

**Advanced Control Methods for Particle Accelerators  
ACM4PA 2019**

Hosted by Los Alamos National Laboratory and the  
Los Alamos Neutron Science Center (LANSCE)

Hotel Santa Fe, in Santa Fe, NM, USA

August 19 - August 22, 2019

# Advanced Control Methods for Particle Accelerators (ACM4PA) 2019\*

Alexander Scheinker,<sup>1,†</sup> Claudio Emma,<sup>2,‡</sup> Auralee L Edelen,<sup>2,§</sup> and Spencer Gessner<sup>2,¶</sup>

<sup>1</sup>*Los Alamos National Laboratory, Los Alamos, NM, 87544, USA*

<sup>2</sup>*SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA*

(Dated: December 11, 2019)

Los Alamos is currently developing novel particle accelerator controls and diagnostics algorithms to enable higher quality beams with lower beam losses than is currently possible. The purpose of this workshop was to consider tuning and optimization challenges of a wide range of particle accelerators including linear proton accelerators such as the Los Alamos Neutron Science Center (LANSCE), rings such as the Advanced Photon Source (APS) synchrotron, free electron lasers (FEL) such as the Linac Coherent Light Source (LCLS) and LCLS-II, the European X-ray Free Electron Laser (EuXFEL), the Swiss FEL, and the planned MaRIE FEL, and plasma wake-field accelerators such as FACET, FACET-II, and AWAKE at CERN. One major challenge is an the ability to quickly create very high quality, extremely intense, custom current and energy profile beams while working with limited real time non-invasive diagnostics and utilizing time-varying uncertain initial beam distributions and accelerator components. Currently, a few individual accelerator labs have been developing and applying their own diagnostics tools and custom control and ML algorithms for automated machine tuning and optimization. The goal of this workshop was to bring together a group of accelerator physicists and accelerator related control and ML experts in order to define which controls and diagnostics would be most useful for existing and future accelerators and to create a plan for developing a new family of algorithms that can be shared and maintained by the community.

## Organizing Committee

Alexander Scheinker	LANL
Spencer Gessner	CERN
Claudio Emma	SLAC
Auralee Edelen	SLAC

## Invited Speakers

Sam Barber	LBNL
Kip Bishofberger	LANL
Joseph Duris	SLAC
Sandy Easton	CERN
Jonathan Edelen	RadiaSoft
Adi Hanuka	SLAC
John Lewellen	LANL
Jerry Ling	UCSB
Quinn Marksteiner	LANL
Brendan O'Shea	SLAC
Tor Raubenheimer	SLAC
Samuel Schoenholz	GOOGLE
Anna Solopova Shabalina	JLAB
Yine Sun	ANL
Faya Wang	SLAC

## Workshop Website

[www.lanl.gov/advanced-control-accelerators](http://www.lanl.gov/advanced-control-accelerators)

---

\* This workshop was hosted by Los Alamos National Laboratory at Hotel Santa Fe, in Santa Fe, NM, USA, from Tuesday, August 19 to Thursday, August 22, 2019.

† [ascheink@lanl.gov](mailto:ascheink@lanl.gov)

‡ [cemma@slac.stanford.edu](mailto:cemma@slac.stanford.edu)

§ [edelen@slac.stanford.edu](mailto:edelen@slac.stanford.edu)

¶ [sgess@slac.stanford.edu](mailto:sgess@slac.stanford.edu)

## I. 2019 ACM4PA



FIG. 1. Group photo of participants at the 2019 ACM4PA.

## II. SCHEDULE

Monday, August 19	Topic	Speaker	Affiliation
4:00 – 5:30 PM	Registration		
Tuesday, August 20	Topic	Speaker	Affiliation
8:00 – 8:30 AM	Registration/Breakfast/Coffee		
8:30 – 9:15 AM	Opening remarks/feedback control overview	Alex Scheinker	LANL
9:15 – 10:00 AM	Plenary, FELs overview	Tor Raubenheimer	LCLS
10:00 – 10:15 AM	Coffee		
10:15 – 11:00 AM	Plenary, PWFAs overview	Vitaly/Mark/Claudio	FACET-II
11:00 – 11:30 AM	ML for accelerators overview	Auralee Edelen	SLAC
11:30 – 1:00 PM	Lunch		
1:00 – 1:30 PM	PWFAs dynamics	Spencer Gessner	CERN
1:30 – 2:00 PM	PWFAs non-invasive diagnostics overview	Claudio Emma	FACET-II
2:00 – 2:30 PM	Diagnostics	Brandon O’Shea	SLAC
2:30 – 3:00 PM	Phase space manipulation	Joe Duris	SLAC
3:00 – 3:30 PM	Coffee		
3:30 – 4:00 PM	GP-based optimization/virtual diagnostics	Adi Hanuka	SLAC
4:00 – 4:30 PM	Diagnostics	Quinn Marksteiner	LANL
4:30 – 5:00 PM	Phase space manipulation	Yine Sun	APS
5:00 – 5:30 PM	Discussions/Summary		
Wednesday, August 21	Topic	Speaker	Affiliation
8:00 – 8:30 AM	Breakfast/Coffee		
8:30 – 9:15 AM	Plenary, ML overview	Sam Schoenholz	GOOGLE
9:15 – 10:00 AM	General Discussion of problems that people are currently working on: suggestions/approaches		
10:00 – 10:30 AM	Coffee		
10:30 – 11:00 AM	Diagnostics	Kip Bishofberger	LANL
11:00 – 11:30 AM	Autonomous Accelerators	John Lewellen	LANL
11:30 – 1:00 PM	Lunch		
1:00 – 1:30 PM	Modeling + Control	Jonathan Edelen	Radiasoft
1:30 – 2:00 PM	NN machine optimization	Faya Wang	SLAC
2:00 – 2:30 PM	Laser Plasma beam measurements/optimization	Sam Barber	LBL
5:00 – 5:30 PM	ML techniques	Mike McKerns	LANL
3:00 – 3:30 PM	Coffee		
3:30 – 4:00 PM	ML RF Cavities	Anna Solopova	JLAB
4:00 – 4:30 PM	Beam Schedule Automation	Sandy Easton	CERN
4:30 – 5:00 PM	Fast Emittance Measurements	Jerry Ling	UCSB
5:00 – 5:30 PM	Discussions/Summary		
Thursday, August 22	Topic	Speaker	Affiliation
8:00 – 8:30 AM	Breakfast/Coffee		
8:30 – 9:00 AM	Diagnostics Discussion/Summary	Claudio Emma	FACET-II
9:00 – 9:30 AM	ML/Control Discussion/Summary	Auralee Edelen	AWAKE
9:30 – 10:00 AM	Phase Space Control Discussion/Summary	Spencer Gessner	SLAC
10:00 – 10:30 AM	Feedback Control Discussion/Summary	Alexander Scheinker	LANL
10:30 – 11:00 AM	Coffee		
11:00 – 12:00 PM	Discussions/Summary		

FIG. 2. Schedule of talks and discussions at the 2019 ACM4PA.

### III. MOTIVATION

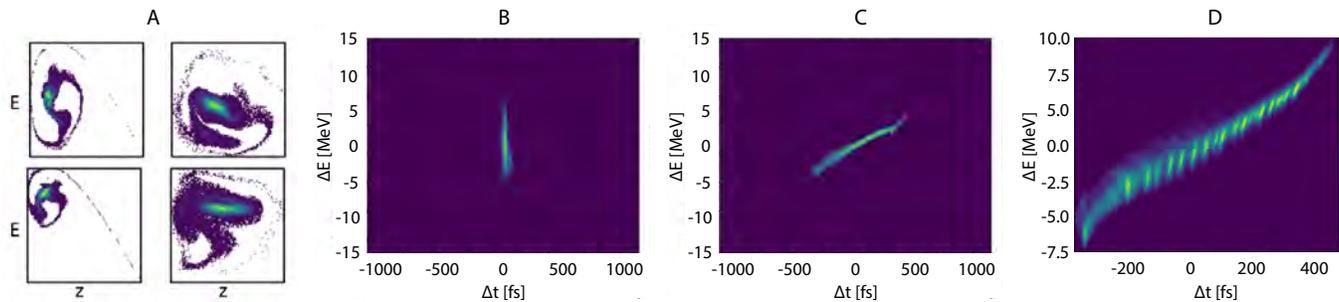


FIG. 3. Simulation of a proton bunch passing through LANSCE is shown in (A). Nonlinear space charge forces twist the beam, create halos and tails of the bunch particle distribution in energy vs time phase space. Results of the first stage of bunch compression at the EuXFEL is shown in (B-D). The bunch in (B) is nicely compressed, creating a short, intense beam. The bunch in (C) is not as tightly compressed as the one in (B), and is suffering from microbunch instabilities where local density fluctuations naturally occurring in the bunch are amplified by the compression process, as seen in more detail in (D).

Precise control of the longitudinal phase space (LPS) (time vs energy: bunch length, current profile, energy spread) of charged particle bunches is required for all particle accelerators. The following outlines some accelerator capabilities and challenges (from [1]):

*Free electron lasers (FEL) are incredibly powerful scientific tools for studying physics at previously inaccessible length (nanometers) and time (femtoseconds (fs)) scales for high energy physics, biology, chemistry, material science, and accelerator physics experiments [2-4]. FELs can produce extremely short ( $\sim$ fs) coherent X-ray bursts with tunable wavelength which are many orders of magnitude brighter than traditional sources such as synchrotrons. For example, the Linac Coherent Light Source (LCLS) FEL, the first hard X-ray FEL, provides users with photon energies ranging from 0.27 keV to 12 keV based on electron bunches with energies from 2.5 GeV to 17 GeV. Operating electron bunch charge can range from 20 pC to 300 pC and the bunch duration from 3 fs to 500 fs to suit experimental needs [5-7]. The European X-Ray FEL (EuXFEL), one of the newest and most advanced FELs in the world, is capable of producing 27000 pulses of bright, coherent light per second utilizing electron bunches with energies of up to 17.5 GeV, with charges ranges from 0.02 to 1 nC per bunch, and photon energies from 0.26 keV up to 25 keV [8-10]. Precise control of bunch lengths, current profiles, and energy spreads of increasingly shorter electron beams at femtosecond resolution is extremely important and challenging for both the LCLS and the Eu-XFEL [11, 12].*

*The extremely bright and short X-ray bursts that make FELs such powerful instruments also makes them incredibly challenging to control. High power FELs are driven by few kilometer long high power particle accelerators composed of thousands of interacting electromagnetic components including radio frequency (RF) accelerating cavities and magnets. The performance of all of these components is susceptible to drift, e.g. such as thermal drifts. A major challenge is the FEL lasing process itself, self-amplified spontaneous emission (SASE) is stochastic in nature and extremely sensitive to the beam's initial conditions including charge density and energy spread, therefore there is a large variance in the output power of FEL light even when all machine set points are fixed and properly timed because of uncertainty in and time variation of the electron distribution coming off of the photo cathode and entering the accelerator. Therefore, the output power of an FEL is a highly nonlinear, time-varying, noisy and analytically unknown function of all of the FELs thousands of components. Traditional model-based approaches are severely limited by such uncertainties and time variation of both the accelerated beam's phase space distribution and the accelerator's components as well as misalignments, thermal cycling, and collective effects such as space charge forces, wakefields, and coherent synchrotron radiation emitted by extremely short high current bunches.*

*One example of such difficulties is the process of reconfiguring the LCLS to a low charge mode to provide 3 fs bunches, a process which may require many hours of expert hands on tuning. These difficulties will only grow for future facilities, such as the LCLS-II [13] for complex schemes such as multi-color operation [14], multi-stage amplification [15] or self-seeding [6, 16]. Plasma wakefield accelerators (PWFA) are another class of particle accelerators which require extremely intense, high current and sometimes extremely short charged particle bunches with complex beam dynamics and phase space manipulations [17]. For example, the facility for advanced accelerator experimental tests (FACET-II) is being designed to provide custom tailored current profiles for various experiments with bunch lengths as low as 1 or 2 fs [18, 19].*

One illustrative problem is that of creating tightly packed charged particle bunches and maintaining these bunches while accelerating them. Accelerators utilize radio frequency (RF) resonant cavities to create large (tens of MV/m)

electric field gradients with energy gain proportional to  $V \cos(\phi)$  with  $\phi$  the phase of the electromagnetic wave at the time of a particle's arrival. Non-zero  $\phi$  creates a precise energy deviation along the electron bunch. For a linear proton accelerator such as LANSCE, energy difference translates to significant velocity difference causing beam bunching. For highly relativistic electron accelerators, such as PWFAs and FELs, the velocity difference is insignificant ( $v \approx c$ ), so a beam is passed through a magnetic chicane translating energy difference to path length difference, compressing the bunch by orders of magnitude. For an RF field of frequency  $\omega_{\text{rf}}$ , the phase  $\phi$  relative to the RF at position offset  $\Delta z_0$  is given by  $\phi = -\omega_{\text{rf}} \Delta z_0 / c$ . For small  $\phi$ , for an initial bunch length of  $\sigma_{z0}$  has approximate final bunch length

$$\sigma_{zf} \approx \sqrt{\left(1 + R_{56} \frac{eV_{\text{rf}} \omega_{\text{rf}}}{cE}\right)^2 \sigma_{z0}^2 + R_{56}^2 \sigma_{\Delta E_0}^2}, \quad R_{56}(z) = \int_{z_0}^z \frac{R_{16}(z')}{\rho(z')} dz', \quad (1)$$

where  $\sigma_{\Delta E_0}$  is initial beam energy spread,  $R_{16}$  is transverse displacement resulting from energy offset in a dispersive region of the accelerator, and  $\rho(z)$  the local radius of curvature for a beam passing through bending magnets down the length of the accelerator ( $z$ ). In practice bunch compression is a much more complex process than (1). Intense particle bunches undergo collective effects such as space charge forces resulting in collective phenomenon such as halo formation and micro-bunch instability creating local density fluctuations. Precise setup and control of bunch compression requires non-invasive diagnostics to ensure that beam quality is preserved. Figure 3 shows LPS simulations of proton bunches in LANSCE and TCAV measurements of the LPS of electron bunches in the EuXFEL after the first bunch compressor.

