AUTOMATED ANOMALY DETECTION ON EUROPEAN XFEL KLYSTRONS

A. Sulc, Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany
A. Eichler, T. Wilksen, Deutsches Elektronen-Synchrotron, Hamburg, Germany

Abstract

High-power multi-beam klystrons represent a key component to amplify RF to generate the accelerating field of the superconducting radio frequency (SRF) cavities at European XFEL. Exchanging these high-power components takes time and effort, thus it is necessary to minimize maintenance and downtime and at the same time maximize the device's operation. In an attempt to explore the behavior of klystrons using machine learning, we completed a series of experiments on our klystrons to determine various operational modes and conduct feature extraction and dimensionality reduction to extract the most valuable information about a normal operation. To analyze recorded data we used state-of-the-art data-driven learning techniques and recognized the most promising components that might help us better understand klystron operational states and identify early on possible faults or anomalies.

INTRODUCTION

European XFEL is currently operating 25 klystrons. They play a crucial role in the acceleration and operation of the European XFEL. They function as radio frequency (RF) power amplifiers, providing high-power RF signals that are used to accelerate the charged particles to high energies as they travel through the linear accelerator structure. The performance and reliability of the klystrons directly impact the accelerator's ability to deliver the desired beam energy, intensity, and stability to the experiment. Klystrons that are not functioning optimally can lead to fluctuations in the RF power, resulting in beam energy variations, reduced beam intensity, or even beam loss. Failure of a klystron can cause a complete interruption in the accelerator's operation, as the loss of RF power will prevent the particle beam from being accelerated, leading to downtime for the accelerator facility and disrupting ongoing experiments.

At the same time, klystrons are highly complex devices, with many interdependent components and parameters. Diagnosing and predicting potential failures in them can be very challenging because they have relatively high stability and mostly operate as black boxes from a signal viewpoint.

In this work, we leverage machine learning techniques for anomaly detection to better understand klystron operational states and identify potential faults before they lead to downtime. We then employ a specialized one-class anomaly loss [1] which is an unsupervised machine learning approach that reduces the dimensionality of the inputs to isolate key characteristics distinguishing normal functions. By training models on this reduced feature set, we can recognize anomalies and incipient failures based solely on klystron signals that we retrieve from DOOCS [2] and collect with DxMAF [3], providing a data-driven means of maximizing uptime for these delicate and expensive devices.

RELATED WORK

The first fully digital version of the Klystron Lifetime Management (KLM) system was developed by Butkowski et al. [4] to maximize the lifetime of klystron tubes used in the X-Ray Free Electron Laser (XFEL) at DESY and has been successfully running since then. In their setting at European XFEL, klystrons are linear-beam vacuum tubes that operate at 1.3 GHz and 10 MW power to accelerate electron bunches for the European XFEL. A crucial component is the continuous operation of klystrons for at least 20 years, thus the klystron lifetime needs to exceed 60,000 hours. The KLM itself is a digital system that detects exceptional events like arcing, RF breakdowns, etc., and takes preventive actions (interrupt the drive) to avoid potential complications leading to longer downtime. In the implementation of [4], it is crucial to have a fast and reliable FPGA implementation that handles protection functions such as reflection limitation, forward/input power correspondence, and energy monitoring within 300 nanoseconds. The KLM system serves as an effective preventive measure to avoid klystron damage from exceptional events during operation. Our research aims to go a step further; by detecting potential issues at an even earlier stage before they manifest as urgent events, we can provide more lead time to anticipate problems and take preventive action.

In [5] authors propose an anomaly detection approach using a neural network model to predict breakdowns on superconducting radio frequency (SRF) cavities at the European XFEL. They experiment with two models: one trained with a semi-supervised anomaly loss (SAL) [6] and another with binary cross-entropy loss (BCE). The SAL model uses a small set of labeled anomalous data along with a larger set of normal data, while the BCE model is trained in a fully supervised manner on the labeled data and requires balance in training data.

