# Application of deep learning methods for beam size control during user operation at the Advanced Light Source

Thorsten Hellert,* Tynan Ford, Simon C. Leemann,
Hiroshi Nishimura, and Marco Venturini
*Lawrence Berkeley National Laboratory, Berkeley 94720, California, USA*

Andrea Pollastro
*Department of Electrical Engineering and Information Technology (DIETI), University of Naples Federico II, Naples, Italy and Instrumentation and Measurement for Particle Accelerator Laboratory (IMPALab), Naples, Italy*

(Received 1 April 2024; accepted 10 July 2024; published 30 July 2024)

Past research at the Advanced Light Source (ALS) provided a proof-of-principle demonstration that deep learning methods could be effectively employed to compensate for the significant perturbations to the transverse electron beam size induced by user-controlled adjustments of the insertion devices. However, incorporating these methods into the ALS' daily operations has faced notable challenges. The complexity of the system's operational requirements and the significant upkeep demands has restricted their sustained application during user operation. Here, we introduce the development of a more robust neural network (NN)-based algorithm that utilizes a novel online fine-tuning approach and its systematic integration into the day-to-day machine operations. Our analysis emphasizes the process of NN model selection, demonstrates the superior performance of the NN-based method over traditional feedback methods, and examines the effectiveness and resilience of the new algorithm during user-operation scenarios.

DOI: 10.1103/PhysRevAccelBeams.27.074602

## I. INTRODUCTION

The performance of storage ring light sources is critically reliant on the stability of the radiation output in terms of source position/angle and intensity. A major advance toward improving the radiation intensity long-term stability was achieved with the adoption of "top-off" (or "top-up") injection [1,2]. The radiation source position/angle stability is achieved by control of the electron beam orbit through a combination of local and global orbit feedback (FB) and feed-forward (FF) systems. Orbit stability at the sub-μm level, maintained over several hours, is now typical (see, e.g., [3,4]).

On the short timescale, the stability of the radiation intensity is primarily affected by the electron beam transverse size response to changes in the insertion device (ID) parameters (gap and phase) that occur during user operation. While in general the horizontal beam size is largely independent of the exact ID settings (as long as the natural emittance remains dominated by the radiation losses in the bending magnets), the vertical beam size tends to be sensitive to the normal and especially skew quadrupole-field errors originating from the IDs. To compensate for these errors, the Advanced Light Source (ALS) [5], like other storage-ring light sources, employs quadrupoles, and skew quadrupole correctors in an FF configuration [6].

Because of the difficulty of developing an accurate physics model for the ID errors, the FF corrections are best defined based on beam measurements. A distinct set of measurements is conducted for each individual ID resulting in the creation of lookup tables that, in correspondence to the given ID gap and phase configuration, specify the lattice corrections necessary to remove beta beat and linear coupling. Linear superposition is then invoked to combine the corrections originating from all the IDs. Since these measurements are time-consuming, the quality of the lookup tables can be negatively affected by the machine short-term drifts during the measurement process. Moreover, long-term drifts (due to factors, such as ground motion and radiation aging of magnets) will eventually compromise their effectiveness in driving the FF correction, and violation of the linear superposition assumption will invariably result in imperfect compensation even in the absence of drifts.

A conceivable way to remedy these shortcomings is to add a conventional FB system, whereby the vertical beam size is monitored in real time at a diagnostic beamline and the control variable driven by the beam size measurement

*Contact author: thellert@lbl.gov

excites a vertical dispersion wave through a suitable combination of skew quadrupole correctors. However, as the underlying measurement of the beam size is prone to noise, thus limiting the range of usable FB gain, we found that an FB system falls short of achieving the required closed-loop bandwidth to dampen disturbances throughout the entire desired operational range.

Previous investigations [7] successfully demonstrated that a more effective approach is to supplement the lookup-table system with an additional FF system layer based on a neural network (NN) trained on beam size measurement, ID parameters, and dispersion wave parameter (DWP) data. Unfortunately, the implementation in day-to-day ALS operations of the algorithm proposed in [7] encountered significant barriers, which have effectively prevented its sustained deployment beyond initial applications. This paper addresses the limitations of the NN approach of Ref. [7] and presents an improved NN algorithm, which is now successfully applied to routine ALS user operation.

Though discussed in detail later in this paper, Fig. 1 illustrates the method's effectiveness, showing data from about a week of ALS user operation following its implementation in the fall of 2023. As the users continually adjust the ID parameter setpoints to accommodate their experiments (the traces in the two top graphs), the measured rms vertical beam size is seen to remain stabilized within a band that is very close to the estimated ∼0.3 μm rms noise floor (red trace in the bottom graph). For comparison, the plot also shows the inferred beam size (blue trace) that would have been observed with the NN FF system turned-off.

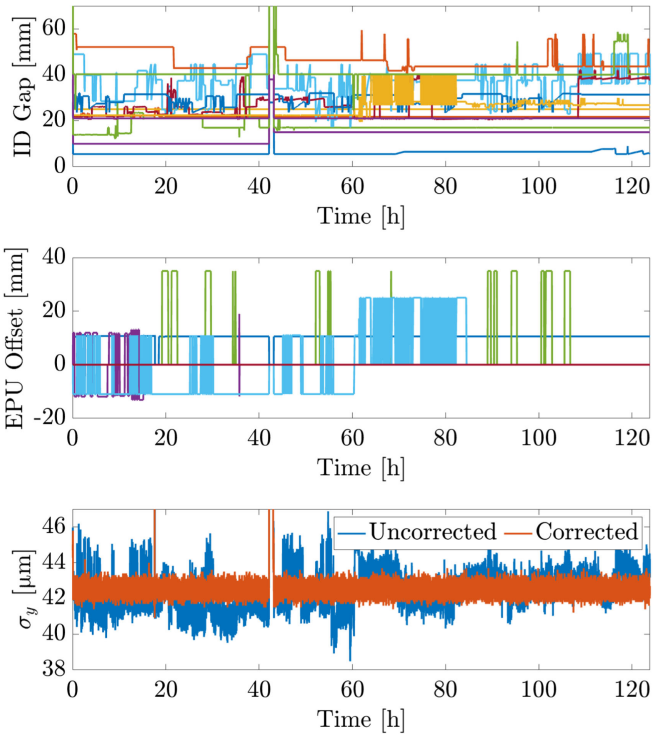


FIG. 1. Operational performance of the NN-based ID FF system during a user run starting on November 7, 2023. Shown are the vertical ID gaps (top), the elliptical polarized undulator (EPU) phase or longitudinal offsets (center), and the vertical electron beam size (bottom) as measured at ALS diagnostic beamline 3.1 (red) and as inferred (blue) if no correction had been applied as calculated using Eq. (1). One beam outage occurred at hour 42 during that 5 day window; notably, the beam size control algorithm dis- and re-engaged automatically without human intervention (see Sec. IV B for a detailed discussion).

Our approach entails a comprehensive analysis of long-term operational data, which led to the identification of methods to make the NN model more robust, including the removal of DWP from its input parameters and instead using only the ID configurations to predict beam size change.

To ensure the NN model's adaptability and performance over time, we have implemented a novel online fine-tuning technique, which is orders of magnitude faster than the previously reported method, thereby allowing for a much quicker response to changing operational conditions while the NN model continuously learns from new data. The refined algorithm can regulate the beam size to nearly the noise threshold of our measurements, achieving an enhancement of more than a factor of 4 in the signal-to-noise ratio compared to the previously reported results [7].

A distinctive feature of our system is its seamless integration with the Experimental Physics and Industrial Control System (EPICS) [8], achieved through a dedicated input/output controller (IOC) that is configured with over 600 process variables (PVs). These PVs are instrumental in providing the extensive monitoring and control capabilities necessary for the detailed manipulation of the NN-based ID FF system, ensuring that the system can respond effectively to a wide range of operational scenarios without the need for manual adjustments. A Control System Studio (PHOEBUS) [9] interface serves as the primary platform for the deployment of the tool, presenting machine operators with a highly intuitive and reliable means of interacting with the system.