Linear proton machines, such as the Los Alamos Neutron Science Center (LANSCE), also face tuning challenges because of highly space charge dominated beams, limited diagnostics, and the complexity of many coupled components, as shown in Figure 4.

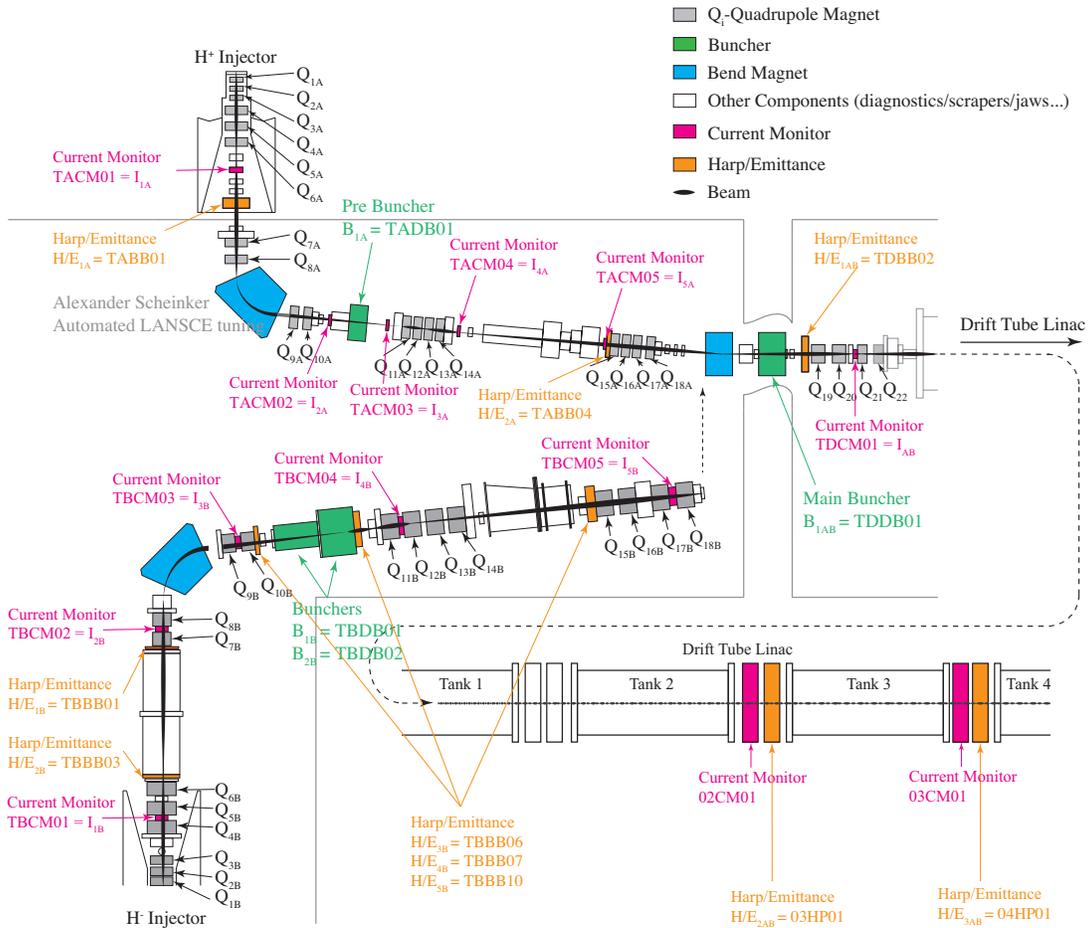


FIG. 4. Overview of the injector and drift tube linac sections of the LANSCE proton linac at Los Alamos National Laboratory.

## IV. SUMMARY

New algorithms for non-invasive diagnostics and control based on adaptive feedback (AF) and machine learning (ML) methods can greatly benefit existing accelerators as well as enable future facilities such as the planned MaRIE, in particular, new adaptive machine learning methods combining AF and ML for complex time-varying systems [1].

### Diagnostics

New diagnostics requirements especially emphasize the need for non-invasive diagnostics that can provide real time information about extremely short and intense charged particle bunches. Many new approaches for non-invasive diagnostics that are being discussed are planning to use the spectrum of light emitted from accelerating electrons beams in order to attempt to reconstruct their current profiles. One such approach is to pass a high energy electron beam through a gentle undulator and then utilize the off-axis radiation to reconstruct the electron density. This has not yet been demonstrated, especially the unique iterative phase retrieval problem needs to be considered carefully. Another technique being considered is optical diffraction radiation created when a beam grazes a metal edge, giving a 1D projection of a slice of a beam's 2D phase space. Numerical methods show promise in using several independent 1D projections to reconstruct a 2D phase space, giving transverse beam diagnostics. There are also computational approaches being considered for diagnostics, one of which is to attempt to train a neural network to map accelerator parameter settings to longitudinal phase space (LPS) distributions, this approach has been demonstrated in simulation and at the LCLS [20]. Another approach is to adaptively tune accelerator models based on information non-invasively gathered about the beam to get an actual predictive match between model and machine. A preliminary version of such an approach was demonstrated at FACET where a non-invasive energy spread spectrum was used to adaptively tune a model in real time to give predictions of the LPS of the electron beam which was confirmed with TCAV measurements [21]. This method was demonstrated to be extremely robust to time-variation of the system, further studies are being planned at LANSCE and FACET-II.

### Optimization for learning

Advanced algorithms have a role to play both in optimization experiments and learning underlying physics. For example, if an optimizer is given the freedom to adjust many parameters at once in order to, for example, maximize the energy gain while minimizing the energy spread of the beam exiting a plasma wakefield accelerator, the optimizer may find an unexpected and not analytically predicted configuration which achieves better performance than model-based methods. Studying this configuration will then teach us either about the machine and what components are not operating in an ideal way and therefore requiring non-ideal component settings, or, more interestingly it could possibly find a new configuration based on physical affects that were not considered in simplified models, which, when understood may lead to the discovery of new methods and approaches to the problem being studied.

### Differentiable programing

Many machine learning codes, such as the Tensorflow environment created by Google for the development of neural networks, are differentiable environments. In these codes, once equations are written, they are implemented as directed graphs with analytically determined derivatives computed for each graph segment, resulting in input-output maps which are automatically differentiable. One very useful feature of this that should be taken advantage of is, for example, the creation of a neural network to learn the mapping between various accelerator components and beam characteristics such as emittance or energy spread. This mapping will then be differentiable and therefore it will be possible to directly differentiate, for example, energy spread versus a particular magnet settings, for a very large complex system based only on data gathered from a real machine. This approach would then provide a differentiable map which would be very useful for everything from parameter sensitivity studies to optimization.

### Time-varying systems (distribution shifts)

Although ML techniques such as deep convolutional neural networks (CNN) have recently become popular due to a growth in computational power, their implementation by the accelerator community so far has been limited to a few test cases. One of the main challenges of applying ML techniques to accelerator applications stems from the fact that the performance of model/data-based ML techniques, such as CNNs, suffers when the system for which they have been trained quickly changes with time. Accelerators suffer from time-varying distribution shifts of, for example, the phase space of the particle bunches entering the accelerators, and also of the performance of the hundreds of magnets and RF systems, which drift with time. Approaches for handling distributions shifts include:

**1). Adaptive ML:** To supplement ML approaches, local, model-independent feedback algorithms exist in the feedback control theory community, that can be extremely helpful for optimizing and tuning noisy complex systems and to adapt to time-varying features and distributions [22–29]. One application of this was demonstrated at the

LCLS for automatically tuning the FEL to achieve desired longitudinal phase distributions [1].

**2). Gaussian processes (GP):** When statistically significant amounts of data can be gathered relating a set of parameters to each other and to a particular objective function (such as quadrupole magnets to X-ray pulse energy), GP methods may be used to give predictions for component settings as well as uncertainty bounds. Such methods are being developed and tested at the LCLS greatly outperforming standard methods such as simplex and robust conjugate direction search (RCDS) [30, 31].

**3). Re-training:** Whether using a particular data set collected from an accelerator over a certain period of time or if learning based on simulation data, it may be possible to re-train only a small final section of a pre-trained NN to give better predictions relative to the latest collected data in order to re-learn continuously as things vary with time. However, for very large numbers of parameters and large networks this may be infeasible as it would take too much data and too long to re-train relative to how fast the system is actually changing. Currently such methods are used for static problems, such as image identification and classification.

**4). Domain transfer:** Another approach for NNs is to initially train based on a very large amount of simulation or experimental data and then quickly train a much smaller network based on new data, a new network which will map readings to their simulation-based counterparts to be fed into the NN. So, for example, if the NN was trained to map a particular magnet setting to a phase space distribution, but that relationship later changed, a smaller domain transfer NN, such as a U-net, could be trained to map the new magnet setting to the old simulation-based setting, to be fed into the NN to predict the new phase space. Such a domain transfer NN would have to be continuously re-trained on new data. Again, this may become infeasible for a NN handling a large number of parameters because the retraining may take too long or would require too large of a manual grid-scan of the parameter space. Currently such methods are used for static problems, such as image identification and classification. One particularly interesting recent use of this method, which has great potential for particle accelerator applications, was to utilize large amounts of simulation data to initially train a CNN to map diffraction images to crystal orientations and then apply both re-training and domain transfer to tune the CNN or to adjust its inputs for it to accurately work on actual experimental data which was not exactly matched by the simulations [32].

### Workshop-based Collaboration

One direct result of this workshop was the establishment of a collaboration between Los Alamos and CERN. CERN invited A. Scheinker to perform automatic optimization and tuning studies on the electron beam line of the CERN AWAKE plasma wakefield accelerator experiment.

Working together with CERN scientists for one week they were able to implement a LANL-developed adaptive extremum seeking (ES) tuning algorithm on the CERN machine. In this work, two ES algorithms were run simultaneously at two different time scales. One ES algorithm, ES<sub>1</sub> slowly tuned 5 parameters: 2 solenoid currents and 3 quadrupole magnets to adjust the transverse phase space of the electron bunch, in order to minimize the size of the electron beam at the end of the beam line. However, while this algorithm was running, it not only affected the transverse phase space, but also the longitudinal: the energy of the beam and therefore its orbit was being modified and the beam was moving away from its design orbit and off of the measurement screen at the end of the beam line. A second ES algorithm, ES<sub>2</sub> was also run, operating  $\sim 3\times$  faster than ES<sub>1</sub>. The second algorithm, ES<sub>2</sub> adjusted 9 steering magnets based on 9 beam position monitor readings to continuously maintain the desired orbit, while ES<sub>1</sub> worked to minimize the beam size. The overall result was a beam size approximately  $2\times$  smaller than what was previously achieved. The experimental results are shown in Figure 5 and are being prepared for a journal publication.

