Title
Demonstration of Machine Learning-Based Model-Independent Stabilization of Source Properties in Synchrotron Light Sources

Journal
Physical Review Letters, 123(19)

ISSN
0031-9007

Authors
Leemann, SC
Liu, S
Hexemer, A
et al.

Publication Date
2019-11-08

DOI
10.1103/physrevlett.123.194801

Peer reviewed

Demonstration of Machine Learning-Based Model-Independent Stabilization of Source Properties in Synchrotron Light Sources

S. Liu*

Department of Chemistry, University of California, Berkeley, CA 94720, USA

S.C. Leemann,† A. Hexemer, M.A. Marcus, C.N. Melton, H. Nishimura, and C. Sun

Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

(Dated: August 23, 2019)

Abstract

Synchrotron light sources, arguably among the most powerful tools of modern scientific discovery, are presently undergoing a major transformation to provide orders of magnitude higher brightness and transverse coherence enabling the most demanding experiments. In these experiments, overall source stability will soon be limited by achievable levels of electron beam size stability, presently on the order of several microns, which is still 1–2 orders of magnitude larger than already demonstrated stability of source position and current. Until now source size stabilization has been achieved through corrections based on a combination of static predetermined physics models and lengthy calibration measurements, periodically repeated to counteract drift in the accelerator and instrumentation. We now demonstrate for the first time how application of machine learning allows for a physics- and model-independent stabilization of source size relying only on previously existing instrumentation. Such feed-forward correction based on a neural network that can be continuously online-retrained achieves source size stability as low as $0.2\,\mu$m (0.4%) rms which results in overall source stability approaching the sub-percent noise floor of the most sensitive experiments.

PACS numbers: 29.20.db, 29.27.Fh, 41.85.-p, 41.85.Lc

INTRODUCTION

Synchrotron radiation sources, specifically 3rd-generation storage ring light sources, have been tremendously successful tools of scientific discovery since the early 1990s [1]. As these facilities mature, a new era of 4th-generation storage rings (4GSRs) based on *diffraction-limited storage rings* (DLSRs) [2–8] is being ushered in. These sources will increase average brightness by 2–3 orders of magnitude while also delivering high degrees of transverse coherence, for the first time even for x-rays. High coherent flux will enable scientists to understand material compositions and dynamics ranging in length from microns to nanometers and in time from minutes to nanoseconds. The most notable and direct result of the new beam properties will impact two techniques in particular. Ptychography [9] will take direct advantage of an increase in coherent flux to decrease measurement times by orders of magnitude. This will allow for the collection of complex 3D chemical maps with unprecedented resolution and will lead to deeper understanding of electrochemical systems such as batteries and fuel cells. The measurement of dynamics and kinetics to study chemical systems is another category that will be directly impacted by the new sources. An emerging technique to study this is X-ray Photon Correlation Spectroscopy (XPCS) [10]. Ptychography as well as XPCS rely heavily on high beam stability over extended periods of time.

To large extent the success of storage ring light sources lies in their stability, resulting in constant position, angle, and intensity of radiation delivered at a tunable wavelength with narrow width. In order to maintain constant intensity, a combination of top-off injection (maintaining constant beam current) [11, 12] and precise control over source position and size is required. In 3rd-generation light sources (3GLSs) the latter usually called for transverse beam size stability within 10% of the rms electron beam size [13, 14]. Now however, first experiments at these sources are starting to show limitations arising from such levels of source size control and it is evident that DLSRs, operating at much smaller source sizes, will call for significantly tighter control over source size stability in order to exploit ultra-high brightness and transverse coherence.

STATE-OF-THE-ART STABILIZATION EFFORT AND ITS LIMITATIONS

A typical example for the above mentioned source size stabilization challenge is shown in Fig. 1. The vertical electron beam size as measured at diagnostic beamline 3.1 [15] of