Our approach is similar to [5], but we use the unsupervised one-class loss [1] (OC) instead of the SAL [6], as we lack labeled anomalous cases. The model is trained solely on standard waveform modes, allowing it to learn normal patterns and detect deviations as anomalies without labels. The one-class loss's sensitivity can potentially enhance the klystron lifetime and reliability by alarming operators based on training to sensitive modes of klystron signals characterizing the state. Additionally, we obtain a latent variable en-

coding the klystron state and potential drift(s) as the model's output embedding of the klystron state changes.

METHOD

One Class Loss

Deep OC classification [1] is a loss function that aims to train a model to inputs belonging to a specific class(es) by learning solely from training data containing examples of that class. This approach differs from traditional classification problems where the training data includes examples from all classes, and where the goal is to distinguish between the classes. The objective of OC loss is to model the characteristics of the single class of interest which is dominating the inputs, rather than distinguishing it from other classes.

Consider a function $f_\theta : \mathbb{R}^N \mapsto \mathbb{R}^M$ (a neural network). The objective of OC is to minimize the distance between a fixed (randomly chosen) hypersphere center $\mathbf{c} \in \mathbb{R}^M$ and the projection f_θ of most (non-anomalous points) $\mathbf{x}$,

$$\arg \min_\theta \underbrace{\| f_\theta(\mathbf{x}) - \mathbf{c} \|_2}_{\text{anomaly score } s(\mathbf{x})}, \qquad (1)$$

where parameters θ of f are optimized to project $\mathbf{x}$ to $\mathbf{c}$. The OC loss is an anomaly score s. If f_θ has bias parameters, fix biases and $\mathbf{c}$ to avoid trivial solution [1], where biases converge to $\mathbf{c}$.

Furthermore, the $f_\theta(\mathbf{x})$ encodes $\mathbf{x}$'s state based on distance and position, thus outputs can be visualized and analyzed via T-SNE [7] as a state embeddings.

Architecture

The architecture consists of an input layer that takes a sequence of 205 features, followed by an LSTM layer with either 32 or 64 hidden units (M) to capture temporal dependencies in the input sequence. The LSTM layer's output is then mapped by a linear layer to real numbers, which is the final model embedding, see Eq. 1.

Inputs

States of the klystrons are expressed by the following waveforms: `FD.F1` is forward power at the first klystron arm, `FD.F2` is forward power at the second klystron arm, `FD.FI` is forward power at klystron input, `FD.R1` is reflected power at first klystron arm, `FD.R2` is reflected power at second klystron arm, and `FD.RI` is reflected power at klystron input. Each waveform consists of 2048 amplitude and phase values.

Pre-processing First, we transform the amplitude and phase into in-phase and quadrature components (IQ). Then we sub-sample waveforms, taking only every 10th value to lower computational requirements. Our experiments show only `FD.RI` is essential without losing too much recognition ability. Each waveform is normalized to zero mean and unit standard deviation.
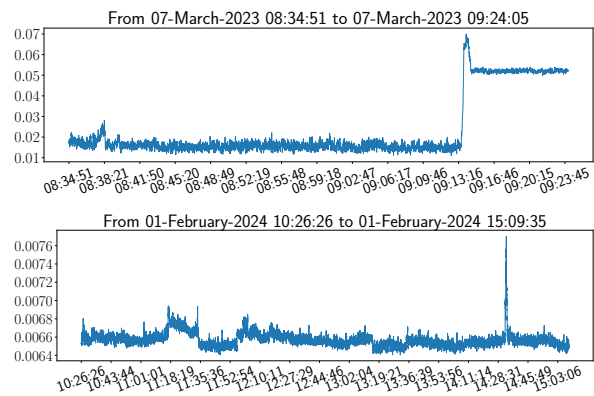

Figure 1: Anomaly scores s of the event that took place on Mar. 7th 2023 (top) and Feb. 1st 2024 (bottom) Top: Notice the bump around 9:13:16 which increases the s until the klystron is shut down (after 9:23:45). Right: The anomaly score fluctuates quite significantly over an extended period (10:24:26 - 15:03:06). There is a peak around 14:30, which precedes severe disruption of observed signals.

