Journal of Physics: Conference Series **2420** (2023) 012074 doi:10.1088/1742-6596/2420/1/012074

Online correction of laser focal position using FPGA-based ML models

J Einstein-Curtis[1], S J Coleman[1], N M Cook[1], J P Edelen[1], S Barber[2], C Berger[2] and J van Tilborg[2]

[1] RadiaSoft LLC, Boulder, CO, USA
[2] Lawerence Berkeley National Lab, Berkeley, CA, USA

E-mail: joshec@radiasoft.net

Abstract.

Ultrafast lasers play an increasingly critical role in the generation, manipulation, and acceleration of electron beams for High Energy Physics applications. Laser plasma accelerators enable order of magnitude improvements in accelerating gradient and promise compact tunable GeV electron beam sources, while novel photocathode systems permit fundamental advances in electron beam manipulation for accelerator and radiation applications Advances in fast feedback systems are required to stabilize laser performance at kHz repetition rate operation against environmental fluctuations. A field programmable gate array (FPGA) based digital control system, coupled with responsive optics, can provide rapid and precise stabilization of ultrafast lasers. A collaboration between RadiaSoft and the Lawrence Berkeley National Laboratory BELLA Center to develop, test, and deploy these systems across a range of beamlines operating at >1 Hz repetition rate, including 1 kHz systems, was created.

1. Introduction

Laser plasma accelerators (LPAs) rely upon accurate control of ultrafast lasers, typically Ti:Sapph and Nd:Yag amplifier systems [1, 2]. The BELLA Center at Lawerence Berkeley National Laboratory (LBNL) features several ultra-short pulse, high-energy beamlines to develop LPAs. These accelerators require highly repeatable, stable interaction points to generate high-quality electron beams, which necessitates a collection of active and passive controls to mitigate environmental, mechanical, and component variations.

Recent work has primarily focused on enhancing transverse beam stability [3–5]. This paper describes a a strategy to address focal position stability, leveraging a machine learning (ML) enhanced wavefront diagnostic in tandem with a Field Programmable Gate Array (FPGA) controller to correct focal position at a kHz-scale rate. By building a model of wavefront at the interaction point, it is possible to use a non-perturbative measurement to calculate the focal position [6].

2. Facility and equipment

The initial model was created for the BELLA HTU laser system, shown in figure 1. This beamline operates with 1 kHz seed pulses and a 1 Hz full-power pulse. A HASO FIRST Shack-Hartmann wavefront sensor was used as the ground-truth imaging device of the interaction and post-interaction region, with the pre-interaction region sensor a Thorlabs WFS20-7AR. A Xilinx Zynq ZCU104 FPGA evaluation kit was used for testing to provide flexibility during the