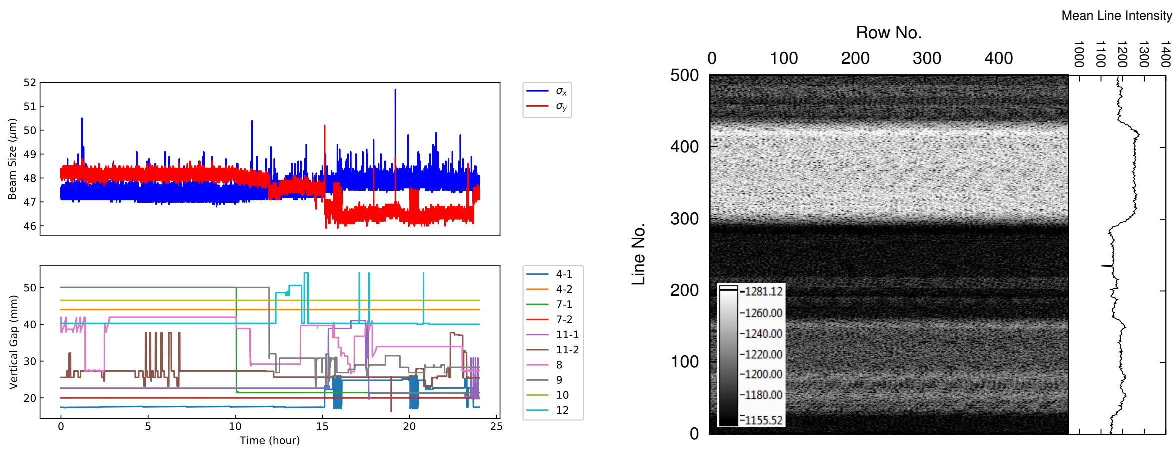

FIG. 1. Left: Electron beam size as measured the ALS diagnostic beamline 3.1 during a user run (top) showing $> 2\,\mu$m variation (4%) in the vertical caused by changes in the ID gaps (bottom). Right: STXM image from ALS beamline 5.3.2.2 showing banding (3.2% rms intensity variation) as a consequence of various ID configuration changes over the course of the scan.

Lawrence Berkeley National Laboratory's Advanced Light Source (ALS) is displayed during a typical user run. While the horizontal beam size is maintained constant (spikes observed in both planes at the same time are perturbations from top-off injection occurring roughly twice a minute), the vertical beam size fluctuates due to changes in the magnetic field configuration of the various insertion devices (IDs), e.g. variable field undulators and wigglers. Although such vertical beam size fluctuations are below typical stability requirements of 3GLSs, already today, at experiments that are highly sensitive to intensity fluctuations, such as scanning transmission x-ray microscopy (STXM) [16–19], scans that typically take several minutes at a single energy, will show both banding and pattern noise. The former, clearly visible in Fig. 1 (right), is caused by low-frequency variations in intensity (due to electron beam size changes at the source point) while the latter is the consequence of high frequency perturbations (e.g. vibrations of optical elements in the beamline). A typical STXM experiment involves quantifying contrast changes across several images acquired at different x-ray energies, but without a concurrent source intensity measurement, normalization within a single image is not possible and normalization across several images is less precise. Likewise, since data acquisition time per pixel (≈ 1 ms) is very short compared to typical perturbations from ID configuration changes, such effects cannot be averaged out during the scan. Thus, banding will effectively determine the noise floor of the experiment. While tight control over the beamline and end station equipment along with advances in detector technology enable a noise floor below 1%, the data shown here indicates substantially larger noise caused by low-frequency electron beam size variations resulting from ID gap/phase motion which changes the magnetic field configuration in the ID.

Common practice in state-of-the-art 3GLSs is to counteract ID gap/phase motion-induced perturbations on the electron beam through a two-pronged approach involving both local and global corrections: orbit correction (e.g. [20–22]) and optics correction whereby the latter usually comprises linear optics correction (e.g. [21, 23–25]), correction of the coupling between horizontal and vertical planes (e.g. [24, 26–29]), and in some cases also nonlinear correction (e.g. [24, 30, 31]). Orbit and linear optics corrections are often implemented as both feedbacks (FBs) and feed-forwards (FFs) because static model based FF corrections alone are usually not capable of sufficiently correcting transient behavior arising from comparably fast ID gap/phase motion. Feed-forward corrections usually rely on a physics model (for which linear approximations are used and linear superposition is assumed) and/or beam-based measurements rendering look-up tables that describe required corrections for a specific ID gap and phase setting. Recording a look-up table has to be performed for each ID individually, requires ample dedicated machine time, and, because it is usually a lengthy measurement, is also susceptible to drift. Because of the large number of IDs in most 3GLSs and the scarcity of dedicated machine time, these look-up tables cannot be frequently re-measured. Hence, as the machine drifts (temperature, ground motion, tidal effects, etc.) the fidelity of the look-up table and thus of the FF correction tends to deteriorate. Feedback corrections attempt to counteract such drift, but often do not offer sufficient closed-loop bandwidth to remove perturbations over the entire desired range.

