GLOBAL MOMENTUM COMPENSATION SCHEME ENABLED BY SEQUENTIAL INFERENCE IN LARGE-SCALE LINACS

Y. Liu*, K. Futatsukawa, E. Cicek, T. Miyao, Z. Fang, Y. Fukui, K. Seiya, Y. Nakazawa,
High Energy Accelerator Research Organization, Tsukuba, Japan
R. Sharankova, J.-F. Ostiguy, Fermi National Accelerator Laboratory, Batavia, United States
S. Wagner-Rossel, Goethe University Frankfurt, Frankfurt am Main, Germany

Abstract

A fundamental challenge in large-scale accelerator operation is to infer internal machine states from limited monitoring. In the J-PARC linac, Radio-frequency (RF) cavity errors accumulate and manifest as momentum drifts observable through beam phase monitors. Although the number and placement of monitors may formally satisfy solvability conditions, the sequentially coupled nature of beam dynamics renders the inverse problem numerically ill-conditioned.

We show that a multilayer perceptron (MLP) can accurately model the forward mapping but fails in inverse reconstruction under non-redundant conditions, revealing a limitation of static regression models rather than insufficient capacity. We reformulate the problem as a sequential inference task using an encoder–decoder architecture. Accurate reconstruction is achieved even in the non-redundant case, enabling global momentum compensation through direct correction of distributed RF parameters, without relying on iterative local tuning.

This work demonstrates that reliable global compensation becomes achievable when the inverse problem is formulated consistently with the sequential structure of beam dynamics.

INTRODUCTION

Accelerator facilities such as J-PARC and those at Fermilab operate at increasingly high beam power, placing stringent requirements on beam energy and phase stability across the accelerator chain [1–4]. In large-scale linear accelerators, small amplitude or phase deviations in individual RF cavities accumulate along the beamline and lead to measurable momentum drifts. Accurate identification and correction of distributed cavity offsets are therefore essential for stable and high-performance operation. In practice, cavity conditions cannot be directly measured with sufficient precision during beam operation. Instead, they must be inferred from limited beam-based diagnostics, such as phase monitors and longitudinal phase-space measurements [5–8]. This leads to an inverse problem of reconstructing distributed cavity offsets from downstream observations, which, although formally solvable in principle, is numerically ill-conditioned due to the cumulative nature of beam dynamics and thus highly sensitive to measurement noise and modeling uncertainties [9].

Machine-learning (ML) techniques have recently been widely explored in accelerator systems [17], including our previous works [10, 11]. An MLP is found to accurately reproduce the forward mapping from cavity offsets to beam observables. However, the inverse mapping—predicting cavity offsets from monitor signals—remains challenging. While accurate reconstruction is achieved under redundant (i.e., with overlapping sensitivity to upstream cavities, as shown in Fig. 1) monitoring configurations, the MLP fails under non-redundant conditions, where each cavity is effectively observed by a single downstream probe, where each cavity is effectively observed by a single downstream probe.

The success of the forward model and the failure of the inverse mapping reflect a fundamental limitation of static regression models (e.g., MLPs) when applied to inherently sequential systems such as linacs. While MLPs possess universal approximation capability, they lack an appropriate representation of the underlying sequential structure, as recognized in representation learning [12] and emphasized in scientific machine learning [13].

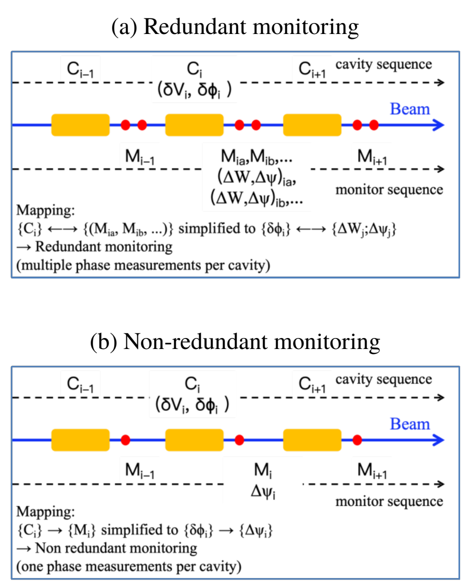


Figure 1: Schematic of cavity–beam–monitor configuration in a linear accelerator.

