NEURAL NETWORK-BASED AMPLITUDE FEEDFORWARD COMPENSATION ALGORITHM FOR LLRF SYSTEMS *

Yuchen Wang, Xiaofang Hu†, Shenghua Yang, Jian Pang,
Fangfang Wu, Kai Zhang, Shancai Zhang
National Synchrotron Radiation Laboratory, University of Science and Technology of China,
Hefei, Anhui, 230029, China

Abstract

In free-electron laser facilities, the amplitude-phase flatness of the Radio Frequency (RF) pulses driving the electron beam is a key factor determining beam energy spread. Imperfections in the RF driving chain, such as the static nonlinearities of the high-power amplifier, can significantly increase the intra-pulse flattop amplitude error and degrade the pulse flatness. To address this, this paper proposes a U-Net deep neural network-based amplitude feedforward compensation algorithm for Low-Level Radio Frequency (LLRF) systems to suppress intra-pulse amplitude fluctuations. The algorithm has been validated at the output of a Solid-State Amplifier (SSA): under four randomly selected LLRF output configurations (amplitude, pulse width, and pulse delay), the average intra-pulse amplitude flatness (RMS) was reduced from 1.208 % to 0.398 %, and the average peak-to-peak variation was reduced from 4.683 % to 1.353 %, demonstrating a significant compensation effect.

INTRODUCTION

In the Infrared Free-electron Laser (IR-FEL) facility, the Radio Frequency (RF) system must provide power to a 476MHz room-temperature subharmonic pre-buncher, a 2856MHz room-temperature buncher, and two 2856MHz room-temperature accelerators [1]. Solid-state amplifiers (SSA) and klystrons are core components of the RF driving chain. The 476MHz pre-buncher is powered by an SSA, while the 2856MHz buncher and accelerators are powered by two independent SSA–klystron chains [2]. However, the nonlinear amplification characteristics of SSAs and klystrons introduce severe waveform distortion of the accelerating field. This distortion leads to intra-pulse flattop fluctuation and rising-edge deformation of RF pulses, further resulting in inconsistent energy gain for electron bunches within the 5–10 μs macro-pulse and eventually deteriorating overall beam quality. To maintain uniform acceleration conditions for individual bunches, the Low-Level Radio Frequency (LLRF) system is required to realize high-stability flat-top amplitude and phase of RF pulses. A range of linearization and RF pulse flattening methods have been reported for accelerator RF driving chains. Analytical modeling approaches combining fifth-order polynomial and modified Saleh functions have been employed to characterize amplifier nonlinearity and realize full-range predistortion compensation [3]. Various FPGA-implemented predistortion algorithms based on polynomial fitting and lookup tables have also been investigated for klystron linearization [4]. Furthermore, mechanistic klystron modeling and model-free iterative learning control have been proposed for operating point optimization and intra-pulse amplitude-phase flattening, respectively [5, 6]. Nevertheless, these conventional methods rely heavily on fixed analytical formulas, polynomial fitting, lookup tables or iterative rules, suffering from limited generalization and poor adaptability to complex time-varying operating conditions, which limits their performance in high-precision pulse flattop compensation. As an efficient deep learning architecture for feature extraction and waveform reconstruction, U-Net has been widely used in image processing and time series prediction tasks [7]. This paper proposes a novel data-driven feedforward compensation method by embedding 1D U-Net into the LLRF control framework. The network is trained to directly learn the inverse response mapping from the measured accelerating field waveform to the LLRF Vector Modulator (VM) output, and serves as a data-driven predistortion model for intra-pulse flattop compensation.

MODEL AND DATASET DESIGN

This study aims to build a deep neural network that learns the nonlinear inverse response of the RF driving chain, enabling feedforward waveform pre-distortion. The framework includes offline training and online deployment. Offline, a dataset of measured VM output waveforms and corresponding controlled-location waveforms is collected. The network is then trained to map the controlled-location waveform (input) to the VM output waveform (label), thereby learning the inverse model of the RF driving chain. Online, the trained pre-distortion network runs on a GPU-equipped computer. The desired waveform at the controlled location is fed into the network to generate pre-distorted waveform data, which is written directly to the LLRF DAC via Direct Memory Access (DMA) and output through the VM, achieving amplitude compensation.