This paper begins in Sec. II A with an in-depth analysis of the impact of the dispersion wave parameter on beam size. This foundational understanding informs our methodology for collecting training data, as described in Sec. II B. The process of selecting the most suitable model architecture is thoroughly investigated in Sec. II C, ensuring that the chosen model is optimally configured for the task at hand. The sensitivity of the model to the training dataset size is discussed in Sec. II D, setting the stage for the implementation of an online fine-tuning strategy, which ensures the model remains responsive to changing operational conditions, as explained in Sec. II E. Section III focuses specifically on the practical aspects of deploying our model into routine user operation. Finally, in Sec. IV, we review the performance of the system during user operation (Sec. IV A), show its robustness to beam outages

(Sec. IV B), and include a comparative analysis against an FB system in Sec. IV C, highlighting the advantages of our NN-based FF approach.

## II. MODEL DEVELOPMENT

### A. Dispersion wave parameter

The method we adopted to regulate the electron beam size leverages a common practice employed in storage-ring light sources during the lattice tuning for machine setup. The lattice tuning proceeds by adjusting the skew quadrupoles first to correct for betatron coupling and spurious vertical dispersion and then to introduce a deliberate vertical dispersion wave [10], as a means to generate vertical emittance in order to enlarge the vertical beam size in a controlled manner and thereby extend the beam lifetime. At the ALS, the excitation of the dispersion wave involves the tuning of 32 skew quadrupoles. In the context of this effort, we define the dispersion wave parameter (DWP) as the dimensionless scaling parameter that quantifies our adjustments relative to the excitation pattern of the skew quadrupoles established after completion of a machine setup.

Previous work [7] included DWP data in the training of the NN model, on the assumption that the beam size sensitivity to the DWP could depend on the specific ID configuration and be impacted by machine drifts. We tested these assumptions by carrying out repeated measurements of the beam size response to DWP excitation for various ID configurations and lattice conditions over an extended stretch of time as in Fig. 2. Specifically, this figure presents vertical beam size measurements at ALS diagnostic beamline 3.1 under a zigzag-like DWP excitation taken during accelerator physics (AP) shifts over about a month. The data were averaged over 20 beam size measurements at each DWP step. This dataset encompasses observations made both prior to, and following, a machine setup, including cycling of the magnets. The analysis shows no discernible impact of the ID configurations on the beam size sensitivity to DWP changes and no evidence of strong dependence on machine drifts, at least on the time scale of these measurements.

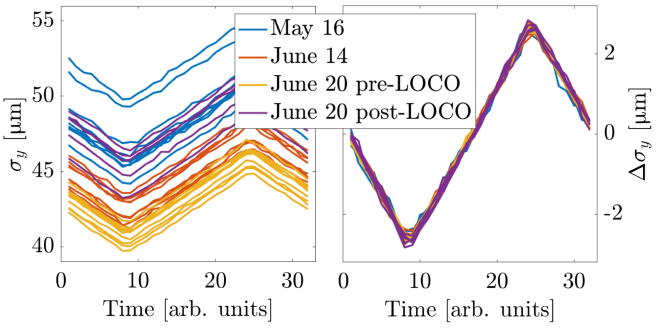


FIG. 2. Vertical beam size measurements during a zigzag-like DWP excitation in the ±0.08 range. The left graph depicts the observed beam size, monitored at various times (indicated by different color coding) over a span of about a month and encompasses several distinct ID configurations, as denoted by the multiple lines within each color group. After subtracting the time averages, the traces are seen to overlap remarkably well (right graph), indicating that the beam response to the DWP excitation is not sensitive to the ID configuration. Note that the datasets include examples of measurements taken immediately before ("pre-LOCO") and after ("post-LOCO") a machine setup.

The measurements did, however, show evidence of hysteresis effects, which had been previously overlooked. Fig. 3 displays the beam size measurements conducted during periodic zigzag-like DWP excitations with two distinct 0.12 and 0.24 amplitudes, the first being close to the value we later determined to be required by the NN FF system during user operation. In this case, the observed vertical width of the hysteresis loop is 0.15 μm on average; it is noticeable but remains significantly smaller than the 0.3 μm rms measurement noise and is going to be neglected in this study. In summary, our findings are consistent with the following linear dependence of the vertical beam size $\sigma_y$ on the DWP

$$\sigma_y(\boldsymbol{L},\boldsymbol{u},\mathrm{DWP}) \simeq \sigma_{y,0}(\boldsymbol{L},\boldsymbol{u}) + \sigma_{y,1}\cdot\mathrm{DWP}, \qquad (1)$$

where the coefficient $\sigma_{y,1}$ is essentially independent of $\boldsymbol{u}$, the vector representing all the ID parameters, and all other relevant lattice parameters $\boldsymbol{L}$ known to affect the beam size. Correspondingly, $\sigma_{y,0}$ is a function describing the relationship between the beamsize and the lattice and ID parameters

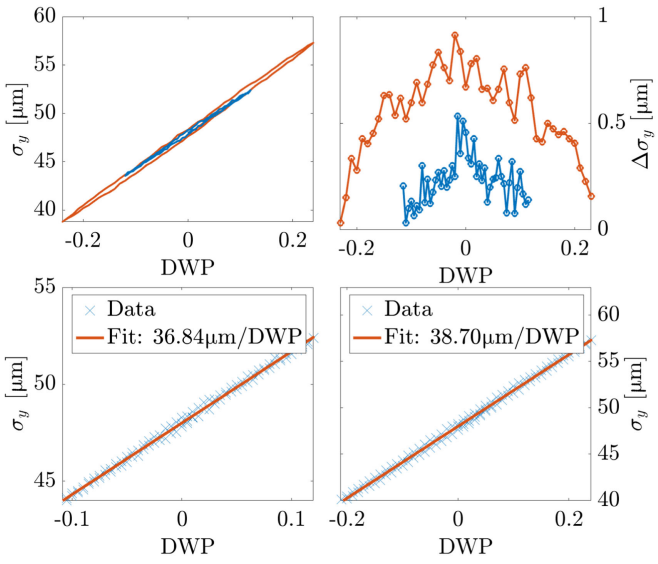


FIG. 3. Top left: hysteresis loops showing the response of the measured vertical beam size to the DWP, as the DWP undergoes a zigzag excitation with ±0.12 (blue) and ±0.24 (red) ranges. Top right: difference between upper and lower branches of the two hysteresis loops. Bottom: linear fits of the data in the two cases [see $\sigma_{y,1}$ in Eq. (1)].

$\boldsymbol{L}$ and $\boldsymbol{u}$, respectively, with DWP $= 0$. We refer to this quantity as the *uncorrected beamsize* (see also the blue line in Fig. 1). Based on these findings, it is justified to exclude the DWP data from the NN model, with an obvious benefit in terms of dimensionality reduction. As a result, dedicated machine time will be required to measure and maintain $\sigma_{y,1}$ up to date, but these measurements are very quick and they will not be required often, as the $\sigma_{y,1}$ dependence on machine drift is apparently very limited.

We should add that while neglecting hysteresis effects is presently justified, this may change if operational conditions and the available diagnostics evolve over time, causing for example a significant reduction of the measurement noise floor or a requirement for larger DWP amplitudes. In these cases, the performance of the NN FF system would likely be improved by considering a more advanced NN-model that incorporates the histories of the skew quadrupole settings (see, e.g., [11]).

### B. Training data acquisition

The training data for our study are gathered during dedicated AP shifts, with the machine, otherwise, operated as under user operation conditions. The dataset was acquired over a 12-h period, representing the maximum duration typically allocated within AP shifts.

Each ID is independently exercised to collect relevant data. For elliptically polarized undulators (EPUs), the vertical gap, horizontal offset, and polarization are considered dependent parameters and are treated as such during data acquisition. Whenever an ID reaches its designated setpoint, the next setpoint is established, allowing for the continuous acquisition of relevant data in a time efficient manner, see Fig. 4 for a graphical representation. To mitigate undue strain on the ID amplifiers, the control script incorporates brief pauses whenever a new setpoint is assigned to an ID.

The setpoints for the IDs are derived from a comprehensive analysis of records spanning 2 years of user operation. This approach ensures that the setpoints accurately represent the operational conditions and requirements. However, it is important to recognize that during regular user operation, IDs predominantly maintain their setpoints. In contrast, during AP shifts, IDs are more frequently adjusted as part of the data acquisition process in order to maximize the sampled volume in the input parameter hyperspace. Consequently, this leads to a distinguishable disparity in the data distribution between the training datasets and the conditions encountered during user operation.

