern of hits is used to reconstruct the particle's path through the detector and are used for particle identification via the deposited energy per unit length (dE/dx). Gas density affects the gain, the ratio of final to initial number of electrons in the avalanche. Since the CDC is operated at atmospheric pressure, the variation in gain due to changes in atmospheric pressure throughout data taking is accounted for via calibration.

3 Data and Model Selection

The atmospheric pressure, gas temperature, and the current drawn by the CDC HV boards are measured throughout the duration of an experiment and are readily accessible via the Experimental Physics Industrial Controls System (EPICS) archive [3]. For training, 601 production quality runs from the 2020 and 2021 run periods were used. Multiple run periods were selected so that our model could learn from a diverse data set. For the 2020 run period, the mean high voltage board current, mean gas temperature, and mean atmospheric pressure were 9.0 uA, 100.5 kPa, and 299.2 K, respectively. For 2021, the mean high voltage board current, maximum gas temperature, and mean atmospheric pressure were 0.9 uA, 99.9 kPa, and 299.8K, respectively. The training set consists of 480 (2020: 430, 2021: 50) runs, and the test set contains 121 runs (2020: 106, 2021: 15). The training and test set distributions are shown in Fig. 2.

The US Department of Energy has identified machine learning model uncertainty quantification as a priority research area [4]. Moreover, unlike a physics-driven algorithm, the novel use of a data-driven approach requires the GlueX collaboration and the CDC expert to trust that the control system will not interpolate or extrapolate in unexpected ways. Addressing these concerns, Gaussian Process Regression (GPR) has been extensively applied

https://doi.org/10.1051/epjconf/202429502003

Table 1. FCAL CNN performance for masked and unmasked data sets.

Dataset	Fold Idx	Avg. Residual	MAPE	MSE
unmasked	average	0.258	23.848	5.183
hline masked	average	0.027	2.370	0.004

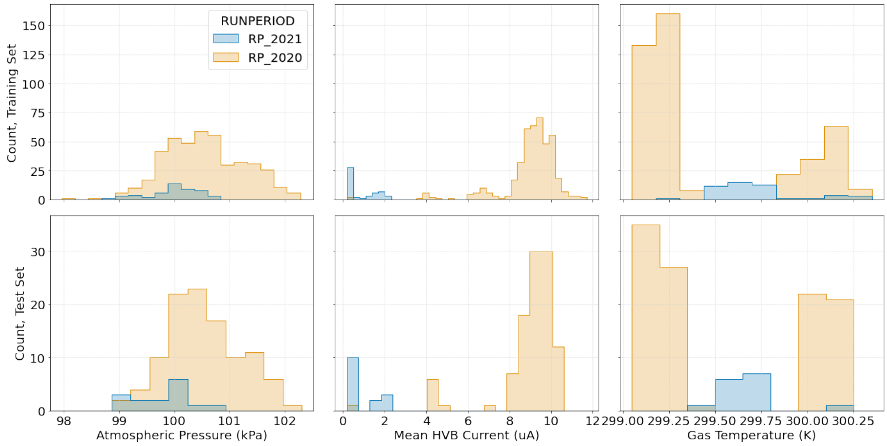

Figure 2. Input feature distributions used for training and testing the Gaussian process. The 2021 run period is shown in blue and the 2020 run period is shown in orange.

in high-stakes decision applications such as aviation design and healthcare. This is due to the intrinsic property of providing distance-aware uncertainty estimates [5]. Using a GPR, a prediction and the associated uncertainty are derived by conditioning a prior with the training data [6, 7]. A well-calibrated GPR will provide high uncertainty estimates when extrapolation is required.

A control system can then act based on uncertainty thresholds. While unlikely due to operational constraints of the Thomas Jefferson National Accelerator Facility accelerator, if environmental or experimental conditions stray into parameter regions resulting in a high or low predicted GCF with high uncertainty, it is critical that we do not adjust the high voltage of the CDC to a correspondingly uncertain, or *unsafe*, value. In addition, the use of a data-driven approach to control a particle detector requires physicists to trust the data-driven algorithm; otherwise, it would not be implemented.

A Gaussian process (GP) with a multilayer perceptron (MLP) learned prior was chosen to predict the calibration constants. A White Noise and Radial Basis Function kernel was used to model the underlying distributions of the input data. Because of the small number of input features, the GP is able to make inferences in milliseconds. The mean absolute percent error was 4.8% on the test set.

4 Using ML to Calibrate and Control the GlueX CDC

We developed a modular control system that utilizes a MySQL database to store model information as well as the input features, inferences, and control decisions throughout data taking. A Grafana dashboard displays this information such that the shift crew and experts can monitor the system in real time. Additionally, the shift crew can toggle the control aspect of the system on or off as deemed necessary by detector experts. In certain circumstances, such as high beam current tests, the system can default to "trusting the humans" and will not try to adjust the HV setting based on the input data it is receiving. Default input feature settings, control windows, look-back periods, and the use of Uncertainty Quantification (UQ) based corrections can be changed via a configuration file without interrupting data taking. A diagram of the system is shown in Fig. 3.