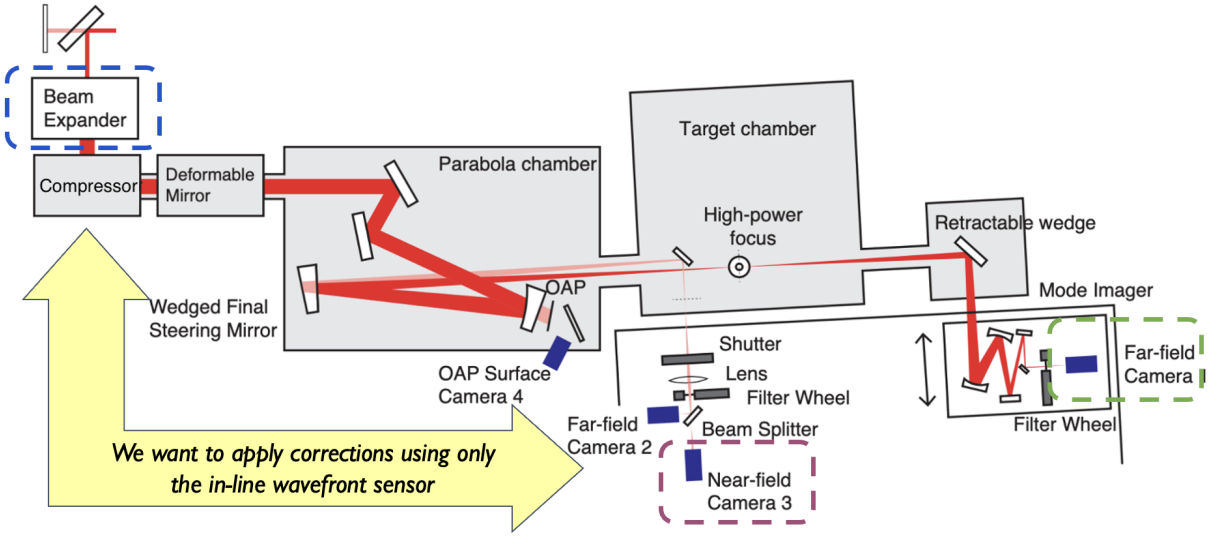

Figure 1. Diagram of HTU laser system at LBNL, highlighting the proposed correction scheme. Machine learning techniques are used to correlate a fast, non-perturbative sensor (2) with a high-quality, but perturbative wavefront sensor (1) which cannot be used for online correction. The resulting online diagnostic is used to deduce variations away from the desired focal position, which is then corrected for prior to the next shot by changes made to a transmissive lens beam expander (3).

Table 1. Optimal lens movement vs focal shift and beam size change. Focus shift is per mm lens translation. Beam size change is change per mm lens translation.

Element	**Shift**	**Size Change**
Transmissive Amp3-in	2mm	x1.348
Transmissive Amp4-in	2mm	x1.046
Reflective Amp4-out	1mm	x1.002

prototype phase, including a variety of customizable I/O, well-supported manufacturer-provided software, and a variety of processing options in support of ML operations.

3. Focal position investigations

To determine the optimal lenses to move for a focal shift, we looked at the magnitude of the shift at final focus and the (unwanted) increase in beam size throughout the optical chain. Table 1 summarizes these parameters for three different lenses in the telescope.

From these simulations we determined that the reflective Amp4-out is not ideal as a motorized correction optic for focal location because it is more weakly responsive, shifting the focus by only 1 mm per mm translation. Moreover, the off-axis reflective geometry introduces beam centroid kicks, even in response to relatively mild beam size variations. Ultimately, we determined the Amp4-in telescope is the best choice.

To verify our model, we measured the focal location vs lens separation at high power. Our measurement used a comparable method of capturing leakage from the final steering mirror thus measuring raw focal location without the need for further calibration or renormalization. The inset of figure 2 provides details of the measured focal position and radius of curvature taken from the wavefront sensor.

When comparing measurements to the simulation, we note that the focus shift per mm stage

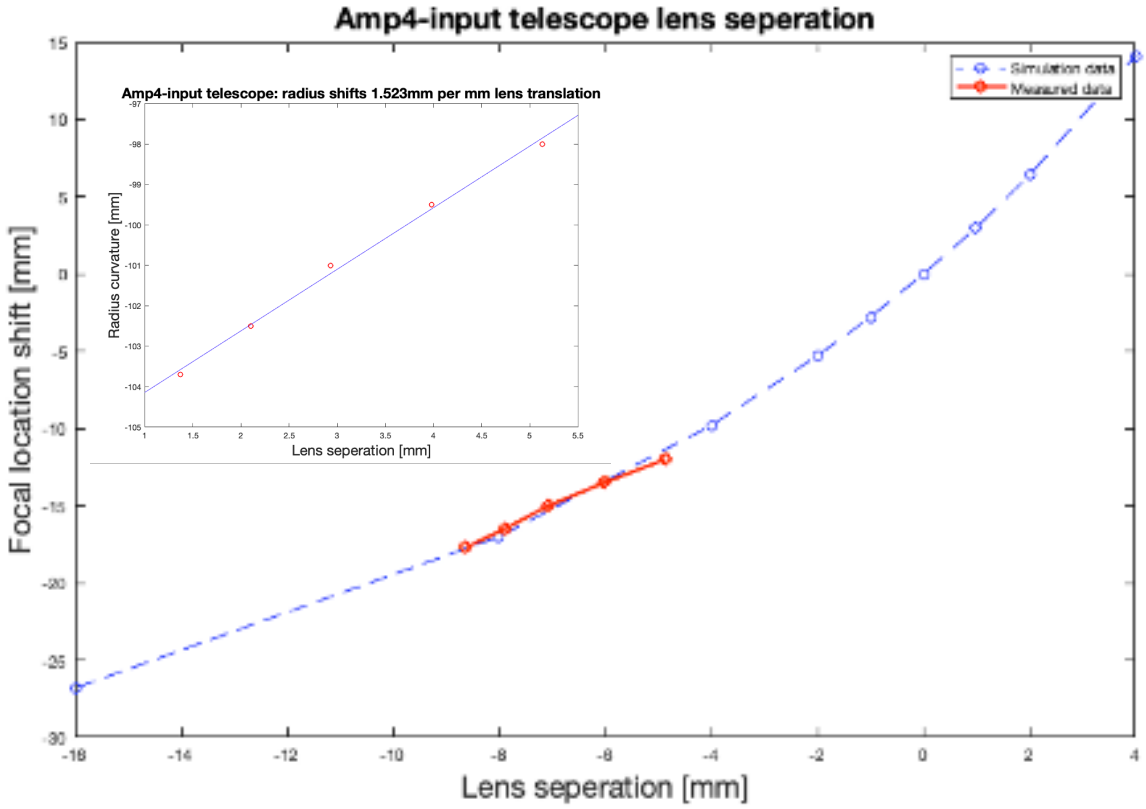

Figure 2. Focal location vs lens separation.