Throughout the data acquisition process, all ID readback values, as well as the vertical beam size as measured at the diagnostic beamline 3.1, are recorded synchronously at a sampling rate of 10 Hz through the EPICS-based archiver appliance [12]. The dataset employed in the subsequent sections to illustrate the model development comprises a

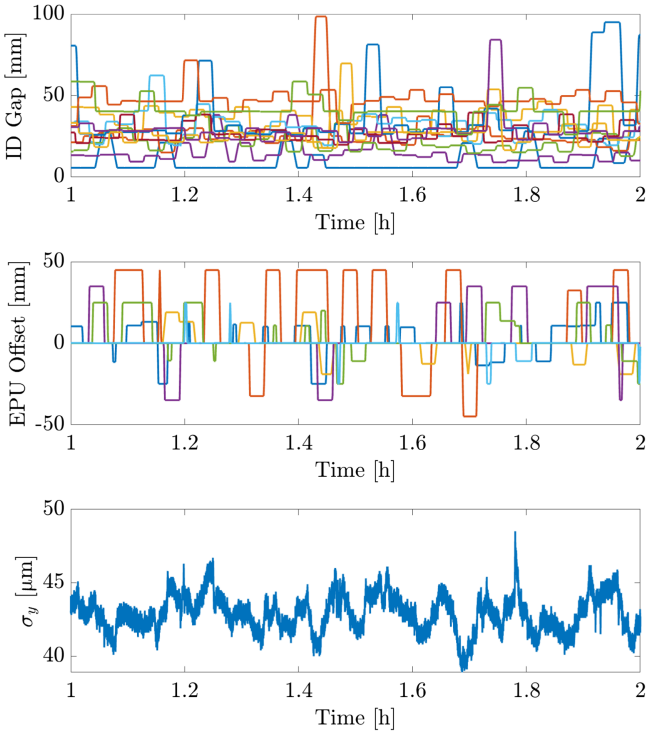


FIG. 4. Training data acquisition depicting the relationship between ID parameters and beam size fluctuations. Plotted are the vertical gaps (top), the horizontal offsets for the EPUs (center), and the vertical beam size as recorded at diagnostic beamline 3.1 (bottom). The overall training data collection lasted 12 h.

12-h recording of 27 ID parameters and the resulting changes in beam size. Overall, 432k data points were recorded for each channel.

It should be acknowledged that although the data sampling process is conceptually straightforward, the actual collection of data for this study encountered significant operational challenges. Due to the high value and limited availability of accelerator physics time, the automated collection of training data is primarily scheduled during nighttime hours. However, the nature of the ID setup, which is not designed for extensive ramping, presented frequent issues. A common scenario involved the tripping of one or several ID amplifiers, which are driving the ID gap or shift changes. These amplifiers often became unresponsive, at times effectively flat lining the corresponding ID for several hours until it was noticed by operators and the amplifier reinitialized. Additionally, the local ID skew quadrupole correctors, essential for proper local coupling correction thereby enabling tight beam size control, tripped on several occasions as well.

The impact of these trips is even more detrimental as it effectively alters the impact of the ID configuration on the vertical beam size. The skew power supply trips were caused by actual insulation damage of the skew quadrupole corrector coils on an EPU and subsequent ground fault in certain extreme gap position. Although this insulation has

been repaired and caused no power supply trips since, the possibility of similar failures has to be taken into account.

These technical setbacks posed substantial obstacles to the smooth collection of training data and required a careful measurement strategy that balances very fast sampling of a large amount of ID configurations with a very low failure rate during data collection. Notably, the implementation of watchdog mechanisms was crucial. These watchdogs are designed to provide auditory FB to operators in the event of the emergence of known issues, thereby enhancing operational oversight and response efficiency.

### C. Model selection

We conducted investigations into various models and determined that, consistent with findings in [7], multilayer perceptron (MLP) NNs [13] as a model yields the most favorable outcomes. Therefore, our subsequent discussion will focus exclusively on MLPs.

To explore and evaluate different hyperparameters sets, we employed *grid search* as automatic hyperparameter optimization algorithm [13]. Details related to the search space of each hyperparameter are outlined in Table I. Our exploration covered model architectures ranging from a simple single layer with two neurons to a complex three-layer structure with a 512-256-128 neuron configuration. We evaluated each setup with three distinct activation functions, totaling 387 different model configurations in our search.

While various approaches exist in the literature for evaluating machine learning models on training data with a given set of hyperparameters, we focused on employing *k-fold cross validation* due to its well-established statistical significance [14]. Specifically, we set the number $k$ of folds to $k = 10$. Also, the data were partitioned without shuffling to preserve chronological ordering within each partition. This consideration is critical given that our data are sampled at 10 Hz, whereas ID configurations vary on a timescale of seconds, resulting in slightly oversampled data. If the selection of the training and test samples was done randomly, it would increase the risk that the two datasets become too similar, which would undermine their use in evaluating the model's generalization capability on new data.

TABLE I. Details of the hyperparameters search space for MLP architecture tuning. To mitigate the number of experimental permutations, only network settings with decreasing number of nodes per layer are considered. Dropout and weight decay are varied only for the best performing MLP architecture.

| Hyperparameter | Search space |
|---|---|
| Number of hidden layers | $\{1, 2, 3\}$ |
| Number of neurons per layer | $\{2^n\}, 1 \leq n \leq 9$ |
| Activation function | {ReLU, Tanh, Sigmoid} |
| Weight decay | $\{10^{-n}\}, 1 \leq n \leq 5$ |
| Dropout rate | $\{0.2, 0.4, 0.6, 0.8\}$ |

For each fold, 20% of the data were randomly sampled from the training set and considered as validation set, to prevent overfitting. Before each training stage, the data are normalized using $Z$-score normalization [15], where means and standard deviations are computed over the training set. The best set of hyperparameters is selected based on the lowest average test rms error (RMSE) across the ten folds. Parameter optimization is done using Adam optimizer [16] with a learning rate set to $10^{-3}$. The number of epochs is set to 1000, and Early Stopping [13] is considered as convergence criterion with a patience of 5 epochs. The MLP is implemented using PyTorch 2.0.0, and the training was conducted on an NVIDIA GeForce RTX 2060 GPU.

The results decisively indicate that models with fewer than three layers underperformed, leading us to eliminate smaller models from consideration. We observed a performance plateau beginning at the model size of 128-64-32, with negligible gains from larger models. Consequently, we opted for this particular model size. The rationale behind this decision is outlined in Sec. II E; the incremental benefits of a larger model are overshadowed by the anticipated impact of online fine-tuning. For fine-tuning purposes, a smaller model size is preferable to expedite the convergence of training. Hence, the model size of 128-64-32 is chosen to strike an optimal balance between the performance and fine-tuning efficiency.

The Tanh activation function exhibited the most optimal test performance when averaged over the ten folds, leading to $0.64 \pm 0.03$ μm test RMSE. With training and validation RMSE of 0.30 μm, which aligns with the noise level of the beam size measurement, it suggests that this can be considered a well-fitting model for the underlying training data.

Then, weight decay [14] and dropout technique [17] were introduced to explore potential improvements in the generalization capabilities of the best architecture identified in the aforementioned stage. Their respective values (i.e., amount of weight decay and dropout probability) underwent further optimization through a grid search approach with tenfold cross validation, the details of which are provided in Table I. Results are shown in Fig. 5. The optimal configuration was found to have a weight decay of $10^{-4}$ and dropout probability $p = 0.2$, leading to an averaged test RMSE of $0.58 \pm 0.06$ μm.