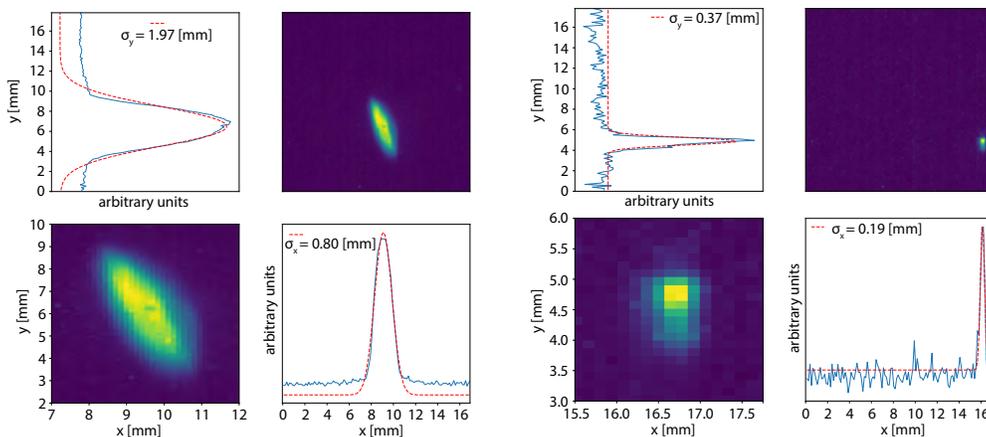


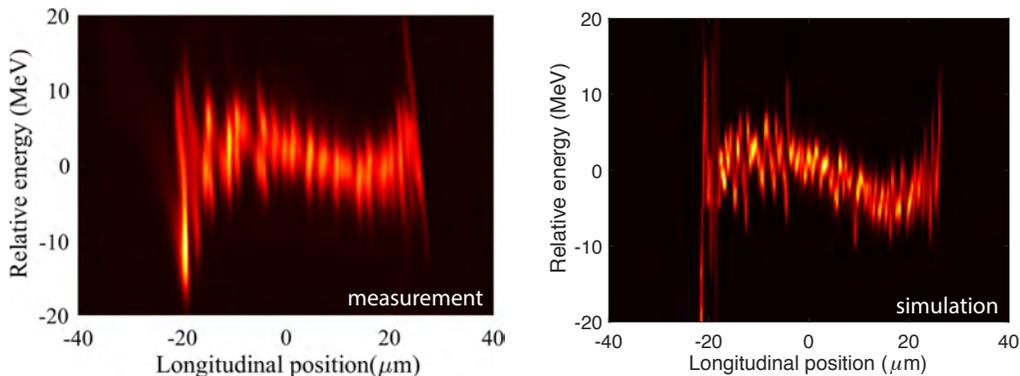
FIG. 5. Initial beam shown on the left, final beam to which the algorithm converged is shown on the right.

## V. BEAM DYNAMICS

- Dynamics in high brightness, low energy electron beams is a challenge for extremely bright electron sources. For example, the LCLS-II injector is being designed to operate at a 1 MHz repetition rate with bunch parameters summarized in the table below. Such an injector capable has not yet been demonstrated [33]:

Bunch Charge [pC]	Peak Current [A]	Normalized Slice Emittance (95%) [ $\mu\text{m}$ ]
20	5	0.25
100	10	0.4
300	30	0.6

- Emittance improvement is extremely important and would allow for the generation of higher energy light for the same energy beam. When the emittance is too high, particles with large amplitude motion slide out of the bucket. At LCLS-II they expect to generate 12.8 keV light with an 8 GeV beam. If they can decrease the electron beam's emittance by 50% they may be able to use the same energy electrons to create 18 keV light. High quality factor (Q) high efficiency super conducting cavities have extremely narrow bandwidths and therefore slow response times to control signals and have extremely demanding resonance control issues. For example, the LCLS-II accelerator will utilize cavities with  $Q \approx 3 \times 10^{10}$ . A pulsed machine with such high Q cavities would require advanced adaptive feed-forward iterative learning controllers to compensate for Lorentz force detuning. High current beam quality needs to be preserved over kilometer length scales despite wakefields, space charge forces, and coherent synchrotron radiation (CSR).
- Bunch compression is extremely important for all accelerators. Improved diagnostics for feedback-based tuning could potentially help achieve higher compression ratios while maintaining beam quality. Major beam dynamics challenges include optimal matching into bunch compressors, microbunch instability and beam halo. Detailed simulations are too slow. Although simulations have now been performed with the same number of macroparticles as particles in accelerators, they require tens of hours even when utilizing very large super computers and therefore cannot aid in online accelerator tuning and optimization [34]:



- Variable gap undulators add flexibility to FELs, allow for keeping beam energy constant while tuning X-ray wavelength. These devices introduce large numbers of coupled components that require advanced tuning algorithms for automated adjustment and optimization.
- Modern injectors are good places for developing and testing adaptive tuning techniques with few knobs and many diagnostics. Superconducting RF (SRF) systems have only local diagnostics, global tuning is difficult and must handle slow variations throughout facility, could greatly benefit from ML approaches. More precise and automated pulse-length controls are desirable for all facilities. More complex tuning procedures, such as multi-bunch/multi-color modes at FELs require advanced tuning algorithms for multiple accelerator subsystems simultaneously including the injector laser, injector, magnets, and RF systems.
- Plasma wakefield accelerators (PWFA) face several major challenges, the most general one is finding the an optimal approach to plasma wakefield acceleration of electron beams. There is also the unsolved problem of positron acceleration, which turns out to be much more complicated than electron acceleration. The goal is large accelerating gradients, the processes need to be efficient ( $> 50\%$ ), need low energy spread ( $< 1\%$ ), and need emittance preservation. All of these methods have extremely large parameter spaces over which it is impossible

to do a brute force search. Detailed physics simulations are too slow, experiments are too few and have limited diagnostics. There are analytic models for: charge, bunch radius, bunch length, emittance, long. profile, bunch separation, and plasma density, but advanced algorithms have a role to play because all of the analytic models make simplifying and idealized assumptions. Advanced algorithms can be used to generate general longitudinal profiles for PWFA experiments. An initial guess based on theory for parameters may be used to give gaussian beam + trapezoid, then fine tuning with adaptive methods.

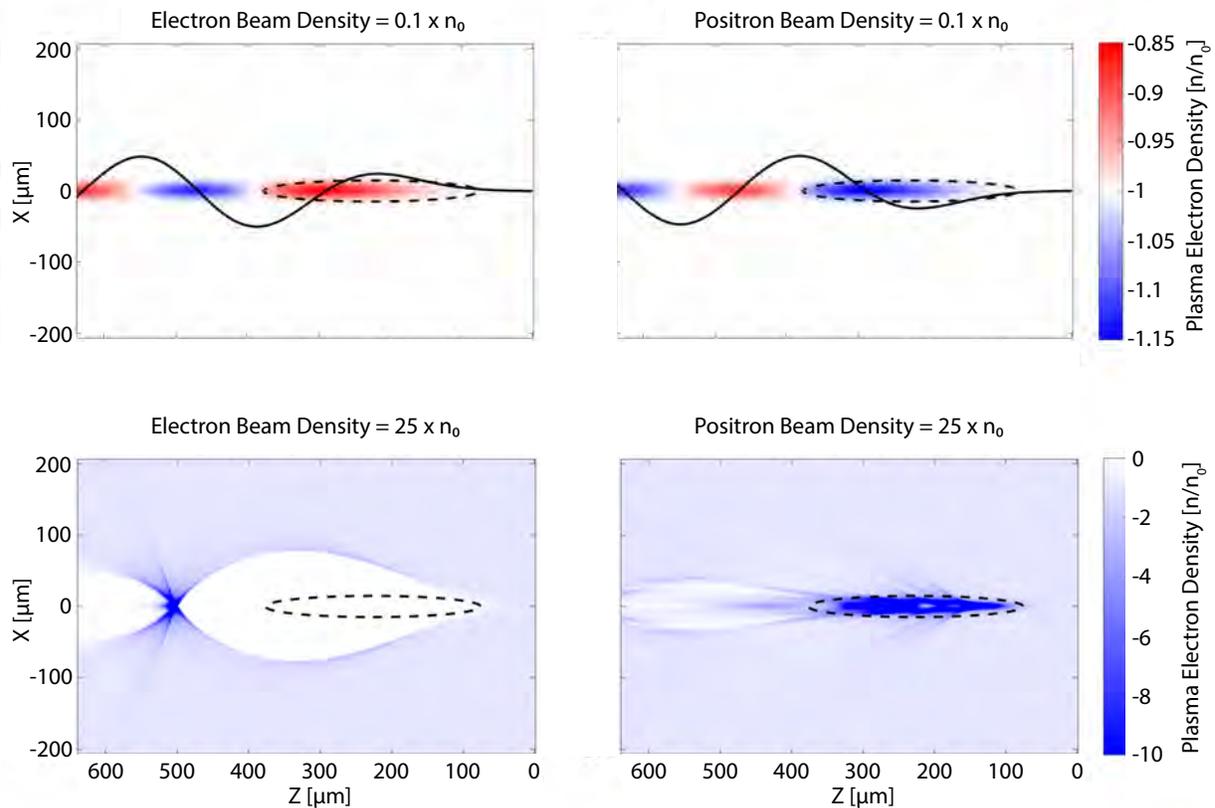


FIG. 6. Figure from Spencer Gessner's talk, at low space charge (top) a plasma's response to electrons is symmetric relative to its response to positrons. With higher beam charge, a large difference between accelerating electrons and positrons in a plasma becomes apparent (bottom).

- Furthermore, the analytical methods are limited because things are extremely complicated... e.g. focusing and matching conditions + preserving beam emittance (parameters are all wrapped up together). One possible solution is to just optimize on the final signal (energy gain and energy spread). Sometimes the algorithms might point us toward interesting places to look to understand the physics better (e.g. history with wakefield sims / new regimes).
- There is resistance to using learning and optimization algorithms to achieve core physics goals...but using optimization as a sub goal is acceptable. **Optimization gives us an opportunity to learn if the physics is missing from the model if the optimal settings do not correspond to the optimal experimental outcome.** It is essential to come up with good tests and show convincingly where these approaches work. For example, there is a claim that beam offset and emittance cannot be simultaneously optimized based on analytic description, but solutions were found using more expressive physics models. Warm start approaches are necessary to get close to expected optimum, then explore from there.
- Given some cost function can we run multiple PIC simulations with an optimizer to learn from the results what makes a good plasma accelerator? Use simulation + optimization to guide our understanding of why those optima work. Simulations are really slow for 3D pic, plus few opportunities to run beam (e.g. AWAKE). How to choose the next setting for the experiment? One approach is to use a very simplified model to give a rough estimate, then do detail simulations or experiments on fewer cases to zoom in on optima. Advanced algorithms have a role to play both in opt. experiments and learning underlying physics.

- The Plasma Photocathode is an advanced PWFA technique in which two gases are used with different ionization thresholds. One laser pulse then creates a plasma bubble, exciting plasma wakes in the lower threshold gas, once the bubble is excited and moving, a second laser pulse with higher energy excites electrons in the second gas, creating a plasma cathode with extremely high electric field gradients (GV/m) with the potential of creating beams with much higher 6D brightness because a relatively cool beam is born directly inside of the accelerating bubble. A beam driven version of this approach, where a driving particle bunch creates the plasma bubbles, requires extremely precise control of the 6D phase space of the driving beam.
- Phase Space control is extremely important for FELs, but there are major limitations on bunch lengths for FELs. While LCLS has 10 fs resolution the goal is 100 as resolution. Soft x-rays can't be shorter than 1 fs. But hard x-rays can and with emittance spoiling can get down to 200 as. Enhanced SASE technique allows you to go to short pulses (100as) even at soft X-rays. The basic principle is to make a very high current spike by modulating the electron beam with a laser in a wiggler then sending it through a chicane to compress. With external laser modulator you get 10s of uJ per pulse with at least 2 spikes lasing. Alternatively you can use coherent radiation generated by the tail of the electron beam to modulate the energy in the wiggler. The cool thing about that is that it's stable (modulation locked to the tail) and you can get only a single spike lasing 350 as pulse. Attosecond FELs improve peak power by 4-5 orders of magnitude compared to HHG in gas. Can do two-color as well, will upgrade XLEAP-II to do high power two-color including near TW level peak power. **Opportunity:** Challenges for tuning XLEAP - very nonlinear phase space and FEL dynamics, can we use ML models to improve the performance/design the FEL architecture (chicane R56, tune the space charge chirp etc.) for XLEAP?
- In general, the e-beam out of injector doesn't always have the distribution needed for a given application. Beam Shape control and phase space manipulation are required, some examples of beams that might be wanted are: flat beams, compressed beams etc. Transverse to longitudinal phase space exchange techniques exist, such as emittance exchange (EEX) for precise beam profile control and longitudinal bunch shaping [35]. These techniques work by inserting masks in the beam path and utilizing quadrupoles for focusing / defocusing. There is a dedicated EEX experimental beam line at the Argonne Wakefield Accelerator (AWA). Laser shaping has also been combined with EEX for precision control of the electron longitudinal bunch shape [36].