In spite of the above mentioned correction schemes, residual ID-induced skew quadrupole errors (spurious focusing fields that render undesired coupling of motion in the transverse planes) result in vertical beam size variations in the storage ring (cf. Fig. 1, left). Low and medium energy light sources are especially susceptible to these errors due to the low beam rigidity and the prevalence of strong elliptically polarizing undulators (EPUs) [32].

As in most 3GLSs, the ALS performs coupling corrections for ID-induced skew quadrupole fields in a FF configuration whereby a large number of skew quadrupole coils can be excited to compensate for ID-induced skew components [33]. Look-up tables are on average re-recorded at most twice a year and for a typical EPU require an entire 8-hour machine shift. Furthermore, as DLSRs come online source beam sizes will shrink dramatically, while ID technology is advancing at comparably slower pace. We can assume that residual errors will remain comparable to present-day levels and therefore, size stability will deteriorate dramatically if a new approach to minimizing residual errors is not developed.

A NEW APPROACH: MACHINE LEARNING AND NEURAL NETWORKS

Recently, data driven methods have been applied to many different research areas. Specifically, neural networks (NNs) have proved to be most effective for nonlinear function fitting, both theoretically and empirically [37, 38]. Here, we propose a NN approach to predicting electron beam size as a function of arbitrary ID gap/phase configurations and employing this prediction to correct for perturbations thereby suppressing source size fluctuations. The NN can learn complex nonlinear relationships between ID settings and vertical beam size using large amounts of training data and advanced optimization techniques, which is a substantial improvement compared to the current approach based on linear optics and superposition.

Control of the electron beam size exploits the fact that commonly 3GLSs use skew quadrupoles to correct betatron coupling and spurious vertical dispersion first, and then to excite a vertical dispersion wave which improves beam lifetime within the boundaries of the diffraction limit [26, 34–36]. Such a dispersion wave generates vertical emittance (a global and conserved quantity) which results in a dominating contribution to the vertical source size at most source points. For these studies we can therefore slightly adjust the excitation of this vertical dispersion wave to control the vertical emittance and thus the vertical size at the source points[1]. At the ALS, 32 skew quadrupoles are included in the generation of the dispersion wave. We shall refer here to the dispersion wave parameter (DWP) as the scaling parameter describing our small relative adjustment of the standard skew quadrupole excitation pattern ($\leq 15\%$ of the overall vertical dispersion wave amplitude).

[1] As an example, Fig. 5 in [33] shows various vertical beam size contributions in a typical ALS ID source point. The contribution from ID-induced betatron coupling (canceled by skews) is smaller than that generated by the dispersion wave (excited by skews).

We demonstrate here that, given the ID gap/phase settings and DWP, the vertical source size can be predicted to within 0.4% rms (0.2 μm at the diagnostic beamline) with NNs. To train the NN model, high quality input data needs to be collected. For this purpose, beam sizes (as measured at e.g. a diagnostic beamline) along with all relevant beam parameters and ID settings have to be captured at high data rates. At the ALS we have so far chosen an acquisition rate of 10 Hz (faster than beam size measurement update rates and typical ID gap/phase variations) at which we collect data for roughly 35 independent channels. Input and output data is normalized with mean 0 and standard deviation 1. The NNs are implemented using Keras with Tensorflow backend [39] using mean squared error as the loss function. The models are trained using the back-propagation method [40] employing the Adam optimizer [41] for 50 epochs. The learning rate is set to 10^{-3} with a decay multiplier of 0.8 after each epoch for convergence. We screened a variety of NN architectures, regularization methods and activation functions. Deeper (i.e. more hidden layers) and wider (i.e. more nodes per layer) neural networks can generally provide better fitting on training data; however, a larger model is prone to overfitting and requires larger computational resources for both training and correction stages. We choose a NN containing three hidden layers with sizes 128, 64, 32, respectively, with the ReLU activation function [42]. A small L_2 regularization with $\lambda = 10^{-4}$ and a dropout with rate 0.2 was also used. The L_2 regularizer penalizes the large weights in neural networks and the dropout reduces the "co-adapting" between the weights [43], which is helpful to improve the generalizability of the model. These parameters are optimized through cross-validation [44], which is commonly used for model selection. The training takes 20 minutes on a single desktop-class CPU. The root mean squared error (RMSE) for training data is 0.16 μm while the validation RMSE is 0.20 μm. We also implemented a conventional linear and quadratic regression model by assuming that beam size can be approximated by linear or quadratic functions of the ID settings. The best training and validation RMSEs are 0.57 μm and 0.62 μm, respectively. The RMSEs appear to saturate towards orders 5–6 indicating further increase of polynomial order cannot improve the prediction. Figure 2 shows a visualization of the prediction on a segment of the validation dataset. The NN approach clearly outperforms the polynomial regressions. One of the possible reasons is that the NN can capture the interactions between IDs much more flexibly compared to the conventional regression method. The NN model has been proven to be effective for beam size prediction with RMSE 0.2 μm. Given a target beam