This motivates reformulating the inverse problem. One effective approach is to cast it as a sequential inference task, closely related to state-space formulations such as Bayesian filtering and smoothing [14], where information

* yong.liu@kek.jp

is propagated across an ordered system. In this work, an encoder–decoder architecture is adopted to represent the ordered system and capture the hidden correlations in beam evolution, enabling reconstruction of cavity offsets even under non-redundant monitoring conditions.

The reconstructed offsets are used to implement global momentum compensation by correcting distributed RF offsets across the linac. Here, “global” refers both to inferring all cavity offsets within a unified model, avoiding iterative local corrections, and to compensating errors throughout the accelerator rather than only matching the final beam energy.

This paper presents the conceptual framework and its validation in a realistic accelerator setting. The focus is on establishing the methodology and demonstrating its effectiveness; detailed theoretical analysis and experimental validation will be reported in future work.

FORMULATION OF THE LINAC ERROR RECONSTRUCTION AND DATA-DRIVEN FRAMEWORK

In a linac, the beam centroid phase evolves as the beam traverses a chain of RF cavities. Each cavity introduces beam centroid energy and phase changes $(\Delta W, \Delta\psi)$, determined by its field amplitude and phase offsets $(\delta V, \delta\phi)$, which propagate downstream and are observed at beam phase monitors (in the J-PARC linac, fast current transformers, FCTs), as illustrated in Fig. 1. RF amplitude and phase offsets both manifest predominantly as beam energy deviations, and their individual contributions are difficult to disentangle. In this work, cavity offsets are therefore represented equivalently as effective phase deviations. This choice is also consistent with operational practice, where phase adjustments are preferred, while amplitude corrections are applied only when necessary.

The monitoring configuration plays a central role in this inverse problem. We illustrate the formulation using the J-PARC linac as a representative example, which consists of $N_c = 46$ independently controlled RF cavities. In the original design, a redundant monitoring scheme with $N_{m0} = 112$ FCTs was employed. In the present configuration, however, the number of monitors is reduced to $N_{m1} = 87$ due to aperture constraints in the 191–400 MeV section. This results in partially non-redundant conditions and provides a realistic test case for reconstruction under limited diagnostics.

Data used in this study are generated using high-fidelity simulations with the code IMPACT [18]. The simulations accurately reproduce longitudinal beam behavior and have been extensively benchmarked against experimental observations. To ensure generality, simulations were cross-checked using multiple beam dynamics tools. The training datasets are constructed from large-scale simulations with randomized cavity offsets, yielding more than 10^7 samples. Although generating such data is computationally expensive, particularly for high-fidelity models, the trained machine learning models enable rapid inference once deployed.

The reliability of the trained ML models is therefore tied to the accuracy of the underlying simulations, which can be further calibrated using beam-based measurements to adapt to actual machine conditions. The overall approach follows a hybrid workflow: a high-accuracy model is first trained offline using simulation data, then calibrated with experimental measurements, and finally deployed for fast online diagnosis and global compensation.

BASELINE RESULTS WITH STATIC REGRESSION MODELS USING MLP

To establish a baseline, forward mapping and inverse reconstruction were modeled using an MLP consisting of two hidden layers with 2048 neurons each and rectified linear unit (ReLU) activation. We note that the results are not sensitive to this specific architecture: similar behavior is observed across a wide range of network sizes and hyperparameters. The forward mapping is accurately captured, as shown in Fig. 2, demonstrating that the response of beam phase monitors to cavity perturbations can be effectively learned within a static regression framework. All results reported in this section are evaluated on independent test datasets generated with randomized cavity offsets distinct from those used for training.

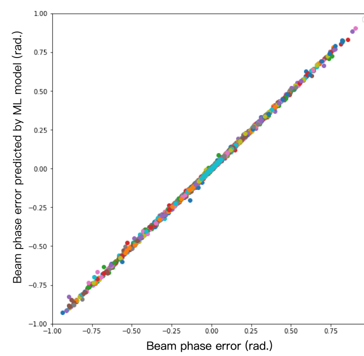
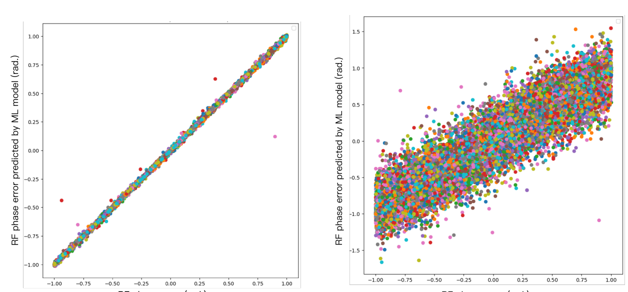


Figure 2: Performance of the MLP-based forward model for the J-PARC linac, showing accurate prediction of beam phase responses from cavity perturbations.