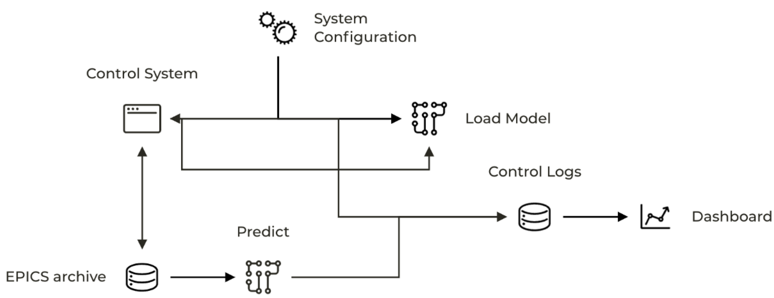

Figure 3. Diagram outlining the main processes for HV control.

When the shift crew starts a run via the data acquisition software (DAQ), the system gathers the relevant input features from the EPICS archive. These are given to the model and a prediction is obtained in milliseconds. From this prediction, the HV is extracted using a polynomial fit obtained from data taken during HV scans. The HV of each of the 72 boards is adjusted before data taking begins.

5 Deployments

5.1 Cosmic Ray Tests

The first test of the HV control system took place during a two week "down time" period. The CDC was divided in software such that the HV setting on one side of the chamber was fixed at 2130 V, and the other side was allowed to vary in the range [2113, 2140] V, as determined by the Gaussian process. The HV was adjusted every 5 minutes during the test, and experts monitored the system remotely. A schematic of the CDC and the results are shown in Fig. 4. The changes in the GCF throughout the run are reduced when adjusting the HV (blue points) as compared to the traditional operation of the CDC (orange points). The large variation in the GCF for the constant HV side was due to a thunderstorm, causing a drop in the atmospheric pressure.

5.2 Charged Pion Polarizability Experiment

The second deployment was during the Charged Pion Polarizability (CPP) experiment that took place in GlueX in 2022. During this experiment, the CDC was not a critical detector for particle tracking or identification. In addition, the target position was adjusted from its usual location, and the experiment required a low beam current. The system was run automatically at the start of each two hour data taking session (run). If the predicted GCF was less than 3% of the ideal GCF (set at the start of data taking under ideal conditions), the HV would be adjusted to the recommended setting from the known relationship between the HV and GCF. Otherwise, the closest HV setting in Euclidean distance on the uncertainty surface was used. This surface is shown in Fig. 5. For empty target runs, the voltage was set back to the standard operating voltage of 2125 V.

https://doi.org/10.1051/epjconf/202429502003

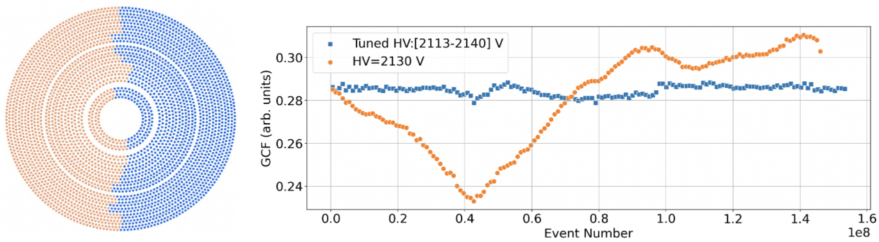

Figure 4. Left: Schematic of CDC with regions for fixed HV (orange) and varied HV (blue). Right: Gain Correction Factor (GCF) as a function of event number for a single run (3.5 days of data taking).

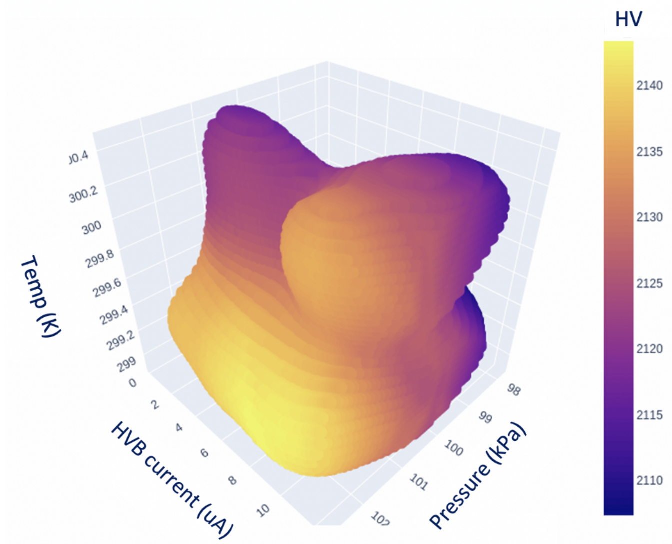

Figure 5. Surface plot indicating the recommended HV setting based on our three input feature values.