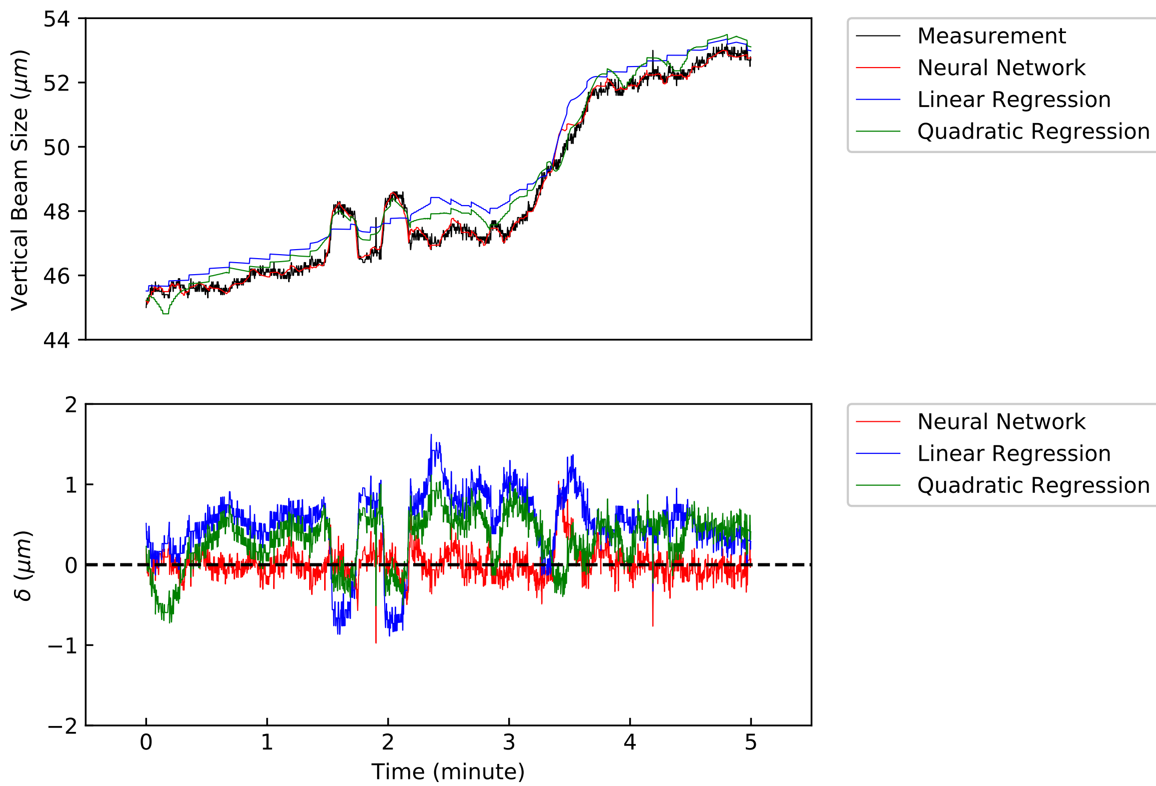

FIG. 2. Measured vertical beam size and predictions from polynomial regression and NN (top). Difference between predicted and measured vertical beam sizes (bottom). In terms of RMSE, the NN outperforms the regression models by roughly a factor 3.

size and the current combination of ID settings, we pre-screen 100 possible DWPs between -0.06 to 0.06 uniformly using the NN. Evaluating 100 DWPs takes only milliseconds on a single CPU, which enables us to implement this control at $> 10\,\mathrm{Hz}$. We select the DWP which renders the beam size closest to the target. The selected DWP value is used in a FF manner to adjust the skew quadrupole excitation pattern that generates the vertical dispersion wave. The experimental result is shown in Fig. 3. We turned FF control on and off repeatedly to verify the effectiveness of our beam size stabilization approach. In this example, when the FF is off, the variation of vertical beam size as measured at the diagnostic beamline is $1.5\,\mu$m rms (3%) and $7.5\,\mu$m peak-to-peak (15%). When the NN-based FF is turned on, this variation decreases to $0.2\,\mu$m rms (0.4%) and $1.9\,\mu$m peak-to-peak (4%). For comparison with the NN-based FF, we also implemented a simple FB loop relying solely on beam size measurement as an input and returning a DWP requested for beam size correction. During calm periods with only very slow ID configuration changes, the FB performance was