Network Architecture: 1D U-Net Convolutional Neural Network

To effectively process long-sequence time-domain waveforms and capture their local features along with global context, we adopted a one-dimensional U-Net architecture. This network adopts an encoder-decoder structure as its

* Supported by National Natural Science Foundation of China (12505182)
Supported by National Key Research and Development Program of China (2024YFA1612200)
† Corresponding author, email: huxiaofang@ustc.edu.cn

main body, specifically optimized for the characteristics of RF pulse waveforms.

In the encoder-decoder backbone, the network employs symmetric contracting and expanding paths. The encoder progressively extracts multi-level features through a series of down-sampling blocks while compressing the sequence length. Conversely, the decoder gradually restores the spatial resolution and fuses high-resolution features from the encoder through up-sampling and skip connections, ultimately reconstructing the complete 2048-point pre-distortion waveform. The encoder part of this one-dimensional U-Net architecture is shown in Fig. 1, and the decoder part is shown in Fig. 2.

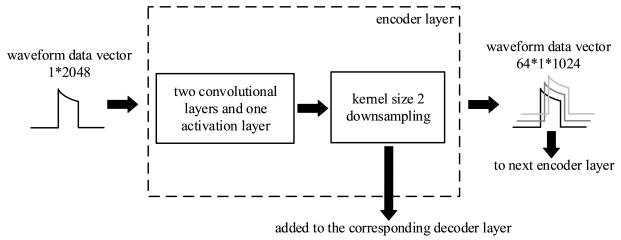

Figure 1: The encoder part of one-dimensional U-Net architecture.

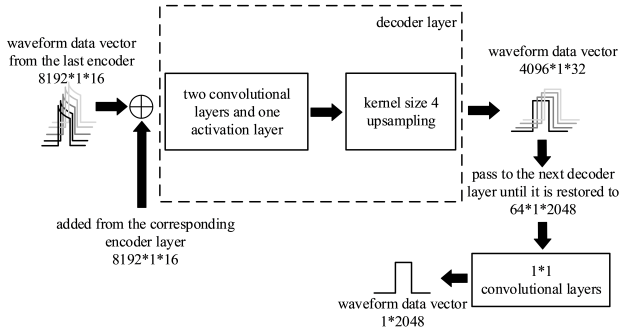

Figure 2: The decoder part of one-dimensional U-Net architecture.

Dataset Processing and Construction

The dataset utilized in this study originates from a test platform comprising a physically implemented LLRF control system and an SSA. By systematically acquiring the system's responses under various excitation parameters, a large-scale collection of waveform pairs was constructed for training and validating the deep inverse model, with the aim of comprehensively covering its amplitude nonlinear characteristics.

The hardware for data acquisition consists of an LLRF controller, an SSA, an attenuator, a trigger generator, an RF signal generator, a frequency synthesizer, and a trigger level converter. During system operation, the LLRF controller generates and outputs the initial RF pulse, which is then power-amplified by the SSA. The signal is subsequently attenuated back to an appropriate level via the attenuator and routed back to the acquisition channel of the LLRF controller. Finally, the Process Variables (PV) of the corresponding channels are recorded by the host computer's Input-Output Controller (IOC) based on the EPICS framework. The setup of the data acquisition system is illustrated in Fig. 3.

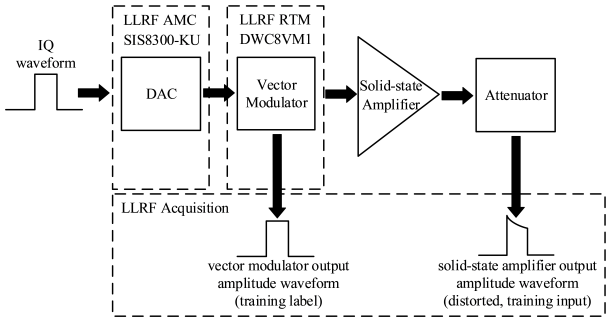

Figure 3: The setup of the data acquisition system.