(a) Redundant (b) Non-redundant

Figure 3: Performance of the MLP-based inverse model for the J-PARC linac, with predicted cavity offsets (normalized) compared with the ground truth.

In contrast, as shown in Fig. 3 a, accurate reconstruction is achieved under redundant conditions. However, under non-redundant conditions [Fig. 3 b], the MLP fails to produce reliable reconstructions.

The impact of reconstruction accuracy on beam correction is illustrated in Fig. 4. In the redundant case, accurate inverse reconstruction enables effective global momentum compensation [Fig. 4 a]. In contrast, under non-redundant conditions, the failure of inverse reconstruction leads to large deviations in the inferred cavity offsets, making direct global compensation infeasible. In this case, only partial correction of the final beam energy can be achieved using forward-model-based adjustment, while distributed cavity offsets remain unresolved [Fig. 4 b]. These results motivate a reformulation that accounts for the sequential structure of the system.

(a)

(b)

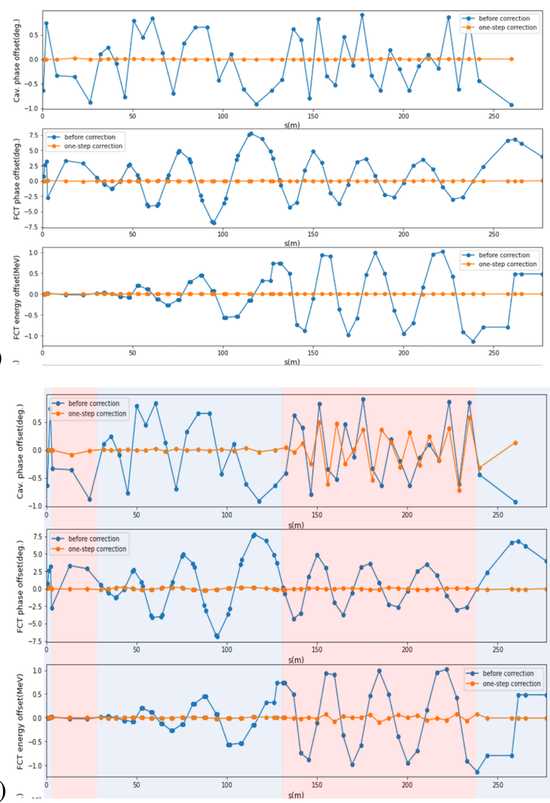


Figure 4: Momentum compensation and reconstruction performance along the J-PARC linac using an MLP model under redundant and non-redundant configurations. (a) Redundant case: accurate reconstruction and full momentum compensation. (b) Non-redundant case: loss of reconstruction accuracy with retained output momentum compensation. Light red and light blue shaded regions denote non-redundant and redundant sections, respectively.

FROM STATIC TO SEQUENTIAL INFERENCE

The inverse problem can instead be formulated as a sequential inference task. The accelerator is represented as an ordered system, and the inverse reconstruction is performed as a step-wise process in which downstream measurements provide constraints that are propagated upstream.

To implement this idea, an encoder–decoder sequential model is adopted. The encoder extracts a latent representation from the sequence of monitor signals, while the decoder reconstructs cavity offsets sequentially along the linac. This structure explicitly captures the causal and cumulative nature of beam transport.

The model is applied to the same non-redundant configuration as in Fig. 3 b, representing the most challenging case. An LSTM-based encoder–decoder [15, 16] architecture with two hidden layers of 256 dimensions is used. We emphasize that the performance is not sensitive to this specific configuration: a wide range of sequence-to-sequence models yield consistent results, suggesting that the key factor is the alignment between model representation and system structure. Accurate reconstruction is achieved across the entire linac, as shown in Fig. 5. The degradation observed for static regression models under non-redundant conditions is no longer present.