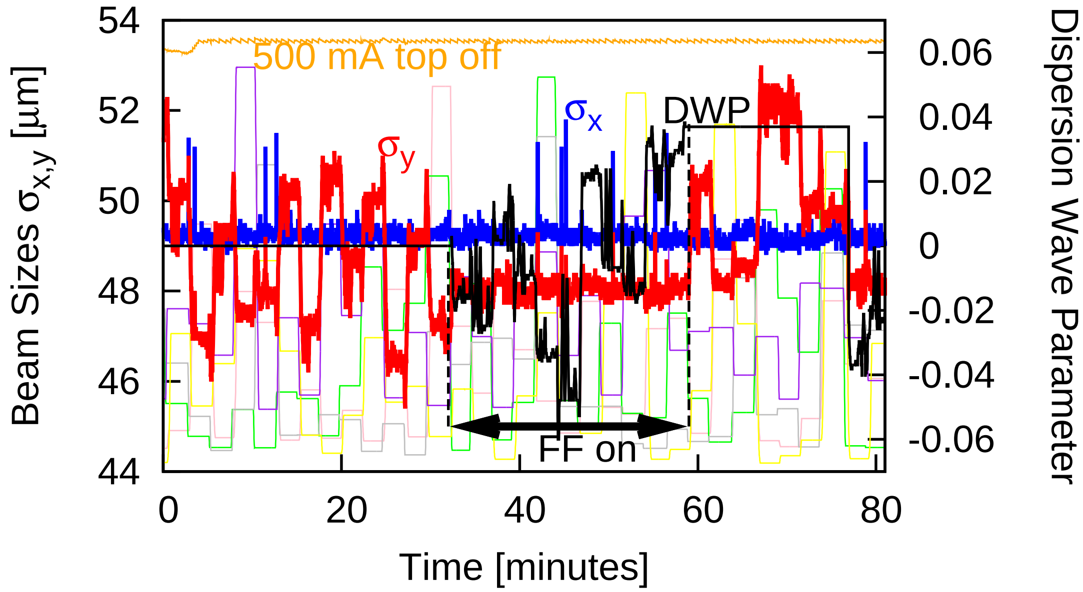

FIG. 3. Beam sizes (blue, red) as measured at the ALS diagnostic beamline 3.1 (spikes are top-off perturbations) along with DWP (black) and various ID vertical gap settings (light colored traces). Labels indicate the period with NN-based FF on.

capable of delivering similar rms stabilization as the NN-based FF. However, as soon as ID configurations changed at rates typically observed during experiments (e.g. 4 mm/s vertical gap motion and 16.7 mm/s horizontal shifts), the FB failed. Depending on PID tuning it was either not capable of stabilizing against transients (leading to 3 μm peak-to-peak vertical beam size variation, i.e. 6%) or it became unstable. The NN-based FF approach outperforms the FB method primarily for two reasons. First, the FF approach is agnostic to the current beam size. Implementing this FF does not require beam size as an input, hence adjusting beam size ahead of the measurement and avoiding measurement delay. Second, the NN-based FF does not have to adjust the DWP in a continuous fashion employing a series of small steps. It can instantaneously adjust the DWP by any large amount required to maintain stable beam size without overshoot.

So far, these experiments have revealed that the NN-based FF can stabilize the vertical beam size at the diagnostic beamline. It is, however, a priori not at all evident that stabilizing the source size at one point in the storage ring is equivalent to stabilizing the beam at the relevant source points. We originally chose to act on the beam size by means of the vertical dispersion wave, since it adjusts the vertical emittance, a global and conserved property, and we can therefore expect it to stabilize globally in spite of training the NN using a localized measurement. In order to demonstrate that this interpretation is correct, we conducted experiments at ALS beamline 5.3.2.2, which is the most sensitive ALS beamline in terms of vertical beam size [18, 19]. Figure 4 shows STXM scan data taken at 5.3.2.2 while ID configurations in the rest of the ALS were continuously changing. The measurement