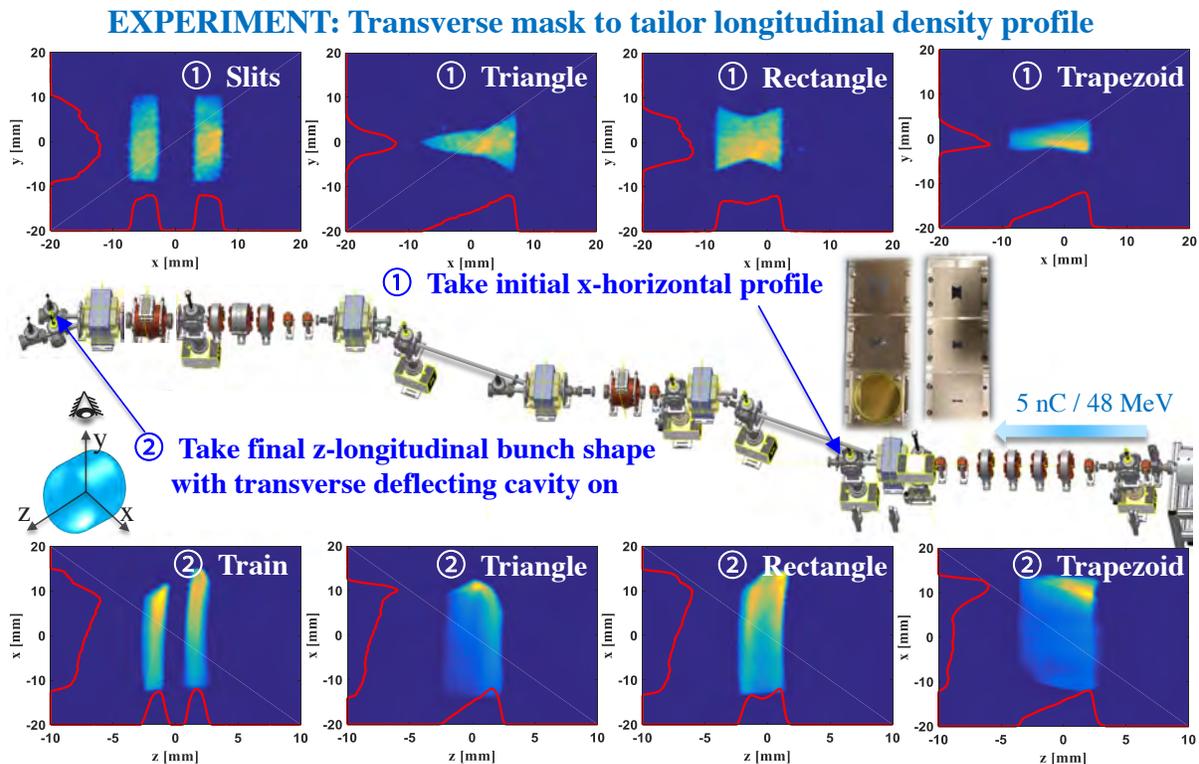


FIG. 7. Figure from [36] on longitudinal phase space bunch control.

- Laser plasma acceleration (LPA) has its own set of unique challenges. The BELLA laser plasma accelerator creates gradients of  $\sim 30$  GV/m. One stage of LPA has been demonstrated to create 10 GeV bunches for few pC to nC bunches with sub micron to 10 micron source size, mrad source divergence, and 1-10 micron bunch length (which depends on bubble size, such short bunches have very high current). However the energy spread of the bunches produced is  $\sim 50$  % for bunches and the shot-to-shot jitter is extremely large. Creating controllable mono-energetic beams depends on controlling and stabilizing the lasers and precisely controlling plasma density with density down-ramping. These techniques are also limited in terms of diagnostics. One method for controlling injection is by using blade in a gas jet. Moving the blade around tunes the energy of the accelerated beam and optimizes steering. It is important to measure emittance of LPA beams - can use butterfly technique. Showed that down-ramp injection gives better emittance. Butterfly diagnostic can be used as a knob to optimize beam delivery. Active plasma lenses (APL) are a useful diagnostic for beams. Radially symmetric focusing, tunable with kT/m gradient. Used in LPA staging experiment and used as a high resolution GeV magnetic spectrometer. Used APL to measure emittance of GeV class beams at BELLA  $\sim 5$   $\mu$ m. Liquid crystal 20nm thick based plasma mirror used to deflect PW drive laser - a good diagnostic that doesn't destroy the emittance of the beam.

VI. DIAGNOSTICS

- Most existing non-invasive diagnostics are limited and provide only bunch-averaged information, such as beam centroids provided by beam position monitors (BPMs). For many accelerator tuning tasks, such as orbit control, BPMs are sufficient and can be used to gain other information, together with higher order mode (HOM) couplers can detect beam being off center in an RF cavity. FAST BPM measurements moving steering magnets work as expected. Work has been done to steer the beam through a section of the accelerator and find the minimum HOM signal, this tells you you're on center. Then, sending a microbunched beam deliberately off-center shows oscillation of subsequent microbunches kicked by the wakefield of the head of the beam. This method can be used to assign where initial deflection occurred to find badly behaving beam line elements. A framing camera can image microbunches individually, this was demod at FAST, plotted centroid position of each microbunch, verified with BPM readings (these are highly correlated, plus framing camera gives much more info about what the profile of the microbunch looks like).
- There are also computational approaches being considered for diagnostics, one of which is to attempt to train a neural network to map accelerator parameter settings to longitudinal phase space (LPS) distributions, this approach has been demonstrated for simulations of the LCLS and FACET-II and in hardware at the LCLS [20] in which a trained NN was able to predict LPS distributions based on accelerator parameter settings, as shown in Figure 15.
- Another approach is to adaptively tune accelerator models based on information non-invasively gathered about the beam to get an actual predictive match between model and machine. A preliminary version of such an approach was demonstrated at FACET where a non-invasive energy spread spectrum was used to adaptively tune a model in real time to give predictions of the LPS of the electron beam which was confirmed with TCAV measurements. This method was demonstrated to be extremely robust to time-variation of the system, further studies are being planned at LANSCE and FACET-II. Results from preliminary work [21] are shown in Figure 8. As the simulation is adjusted to match the energy spread spectrum prediction to the measurement, it is able to track the actual LPS of the electron beam as accelerator components are changed over time, predicting single and double bunch configurations and bunch widths.

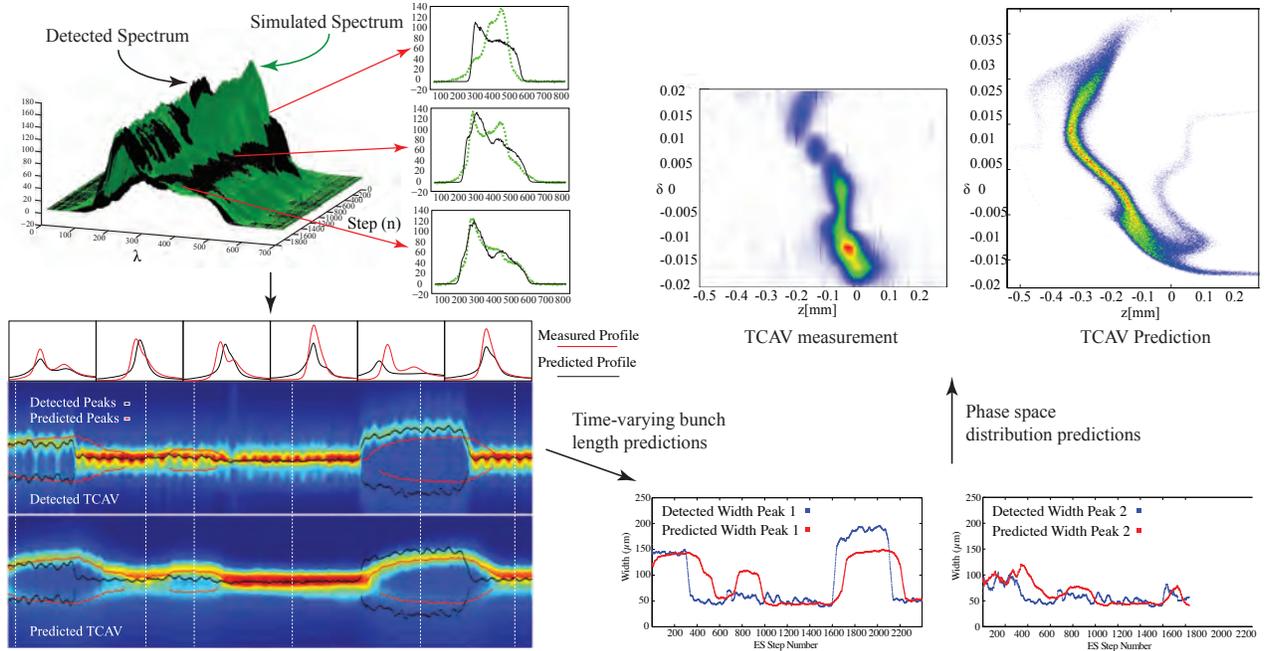


FIG. 8. In this approach the adaptive tuning algorithm adjusted the parameters of a simulation until the non-destructively measured energy spread spectrum of the beam was matched by the simulation’s prediction of this spectrum. Once the simulated spectrum converged to the measured spectrum the adaptive scheme was continuously run in real time while accelerator settings were adjusted and was able to track changes predicting the LPS of the beam as it changed from single to double bunch mode. Figure from [21].

- Optical Diffraction Radiation (ODR) is a non-invasive method for transverse beam profiles measurement. A beam grazes a metal edge, a few projections are all that's needed to reconstruct. Multiple 1D  $x$  profiles at different locations in the beam line are recorded to reconstruct 2D  $(x, x')$  phase space distribution. Multiple diagnostics can corroborate or disagree with each other - understanding the root source of signatures can help with feedback + optimization. Bunch-to-bunch data at  $< 1 \mu\text{s}$  time scales available. If there is enough charge in a micro-bunch, then if you cause the bunch to undergo a phase advance between ODR stations, you can get a reconstruction of a single bunch non-invasively.

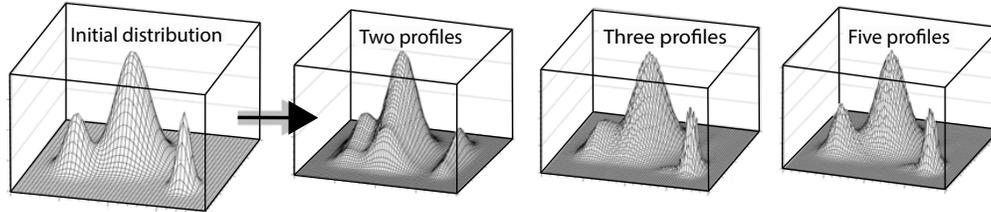


FIG. 9. Figure from talk of Kip Bishofberger on application of ODR for  $(x, y)$  phase space reconstruction.

- Non-invasive measurements for PWFAs. Want transverse and longitudinal measurements non-invasively shot-to-shot. PWFA Diagnostics have several issues:
  - General issue: PWFA has small spatial scale. Diagnostics are hard.
  - Beam alignment: need to separate beams to measure accurately. Jitter tolerances depend on BPM uncertainty?
  - Beam size: What is uncertainty on beam size while still preserving emittance?
  - Beam size measurement: limited by optics. can't do much better than two microns.
  - Butterfly emittance: use energy spread to fit emittance.
  - Uncertainty for butterfly: consider resolution issues. No consideration of matching for this case. "Best possible scenario" is limited by diagnostic resolution.
  - Need to measure beam profiles... TCAV description. TCAV needs stable voltage. Resolution is 0.3 microns.