RESULTS

We selected two events that were detected post mortem as anomalous with our detection system, from which one led to a fatal failure of klystron (Feb. 1st 2024).

Event on Mar. 7th 2023

The probable cause of this issue is multipacting, where electrons emitted from the cathode or other surfaces within the vacuum envelope of the klystron A13 can become trapped in a resonant trajectory, repeatedly striking the cavity walls. This process can begin suddenly and escalate rapidly, creating a high density electron cloud that can absorb energy from the RF field, leading to excessive heating, outgassing, and ultimately RF breakdown in the klystron.

Our algorithm notices the first issue at 08:38:21, where there is a slight bump in s, see Fig. 1 (top). Furthermore, at around 9:13:16, there is a noticeable jump of s followed by a steady increase in score. An unexpected state change is observed in the `FD.FI` signal, as it is also visible from the phase in Fig. 2 at that time. After this bump, which peaks approximately around 9:14:00, there is a noticeable change in the phase. When we take a detailed look at the T-SNE embedding of the vectors f in that time range, we can divide the space into three separate clusters according to Fig. 4 (left): blue states corresponding to normal state(s), green points which are waveforms that are projected by f away from the blue points, and red states which are states that follow after the peak (a klystron drift).

Event on Feb. 1st 2024

The A16 klystron experienced a catastrophic failure, being unable to sustain the required high voltage, likely due to excessive dark current. The reflected power levels fluctuate erratically, another potential symptom of dark current
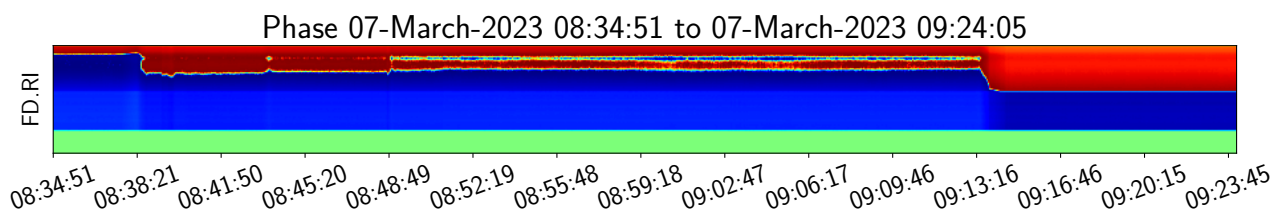
Figure 2: The phase of the event that took place on Mar. 7th 2023. Each column is one waveform `FD.RI` of one pulse.
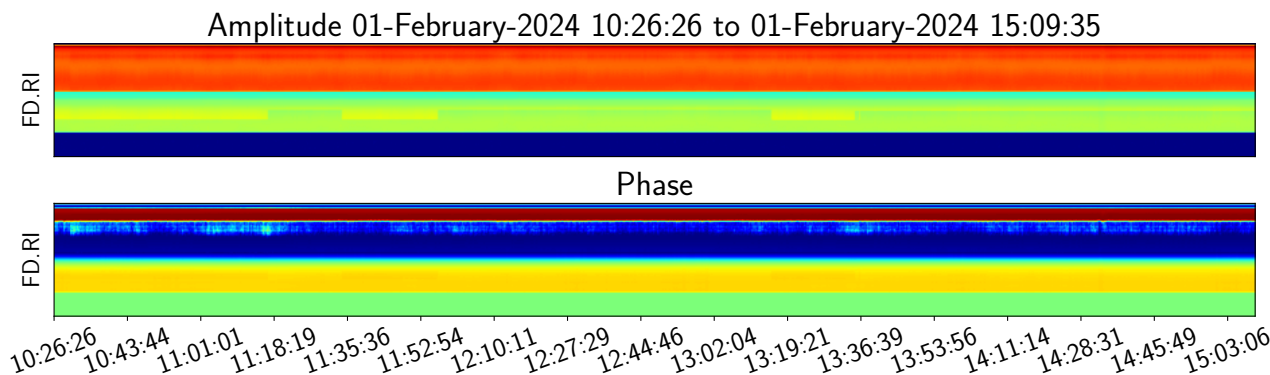

Figure 3: Amplitude and phase of `FD.RI` of the event from Feb. 1st 2024.

issues. Despite attempts to mitigate arcing events, the arcing persisted, eventually leading to permanent damage to the klystron tube. Physical disassembly and inspection of internal components is required for comprehensive failure analysis after klystron replacement.