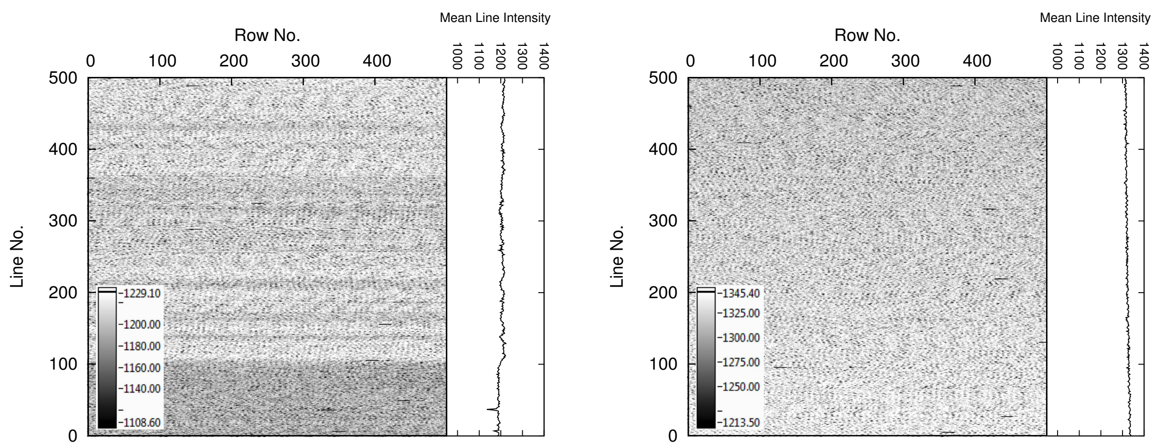

FIG. 4. STXM images from ALS beamline 5.3.2.2 at 390 eV. Left: scan performed while the NN-based FF was on (0.8% rms intensity variation). Right: scan performed without any ID motion in ALS (0.5% rms intensity variation).

data reveals that the stabilization observed at the diagnostic beamline can indeed also be observed in the STXM scans at this sensitive beamline. A comparison of Fig. 4 (left) to Fig. 1 (right) demonstrates a 4-fold reduction in noise at the STXM beamline from the NN-based FF. These STXM measurements have also revealed that this stabilization of low-frequency perturbations does not occur at the expense of exciting any high-frequency noise. Finally, Fig. 4 also reveals that the residual noise from ID configuration changes now lies only 60% above the noise floor of the beam line. We expect to further reduce this residual by increasing the beam size measurement refresh rate and consequently the NN-based FF update rate.

ONLINE STABILIZATION & RETRAINING

With the above determined performance at the most sensitive experiments, the NN-based FF can be put into operation during regular user experiment runs. Several user runs employing the NN-based FF so far have demonstrated that the vertical beam size can be stabilized to the sub-micron ($< 2\%$) rms level over the course of many days. One key advantage of this NN-based stabilization approach lies in the fact that data acquisition for retraining of the NN can be continuously carried out even while the NN-based FF is active during a regular user run. Under *online retraining* we understand continuous retraining of the NN (with machine data affected by the online NN) effectively allowing the NN to constantly adapt to a drifting machine, but also to changes in the ID configuration space occupied by experimenters during run periods.

Here, we demonstrate online retraining by combining data collected during a dedicated machine shift (for which the initial NN had been trained) with data collected during a 3-day user period with NN-based FF running. For online retraining, the user run data was randomly down-sampled to 1/15 of its original size to balance sample sizes. Retraining the NN using both data sets requires just 15 minutes on a desktop-class CPU. After verifying that predictions of the online-retrained NN better matched measured beam sizes than those coming from the original static NN, the FF was reconfigured to thenceforth rely on the online-retrained NN. An example of such a run is shown in Fig. 5. The observed level of