### D. NN model performance vs dataset size

In this section, we investigate the dependence of the model performance on the training dataset size. The training data utilized for this purpose are as detailed in Sec. II B. It is important to remark that the dataset exhibits a degree of oversampling, which implies that employing a straightforward random partitioning strategy for dividing the data into training, testing, and validation sets would result in highly

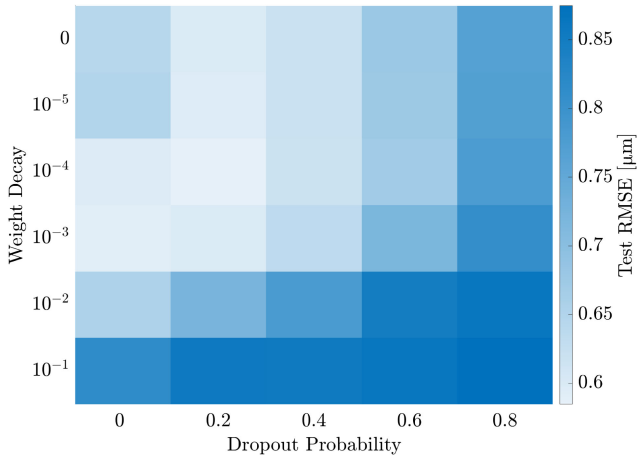


FIG. 5. Graphical representation of tenfold cross-validation grid search results for the optimization of the weight decay and dropout probability hyperparameters. For each configuration, the test performance averaged across the ten folds is shown.

similar subsets. As a result, a more nuanced method for data partitioning is required.

Our approach starts by dividing the initial 12-h acquisition dataset into nonoverlapping, 1-h test segments. Random samples of these segments are then used to evaluate every training set. The training set is sampled chronologically from the remaining 11 h, with its size incrementally expanded from 1 to 11 acquisition hours for the analysis in a bootstraplike approach [14]. In particular, for every specified size of the training dataset, we draw five individual training sets from the overall available 11-h periods. The results on predicting the test dataset for each training size are then averaged across all test hours to derive the overall performance metric.

The outcomes of our analysis are shown in Fig. 6. Initially, the model demonstrates a promising enhancement in predicting beam size variations due to the rapidly

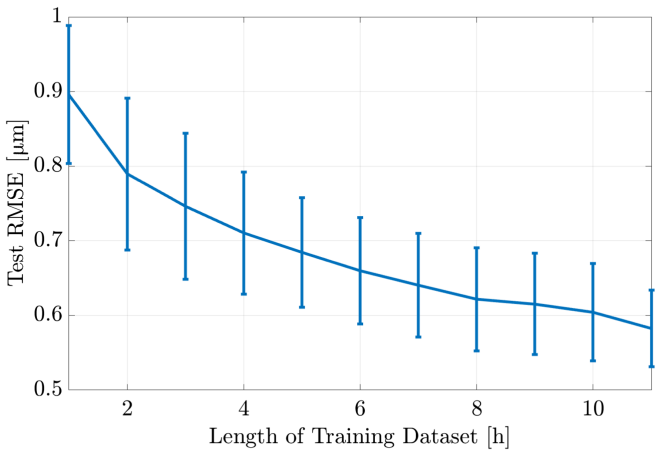


FIG. 6. Convergence analysis of model performance over varying durations of training data size, up to 11 h, using bootstraplike sampling (five samples).

changing ID configurations within the allocated machine time. However, model convergence quickly begins to decelerate, suggesting that to achieve prediction performance at the noise floor of 0.3 μm, an increase in training data acquisition time by around 50 h would be necessary.

We can conclude that the duration of a standard 12 h shift is inadequate for thoroughly probing the parameter space that would be necessary for training a model that is capable of predicting beam size variations with precision down to the noise floor. However, extending the data acquisition time to on the order of 50 h poses significant operational challenges, rendering such an approach impractical within the constraints of available machine time.

### E. Online fine-tuning

In addressing the complexities of the vast parameter space outlined in our study, the training of a base model that can independently and effectively predict the beam size emerges as a formidable challenge. While we are optimistic about attaining this level of predictive proficiency in the future, current limitations compel us to adopt an alternative methodology that continually adapts the model during user operation.

During the proof of principle studies [7], such adaptation has been applied through online retraining by integrating the original training data with a randomly downsampled segment from the ongoing user run, subsequently continuing the training of the active model. However, with hundreds of thousands of samples in the dataset, retraining demanded approximately 15 min to complete on a CPU, which significantly limits the model's reactivity to changes in the ID configuration space.

However, these challenges are widely recognized in the deep learning community. Training a network from scratch requires substantial computational power, memory resources, and large training datasets [18]. Gathering large high-quality training datasets is usually the most felt and challenging task [19], as training a network with a small dataset frequently results in over fitting issues [20]. Fine-tuning [13,21] is a widely used solution for addressing these challenges. The core of this approach lies in the realization that rapid and efficient adaptation to new data—achieved through the adjustment of NN weights—can greatly enhance the model's predictive accuracy while circumventing the extensive data processing typically necessitated by comprehensive training phases [22]. Fine-tuning has been widely used in various fields, such as computer vision [19,23–26] and natural language processing [27–29], where the requirements for both the size of training data and the computational power for training are notably high.

In our approach to fine-tuning the model, we exclusively utilize data acquired during the current user run, which is stored in a first-in-first-out buffer, opting not to incorporate the original training data. To safeguard against any potential runaway scenarios, each training cycle commences with

the original base model. This strategy effectively anchors the model, ensuring it remains closely aligned with our dedicated training dataset, thereby maintaining stability and reliability in the model's predictive performance.

Our observations indicate that a buffer size on the order of 1000 samples optimally balances reduced noise sensitivity with prompt responsiveness to machine conditions poorly predicted by the initial model. During fine-tuning, we generally set 20% as validation set, employ a learning rate of $10^{-3}$ and a cap of 1000 epochs, incorporating an early stopping mechanism activated after 5 epochs without improvement. Typically, convergence is achieved well below 100 epochs, with each epoch taking less than 0.01 s on a CPU. This efficiency results in a model update frequency that typically exceeds 1 Hz. Upon completion of fine-tuning, the actively deployed model for predicting beam size variations is replaced with the newly trained model, initiating a new cycle of fine-tuning from the start.

An ostensible difficulty is that to fine-tune the NN during operation, one would need to know the uncorrected beam size data, whereas only measurements of the beam size after correction are available, since the NN FF system is always active. One method to overcome this difficulty is to make use of Eq. (1) to derive the presumed uncorrected beam size $\sigma_{y,0}$ from knowledge of the measured corrected value $\sigma_y$ and the current DWP read-back value.

It is worth highlighting that the online fine-tuning method employed in our study essentially functions as a form of FB. By adjusting the amount of data in the fine-tuning buffer, we control the noise level in the data, and by tuning the fine-tuning hyperparameters, we modulate the model's responsiveness to new data. The balance between FF and FB elements in such an elegant way is a key feature of our ID compensation algorithm that contributes significantly to its robustness.

While this framework yields very good performance as shown in the following sections, the introduction of FB components into what was previously a pure FF framework introduces new challenges. Specifically, the system must contend with perturbations of the beam size measurement that can now affect the stability of the beam size. As an example, it has become necessary to introduce a region of interest FB mechanism on the CCD crystal of the diagnostic beamline camera. This adjustment ensures the accuracy of the algorithm responsible for calculating beam size from the captured images as the beam wanders on the crystal due to slow drift. Additionally, we verified that the beam size remains consistent regardless of the crystal's impact point, confirming its independence from the beam's position on the crystal.

Additionally, external factors outside our control, such as camera malfunctions, disturbances near the diagnostic beamline table causing vibrations, and beam blowups due to multibunch instabilities, can impact our operations. Our current approach uses a first-in-first-out buffer with a brief duration of approximately 2 min. This approach not only facilitates rapid updates to the model but also ensures that incidents severely impacting beam size measurements only temporarily influence the FF system, thus effectively mitigating potential problems from FB components. Moving forward, we plan to explore the potential of employing anomaly detection methods, such as autoencoders [13,30] to identify anomalies in beam size measurements. This advancement would allow for a substantial increase in buffer size. By utilizing a more compact NN architecture and expanding the buffer for online fine-tuning, we might achieve superior performance compared to our current methodology, all while preserving a rapid model update rate.

In summary, despite facing certain challenges, the advantages of our online fine-tuning strategy are substantial, far outweighing the trade-offs. This method offers exceptional predictive accuracy and operational flexibility, making it highly suitable for deploying our NN-based FF system in a production environment. This is especially noteworthy given the training data constraints highlighted in the previous section.

## III. NN FF SYSTEM DEPLOYMENT

In this section, we discuss specific design choices that were crucial for the successful deployment of our system in day-to-day user operation.

### A. Architecture

The Python backend, which handles the actual machine learning part, communicates to and from an EPICS IOC, equipped with 600 PVs, with a PHOEBUS panel serving as the user interface for controlling the tool. This interface ensures real-time interaction between the system's backend calculations and the user-operated frontend, facilitating the dynamic control of the ID feed forward process within the standard framework utilized in the ALS control system. A graphical representation is shown in Fig. 7.