Need drive and witness beam characterization on every shot (edge radiation, butterfly emit. diagnostic). FACET: major challenge for diagnostic and control - stabilize LPS and compression against drift. LiTrack + adapt parameters to match the spectrum, also match the TCAV (non-invasive estimate of destructive TCAV diagnostic). For all adaptive methods, the convergence rate and accuracy is sensitive to initial guess. These methods will be tried at FACET-II, but collective effects may make this a challenge (plus also won't include CSR, microbunching etc). FACET-II has different challenges: want to measure LPS and stabilize compression with respect to shot-to-shot jitter of linac parameters, and measure emittance shot-to-shot. Virtual diagnostics for LPS studied in simulation with jitter estimates (use ML model based on Lucretia runs). Virtual diagnostic designs need to consider several factors:

- How many shots to feed ML model to get accuracy, it took approximately 600 in a demonstrated simulation-based case (will change depending on the configuration).
- Need to be able to trust the virtual diagnostic, how to flag the shots we trust, also can use redundant diagnostics (need a measure of confidence).
- Sensitivity of model to different input diagnostics, need BC20 peak current measurement (e.g. how close to full compression).
- also looking at for two bunch case including smearing effect from TCAV
- Does TCAV image give us an answer that matches what we expect at the IP: sort of, after peak current of 30kA they don't match - need to augment with correlations with other measurements, or spectral measurements, or simulations.
- Want a THZ spectrometer to give us a second estimate (if difference between the two is very large, could throw out that shot).
- For LCLS case measured data, but also found that sometimes it does not work - sometimes the TCAV and bunch length monitor do not agree, and these give bad shots - also need to flag these measurements.

For emittance measurements, edge radiation would enable doing tuning based on emittance at multiple points along machine (simple fitting not enough to analyze image). Off-axis undulator radiation bunch length diagnostic is another possibility. The off-axis undulator radiation pattern from short electron bunches is proportional to the bunch length. Measuring the 2D radiation distribution and the integrated energy gives a measure of relative bunch length. In principle is non-destructive and can be used to measure very short bunches  $\sim 1\mu\text{m}$  in length, but iterative phase reconstruction methods still need to be proved with real bunches or realistic simulations.

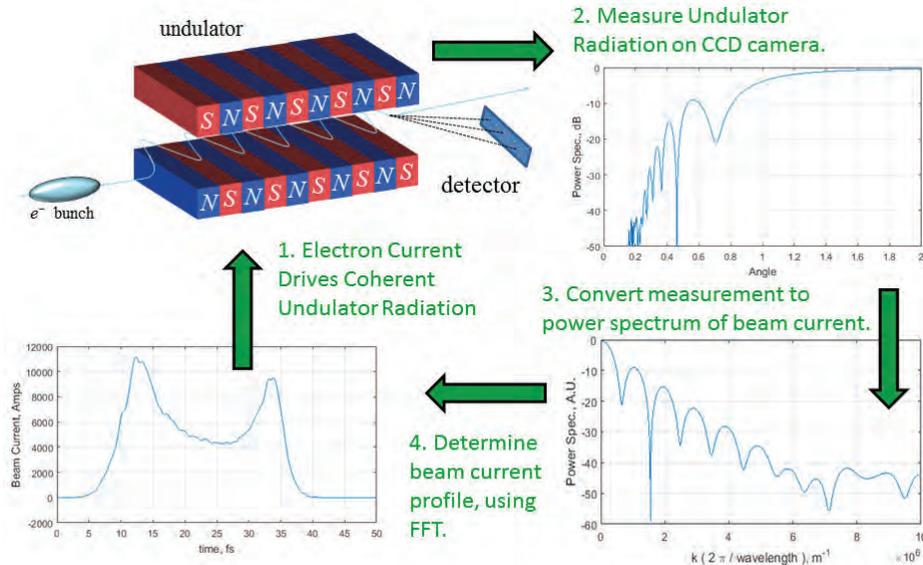


FIG. 10. Figure from talk of Quinn Marksteiner on using off-axis undulator radiation for bunch profile predictions.

- The ability to have autonomous accelerators is growing in importance. An autonomous accelerators: Self optimizes, self-adapts to task, finds workarounds to operating problems, makes go/no-go decisions for operating. Complexity vs consequence - accelerators are getting more complex and the consequences more serious. Example: space-borne linac for magnetotail studies - low data rates  $O(60)$  independent RF structures and can't be run from the ground (1/2 sec radio delay). This linac must operate autonomously. Medical device / food sterilization example: accelerator needs to be aware of factors (what am I sterilizing, how to efficiently set the dose, is it safe for me to run?). Accelerators for cancer therapy: Basic research needs workshop comments: doctors want more more powerful doses to better kill cancer cells (this is dangerous potentially deadly). They want improved collimation systems. Waste treatment plant example: Chicago plant processes  $1.5e9$  gallons/day of wastewater. Requires huge doses (14 kGy) and 500 MW of radiation power to sterilize. Compact, high power complex accelerators are here, specialists not always there to help diagnose a problem/repair a fault. Safety concerns important and have to be addressed.

## VII. ADVANCED CONTROL METHODS

- Machine learning (ML) branches from statistics and regression. ML-based tools, such as neural networks (NN), can be trained to automatically tune and control large complex systems such as particle accelerators [37–40]. ML tools are being developed to provide surrogate models to create diagnostics [41]. A NN model has been designed to predict the resonant frequency of the radio frequency quadrupole (RFQ) in the PIP-II Injector Experiment (PXIE), to be used in a model predictive control scheme [42]. In a preliminary simulation study for a compact THz FEL, a NN control policy was trained to provide suggested machine settings to switch between desired electron beam energies while preserving the match into the undulator and a fast surrogate model was also trained from PARMELA simulation results in order to facilitate the training of the control policy [43].

Powerful NN tools have also been developed for ML-based longitudinal phase space prediction of transverse deflecting cavity readings in particle accelerators, which are some of the most important diagnostics that exist for measuring a beam’s longitudinal phase space [20]. A novel Bayesian optimization framework that uses sparse online Gaussian processes has been applied for quadrupole magnet tuning in an FEL [30, 31]. Various ML tools, including clustering for identifying faulty beam position monitors (BPM) using outlier detection and ML methods for optics corrections has been developed and performed at CERN [44–48].

Although there are many ML architectures, neural networks (NN) are the most successful, in particular convolutional neural networks (CNN). The biggest lie in ML is that you sample things that are independently identically distributed, nature is not really like this, e.g. would be equivalent to saying that there is no machine drift in accelerators. NNs are optimized using (stochastic) gradient descent on your average sampled loss (error between known answers and predictions). There are many hyper-parameters that must be tuned for a NN approach:

- Breaking training data down into several individual batches. Batch size: are small batches better for finding good optima? Originally batch sizes came about because it is easier to parallelize over the batch during training, but max reasonable batch size is larger than previously thought (e.g.  $\sim 1000$ ) - noise scales proportionally to  $1/\sqrt{\text{batchsize}}$  times learning rate (keep in mind that learning rate and batch size are correlated, so don’t optimize these together during hyper-parameter tuning).
- Generalization error typically defined as the difference between the error on the test set and the error on the training set (you want this to be small).
- To get a good general model you want to sample your parameter space appropriately (i.e. not just over-sample the bulk of the distribution).
- There is a trade-off between network complexity (depth and width) and generalization (ICLR 2016 paper on over parameterization and 2018 Belkin paper explaining how networks can generalize despite being over-parameterized).
- If you have a fully connected network you don’t gain a lot from having deeper networks (e.g. not much difference between 3 and 10 layers). This is different if you have convolutional layers, then adding depth really helps.
- Very large NNs do not have local minima to be trapped in... because NNs are similar to interpolating functions (see Belkin et al arXiv 2018 and Geiger et al 2019).
- Infinitely wide neural nets are Gaussian Processes (Lee et al 2017 ICLR) fully connected networks reach the ‘infinite’ limit before CNNs.

When systems are changing with time, when there are distribution shifts one common approach is to retrain just a few of the layers of the NN on the new data, thereby using a warm start for network weights based on previously trained models. Furthermore, multiple networks, trained at different times, can be used simultaneously to give some uncertainty (check when network predictions disagree) and some may perform better than others depending on how the system has shifted. In addition to tracking mean of model ensemble, can also track the variance. Transfer learning: feature learning happens in large nets like Inception V3 trained on ImageNet -can use layers from a previously trained net and then re-train only on last layers to transfer to problems with smaller training set sizes. Robustness: models often do poorly in generalizing close to the distribution (e.g. poor performance when add noise to images), there will be classes of noise your model has not seen and it will not be robust to that. Data augmentation: very essential for improving accuracy and robustness (e.g. symmetries, random shifts and crops, cutouts / random erasing, mixup), but note some things like mixup have not been used for regression, but in some cases worth just trying for your specific case. Incorporating Symmetries: distance of test data from training manifold is important (e.g. can incorporate symmetry) - use components that commute with symmetries (equivariant) - convolutions are translation equivariant, but not scale and rotation equivariant.

Incorporating other Structures (e.g. info from domain science) can be very helpful. Differentiability - can write simulations in ML languages and then can differentiate these, can also replace pieces with NNs.

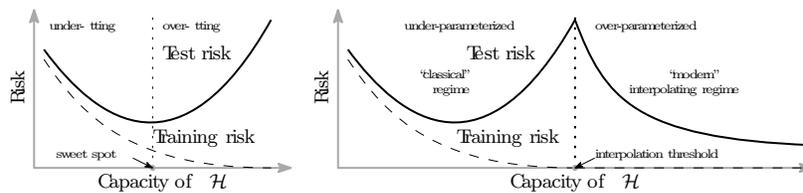


FIG. 11. Figure from [49] showing that unlike traditional regression approaches, once NNs become sufficiently wide and deep they can generalize and do not overfit despite such a large number of parameters.

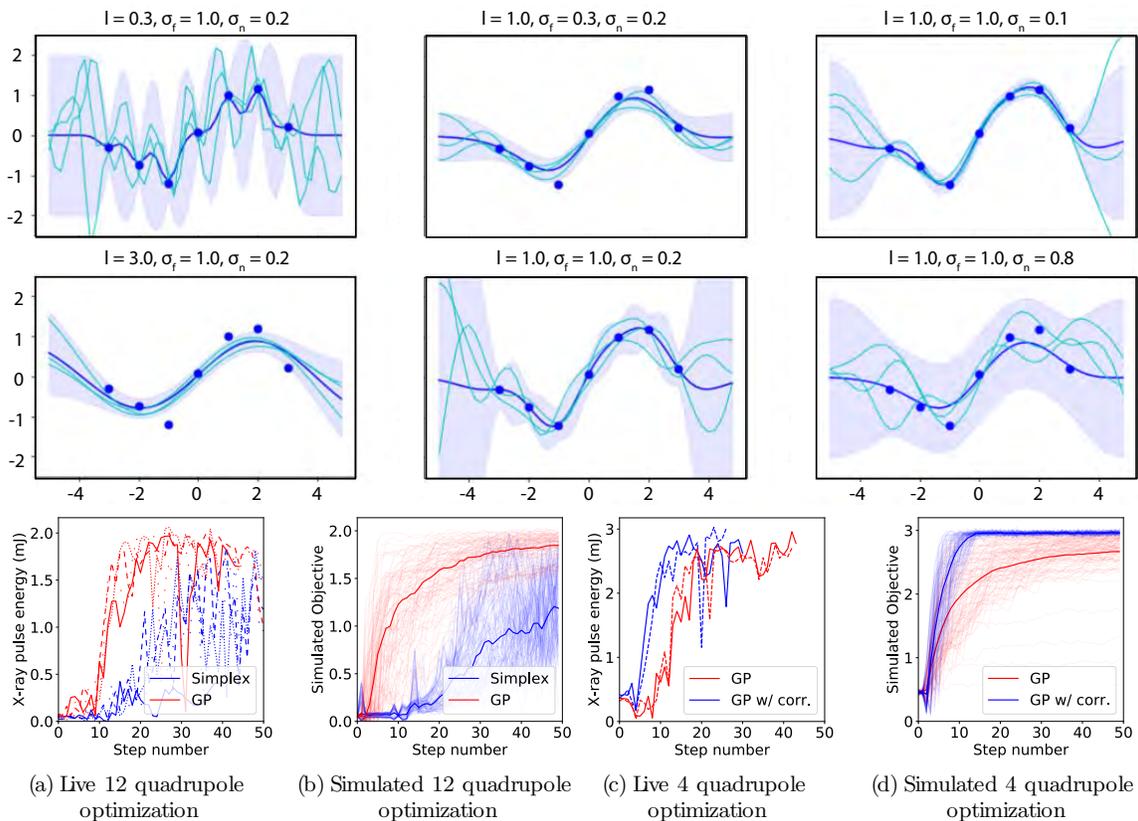


FIG. 12. Figures from work being done at LCLS [30] and presented by Adi Hanuka at the workshop. The Gaussian Process approach is much faster than Simplex and reaches a higher maximum.