motion depends on the nominal Amp4-in lens separation. For a perfectly collimated beam entering the $-f_1/+f_2$ telescope, and for a perfectly collimated beam leaving the telescope (lens separation is $f_2 - f_1$), the slope change is 2mm focus shift per 1mm change in lens separation.

However, for the situation where the lens separation is NOT equal to $f_2 - f_1$, for example because the input beam has a divergence or the output beam is not perfectly collimated, this slope will have a different value.

By overlapping the experimental data (red circles) with the simulation (blue circles), we find a good agreement for one very specific initial lens separation offset (circa -6 mm). The slope at this separation is 1.52mm focal position shift for every 1mm lens motion. Figure 2 confirms this result.

This validates the use of a telescoping optic configuration for making controlled adjustments to the laser focal position. This design was validated through simulation and experimental measurement.

4. Model development

Several datasets were collected to examine changes in focal position on a shot-by-shot basis. The intra- and inter-shot variation over time, as shown in figure 3, show millimeters of variation in the calculated radius of curvature, highlighting the need for correction schemes.

Examining the extrapolated focal positions from each dataset reveals significant discrepancies between the two sensor measurements. Figure 4 shows a correlation plot between the two sensors, for which the raw correlation, as measured by the Pearson's coefficient between computed radius

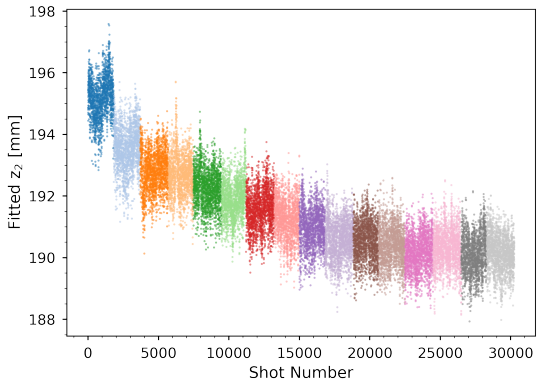

Figure 3. Representative dataset variation in fitted Z_2 calculation.

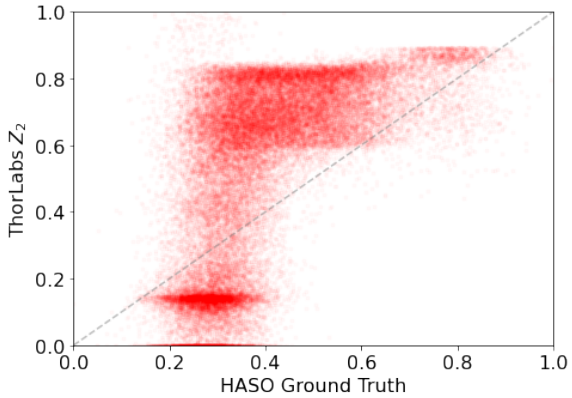

Figure 4. Raw HASO/ThorLabs correlation.

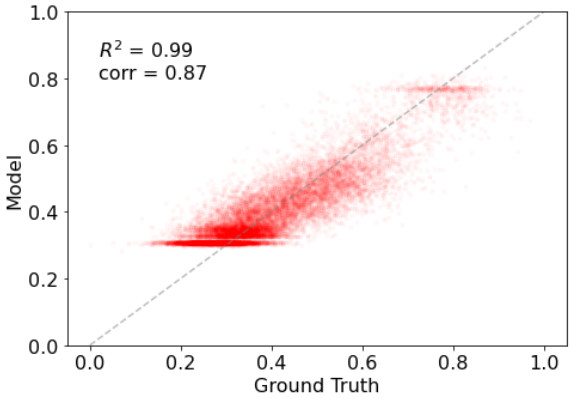

Figure 5. Feed-forward network accuracy.

of curvature, is only 0.45.