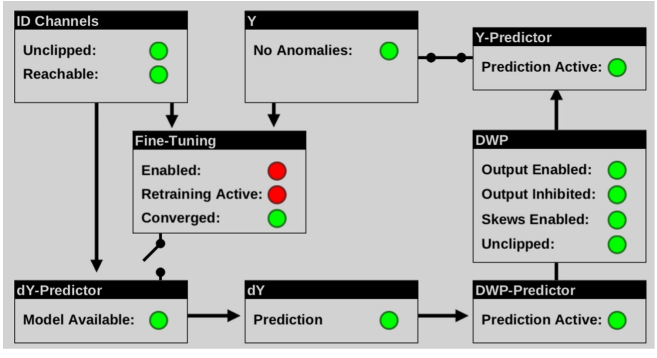


FIG. 7. Screenshot of the PHOEBUS control panel depicting the workflow during operation. In this example, the beam size correction is active while online fine-tuning is not active.

The Python backend operates at a frequency of 10 Hz, executing a series of tasks in a continuous main loop without interruptions. For enhanced flexibility, it is currently implemented in Python and operates on a virtual machine within the control system framework. Looking ahead, we plan to transition the machine learning backend to a dedicated IOC in the near future.

Each iteration initiates by reading the model inputs and control parameters (such as the main loop frequency or fine-tuning parameters) from the PVs, followed by a preprocessing step wherein the procedure involves (i) verifying the readings for any errors and (ii) implementing $Z$-score normalization [15] on the model inputs. In case of a PV read error (for example, because a PV is not responding within the timeout limit), the variable is assigned the last recorded valid value.

With reference to the workflow panel in Fig. 7 and notation of Eq. (1), the following steps outline the backend data processing:

(i) The accelerator is in some state given by the lattice $\boldsymbol{L}$ and the ID configuration $\boldsymbol{u_1}$, with measured beam size $\sigma_y = Y$.

(ii) Users request a change to the ID parameters to $\boldsymbol{u_2}$.

(iii) The $dY$-predictor outputs an expected beam size change $dY = \sigma_y(\boldsymbol{L}, \boldsymbol{u_2}, 0) - \sigma_y(\boldsymbol{L}, \boldsymbol{u_1}, 0) = \sigma_{y_0}(\boldsymbol{L}, \boldsymbol{u_2}) - \sigma_{y_0}(\boldsymbol{L}, \boldsymbol{u_1})$ using the current NN model.

(iv) The DWP predictor calculates the change in DWP needed to keep the beam size constant using Eq. (1), $\Delta\mathrm{DWP} = dY/\sigma_{y,1}$.

(v) If requested by the control system and no exceptions are found (see Sec. III B), $\Delta$DWP is applied to the machine.

(vi) The uncorrected beamsize needed for online fine-tuning is calculated from the measured beamsize $\sigma_y(\boldsymbol{L}, \boldsymbol{u_2}, \mathrm{DWP})$ using Eq. (1) and the DWP read-back value as $\sigma_y(\boldsymbol{L}, \boldsymbol{u_2}, 0) = \sigma_y(\boldsymbol{L}, \boldsymbol{u_2}, \mathrm{DWP}) - \sigma_{y,1} \cdot \mathrm{DWP_{rbv}}$, where $\mathrm{DWP_{rbv}}$ is the DWP read-back value as described below.

(vii) The ID configuration $\boldsymbol{u_2}$ and uncorrected beamsize $\sigma_y(\boldsymbol{L}, \boldsymbol{u_2}, 0)$ are used as input for the online fine-tuning.

(viii) After each fine-tuning cycle, the current NN is updated, resulting in an updated $dY$-predictor.

As mentioned above, the model input parameters are stored in a first-in-first-out buffer to allow the online fine-tuning at user-defined time intervals. We opt to integrate this buffer using Python to enhance flexibility during the initial deployment phase, given that EPICS PVs suitable for storing such a buffer cannot alter their length without rebooting the IOC.

A dedicated thread is executed for fine-tuning the original base model, ensuring simultaneous operation with the primary loop and allowing for the uninterrupted application of the model in predictions. Finally, after the fine-tuning convergence criteria are reached, the model used in the main loop is replaced with the new fine-tuned model.

The read-back value of the DWP is recorded at the EPICS level. To facilitate the calculation of DWP values for each skew quadrupole, the golden values have been stored as EPICS PVs. This allows deriving DWP values from the actual power supply read-back values in combination with the corresponding golden values.

It is worth noting that a critical design decision was to centralize all control logic within EPICS instead of dispersing it between PHOEBUS or the Python backend. This choice was informed by EPICS's superior speed and its archival capabilities, ensuring that all changes and updates remain swift and traceable. Such centralization bolsters the system's robustness and flexibility, allowing for all logic changes to be managed and implemented within a singular, regulated framework, thereby streamlining system updates and maintenance. An example is the inhibitor chain as described in the following section.

### B. Inhibitor chain

The implementation of a dedicated EPICS-based inhibitor chain is integral to the secure deployment of the NN-based FF system during user operation at the facility. This inhibitor chain, by design, does not initiate the NN-based FF but acts as a safeguard, preventing closing of the FF loop (i.e., adjusting skew quadrupole magnet setpoints) unless specific operational criteria are satisfied, see Fig. 8 for a graphical representation.

The chain's conditions are outlined as follows: the minimum beam current must be met, the fast orbit FB (FOFB) system must be operational, the wiggler must be in a closed position (as a proxy for accelerator in user

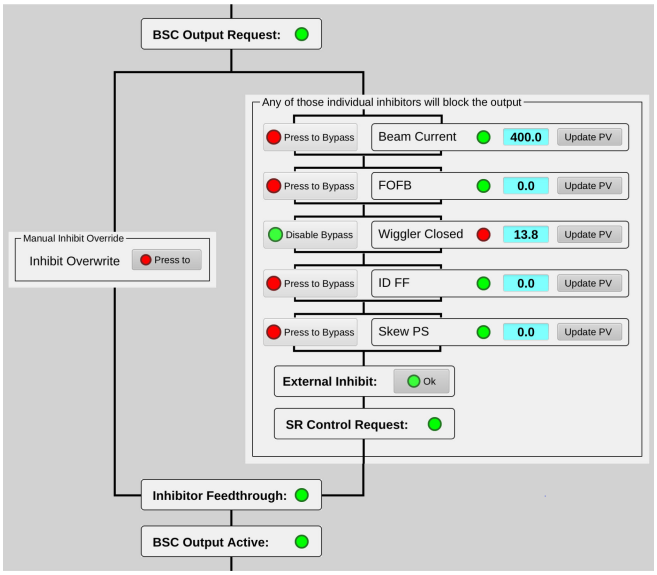


FIG. 8. Screenshot of the PHOEBUS inhibitor pop up indicating which channels of the inhibitor chain are enabled (green) or disabled (red). Manual override options allow for flexible use in specialized experimental environments.

operation conditions), the local ID FFs need to be active, and the skew quadrupole power supplies must be functioning in their regular capacity. For each condition, there is a provision for manual override, allowing individual components of the inhibitor chain to be selectively bypassed if necessary. Moreover, the entire inhibitor framework is designed with an option for complete deactivation, offering flexibility to accommodate specific scenarios, such as during specialized experimental runs, thus providing a balance between stringent safety protocols and operational convenience. A figurative example of the inhibitor chain in operation during a beam outage is shown later on in Fig. 11.

## IV. NN FF SYSTEM PERFORMANCE

In this section, we evaluate the capabilities of the neural network-based ID FF algorithm to stabilize the vertical beam size at the ALS. As detailed in [7], scanning transmission x-ray microscopy (STXM) [31] beamlines are very sensitive to variations of the transverse photon distribution, and the quality of their experiments can be significantly impacted by such fluctuations.

Our measurements have established a linear relationship between variations in the vertical beam size at the diagnostic beamline 3.1 and subsequent changes in intensity observed in STXM scans taken at beamline 5.3.2.2 (consistent with resulting vertical beam size changes being driven by perturbations of the vertical dispersion). Specifically, we observed that a 10% change in the vertical beam size resulted in approximately a 9% intensity change in the STXM scan. Prior to the deployment of our beam size correction algorithm, such fluctuations have been common during user operations over the duration of a STXM measurement, representing a tenfold enhancement relative to the intrinsic noise floor of this STXM beamline.