- Large data, complex problems can benefit from automated feedback and Machine Learning (ML) approaches to detect and predict unwanted changes, extract more information from waveforms, signals, and images, improve the understanding of large, complex machines and automatically, quickly tune. For example, LCLS tuning requires  $\sim 500$  hours of beam time per year. Quadrupole magnets are tuned 2-5 times/day for 15 minutes at a time. Examples of ML techniques applied to accelerators:
  - Anomaly detection: CERN. Used clustering algorithms to find bad BPMs. “Isolation forests”, “Local Outlier Factor”, “DBSCAN”.
  - Extract information from diagnostics: LCLS. Use CNN to correlate images with x-ray power simulations.
  - Virtual diagnostics: FAST. Real diagnostics not always available.
  - Virtual diagnostics: SLAC/LCLS-II/FACET-II. Correlate fast parameters with XTCAV measurements.
  - Control and tuning: LCLS. Bayesian Optimization for FEL tuning at LCLS.

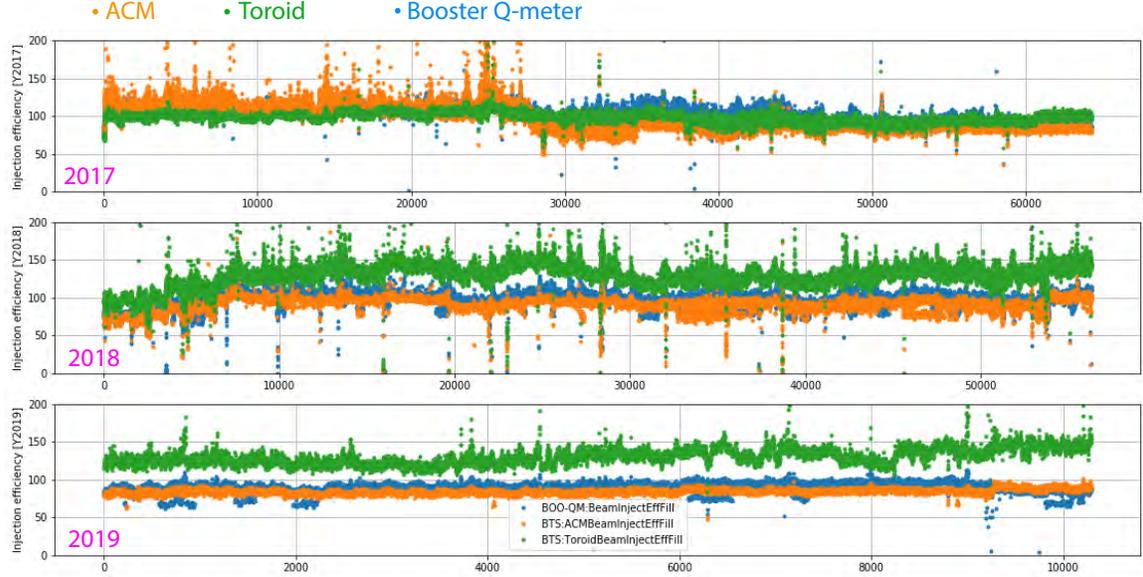


FIG. 13. Figure from presentation of Faya Wang at the workshop showing three years of three different measurements of injection efficiency. This data was analyzed with ML methods to identify optimal trajectory settings as a function of temperature.

- Control and tuning: LCLS. “Warm start”. A trained neural network gives an initial guess for parameter settings to achieve a desired LPS and then a local model-independent feedback, such as extremum seeking (ES) tunes based on real time machine data and performs feedback to continuously optimize the system despite uncertainty and time variation.
- Accurate machine models: LANSCE. GPU accelerated PARMELA.
- Online modeling: UCLA/Pegasus. Found new beam distributions.
- Online FEL tuning based on Bayesian optimization not only automatically adjusts parameters, but also gives uncertainty measures on prescribed settings. Bayesian optimization builds a probabilistic model and chooses sample points in parameter space in a way that minimizes uncertainty. One very powerful Bayesian approach is based on Gaussian Processes (GP), in which a kernel function is used, typically of the form:

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2}(x - x')^T \Sigma (x - x')\right) + \sigma_n^2 \delta(x - x'), \quad (2)$$

where  $\sigma_f$  controls the amplitude of fluctuations of the unknown function being modeled,  $\Sigma$  is a matrix of correlation and length scales which determines both the smoothness and coupling between separate components, and  $\sigma_n$  is the standard deviation of the measured noise in the experiment. GPs combine experience/data with fast decision making, and have been shown to perform much faster than simplex methods in work being done at LCLS. The figure below shows the difference in predictions based on the same 7 data sample points, given varying levels of noise,  $\sigma_n$ , amplitudes,  $\sigma_f$ , and length scales,  $l$ , which are arguments of the matrix  $\Sigma$  and also the results of applying GPs at LCLS from [30], as shown in Figure 12.

- Individual accelerators and companies are starting to develop ML tools that can be shared across the community, such as OCELOT. Radosoft is also working on several ML projects, providing “Sirepo,” an online platform for doing accelerator physics simulations. Can use ML through their Jupyter servers and reference accelerator simulations. For example, multi-slit emittance VD at FAST – updating with measured data (sim / training done on radosoft servers). Optimization of Thermionic Energy Converters - want to use ML surrogate models for optimization. Web-based toolkit for ML for accelerators (link to accelerator simulations + control systems including ACNET and EPICS) – data import/visualization – testing out in collaboration with fermilab (modular surrogate modeling + anomaly detection). Anomaly detection – model-based and clustering based – testing out at FNAL for beam loss monitors. Web-based toolbox to test accelerator control algorithms, directly interface with EPICS, improving data analytics tools for operators (dummy epics server running at radosoft right now).
- ML example use case at LCLS: SPEAR3 Injection Efficiency. At SPEAR3 the injection efficiency varies over time, going up and down by as much as 15%, and it is unclear why. The trajectory is frequently re-optimized,

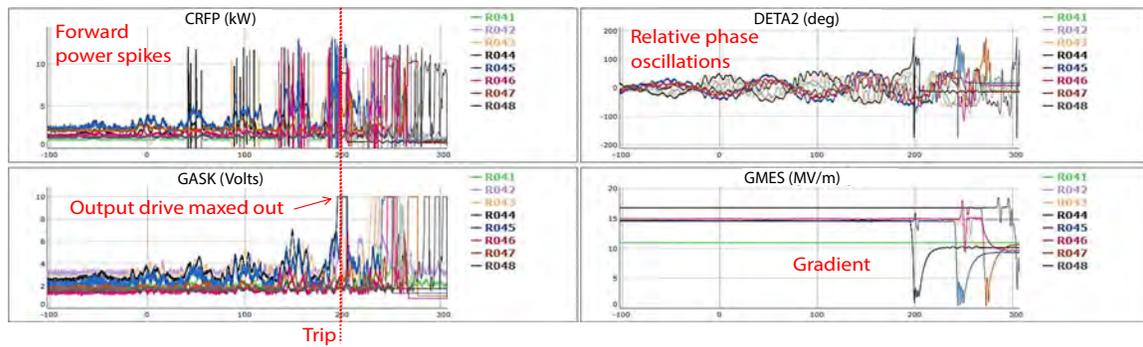


FIG. 14. Figure from presentation of Anna Solopova at the workshop showing various SRF fault types that need to be automatically identified.

it would be good to understand what drives trajectory changes. A look into historic data was taken to try to find out. The data set included all trajectory knobs, BL05 gap and phase, vacuum undulator gap, outdoor ambient air temperature and ground temperature (data from 2017, 2018, 2019) - 60k data points in total, FY19 10k points. Used three ways of measuring the injection efficiency (some noisier, some with calibration drift). High frequency jitter in Q meter reading – processing to remove some of the major jitter. Analyze data by year as well before applying ML (e.g. look at trajectory target during different years). NN model - related control parameters and environmental parameters to injection efficiency (then tried to examine sensitivity). Once NN model was trained, it was possible to take partial derivatives of the NN output with respect to input parameters and found the largest sensitivity to ground temperature. In each temperature zone ( $\pm 0.5$  deg C) they found the top 10 % injection efficiency and the corresponding ideal orbits.

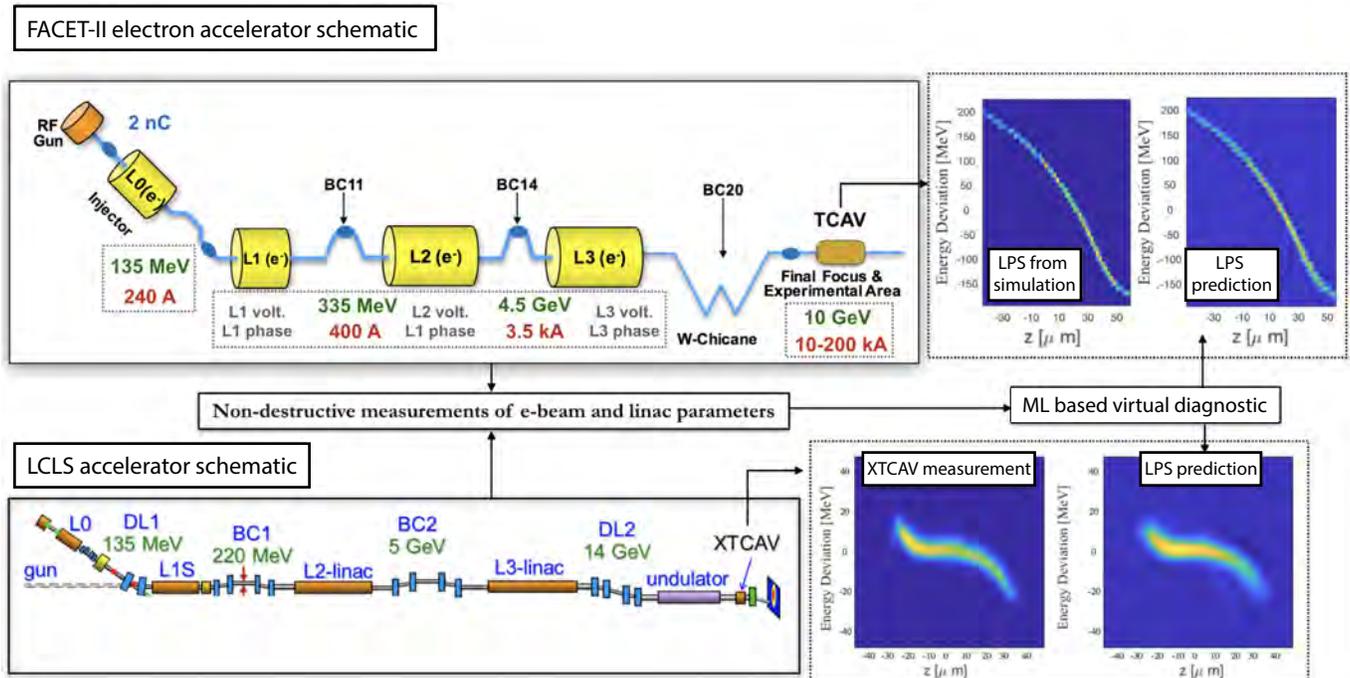


FIG. 15. Figure from [20] showing NN-based predictions of the LPS of FACET-II, compared to simulations, and of LCLS, compared to TCAV measurements.

- Model validation, optimization and uncertainty quantification using the Mystic package. Just using most ML methods, such as NNs, it is fundamental difficulty for ML to produce rigorously validated models of complex physics. Difficult to decide “How good is my ML model?” Usually we decide this post-learning but you would like to incorporate this in learning process. Would like to do this as a game-theoretic optimization. It is

possible to use properly chosen kernels to transfer nonlinear data into higher dimensional space and linearize your problem. Important to also use physics constraints to make penalty functions for your optimization. *mystic* has packages that easily incorporate many constraints for optimization, including Lipschitz cones in hyperspace and the ability to generate predictors that minimize the likelihood of misclassification. One example of *mystic* performing non-convex global optimization is shown in Figure 16.

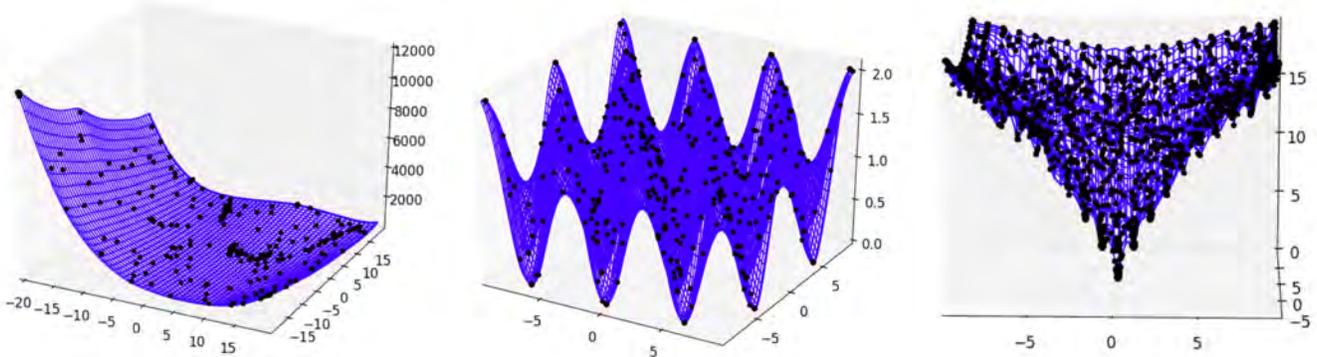


FIG. 16. The *mystic* framework provides way to incorporate physics-based constraints for non-convex global optimization.

