6th North American Particle Accel. Conf. (NAPAC2025), Sacramento, CA, USA JACoW Publishing
ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-THAN02

FERMILAB BOOSTER BEAM LOSS MODELLING AND REBALANCING USING BAYESIAN METHODS*

N. Kuklev†, J. Eldred, M. Balcewicz
Fermi National National Laboratory, Batavia, IL, USA

Abstract

To meet PIP-II upgrade requirements, Fermilab Booster losses need to be reduced by 50% compared to present levels. So far, simulations are not good enough to predict loss patterns. Thus, an extensive Booster tune up will be necessary to achieve required performance. In this paper we present an effort to build a data-driven loss model using Bayesian techniques, and subsequently to rebalance losses for higher trip margins. We first created several sets of spatially and temporally isolated orbit and optics knobs, and trained Gaussian process models for each beam loss monitor as well as beam current. Novel techniques of uncertainty constraints and approximate GP fitting were introduced to handle safety and timing requirements. We then performed single and multi-objective tuning using scalarized objectives comprised of critical beam loss locations. We achieved significant rebalancing of losses, increasing margins by 25%, as well as an overall improvement in transmission efficiency of 0.4%. Automated data collection is being developed so that more accurate surrogate models can be trained over time.

INTRODUCTION

Fermilab Booster is a key component of the Fermilab accelerator complex. It provides 8 GeV protons to various users directly, as well as sending beam to the Recycler/Main Injector. For the upcoming LBNF/DUNE experiments, PIP-II upgrade will increase the power on target to 1.2 MW. It includes a new superconducting linac, a new BTL transfer line, and various improvements to the complex including the Booster. Key changes for Booster are the increase of injection energy to 800 MeV, higher beam intensity (from 4.8×10^{12} to 6.7×10^{12}), a new painting scheme, new collimation and RF systems, and a faster ramp rate of 20 Hz. Given that the Booster is a 50-year old machine with aging hardware, areas of concern have been identified in terms of machine reliability and capability to operate at PIP-II parameters. In terms of beam dynamics, significant challenges are expected due to extreme levels of space charge at injection, as well as new instabilities at transition crossing.

Maximum Booster power is determined primarily by administrative beam loss limits, with PIP-II requiring a twofold reduction in relative losses as compared to the current levels. An extensive simulation campaign is ongoing to better understand and control beam dynamics, but so far it has proven difficult to obtain quantitative agreement with experimental data, especially for local loss distribution. Thus, it is expected that an extensive experimental tune-up campaign will be necessary to commission the Booster for PIP-II.

Driven by PIP-II requirements as well as challenges in the DOE GARD roadmap [1, 2], several AI/ML projects were recently started at Fermilab, focusing on development of advanced optimization algorithms, fast differentiable simulations, and integrated facility-scale virtual accelerators (digital twins) ([3]. Using virtual accelerators for algorithm and application testing has proven effective in ensuring smooth commissioning [4, 5]. To create a virtual accelerator for loss optimization, we started developing a data-driven surrogate model (in place of a standard simulation). A key concern was training data availability. Recently, an electrical substation fire has taken down Main Injector and freed up majority of cycles for parasitic beam studies - a unique opportunity we took advantage of. In this paper, we describe our results in applying Bayesian methods to experimentally explore, learn, and manipulate the beam loss distribution in the Booster.

BOOSTER DETAILS

Fermilab Booster is a rapid cycling synchrotron with 24 sectors (short (S) and long (L) subsections). Main combined function magnets are connected to a shared resonant power supply. In addition, there are 48 independently powered correction elements with dipole, quadrupole, skew quadrupole, sextupole, and skew sextupole components [6]. Booster diagram is shown in Fig. 1.

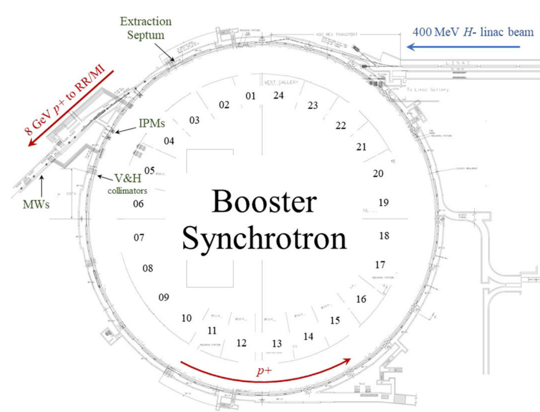

Figure 1: Booster synchrotron with sectors marked.