### A. Performance during user operation

At the time of writing, the NN-based FF system had been continuously operational for 2 months. The performance over about 1 week was showcased in the introduction in Fig. 1 and was typical. The vertical beam size stability has been remarkably consistent, to within an average of 0.32 μm rms per user run (or 0.75%), closely approaching the measurement noise floor at 0.3 μm rms.

This can be seen in more detail in Fig. 9: data points to the right of the shaded area, where the 2-month period with operating NN FF system is segmented into seven uninterrupted user operation intervals. For each interval, we report the rms beam size fluctuations before correction (blue), and as corrected using the base model without fine-tuning (red), and finally as measured with correction by the fine-tuned system (crosses). The blue and red data points are inferred quantities; specifically, the blue data were obtained by subtracting the contribution due to the DWP adjustments from the measurement of the stabilized vertical

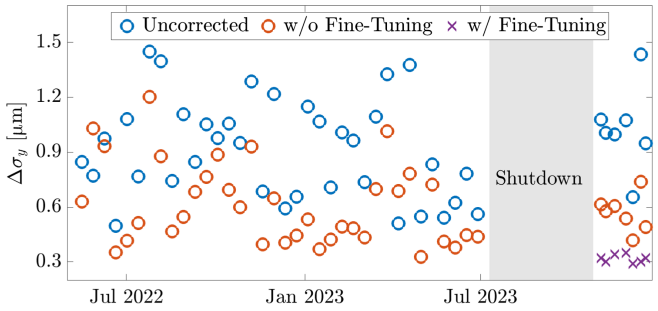


FIG. 9. Data on the right of the shaded area are with the NN FF system fully deployed: crosses represent the vertical beam size rms fluctuations as corrected by the fine-tuned NN model and measured. They are compared to the uncorrected (blue circles) and partially corrected (red circles) beam size fluctuations, the latter representing the correction made by the NN FF system without fine-tuning; the values for both of these datasets are inferred estimates (see body text). Each data point is a time average over about 1 week of user operation. The data points on the left of the shaded area represent a backward-test of the NN model based on archived data.

beam size. Barring the small hysteresis effects not accounted for in our simplified model (see Sec. II A), we believe that the uncorrected beam size so calculated should be a fairly accurate estimate of the actual beam size that would have been observed without correction. Note that in this figure, the data points are time averages during the operation period (about a week).

The red data points preceding the shaded area are the result of a study, in which the NN model trained on June 2024 was retroactively applied to archived data of beam and ID parameters from the preceding year.

It should be noted that at the time, the ALS archived data had two important limitations. First, the vertical beam size in the absence of closed IDs was neither routinely measured nor archived, meaning that only relative beam size changes could be evaluated (therefore, in the figure, the preshutdown data points represent rms deviations from the average beam size measurement over a week). Second, the downsampling of the long-term storage of the PVs had the unfortunate consequence of causing a loss of synchronization between the ID configuration and beam size measurement data streams, with time errors up to 20 s (incidentally, for these reasons, using archived data for training a base model proved unfeasible).

In spite of these limitations, the backward-test results are instructive. One can observe that the distributions of the uncorrected and corrected (without fine-tuning) data points both before and after the shutdown appear to remain roughly similar over time, suggesting that whatever machine drifts may have been present they did not compromise the accuracy of the NN model significantly. This suggests two positive practical consequences: there is likely no need to refresh the base model frequently and it could be possible to improve accuracy by accumulating data gathered from user operation over extended periods of time.

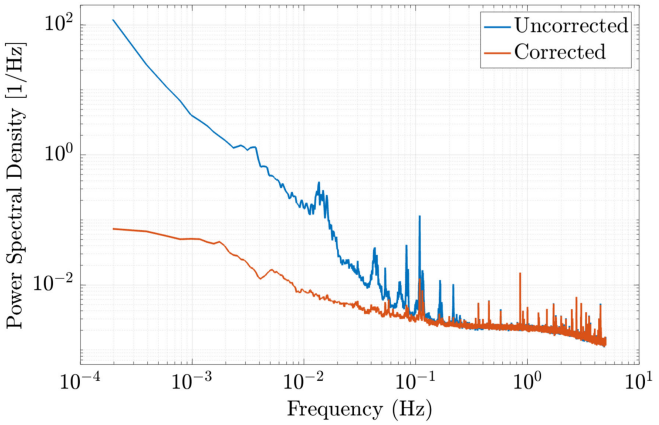


FIG. 10. Frequency spectrum of the vertical beam size during 760 h of user operation with fully functional NN FF system (red) and the inferred beam size without correction (blue). The features at around 0.1 Hz are attributed to EPUs, the spikes above 1 Hz are associated to beam injection transients during top off.

Additional insight can be gained by data analysis in the frequency domain. In Fig. 10, a discrete Fourier transformation was applied to analyze a combined total of 760 h of user operation data with the NN FF system on. The graph confirms that the system is effective over a broad range of frequencies. The spikes observed at approximately 0.1 Hz, linked to EPU phase switching, highlight areas of potential further improvement, likely to be achieved with the acquisition of better training data. The loss of correction effectiveness seen at the lower end of the frequency range is also likely to be impacted by the EPU phase switching. While the switching occurs on a time scale on the order of seconds, this repetitive motion often extends over many hours or even days, resulting in a low frequency modulation of the data. However, the total integrated spectral power within in the lower frequency range is small, minimally impacting the system's overall performance.

This context emphasizes the robust performance of the NN-based FF system across the evaluated time frame, demonstrating its effectiveness in maintaining beam stability despite the perturbations caused by EPU switching.

### B. Recovery after beam outage

During the reported 2-month operation period, the facility experienced 12 instances of beam outages, each followed by subsequent recovery. The intervals between these events ranged from a few minutes to several hours. Notably, in each instance, the NN-based FF disengaged at the trip and autonomously closed its loop again shortly before the accelerator resumed user operation. This seamless re-engagement occurred without the need for manual intervention, highlighting the algorithm's robust predictive capabilities and its substantial contribution to operational automation.

An example of a beam outage event, caused by an rf power trip, followed by a machine refill and closing of the

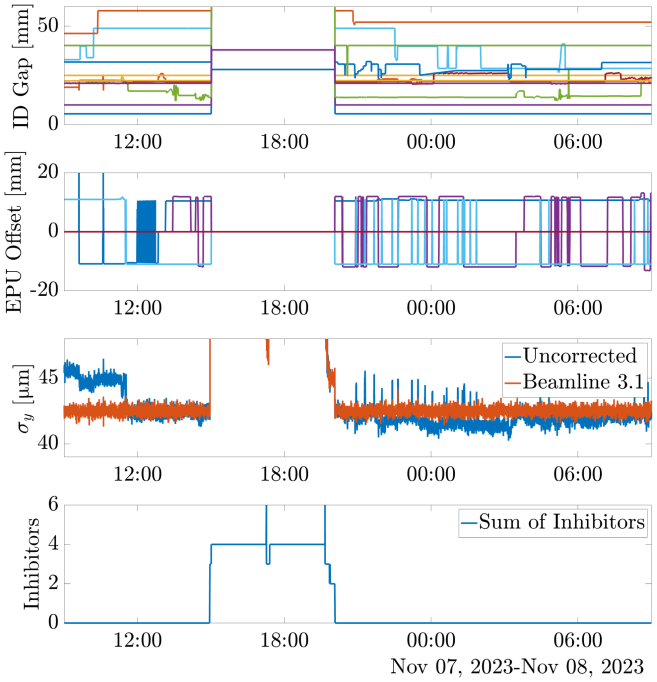


FIG. 11. Example of beam dump with subsequent restart of the NN-based FF without human intervention. The top two plots show ID gaps and EPU offsets, respectively. The third plot shows the vertical beam size and the lower plot the sum of the currently active inhibitor PVs.