- ML example use case at CEBAF: CEBAF - first large high power SRF machine in US, over 25 years operation. Cavities strongly coupled, when one shakes they all shake. Operators don't know which cavity caused the trip so they just reduce the gradient. To keep trips below 10/hr they have to run below design energy, mostly due to microphonics and field emission. There was a need to automate fault detection and classification. Different trips are accompanied by different features, 7 different types identified, some of which are shown in Figure 14. Questions: which cavity caused trip, what type of trip, what should we do? Used deep learning model and got good results for prediction. Want to use deep learning models to not only predict but prevent trips in future.

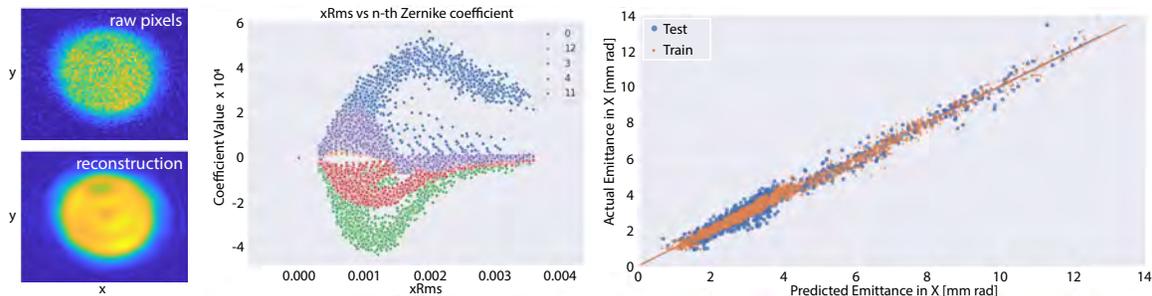


FIG. 17. Figure from talk of Jerry Ling on application of Zernike polynomials and NNs for emittance measurements.

- ML example use case at CERN: CERN has many coupled accelerators (beam prepared in one and sent to another) operating simultaneously and requires scheduled beam delivery to each from injectors. A supercycle is a communal schedule across the injector complex. The SPS, feeding the LHC is top priority, smaller injectors fill space in between. A supercycle has some requirements: each user is given a min num of shots/unit time, each experiment has minimum time between shots, magnet switching constraints. Currently a supercycle is built manually. Sometimes changed many times during MD (90 times/day record). Optimizing this gives users more data and saves money. They have created an optimization algorithm that gives the most beam to everyone. The algorithm is brute force and uses topological equivalence to be complete without being exhaustive. Algorithm uses a method called local 'signposting' to map the permutation space of decision points.
- ML example use case at SLAC: LCLS and FACET-II at LCLS are developing a neural network (NN) approach to predicting the longitudinal phase space (LPS) of the electron beam by training the NN to relate machine parameter settings to recorded LPS images, as shown in the results from [20] in Figure 15.
- ML example use case at CERN: at CERN there is work ongoing to predict measure beam emittance based on machine component settings, both using Zernike polynomial representations and a neural network. While both

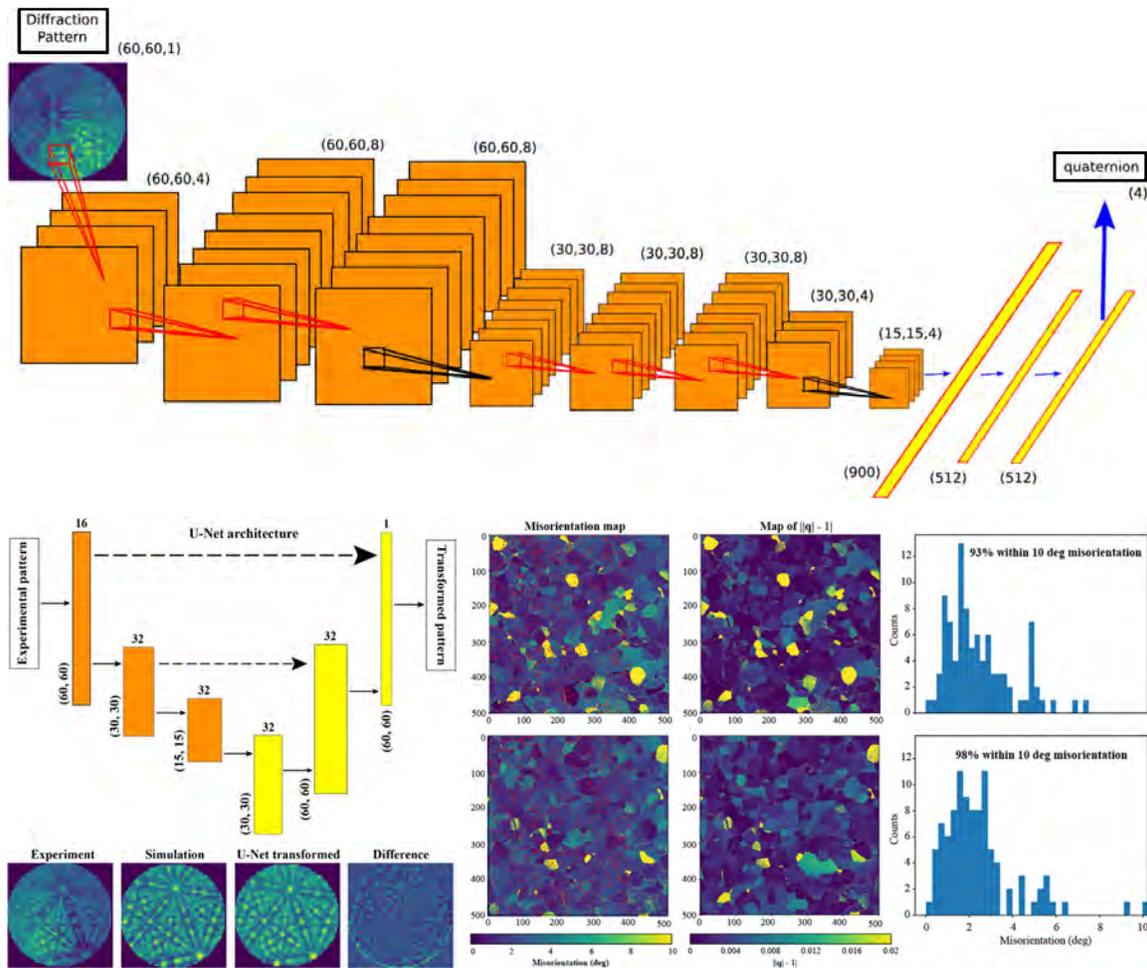


FIG. 18. U-net is trained to map experimental patterns to simulated patterns to be input into a simulation trained CNN. Another approach is to re-train just the last three layers of the simulation-trained CNN using  $\sim 1$  K experimental measurements. Both approaches predicted  $> 93\%$  of orientations within  $< 10$  deg error on experimentally collected data. Figures from [32].

methods have had good results with simulation data, the NN approach is more accurate, preliminary results are shown in Figure 17.

- Retraining and domain transfer for CNN methods: Recently, a novel ML approach to the mapping of diffraction images to crystal orientations for electron backscatter diffraction (EBSD) was demonstrated which sped up the reconstruction process by several orders of magnitude [32]. Because EBSD data is relatively expensive (time consuming, hours per single sample scan) to collect, the researchers started by creating hundreds of thousands (300 K) of diffraction pattern simulations. They then used pairs of diffraction patterns and their corresponding crystal orientations to train a deep CNN to automatically map diffraction patterns to orientations. Once the CNN was trained they tested 2 approaches to make the CNN useful for actual experimental data without having to rely on extremely large experimental data sets: 1). For the domain transfer approach they created a U-net, a smaller neural network which was trained to map experimentally collected diffraction images to their simulated counterparts, the training of this smaller network required only  $\sim 1$  K experimental diffraction patterns (compared to 300 K for the CNN). This approach allowed them to take raw measured diffraction images, map them to simulated counterparts, and then feed them into the CNN to accurately predict their crystal orientation. 2). In the retraining approach they again utilized only  $\sim 1$  K measured diffraction patterns to re-train only the last few layers of the CNN and it could then accurately map experimentally collected diffraction patterns to their corresponding crystal orientations. Such an approach has great potential to be used for particle accelerators for either direction: (A) Training a NN using simulation data if it is easy to generate and then re-training on a few actual accelerator measurements or (B) Training a NN using experimentally collected accelerator data for computationally expensive processes (space-charge dominated beams, CSR, etc...) and then applying re-training

to match it to particular complex simulations that are very computationally expensive to create new, faster, more accurate versions of these simulations.

- Adaptive machine learning: To supplement ML approaches, a local, model-independent feedback algorithm exists that can optimize and tune noisy complex systems and adapt to time-varying features and distributions [22–25]. This approach has been demonstrated for RF buncher cavity phase control [26] and resonance control [27] at the LANSCE proton linac, for minimizing betatron oscillations in a time-varying lattice in the SPEAR3 synchrotron [28], as a non-invasive diagnostic for FACET [21], and for output power maximization in both the LCLS and EuXFEL FELs [29]. Recently an adaptive ML approach has been developed in which a NN was trained to map longitudinal phase space distributions to the accelerator parameters required to achieve them [1]. The NN alone would not work because the system was drifting with time, but it did give a good approximation of the required settings in the global parameter space after which a local model-independent extremum seeking method is able to zoom in on and track the optimal time-varying parameter settings, as shown in Figure 19.

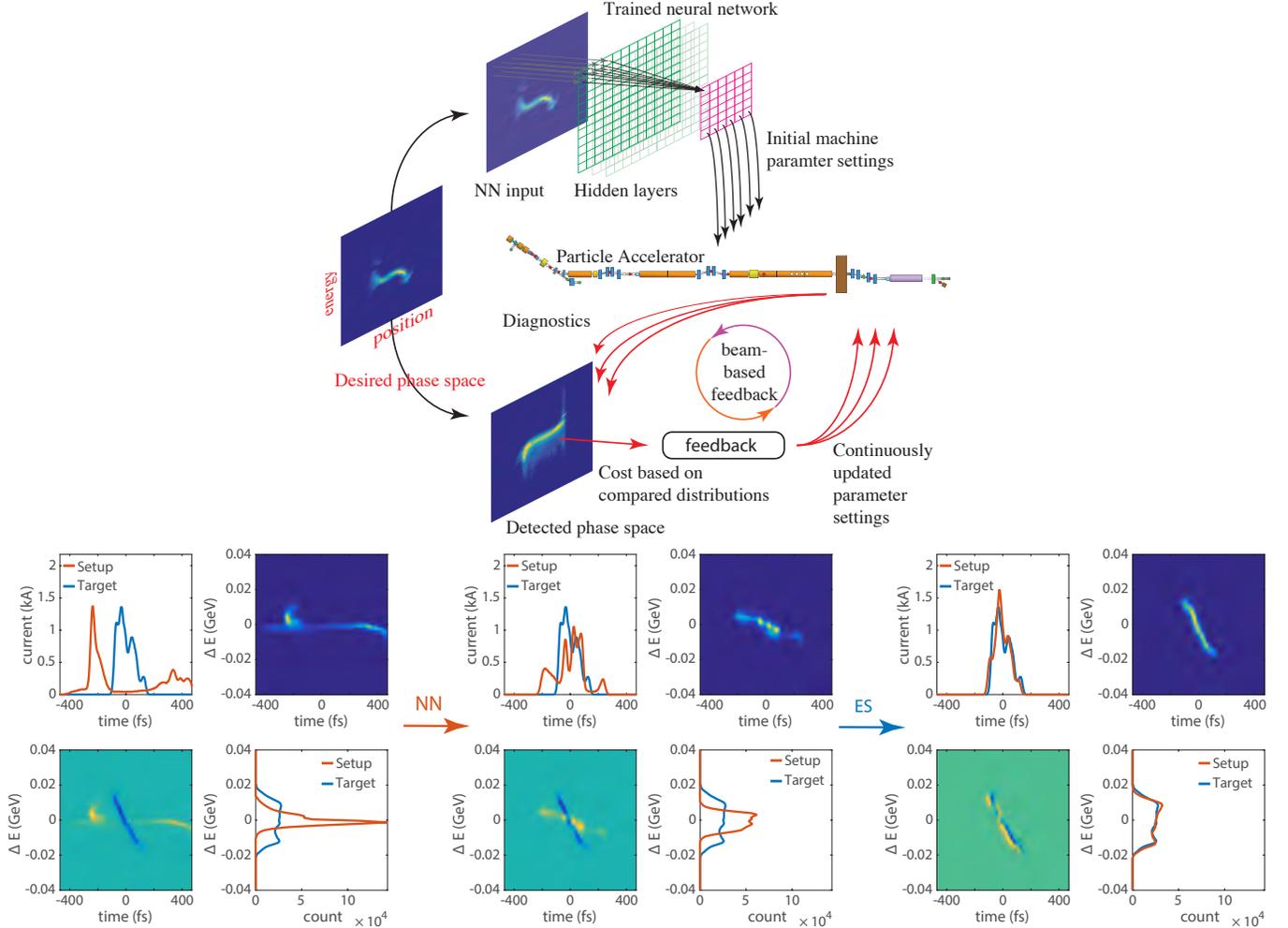


FIG. 19. **Top:** Overview of the adaptive ML approach in which a NN gives an initial guess and then adaptive feedback zooms in on and tracks optimal settings while both the initial beam conditions and accelerator components drift with time. **Bottom:** Experimental results at the LCLS FEL showing the initial guess of the NN after which adaptive feedback was able to zoom in on the optimal solution.