Due to the lack of correlation between the two sensors using raw pixel data, it became necessary to develop a pre-processing flow and system model to accurately capture systematic differences in the two measurements. We thus developed and trained a set of neural network models with the aim of improving the correlation between the two devices. The output for the trained network was a prediction of the radius of curvature, to be compared against the HASO WFS measurement.

Our initial efforts considered two different types of neural networks – convolutional neural nets (CNN) and more general feedforward neural nets (FFNN). Each network was trained using PyTorch, an open source library for developing machine learning models.

CNNs are designed to operate on images as inputs and are useful for computer vision applications. Our CNNs were trained using 12×12 pixel-by-pixel image data of the wavefront from the Thorlabs WFS. We found that the CNN produced only modest improvements in the correlation, to upwards of 0.63 from the initial value of 0.45. We thus transitioned to exploring more general FFNNs.

Our FFNN architecture featured 2-4 fully connected hidden layers, ReLU activation functions, and implemented a robust scaler on inputs and outputs. These features were chosen to be fast-executing and compatible with our FPGA deployment strategy. We concluded that the FFNN consistently produced better results than the CNN, but did exhibit tradeoffs between input space size and network complexity, with implications for performance at high repetition rate.

The best correlation was found by augmenting the pixel data with additional Zernike polynomial fitting terms; we explored several different strategies for generating the fit. Using

Table 2. Correlation values for different data.

Dataset	Pearson Correlation
Zernike fits only	0.45
CNN - pixel data only	0.63
FFNN - pixel data only	0.82
FFNN - pixel data & Zernike fits	0.87

the Thorlabs toolkit to produce a 5th order fit provided an additional 16 terms to include in the input space of the network, and improved dataset correlation to as high as 0.87, using only two hidden layers, as shown in figure 5.

Using external fitting libraries, such as the Mahotas library [7], permitted higher order fits, such as a 28-value, 6th-order fit. However, increasing the fit complexity showed diminishing returns, as correlations did not improve significantly, while speed of execution declined. Using a 6th-order fit does enable a network to be trained using only fitting data (28 inputs), and can result in comparable performance to that of the full set of pixel values. Table 2 summarizes the correlation performance for each of our approaches.

5. Implementation

A full correction implementation was prototyped using the FPGA system and was tested on the bench to meet-or-exceed the operational requirements of the HTU beamline. This implementation utilized the Xilinx Vitis AI toolkit in conjunction with the Xilinx Deep Learning Processor (DPU) to minimize the use of custom FPGA designs and software.

Due to driver limitations of the Thorlabs WFS20 sensor, in particular being limited to a Windows-based platform, the sensor was unable to be directly connected to the processing platform. This necessitated the use of an alternative data communication channel. This channel was created in Python using ZeroMQ, and tested to transfer wavefront data to the processing platform at the limit of the sensor capture rate (about 0.9 kHz). Validation data from model development was used to test the model processing performance, and achieved a better than 5 kHz throughput, with well-understood bottlenecks and limitations. Due to the minimal number of outputs, additional output data processing should not over-burden the system, enabling performance that meets the 1 kHz seed pulse rate on the HTU beamline.

6. Conclusion

We have demonstrated a model of the BELLA Center HTU beamline interaction region and developed a correction method for the focal position. This model, in conjunction with slow controllers, corrects for measured system variations in simulation. This method has been demonstrated in prototype hardware using simulated data and meets-or-exceeds the necessary performance requirements with room for expansion and increased model complexity as needed.

Limitations exist in that variations between seed and full-power pulses might require multiple models for proper correction. Plans exist to continue this work on additional beamlines to develop a flexible, plug-and-play framework for additional LPAs.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0021680.

References

[1] Esarey E, Schroeder C B Leemans W P 2009 *Rev. Mod. Phys.* **81** 1229
[2] Dann S J D *et al.* 2019 *Phys. Rev. Accel. Beams* **22** 041303
[3] Isono F, van Tilborg J, Barber S K, Geddes C G R, Tsai H, Schroeder C B, Leemans W P 2019 *Update on BELLA Center's Free-Electron Laser Driven by a Laser-Plasma Accelerator* (CLEO: Science and Innovations SF31)
[4] Barry M 2015 Evaluation of Laser Stabilization and Imaging Systems for LCLS-II Tech. rep. SLAC-TN-15-012 SLAC National Accelerator Laboratory
[5] Nguyen L *et al.* 2019 *Proc. North American Particle Accelerator Conf. (NAPAC'19)* (Lansing, MI USA) pp 938–41
[6] Shalloo R J *et al.* 2020 *Nat. Commun.* **11** 6355
[7] Coelho L P 2013 *J. Open Res. Software* **1** e3