NN-based FF loop without human intervention, is shown in Fig. 11. The beam was lost at 14:57, immediately triggering three of the six inhibitor PVs designed to prevent the NN-based FF from acting on the skew quadrupoles under conditions that are not operationally safe and reliable. At 17:15, during the process of reloading the lattice, the skew quadrupole power supplies exhibited transient conditions, as indicated by the activation of all six inhibitor PVs. Following this, the machine was refilled. However, a subsequent rf fault caused another beam loss. The machine was successfully filled at 19:56, which was then followed by the closure of the ID gaps. From 20:03 onwards, all conditions for closing the FF loop were met, and skew quadrupole corrections were one again applied. This is evidenced by the vertical beam size returning to its target value of 42.5 µm.

### C. Benchmark against a conventional FB system

While attractive in their conceptual simplicity, the good performance of FB control systems typically extends only to a limited frequency range. At low gain settings, these systems tend to exhibit a sluggish response, insufficient to swiftly counteract the fast perturbations caused by rapid changes in the ID configurations. Conversely, at high gain settings, there is an increased susceptibility to noise amplification. The FB mechanisms, in attempting to correct for beam position or beam size fluctuations, can inadvertently elevate the noise within certain frequency

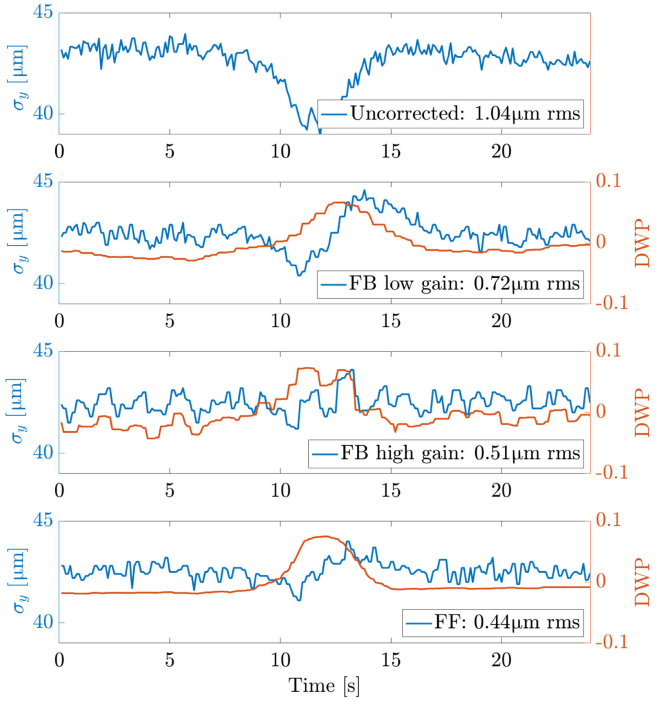


FIG. 12. Vertical beam size (blue) and DWP value (red) during four instances of a phase shift of EPU7-2. Shown are the case without correction (top), with low-gain FB active (second from top), with high-gain FB active (third from top), and with NN-based FF correction (bottom).

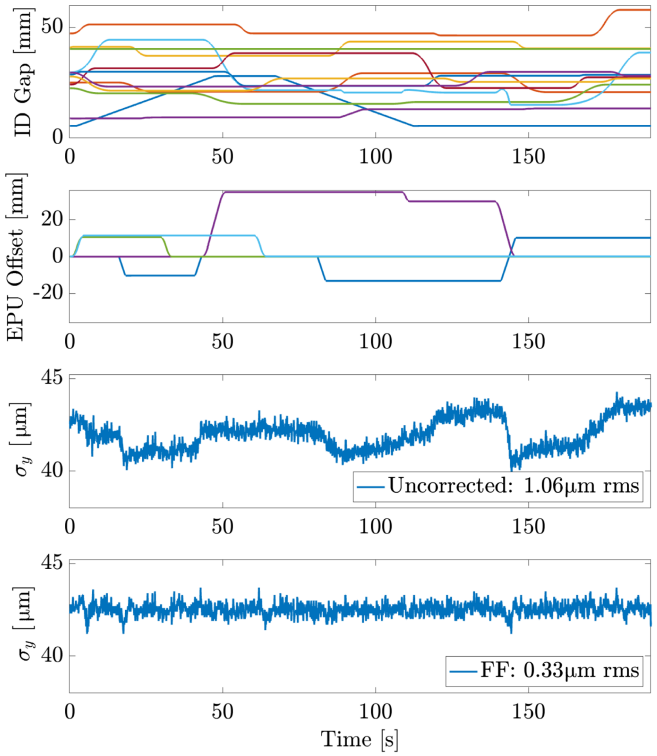


FIG. 13. Illustration of the experimental setup for quantitative evaluation of the beam size correction. The top two plots show the vertical ID gaps and horizontal EPU offsets, respectively. The two bottom plots show the uncorrected beam size and the beam size while running the NN-based FF correction.

bands, thereby exacerbating beam instability rather than mitigating it.

The experimental results presented in Fig. 12 exemplify such limitations, particularly during instances of fast EPU phase shifts, in this example EPU 7-2 (70 mm period, 20 eV–1 keV) from +30 mm to −30 mm. The topmost trace demonstrates the beam size variation without correction, with an rms deviation of 1.04 μm. Subsequent traces illustrate the performance with low and best-performing FB gains, achieving rms deviations of 0.72 and 0.51 μm, respectively, indicating improved stabilization yet not optimal, especially when considering the amplification of noise as seen in the DWP variation. Remarkably, the implementation of the NN-based FF correction (bottom trace) results in a superior rms deviation of 0.44 μm with a smooth DWP.

A quantitative comparison of various FB gain settings and the FF correction is achieved by creating a reproducible scan of ID configurations as depicted in Fig. 13. Drawing from a previously described table of ID setpoints accumulated over 1 year of user operation, five random settings for each ID were selected. This cycling procedure allowed for a dynamic yet controlled environment: each ID would transition to a new setpoint every 4 s. The experiment is initially conducted without any correction to establish a baseline, followed by iterations that included FB with varying gain settings to assess the FB system's performance. Finally, the NN-based FF correction is applied, providing a direct comparative measure of its efficacy against the FB mechanism.

Results are shown in Fig. 14. In this specific pattern of ID configurations, it is feasible to adjust the FB gain to achieve a level of performance similar to that of the NN-based FF correction. However, the FF approach

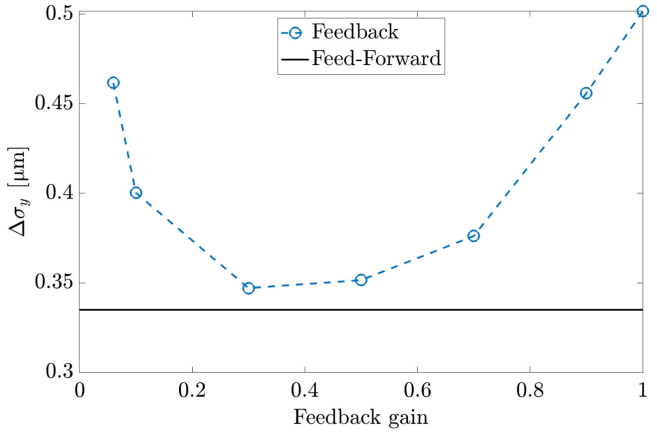


FIG. 14. Vertical rms beam size variation during the controlled setup (see Fig. 13) for various FB gains (blue) and the performance of the NN-based FF correction (black).

consistently yields favorable outcome. It is important to highlight that no FB configuration was able to match the FF results during rapid EPU transitions, as shown in Fig. 12. These frequent and significant EPU alterations, which greatly affect the beam size, are a routine aspect of regular user operation and can persist for extended periods.

## V. SUMMARY AND CONCLUSIONS

In this paper, we detailed the creation and implementation of a NN-based FF algorithm designed to stabilize the vertical electron beam size at the ALS against ID perturbations.

We have outlined our model development process, initiated with the reassessment of existing premises, which guided us to a model architecture that increased both simplicity and accuracy. We have documented the data preparation protocol, the careful selection of the model through hyperparameter tuning, its validation against historical data, the integration of online fine-tuning features, and its demonstrably enhanced performance over traditional FB mechanisms in a variety of operational settings.

This algorithm has operated continuously for 2 months of user operation at ALS with minimal human intervention, even amidst beam outages. Throughout this period, the variation in beam size remained nearly indistinguishable from the measurement noise.