Time-dependent current ramps with up to 64 points can be set, forming thousands of potential knobs. Such high dimensionality is infeasible for Bayesian methods. Based on physics considerations and typical operator tasks, we defined two subsets of interest. First set contained 4D (x/y/px/py) injection closed orbit bumps. Adjusting injection trajectory

---

* This manuscript has been authored by FermiForward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

† nkuklev@fnal.gov

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-THAN02

is a frequent and well understood tuning task. Second set corresponded to quarupole ramp offsets, either in 24D (short or long) or 48D (combined). It was motivated by experimentally observed increase in injection beam losses at lower tunes, likely due to space-charge detuning pushing particles onto (Q_x/Q_y=6.5) resonance [7]. Changing quadrupole offsets enables both working point and low-order resonance strength variation while preserving other ramp features.

Experimental Details

Booster beam losses are measured by beam loss monitors (BLMs), which provide integrated proportional chamber signals at 12.5 kHz. Over 50 BLMs are positioned throughout the ring. Beam current is measured via 12.5 kHz toroid and 20 kHz DCCT signals. We found the new DCCT system to have significantly better noise and nonlinearity behavior. During data collection, a dedicated study event $17 was inserted into the supercycle at a rate of 1 event every 6 s. Power supplies were configured to use an independent ramp slot, and a Python script was used to change the slot settings between events. Machine protection system and linac beam status were monitored, with safeguards to abort if beam permit is lost.

BAYESIAN ALGORITHM

Bayesian optimization (BO) is uniquely suited to tasks where candidate evaluation is relatively expensive, and complications such as constraints and step sizes need to be satisfied. Numerous successful BO applications have been demonstrated in particle accelerators [8–10]. In BO, the output(s) are described by $\mathbf{y} = f(\mathbf{x}) + \varepsilon$ where $f(\mathbf{x})$ is the black-box function of interest and $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ the added noise. Using Gaussian Processes (GP) [11], a surrogate model for f can be parameterized as a multivariate normal distribution with a mean $m(\mathbf{x})$ and kernel $k(\mathbf{x}, \mathbf{x}')$.

Fitted GP model is used for acquisition function optimization to find the best next step(s). For single-objective problems, a typical choice is the upper confidence bound function that balances exploration and exploitation based on GP model mean and variance. If only the variance is used, optimization becomes exploration. More complex variants that integrate variance changes over fantasy models are used for active learning. For multi-objective (MO) problems, there is no single best candidate. The goal is instead to increase the hypervolume - volume relative to reference point covered by the pareto front. We use a recently proposed noisy-EHVI acquisition function [12]. Note that MOBO does not scale well past 3 objectives - outputs or GP models need to be scalarized to reduce dimensionality.

All BO code we used is implemented in a Python toolkit `apsopt` [13], with BoTorch/PyTorch computational backend [14]. For Booster BO, two novel algorithm improvements were necessary.

Uncertainty Constraints

During initial testing, standard BO algorithm had several large violations of efficiency constraints while learning the valid regions. This is expected due to how constraints are implemented. Because a single bad shot can exceed trip limits, additional safeguards were needed for parasitic data collection. Existing methods like trust region and prior mean were unsuitable due to impact on exploration efficiency.

Our solution was to implement a new type of constraint on the variance of the GP model (typically they are applied to the mean). By setting an appropriate noise prior that is higher than the constraint, as well as a lower fitted noise limit, the net effect is a global preference for locations with sufficient model confidence. Typically, those are close to existing measurements, enabling a ‘creeping’ expansion during global exploration. A visualization is shown in Fig 2.

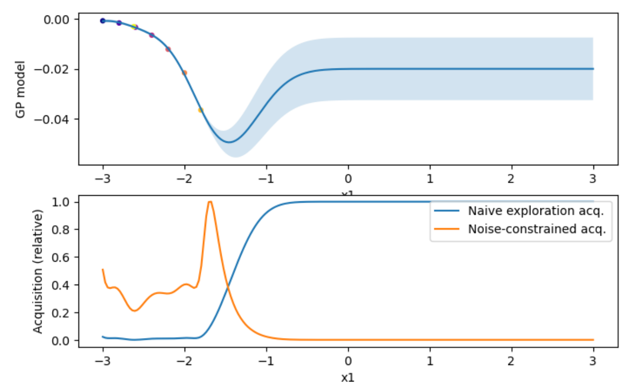

Figure 2: Top: GP model mean and standard deviation. Bottom: uncertainty-constrained exploration acquisition.

Approximate GP Fitting

Due to strict timing requirements, GP fitting and acquisition function search must complete within 4s (for 6s event spacing with read/write margin). Fitting scales poorly (N^3) with dataset size, with wall-time exceeding 1s at ~ 500 points in 48D when using A100 GPU. However, for experimental applications, it is unnecessary to fit GP parameters exactly since readout noise/power supply resolution are relatively large. We modified fitting routines and benchmarked the fit quality with limited number of iterations, wall-time timeouts, higher function tolerance, and higher projected gradient tolerance. We chose 1e-3 gradient termination, no function tolerance, and 2s timeout as final settings. This configuration had identical sampling efficiency but 7-10x faster fitting than BoTorch defaults.