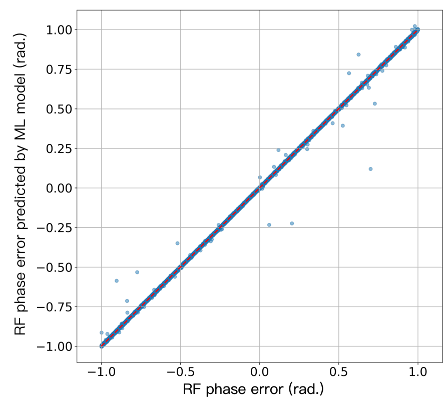


Figure 5: Reconstruction results using the encoder–decoder sequential model under the same non-redundant monitoring conditions as in Fig. 3(b).

DISCUSSION AND CONCLUSION

The central finding of this work is that successful inverse reconstruction depends not only on model capacity, but critically on whether the model representation is consistent with the sequential structure of beam dynamics. In a linac, upstream perturbations propagate cumulatively, making the inverse problem inherently non-local and ill-conditioned when treated as a direct global mapping. By reformulating the problem as a sequential inference task, the reconstruction is aligned with the ordered structure of the accelerator. This enables accurate recovery of distributed cavity offsets even under sparse diagnostics.

The observed weak dependence on model architecture further indicates that the improvement originates from the formulation itself rather than model-specific tuning.

More broadly, these results suggest that inverse problems in sequentially coupled systems benefit from representations that explicitly reflect their causal structure. Aligning model representation with physical dynamics is therefore essential for achieving robust and generalizable performance.

REFERENCES

[1] K. Yamamoto, "Status of J-PARC accelerator chain," in *Proc. IPAC'25*, 2025.

[2] H. Hotchi and Y. Sato, "Progress of beam power upgrade in the J-PARC main ring," in *Proc. IPAC'25*, 2025.

[3] Y. Liu *et al.*, "Progress of J-PARC LINAC commissioning," in *Proc. IPAC'19*, Melbourne, Australia, May 2019. doi:10.18429/JACoW-IPAC2019-TUPTS027

[4] P. Saha *et al.*, "Beyond 1 MW operation of the J-PARC RCS," in *Proc. IPAC'25*, 2025.

[5] K. Futatsukawa *et al.*, "Phase drift correction system of LLRF for humidity at J-PARC linac," in *Proc. PASJ2020*, 2020.

[6] E. Cicek *et al.*, "Development of an integrated monitor system for real-time relative phase measurement between the cavity RF and beam," in *Proc. LINAC'24*, 2024.

[7] E. Cicek *et al.*, "Installation and test of the phase drift monitor at J-PARC linac," in *Proc. PASJ2022*, 2022.

[8] K. Moriya *et al.*, "Longitudinal phase-space direct measurement of intermediate pulses for high-intensity beam accumulation," *Prog. Theor. Exp. Phys.*, 2025.

[9] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, 1998. doi:10.1137/1.9780898719697

[10] S. Wagner *et al.*, "Automation of pulse identification at J-PARC," in *Proc. IPAC'25*, 2025.

[11] M. Sugita *et al.*, "Neural networks approach for controlling a waveform pattern of the paint bump power supply at J-PARC RCS," in *Proc. IPAC'25*, 2025.

[12] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013. doi:10.1109/TPAMI.2013.50

[13] G. E. Karniadakis *et al.*, "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, pp. 422–440, 2021. doi:10.1038/s42254-021-00314-5

[14] S. Särkkä, *Bayesian Filtering and Smoothing*, Cambridge University Press, 2013. doi:10.1017/CBO9781139344203

[15] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735

[16] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, 2014.

[17] A. Edelen *et al.*, "Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems," *Phys. Rev. Accel. Beams*, vol. 23, p. 044601, 2020. doi:10.1103/PhysRevAccelBeams.23.044601

[18] J. Qiang, R. Ryne, S. Habib, and V. Decyk, "An object-oriented parallel particle-in-cell code for beam dynamics simulation in linear accelerators," *J. Comput. Phys.*, vol. 163, pp. 434–451, 2000. doi:10.1006/jcph.2000.6570