The successful deployment of this NN-based ID FF system represents a significant milestone in the operational enhancement of synchrotron light sources. It not only demonstrates the feasibility of employing advanced machine learning techniques in routine user operation within the complex environment of particle accelerators, but also sets a precedent for the utilization of such technologies to improve the performance and reliability of these critical research facilities by highlighting the full lifecycle of deep learning models, including retraining and continuous operation.

## ACKNOWLEDGMENTS

We extend our sincere thanks to Changchun Sun for his assistance with the diagnostic beamline operations and setup. We are also grateful to Miroslaw Dach and Gregory Portmann for their expert guidance and support with the control systems. Our discussions with Fernando Sannibale and Erik Wallen have been insightful and have significantly contributed to the refinement of our research approach. Additionally, one of our authors, Andrea Pollastro, wishes to express his gratitude to the ALS Accelerator Physics Group for their hospitality and support during his stay at Lawrence Berkeley National Laboratory (LBNL). This work supported by the Director of the Office of Science of the U.S. Department of Energy under Contract No. DEAC02-05CH11231.

[1] L. Emery and M. Borland, Top-up operation experience at the advanced photon source, in *Proceedings of the Particle Accelerator Conference, Dallas, TX, 1999* (IEEE, New York, 1999), p. 200.

[2] C. Steier *et al.*, Commissioning and user operation of the ALS in top-off mode, in *Proceedings of the 23rd Particle Accelerator Conference, Vancouver, Canada, 2009* (IEEE, Piscataway, NJ, 2009), p. 1183.

[3] M. Böge, Achieving sub-micron stability in light sources, in *Proceedings of the 9th European Particle Accelerator Conference Lucerne, 2004* (EPS-AG, Lucerne, 2004).

[4] M. Aiba, P. Beaud, M. Böge, G. Ingold, B. Keil, A. Lüdeke, N. Milas, L. Rivkin, Á. Saá Hernández, T. Schilcher, V. Schlott, and A. Streun, SLS: Pushing the envelope based on stability, Synchrotron Radiat. News **26**, 4 (2013).

[5] T. Hellert, S. Borra, M. Dach, B. Flugstad, T. Ford, S. C. Leemann, H. Nishimura, S. Omolayo, G. Portmann, T. Scarvie, C. Steier, C. Sun, M. Venturini, E. Wallén, J. Weber, and F. Sannibale, Status of the advanced light source, in *Proceedings of the 15th International Particle Accelerator Conference, IPAC-2024, Nashville, TN* (JACoW, Geneva, Switzerland, 2024), https://www.jacow.org/ipac2024/pdf/TUPG37.pdf.

[6] C. Steier, S. Prestemon, D. Robin, E. Schlueter, and A. Wolski, Study of row phase dependent skew quadrupole fields in Apple-II type EPUs at the ALS, in *Proceedings of the 9th European Particle Accelerator Conference, Lucerne, 2004*, (EPS-AG, Lucerne, 2004).

[7] S. Leemann, S. Liu, A. Hexemer, M. Marcus, C. Melton, H. Nishimura, and C. Sun, Demonstration of machine learning-based model-independent stabilization of source properties in synchrotron light sources, Phys. Rev. Lett. **123**, 194801 (2019).

[8] L. Dalesio, J. Hill, M. Kraimer, S. Lewis, D. Murray, S. Hunt, W. Watson, M. Clausen, and J. Dalesio, The experimental physics and industrial control system architecture: Past, present, and future, Nucl. Instrum. Methods Phys. Res., Sect. A **352**, 179 (1994).

[9] K. Shroff, T. Ashwarya, T. Ford, K. Kasemir, R. Lange, and G. Weiss, Phoebus tools and services, in *Proceedings of the 19th Biennial International Conference on Accelerator and Large Experimental Physics Control Systems Workshops, ICALEPCS-2023, Cape Town, South Africa* (JACoW, Geneva, Switzerland, 2023), TUSDSC08.

[10] J. Breunlin, S. Leemann, and Å. Andersson, Improving Touschek lifetime in ultralow-emittance lattices through systematic application of successive closed vertical dispersion bumps, Phys. Rev. Accel. Beams **19**, 060701 (2016).

[11] R. Roussel, A. Edelen, D. K. D. Ratner, J. Gonzalez-Aguilera, Y. Kim, and N. Kuklev, Differentiable Preisach modeling for characterization and optimization of particle accelerator systems with hysteresis, Phys. Rev. Lett. **128**, 204801 (2022).

[12] M. Shankar, M. Davidsaver, M. Konrad, and l. Li, The EPICS archiver appliance, in *Proceedings of the 15th International Conference on Accelerator and Large Experimental Physics Control Systems, ICALEPCS-2015, Melbourne, Australia* (JACoW, Geneva, Switzerland,

2015), WEPGF030, https://accelconf.web.cern.ch/DOI/ICALEPCS2015/JACoW-ICALEPCS2015-WEPGF030.html.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).

[14] T. Hastie, R. Tibshirani, J. Friedman, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, Berlin, Germany, 2009).

[15] D. Singh and B. Singh, Investigating the impact of data normalization on classification performance, Appl. Soft Comput. **97**, 105524 (2020).

[16] D. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. **15**, 1929 (2014), http://jmlr.org/papers/v15/srivastava14a.html.

[18] R. Patel and A. Chaware, Transfer learning with fine-tuned MobileNetV2 for diabetic retinopathy, in *Proceedings of the 2020 International Conference for Emerging Technology, INCET-2020, Belgaum, India* (IEEE, New York, 2020), pp. 1–4.

[19] F. Radenović, G. Tolias, and O. Chum, Fine-tuning CNN image retrieval with no human annotation, IEEE Trans. Pattern Anal. Mach. Intell. **41**, 1655 (2018).

[20] S. Mohammadian, A. Karsaz, and Y. Roshan, Comparative study of fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening, in *Proceedings of the 2017 24th National and 2nd International Iranian Conference on Biomedical Engineering, ICBME-2017, Tehran, Iran* (IEEE, New York, 2017), pp. 1–6.

[21] C. Bishop and H. Bishop, *Deep Learning: Foundations and Concepts* (Springer, Cham, 2023), 10.1007/978-3-031-45468-4.

[22] Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris, SpotTune: Transfer learning through adaptive fine-tuning, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2019), pp. 4805–4814.

[23] E. Cetinic, T. Lipic, and S. Grgic, Fine-tuning convolutional neural networks for fine art classification, Expert Syst. Appl. **114**, 107 (2018).

[24] Z. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, Brain tumor classification for MR images using transfer learning and fine-tuning, Comput. Med. Imaging Graphics **75**, 34 (2019).

[25] R. Roslidar, K. Saddami, F. Arnia, M. Syukri, and K. Munadi, A study of fine-tuning CNN models based on thermal imaging for breast cancer classification, in *Proceedings of the 2019 IEEE International Conference on Cybernetics and Computational Intelligence, Cybernetics-Com, Banda Aceh, Indonesia* (IEEE, New York, 2019), pp. 77–81.

[26] Y. Kaya and E. Gürsoy, A MobileNet-based CNN model with a novel fine-tuning mechanism for COVlID-19 infection detection., Soft Comput. **27**, 5521 (2023).

[27] J. Devlin, M. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Association for Computational Linguistics, Minneapolis, Minnesota, 2019), pp. 4171–4186.

[28] N. Ding, Y. Qin, G. Yang, F. Wei, Z. Yang, Y. Su, S. Hu, Y. Chen, C. Chan, W. Chen, J. Yi, W. Zhao, X. Wang, Z. Liu, H. Zheng, J. Chen, Y. Liu, J. Tang, J. Li, and M. Sun, Parameter-efficient fine-tuning of large-scale pre-trained language models, Nat. Mach. Intell. **5**, 220 (2023).

[29] J. Howard and S. Ruder, Universal language model fine-tuning for text classification, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Association for Computational Linguistics, Minneapolis, Minnesota, 2018), pp. 328–339.

[30] A. Pollastro, G. Testa, A. Bilotta, and R. Prevete, Semi-supervised detection of structural damage using variational autoencoder and a one-class support vector machine, IEEE Access **11**, 67098 (2023).

[31] T. Feggeler, A. Levitan, M. Marcus, H. Ohldag, and D. Shapiro, Scanning transmission X-ray microscopy at the Advanced Light Source, J. Electron Spectrosc. Relat. Phenom. **267**, 147381 (2023).