RESULTS

Injection Bumps

Initial tests focused on verifying the operator-tuned injection bumps, and learning a rough model of the losses. After resolving technical roadblocks (as described above), we quickly demonstrated that operational configuration was very close to optimal. One of the scans is shown in Fig. 3

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-THAN02

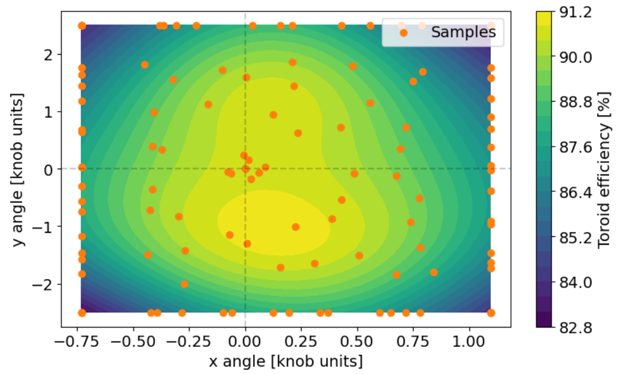

Figure 3: Scan of injection bump - [0,0] is operational value.

Quad Offsets + Efficiency

Quadrupole offset tuning was initially done in 24D (QL) while ramping the beam intensity from 50% to 75% to 100% of the operational one. It was important to perform studies at operational levels due to intensity dependence of beam losses. No beam aborts or other issues were encountered. Eventually, a full optimization run was performed in 48D (QL+QS), demonstrating improvement of 0.4% to transmission efficiency as shown in Fig. 4. While seemingly small, due to already high efficiency this is a notable current loss reduction. Note however that transmission efficiency does not directly correlate with BLM losses because of varying beam energy (and loss impact) during the ramp.

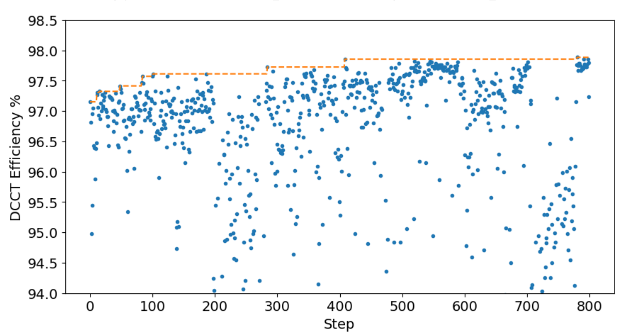

Figure 4: 48D efficiency optimization results.

Quad Offsets + Loss Rebalancing

Using optimization and exploration data, we reviewed the loss/efficiency correlations at key locations. While some areas showed a clear best configuration, others had a pareto front with potentially useful tradeoffs (i.e. where a slight loss increase could be used to improve overall efficiency). An example of such a location, S15, is shown in Fig. 5.

To rebalance losses, we first tried single-objective optimization. Unfortunately, while the model worked, it also ruined overall efficiency/other losses. Moving to MOBO methods (with efficiency as first objective), we initially attempted to use a scalarized output defined as a softplus transform of relative losses above 20% (i.e. a penalty for exceeding 20% of trip limit summed over all BLMs). No progress was observed after 1000 samples. One potential explanation is the extremely complex and non-Gaussian dis-

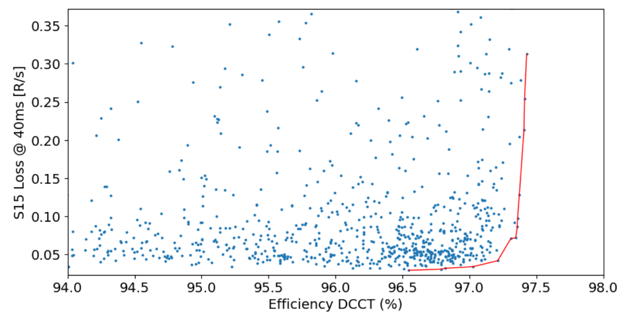

Figure 5: Pareto front for S15 loss vs overall efficiency, demonstrating a potentially useful tradeoff.

tribution of the softplus objective causing poor GP model fit. As mentioned above, a more advanced technique is to model losses as independent GP models but use a scalar objective to combine them. We selected several key locations (L15,L22,L01), and a using scalarized objective successfully demonstrated a transfer of losses to the S06 collimation region without transmission efficiency degradation. Best observed improvement of 25% is shown in Fig. 6.