### VIII. LIST OF ACRONYMS

FEL	Free Electron Laser
linac	Linear Accelerator
LCLS	Linac Coherent Light Source
XFEL	X – ray FEL
EuXFEL	European X – ray FEL
LANSCE	Los Alamos Neutron Science Center
AWAKE	Advanced Proton Driven Wakefield Acceleration Experiment
FACET	Facility for Advanced Accelerator Experimental Tests
MARIE	Matter Radiation Interactions in Extremes
TCAV	Transverse Deflecting Cavity
LPS	Longitudinal Phase Space
fs	femtosecond = $10^{-15}$ seconds
SASE	Self – Amplified Spontaneous Emission
PWFA	Plasma Wakefield Accelerator
LPA	Laser Plasma Acceleration
ML	Machine Learning
NN	Neural Network
CNN	Convolutional Neural Network
ES	Extremum Seeking
CSR	Coherent Synchrotron Radiation
RF	Radio Frequency
SRF	Superconducting RF
EEX	Emittance Exchanger
AWA	Argonne Wakefield Accelerator
APL	Active Plasma Lense
BPM	Beam Position Monitor
HOM	Higher Order Mode
ODR	Optical Diffraction Radiation

- 
- [1] A. Scheinker, “Demonstration of model-independent control of the longitudinal phase space of electron beams in the Linac-coherent light source with Femtosecond resolution,” *Physical Review Letters* 121.4, 044801, 2018.
- [2] H. N. Chapman, “X-Ray Free-Electron Lasers for the Structure and Dynamics of Macromolecules.” Annual review of biochemistry, 2019.
- [3] R. N. Coffee, J. P. Cryan, J. Duris, W. Helml, S. Li, and A. Marinelli. “Development of ultrafast capabilities for X-ray free-electron lasers at the linac coherent light source.” *Philosophical Transactions of the Royal Society A*, 377(2145), 20180386, 2019.
- [4] J. P. Marangos, “The measurement of ultrafast electronic and structural dynamics with X-rays,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 377, 2019, <http://doi.org/10.1098/rsta.2017.0481>
- [5] Y. Ding, et al. “Measurements and simulations of ultralow emittance and ultrashort electron beams in the linac coherent light source.” *Physical Review Letters*. **102.25**, 254801, 2009.
- [6] D. Ratner, et al. “Experimental demonstration of a soft x-ray self-seeded free-electron laser.” *Physical Review Letters*. **114.5**, 054801, 2015.
- [7] P. Emma, et al. “First lasing and operation of an angstrom-wavelength free-electron laser.” *Nature Photonics*. **4.9**, 641, 2010.
- [8] W. Decking and T. Limberg, “European XFEL post-TDR description.” Technical note XFEL. EU TN-2013-004-01, European XFEL GmbH, Hamburg, Germany, 2013.
- [9] S. Serkez, et al. “Overview of options for generating high-brightness attosecond x-ray pulses at free-electron lasers and applications at the European XFEL.” *J. Opt* 20 024005, 2018.
- [10] H. Weise and W. Decking. “Commissioning and first lasing of the European XFEL.” Proceedings of the Free Electron Laser Conference, 2018. APA
- [11] J. Rzepiela, H. Loos, R. Akre, A. Brachmann, F. J. Decker, Y. Ding, P. Emma, A. Fisher, J. Frisch, A. Gilevich, and P. Hering, SLAC-PUB- 16643, 2016.
- [12] T. J. Maxwell, C. Behrens, Y. Ding, A. S. Fisher, J. Frisch, Z. Huang, and H. Loos, “Coherent-radiation spectroscopy of few-femtosecond electron bunches using a middle-infrared prism spectrometer” *Physical Review Letters*. **111.18**, 184801, 2013.
- [13] T. O. Raubenheimer in *Proceedings of the International Particle Accelerator Conference*, Richmond, VA, USA, 2015.
- [14] A. A. Lutman, T. J. Maxwell, J. P. MacArthur, M. W. Guetg, N. Berrah, R. N. Coffee, Y. Ding, Z. Huang, A. Marinelli, S. Moeller, and J. C. Zemella, *Nat. Photonics* **10.11**, 745 (2016).
- [15] A. A. Lutman *et al.*, Femtosecond X-rays from Fresh-slice Multi-stage Free-Electron Lasers, *Phys. Rev. Lett.*, accepted.
- [16] J. Amman, W. Berg, V. Blank, F. J. Decker, Y. Ding, P. Emma, Y. Feng, J. Frisch, D. Fritz, J. Hastings, and Z. Huang, *Nat. Photon.* **6**, 693, 2012.
- [17] M. Litos, et al. “High-efficiency acceleration of an electron beam in a plasma wakefield accelerator.” *Nature* 515, 92-95, 2014.
- [18] V. Yakimenko, N. Lipkowitz, C. Clarke, M. Hogan, G. Yocky, C. Hast, S. Green, Y. Cai, N. Phinney, and G. White, in *Proceedings of IPAC2016*, Busan, Korea, 2016.
- [19] C. Joshi, E. Adli, W. An, C. E. Clayton, S. Corde, S. Gessner, M. J. Hogan, M. Litos, W. Lu, K. A. Marsh, and W. B. Mori, *Plasma Physics and Controlled Fusion* **60.3**, 034001 (2018).
- [20] C. Emma and A. Edelen et al. “Machine learning-based longitudinal phase space prediction of particle accelerators.” *Physical Review Accelerators and Beams* 21.11, 112802, 2018.
- [21] A. Scheinker and S. Gessner, *Physical Review Special Topics Accelerators and Beams* **18 (10)**, 102801, 2015.
- [22] A. Scheinker in *Proceedings of the 2013 International Particle Accelerator Conference*, Shanghai, China, 2013.
- [23] A. Scheinker and D. Scheinker, “Bounded extremum seeking with discontinuous dithers,” *Automatica*, vol. 69, pp. 250-257, 2016.
- [24] A. Scheinker, Alexander, and M. Krstic. Model-free stabilization by extremum seeking. Springer, 2017.
- [25] A. Scheinker and D. Scheinker. “Constrained extremum seeking stabilization of systems not affine in control.” *International Journal of Robust and Nonlinear Control*, vol. 28.2, pp. 568-581, 2018.
- [26] A. Scheinker, et al. “In-hardware demonstration of model-independent adaptive tuning of noisy systems with arbitrary phase drift.” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 756, pp. 30-38, 2014.
- [27] A. Scheinker, “Application of extremum seeking for time-varying systems to resonance control of RF cavities.” *IEEE Transactions on Control Systems Technology* 25.4, pp. 1521-1528, 2016.
- [28] A. Scheinker, X. Huang, and J. Wu. “Minimization of Betatron Oscillations of Electron Beam Injected Into a Time-Varying Lattice via Extremum Seeking.” *IEEE Transactions on Control Systems Technology*, 26.1, pp. 336-343, 2017.
- [29] A. Scheinker, et al. “Model-independent tuning for maximizing free electron laser pulse energy.” *Physical Review Accelerators and Beams* 22.8, 082802, 2019.
- [30] J. Duris, et al. “Bayesian optimization of a free-electron laser.” arXiv preprint arXiv:1909.05963 (2019).
- [31] A. Hanuka, et al. “Online tuning and light source control using a physics-informed Gaussian process Adi.” arXiv preprint arXiv:1911.01538, 2019.
- [32] Y. F. Shen, et al. “Convolutional neural network-based method for real-time orientation indexing of measured electron backscatter diffraction patterns.” *Acta Materialia* 170, pp. 118-131, 2019.

- [33] Schmerge, John F., et al. The LCLS-II injector design. No. SLAC-PUB-16211. 2015.
- [34] Qiang, Ji, et al. “Start-to-end simulation of the shot-noise driven microbunching instability experiment at the Linac Coherent Light Source.” *Physical Review Accelerators and Beams* 20.5 (2017): 054402.
- [35] Nanni, Emilio A., William S. Graves, and David E. Moncton. “Nanomodulated electron beams via electron diffraction and emittance exchange for coherent x-ray generation.” *Physical Review Accelerators and Beams* 21.1 (2018): 014401.
- [36] Ha, Gwanghui, et al. “Precision control of the electron longitudinal bunch shape using an emittance-exchange beam line.” *Physical review letters* 118.10 (2017): 104801.
- [37] M. Buchanan, “Depths of learning.” *Nature Physics* 11.10, 798-798, 2015.
- [38] Y. B. Kong *et al.*, “Predictive ion source control using artificial neural network for RFT-30 cyclotron.” *Nuclear Instruments and Methods in Physics Research Section A*, 806, 55-60, 2016.
- [39] A. L. Edelen, S. G. Biedron, B. E. Chase, D. Edstrom, S. V. Milton, and P. Stabile, *IEEE Trans. on Nucl. Science* **63.2**, 878-897 (2016).
- [40] A. Edelen, et al. “Opportunities in Machine Learning for Particle Accelerators.” arXiv preprint arXiv:1811.03172, 2018.
- [41] A. L. Edelen, et al. “Machine Learning to Enable Orders of Magnitude Speedup in Multi-Objective Optimization of Particle Accelerator Systems.” arXiv preprint arXiv:1903.07759 (2019).
- [42] A. L. Edelen, et al. “Neural Network Model of the PXIE RFQ Cooling System and Resonant Frequency Response.” arXiv preprint arXiv:1612.07237, 2016.
- [43] A. L. Edelen, *et al.* “Using a Neural Network Control Policy for Rapid Switching Between Beam Parameters in an FEL.” 38th International Free-Electron Laser Conference, Santa Fe, NM, USA, August 20 - 25, 2017.
- [44] E. Fol, *Evaluation of Machine Learning Methods for LHC Optics Measurements and Corrections Software*. M.S. Thesis, CERN and Karlsruhe University of Applied Sciences, 2017.
- [45] E. Fol and T. Garcia, “Detection of faulty Beam Position Monitors at CERN,” presented at the Machine Learning Applications for Particle Accelerators Workshop, Feb. 28 - March 2, SLAC National Accelerator Laboratory, 2018.
- [46] E. Fol, J. M. Coello de Portugal, and R. Tomas. “Application of Machine Learning to Beam Diagnostics.” 7th Int. Beam Instrumentation Conf. (IBIC’18), Shanghai, China, 09-13 September 2018. JACOW Publishing, Geneva, Switzerland, 2019.
- [47] E. Fol, J. M. Coello de Portugal, and R. Tomas. “Unsupervised Machine Learning for Detection of Faulty BPMs,” in Proceedings of the 2019 International Particle Accelerator Conference, Melbourne, Australia, 2019.
- [48] E. Fol, J. M. Coello de Portugal, and R. Tomas. “Optics Corrections Using Machine Learning in the LHC,” in Proceedings of the 2019 International Particle Accelerator Conference, Melbourne, Australia, 2019.
- [49] Belkin, Mikhail, et al. “Reconciling modern machine learning and the bias-variance trade-off.” arXiv preprint arXiv:1812.11118 (2018).
- [50] G. Loisch, et al. “Observation of High Transformer Ratio Plasma Wakefield Acceleration.” *Phys. Rev. Lett.* 121, 6, 064801, 2018