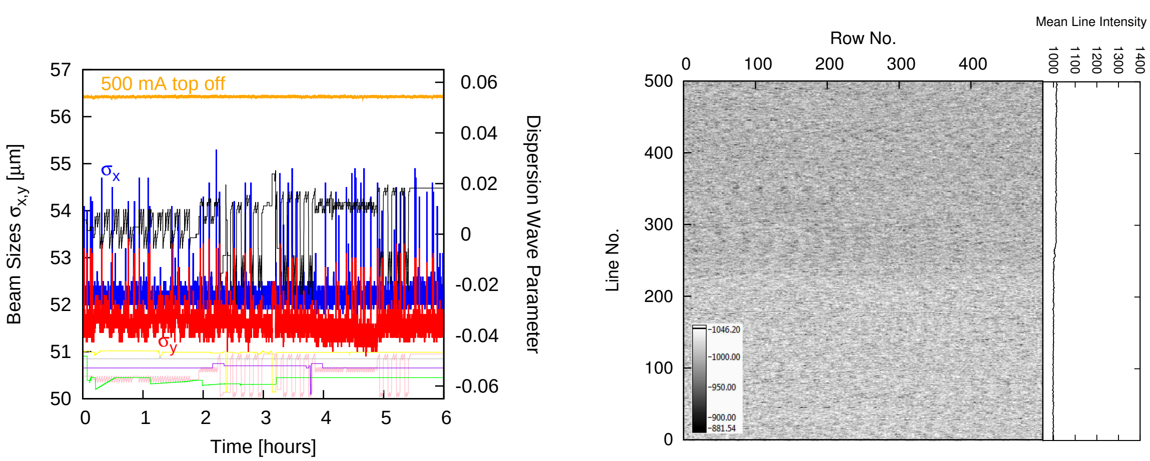

FIG. 5. Left: Beam sizes as measured at ALS diagnostic beamline 3.1 during user operations using a FF based on an online-retrained NN (0.4% rms variation in the vertical). Right: STXM scan from ALS beamline 5.3.2.2 (0.6% rms intensity variation) recorded during the same period.

vertical source size stability at diagnostic beamline 3.1 over the course of several days using the online-retrained NN is $<0.3\,\mu$m rms ($<0.5\%$). This indicates a factor two improvement over the originally applied static NN. In this case again, STXM scans confirm that this also leads to a global stabilization of source points (cf. Fig. 5 right). From these experiments we conclude that online retraining ensures that source size can be stabilized over prolonged periods of time without requiring dedicated machine time to retrain the NN or manual intervention by an expert. Furthermore, we recently demonstrated that even after a several-day interruption (e.g. scheduled maintenance) the previously employed online-retrained NN can upon startup again be put into FF operation without observing a reduction in performance. Online retraining thereafter can continue to ensure the employed NN stays up-to-date. For future operation, we expect to online retrain on-the-fly whenever indicated by a sustained increase in error between NN-based beam size prediction and online measurement beyond a predefined threshold.

CONCLUSION & OUTLOOK

We have demonstrated that machine learning can be employed to render NNs that enable vertical source size stabilization at storage ring light sources without requiring any new instrumentation. This model-independent method ensures levels of stability roughly one order of magnitude better than previously observed using model-based FFs and FB schemes. We have also demonstrated that such a NN-based FF remains effective over prolonged periods of time, including shutdown intervals, by employing online retraining. The achieved level of source size stability results in perturbations at the most sensitive experiments quickly approaching the noise level of the end station. Furthermore, the demonstrated sub-micron/sub-percent level of source size stability already today fulfills requirements for future 4GSRs thereby allowing experiments at these new sources to fully exploit the ultra-high brightness and transverse coherence provided by DLSRs. In the future, we plan to investigate if a NN-based FF can replace model-based FFs entirely, thus freeing up on the order of one hundred hours of dedicated machine time a year, which are nowadays still required to re-record look-up tables. First proof of principle experiments have been carried out and show promising results, including the exciting possibility to extract physics model information from a NN, eg. deriving ID perturbations from a NN trained on a machine without ID FFs.

ACKNOWLEDGMENTS

The authors would like to thank Greg Penn, Thorsten Hellert, and the ALS Operators for their support during measurement shifts. We would like to acknowledge David A. Shapiro for many helpful discussions and his generous support at beamline 5.3.2.2. We are most grateful to Fernando Sannibale, Marco Venturini, and Andreas Scholl for many interesting discussions and their valuable advice. Finally, extensive support is also acknowledged from Daniela Ushizima. This research is funded by the US Department of Energy (BES & ASCR Programs), and supported by the Director of the Office of Science of the US Department of Energy under Contract No. DEAC02-05CH11231.

* liushuai10000@berkeley.edu

† SCLeemann@lbl.gov

[1] W. Namkung, WEXRA01, pp. 2411–2415, Proceedings of IPAC'10, Kyoto, Japan, 2010.

[2] R. Hettel, J. Synchrotron Rad. **21**, 843–855, 2014.

[3] S.C. Leemann, M. Sjöström, Å. Andersson, Nucl. Instr. and Meth. A **883**, 33 (2018).

[4] A.R.D. Rodrigues et al., THXGBD4, pp. 2886–2889, Proceedings of IPAC2018, Vancouver, BC, Canada, 2018.

[5] P. Raimondi, WEXA01, pp. 2023–2027, Proceedings of IPAC2016, Busan, Korea, 2016.

[6] M. Borland et al., WEPOB01, pp. 877–880, Proceedings of NAPAC2016, Copenhagen, IL, USA, 2016.

[7] C. Steier et al., THPMF036, pp. 4134–4137, Proceedings of IPAC2018, Vancouver, BC, Canada, 2018.