The waveforms actually collected after distortion by the SSA are used as the input to the neural network. The corresponding ideal waveforms originally sent by the LLRF system serve as the target output for the network. The model training adopts a supervised learning paradigm, with the objective of enabling the pre-distortion signal predicted by the network to produce the desired ideal waveform after passing through the actual physical system.

The model is trained using the AdamW optimizer with a learning rate of 4×10^{-5} and weight decay of 10^{-4}, combined with a learning rate scheduler that dynamically adjusts the learning rate upon plateaus [8]. To prevent overfitting, in addition to monitoring with an independent validation set, we employ an early stopping strategy, terminating training when the validation loss ceases to decrease for 40 consecutive epochs (patience = 40).

EXPERIMENTAL SETUP AND RESULTS

To evaluate the feedforward compensation performance of the aforementioned model, we conducted a bench-top test. The hardware devices and their interconnections in this test remained identical to those used during the data acquisition process. The model was implemented on an additional computer equipped with a NVIDIA RTX5090 GPU and trained offline using the PyTorch framework. The performance of the model training was characterized by the value of the same loss function employed during training, calculated based on the difference between the model's predictions for the waveform data in the test set and the corresponding ground truth labels of the test set. Upon completion of the model training, the LLRF controller was used once again to acquire waveform data from both the VM readback channel and the SSA output channel. The waveform from the re-acquired VM readback channel was then fed as input into the model. Subsequently, the pre-distortion waveform generated by the model was written into the DAC, thereby enabling the output of the pre-distorted excitation waveform.

With the selected set of optimal training parameters, validation experiments were conducted to evaluate the feedforward compensation performance. Four excitation signals with different LLRF output amplitudes, pulse delays, and

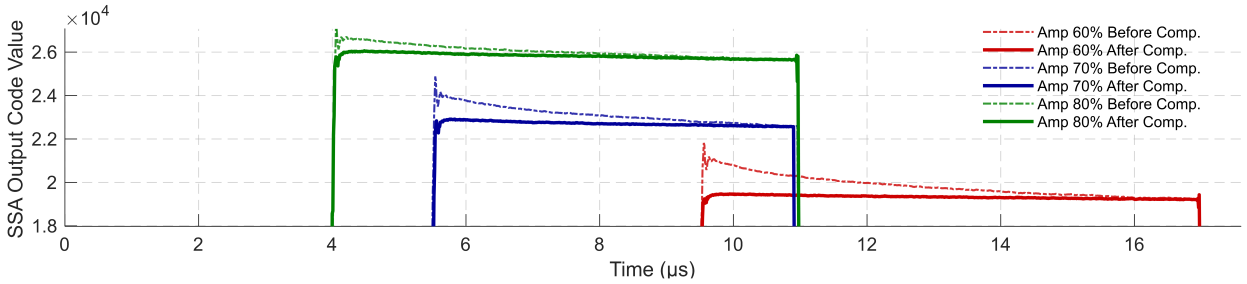

Figure 4: U-Net feedforward performance comparison under different LLRF output pulse settings.