5.3 Primex-η Experiment

For the PrimEx-η run period, the control system was updated to automatically detect an empty target run and adjust the HV to 2125 V. In Fig. 6, the ratio of the GCF to the ideal GCF obtained at the start of the experiment is shown as a function of run number. For the runs where the HV was tuned according to the GP, the gain is more stable even with fluctuating gas density (gray points).

6 Conclusions

The high voltage of the GlueX Central Drift Chamber is now dynamically adjusted throughout data taking in order to stabilize the gas gain in response to changes in the atmospheric pressure, gas temperature, and the current drawn from the CDC high voltage boards. This is in contrast to the traditional operating conditions in which the high voltage was held at a constant 2125 V, requiring the gas gain to be calibrated after data taking has finished. A Gaussian process predicts the gain correction factor using environmental and experimental

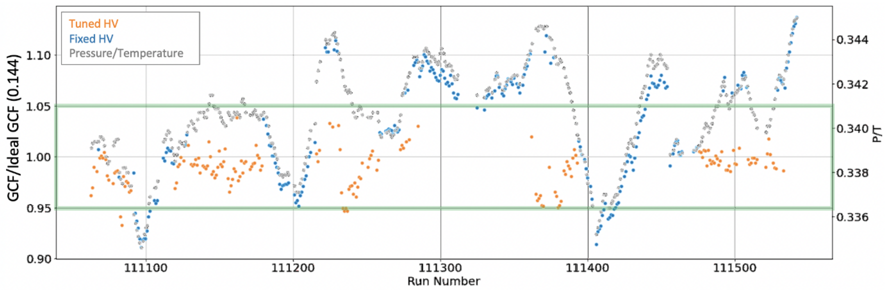

Figure 6. The ratio of GCF to the ideal for tuned HV (orange) and fixed HV (blue) as a function of Run Number. The gray points indicate the gas density as a function of Run Number. The green lines indicate ± 5% of the ideal gain value obtained at the start of the experiment.

conditions as input features. Decision policies, based on the uncertainty of the prediction, are determined in conjunction with detector experts to ensure optimal detector response.

The CPP and PrimEx-η deployments of the CDC HV control system demonstrate a first step in a shift from post-experiment data calibration to real-time calibration through experimental control. Throughout both run periods, control occurred at an approximately two-hour time scale. However, uncertainty-aware machine learning calibration and control should be investigated for additional existing and also future detectors with differing time-scale requirements, and this approach should be considered in the streaming readout regime. To understand the application of these methods to other systems, initial steps utilizing data from the LED monitoring system associated with the GlueX Forward Calorimeter as inputs to a Convolutional Neural Network [8] to predict the gain calibration constants is being explored. For both the CDC and FCAL models, we utilize input variables that are readily available during data taking, as opposed to relying on track reconstruction offline, to perform calibrations in near real-time.

7 Acknowledgements

Jefferson Science Associates, LLC operated Thomas Jefferson National Accelerator Facility for the United States Department of Energy under U.S. DOE Contract No. DE-AC05-06OR23177. The Carnegie Mellon Group is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, DOE Grant No. DE-FG02-87ER40315. This work was supported by the US DOE as LAB 20-2261.

References

[1] S.A. et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **987**, 164807 (2021)

[2] Y. Van Haarlem, C. Meyer, F. Barbosa, B. Dey, D. Lawrence, V. Razmyslovich, E. Smith, G. Visser, T. Whitlatch, G. Wilkin et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **622**, 142 (2010)

[3] A. Kozubal, L. Dalesio, J. Hill, D. Kerstiens, *Experimental Physics and Industrial Control System*, in *ICALEPCS89 Proceedings* (1989), Vol. 288

[4] Energy.gov, *Department of Energy Invests $ 16 Million in Data-Intensive Scientific Machine Learning Research and Analysis*
[5] V. Nemani, L. Biggio, X. Huan, Z. Hu, O. Fink, A. Tran, Y. Wang, X. Zhang, C. Hu, Mechanical Systems and Signal Processing **205**, 110796 (2023)
[6] C.E. Rasmussen, C.K. Williams et al., *Gaussian processes for machine learning*, Vol. 1 (Springer, 2006)
[7] C. Williams, C. Rasmussen, *Gaussian Processes for Regression*, in *Advances in Neural Information Processing Systems*, edited by D. Touretzky, M. Mozer, M. Hasselmo (MIT Press, 1995), Vol. 8, `https://proceedings.neurips.cc/paper_files/paper/1995/file/7cce53cf90577442771720a370c3c723-Paper.pdf`
[8] K. Fukushima, Biological cybernetics **36**, 193 (1980)