[8] A. Streun et al., THPMK029, pp. 4358–4361, Proceedings of IPAC2018, Vancouver, BC, Canada, 2018.

[9] D.A. Shapiro et al., Nature Photonics **8**, 765-769 (2014).

[10] S.-W. Chen et al., Phys. Rev. Lett. **110**, 217201 (2013).

[11] L. Emery, M. Borland, TUCL4, pp. 200–202, Proceedings of the 1999 Particle Accelerator Conference, New York, NY, USA, 1999.

[12] H. Ohkuma, MOZCG01, pp. 36–40, Proceedings of EPAC08, Genoa, Italy, 2008.

[13] G. Decker, ITWM01, pp. 233–237, Proceedings of DIPAC 2005, Lyon, France, 2005.

[14] G. Rehm, 11th International Conference on Synchrotron Radiation Instrumentation (SRI 2012), J. Phys.: Conf. Ser. 425 042001, 2012.

[15] T.R. Renner, H.A. Padmore, R. Keller, Rev. Sci. Instrum. **67**, 3368 (1996).

[16] H. Ade, A.P. Hitchcock, Polymer **49**, 643 (2008).

[17] R. Falcone et al., Contemporary Physics, **52:4**, 293 (2011).

[18] T. Warwick et al., J. Synchrotron Rad. **9**, 254–257, 2002.

[19] A.L.D. Kilcoyne et al., J. Synchrotron Rad. **10**, 125–136, 2003.

[20] C. Benabderrahmane et al., TUPMN008, pp. 929–931, Proceedings of PAC07, Albuquerque, New Mexico, USA, 2007.

[21] J. Chrin et al., Nucl. Instr. and Meth. A **592**, 141 (2008).

[22] E. Plouviez et al., MOPO002, pp. 478–480, Proceedings of IPAC2011, San Sebastián, Spain, 2011.

[23] E. Wallén, S.C. Leemann, TUP235, pp. 1262–1264, Proceedings of 2011 Particle Accelerator Conference, New York, NY, USA, 2011.

[24] S.C. Leemann, H. Tarawneh, TUPJE038, pp. 1696–1698, 6th International Particle Accelerator Conference, IPAC2015, Richmond, VA, USA, 2015.

[25] I.P.S. Martin et al., TUPRI082, pp. 1760–1762, 5th International Particle Accelerator Conference, IPAC2014, Dresden, Germany, 2014.

[26] C. Steier et al., RPPG018, pp. 3213–3215, Proceedings of the 2003 Particle Accelerator Conference, Portland, OR, USA, 2003.

[27] A. Franchi et al., Phys. Rev. ST Accel. Beams **14**, 034002 (2011).

[28] I.P.S. Martin et al., MOPEA071, pp. 249–251, Proceedings of IPAC2013, Shanghai, China, 2013.

[29] W.A. Wurtz, Nucl. Instr. and Meth. A **892**, 1 (2018).

[30] J. Bahrdt, WEPC097, pp. 2222–2224, Proceedings of EPAC08, Genoa, Italy, 2008.

[31] D. Bertwistle et al., 020012, Proceedings of the 12th International Conference on Synchrotron Radiation Instrumentation SRI2015, 2015.

[32] S. Sasaki, Nucl. Instr. and Meth. A **347**, 83 (1994).

[33] C. Steier et al., MOPKF071, pp. 479–481, Proceedings of EPAC 2004, Lucerne, Switzerland, 2004.

[34] Y. Wu et al., TUP3A17, pp. 1098–1100, Proceedings of EPAC 2000, Vienna, Austria, 2000.

[35] C. Montag, J. Bengtsson, B. Nash, FRPMS113, pp. 4375–4377, Proceedings of PAC07, Albuquerque, New Mexico, USA, 2007.

[36] J. Breunlin, S.C. Leemann, Å. Andersson, Phys. Rev. Accel. Beams **19**, 060701 (2016).

[37] K. Hornik et al., Neural networks **2:5**, 359-366, 1989.

[38] Y. LeCun et al., Nature, **521**, 436, 2015.

[39] M. Abadi et al., OSDI, **16**, 265-283, 2016.

[40] Y. LeCun et al., Neural computation **1.4** (1989): 541-551.

[41] D.P. Kingma et al., arXiv preprint arXiv:1412.6980 (2014).

[42] V. Nair et al., Proceedings of the 27th international conference on machine learning (ICML-10), 2010, 807-814.

[43] N. Srivastava et al., Journal of Machine Learning Research, 2014, 1929-1958.

[44] M. Stone et al., Journal of the Royal Statistical Society: Series B, 1974, 111-133.