With the trained algorithm, one can observe mild fluctuations of s, see right Fig. 1. For instance, approximately periods 11:18 - 11:35, 13:02 - 13:19, and finally a noticeable peak around 14:30 which might indicate a bigger problem before the station stops sending signals (after 15:09). By looking at the input signals in Fig. 3, we can see that there are several mode changes (approximately 10:26 - 11:18, 11:35 - 11:52, 13:19 - 13:36), this is also visible in two distinct clusters in the T-SNE embedding in right Fig. 4. What is however quite noticeable is variation in the first third of the phase, where values fluctuate quite significantly.

CONCLUSION

In this work, we demonstrated the application of unsupervised deep one-class classification with an LSTM model for sequential anomaly detection in the operational signals of high-power klystrons at the European XFEL facility. By training solely on normal waveform data, the model learns to characterize standard klystron behavior and identify deviations as potential anomalies or faults. We presented two case studies of actual events - a suspected multipacting issue on Mar. 7th 2023, and a catastrophic klystron failure on Feb. 1st 2024. Our algorithm identified precursors, flagging anomalous waveform patterns before the issues escalated to system failures or downtime which were assessed by the experts.
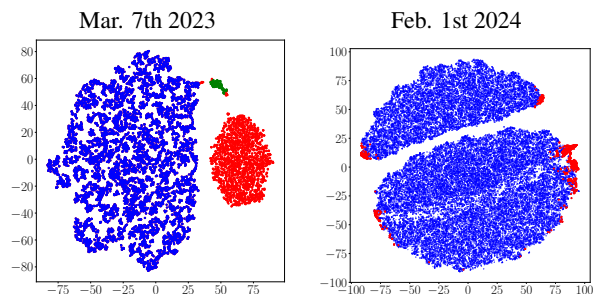
ACKNOWLEDGEMENTS

Figure 4: T-SNE Embedding shows a reduced projection of the network from M dimensions onto 2D while distance is preserved. The colors of the left image encode different s levels (red between $(0.04, 0.06)$, blue between $(0, 0.04)$ and green $(0.06, \infty)$. Similarly on the right figure, red color encodes s above 0.007.

REFERENCES

[1] L. Ruff *et al.*, "Deep one-class classification," in *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, 2018, pp. 4393–4402. `https://proceedings.mlr.press/v80/ruff18a.html`

[2] K. Rehlich, "Status of the ttf vuv-fel control system," in *PCaPAC 2005 conference talk*, 2005.

[3] M. Schütte, J. Georg, A. Grünhagen, and H. Schlarb, "Concept and Design of an Extensible Middle-Layer Application Framework for Accelerator Operations and Development," in *Proc. ICALEPCS'23*, Cape Town, South Africa, Oct. 9–13, 2023, 2024, pp. 30–33.
`doi:10.18429/JACoW-ICALEPCS2023-MO2BCO02`

[4] L. Butkowski, W. Koprek, and K. T. Pozniak, "Klystron lifetime management system," in *2009 MIXDES-16th International Conference Mixed Design of Integrated Circuits & Systems*, 2009, pp. 129–133.

[5] A. Sulc, A. Eichler, and T. Wilsken, "A data-driven anomaly detection on srf cavities at the european xfel," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 2420, 2023, p. 012 070.

[6] L. Ruff *et al.*, "Deep semi-supervised anomaly detection," *arXiv preprint arXiv:1906.02694*, 2019.

[7] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.