Table 1: Comparison of Intra-Pulse Amplitude Flatness (RMS) and Peak-to-Peak Values Before and After U-Net Feedforward Compensation Under Different LLRF Excitation Signals

Amplitude (%)	RMS (before/after)	peak-to-peak (before/after)
60	1.72 %/0.47 %	6.39 %/1.18 %
70	1.54 %/0.40 %	6.25 %/1.52 %
80	1.04 %/0.45 %	3.97 %/1.62 %
90	0.53 %/0.28 %	2.12 %/1.08 %

pulse widths were randomly selected for comparison before and after feedforward compensation. The distorted waveforms at the SSA output corresponding to these LLRF outputs were recorded, and the LLRF output waveforms were used as inputs to the surrogate model to generate the pre-distorted outputs of the LLRF system's VM, thereby demonstrating the effectiveness of the proposed method in flattening the waveform amplitude. The waveforms before and after feedforward compensation are shown in Fig. 4, and the corresponding intra-pulse amplitude flatness (RMS) and peak-to-peak values are presented in Table 1. Since the data for these LLRF excitation signals are not included in the training set or validation set, it can be demonstrated that the feedforward compensation effect is consistently significant over a wide range of amplitude settings, as well as under arbitrary adjustments of pulse widths and delays. However, as shown in Table 1, there remains room for improvement in the feedforward compensation effect corresponding to specific LLRF output amplitudes. Therefore, future work will focus on further optimizing the model and algorithm to address this issue.

CONCLUSION

This study presents a novel neural network-based technical pathway for intra-pulse amplitude nonlinear feedforward compensation in accelerator LLRF systems. The proposed algorithm has been validated at the output of an SSA: under four randomly selected LLRF output configurations (amplitude, pulse width, and pulse delay), the average intra-pulse amplitude flatness (RMS) was reduced from 1.208 % to 0.398 %, and the average peak-to-peak variation was reduced from 4.683 % to 1.353 %, demonstrating a significant compensation effect. Future work will focus on investigating the impact of the proposed method on phase errors, achieving compensation at the accelerator entrance, and validating its improvement on beam quality through beam diagnostics.

REFERENCES

[1] Z. G. He, Q. K. Jia, L. Wang, W. Xu, and S. C. Zhang, "Linac Design of the IR-FEL Project in CHINA", in *Proc. FEL'15*, Daejeon, Korea, Aug. 2015, pp. 46–48. doi:10.18429/JACoW-FEL2015-MOP010

[2] H. Lin, G. Huang, K. Jin, B. Du, and F. Wu, "The RF System of Infrared Free Electron Facility at NSRL", in *Proc. IPAC'17*, Copenhagen, Denmark, May 2017, pp. 4239–4241. doi:10.18429/JACoW-IPAC2017-THPIK063

[3] W. Cichalewski and B. Kosęda, "Characterization and compensation for nonlinearities of high-power amplifiers used on the FLASH and XFEL accelerators", *Meas. Sci. Technol.*, vol. 18, no. 8, pp. 2372–2378, 2007. doi:10.1088/0957-0233/18/8/011

[4] M. Omet, S. Michizono, T. Matsumoto, T. Miura, F. Qiu, B. Chase, P. Varghese, H. Schlarb, J. Branlard, and W. Cichalewski, "FPGA-based klystron linearization implementations in scope of ILC", *Nucl. Instrum. Meth. Phys. Res. A*, vol. 768, pp. 69–76, 2014. doi:10.1016/j.nima.2014.09.007

[5] A. Rezaeizadeh, R. Kalt, T. Schilcher, and R. Smith, "Model-based klystron linearization in the SwissFEL test facility", in *Proc. FEL'14*, Basel, Switzerland, Aug. 2014, pp. 820–823. doi:10.3929/ethz-b-000094989

[6] A. Rezaeizadeh, T. Schilcher, and R. Smith, "RF pulse flattening in the SwissFEL facility based on model-free iterative learning control", in *Proc. FEL'14*, Basel, Switzerland, Aug. 2014, pp. 824–827. doi:10.18429/JACoW-FEL2014-THP044

[7] C.-H. Chuang *et al.*, "IC-U-Net: A U-Net-based Denoising Autoencoder Using Mixtures of Independent Components for Automatic EEG Artifact Removal", *NeuroImage*, vol. 263, p. 119586, Aug. 2022. doi:10.1016/j.neuroimage.2022.119586

[8] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization", in *Proc. ICLR'19*, New Orleans, USA, May 2019. doi:10.48550/arXiv.1711.05101