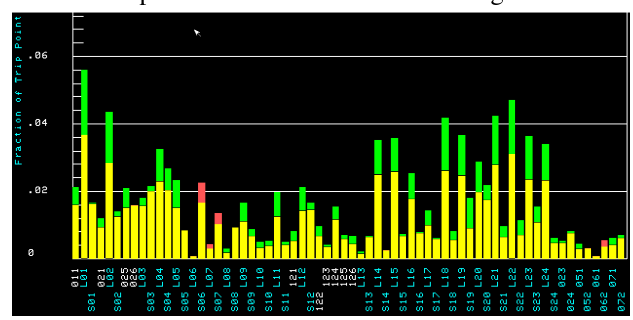

Figure 6: BLM loss rebalancing into S06 collimation region. Green/red indicates lower/higher vs reference values.

CONCLUSION

Improving Booster loss modelling and experimental tuning is critical to achieving PIP-II design parameters. In this paper we highlighted our work on data-driven loss modelling and rebalancing. We developed techniques for safe and fast exploration, and used them to collect extensive datasets with several knob sets. An improvement of 0.4% to transmission efficiency was found, demonstrating that the current operational preset has room for improvement. We then used multi-objective search to intelligently combine objectives and find loss rebalancing opportunities while maintaining overall efficiency. Our best result showed 25% reduction in certain high loss areas, moving them to less sensitive collimation regions without transmission efficiency degradation. Our future work will aim to add chromaticity and collimator knobs, as well as fully automate the process such that it can be continuously run by operators.

REFERENCES

[1] Accelerator beam physics research roadmap. `https://science.osti.gov/hep/-/media/hep/pdf/2022/ABP_Roadmap_2023_final.pdf`

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-THAN02

[2] S. Nagaitsev *et al.*, "Accelerator and beam physics research goals and opportunities", *arXiv*, 2021. doi:10.48550/arXiv.2101.04107

[3] N. Kuklev, M. Borland, L. Emery, Y. Hidaka, H. Shang, and Y. Sun, "Unified differentiable digital twin for the IOTA/FAST facility", presented at IPAC'25, Taipei, Taiwan, Jun. 2025, paper THPM101, unpublished. https://indico.jacow.org/event/81/contributions/8364

[4] V. Sajaev, "Commissioning simulations for the Argonne Advanced Photon Source upgrade lattice", *Phys. Rev. Accel. Beams*, vol. 22, no. 4, p. 040102, Apr. 2019. doi:10.1103/PhysRevAccelBeams.22.040102

[5] V. Sajaev, "Improvements to the commissioning simulations of the APS Upgrade storage ring", in *Proc. IPAC'23*, Venice, Italy, May 2023, pp. 3112–3115. doi:10.18429/JACoW-IPAC2023-WEPL006

[6] E. J. Prebys *et al.*, "New corrector system for the Fermilab booster", *Conf. Proc. C*, vol. 070625, p. 467, 2007. doi:10.1109/PAC.2007.4440247

[7] J. Eldred, V. Lebedev, K. Seiya, and V. Shiltsev, "Beam intensity effects in Fermilab booster synchrotron", *Phys. Rev. Accel. Beams*, vol. 24, no. 4, p. 044001, Apr. 2021. doi:10.1103/PhysRevAccelBeams.24.044001

[8] N. Kuklev, M. Borland, G. I. Fystro, H. Shang, and Y. Sun, "Online accelerator tuning with adaptive 7 optimization", in *Proc. NAPAC'22*, Albuquerque, NM, USA, Aug. 2022, pp. 842–845. doi:10.18429/JACoW-NAPAC2022-THXD4

[9] N. Kuklev, M. Borland, G. Fystro, H. Shang, and Y. Sun, "Robust adaptive Bayesian optimization", in *Proc. IPAC'23*, Venice, Italy, May 2023, pp. 4428–4431. doi:10.18429/JACoW-IPAC2023-THPL007

[10] R. Roussel *et al.*, "Bayesian optimization algorithms for accelerator physics", *Phys. Rev. Accel. Beams*, vol. 27, no. 8, p. 084801, Aug. 2024. doi:10.1103/PhysRevAccelBeams.27.084801

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. MIT Press, Nov. 2005. doi:10.7551/mitpress/3206.001.0001

[12] S. Daulton, M. Balandat, and E. Bakshy, "Parallel Bayesian optimization of multiple noisy objectives with expected hypervolume improvement", in *Proc. NIPS'21*, Virtual, Dec. 2021.

[13] N. Kuklev *et al.*, "High efficiency multi-objective Bayesian algorithm for APS-U nonlinear dynamics tuning", presented at IPAC'25, Taipei, Taiwan, Jun. 2025, paper THPM102, unpublished. doi:10.18429/JACoW-IPAC25-THPM102

[14] M. Balandat *et al.*, "BoTorch: a framework for efficient Monte-Carlo Bayesian optimization", in *Advances in Neural Information Processing Systems*, Virtual, Dec. 2020, pp. 21524–21538. https://proceedings.neurips.cc/paper/2020/hash/f5b1b89d98b7286673128a5fb112cb9a-Abstract.html