ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPS134

ENHANCING QUENCH DETECTION IN SRF CAVITIES AT THE EuXFEL: TOWARDS MACHINE LEARNING APPROACHES AND PRACTICAL CHALLENGES

Annika Eichler[1], Nadeem Shehzad*, Julien Branlard, Lynda Boukela, Burak Dursun, Marco Diomede, Bozo Richter
Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
[1]also at Hamburg University of Technology, Hamburg, Germany

Abstract

Detecting anomalies in superconducting cavities at the EuXFEL is essential for reliable operation. We began with a model-based anomaly detection approach focused on residual analysis. To improve fault discrimination, particularly for quench events, we augmented the detection with a machine learning-based classification. Key challenges are posed by the transition to real-time operation, requiring computational and integration adjustments. For the online application, we deployed two servers at one of the 25 stations to detect and log anomalies with a software implementation. In parallel, we pushed the development of a firmware solution that will counteract critical faults in real-time. At the current stage only the anomaly detection is in online operation, which is planned to be augmented with the online fault classification in the future. The resulting detection system delivers reports across various timescales, supporting both immediate responses and long-term maintenance.

INTRODUCTION

The linac of the EuXFEL operates since 2017 and has demonstrated very high up-time (above 99% during user runs). Typical faults of the radio frequency (RF) system leading to accelerator down-time include klystron gun arcs, coupler sparks or quenches of the superconducting radio frequency (SRF) cavities. A quench detection software, referred to as Quench Detection Server (QDS), based on a statistical analysis of the loaded quality factor of individual cavities has been used since the beginning of operation [1] and has proven quite robust to detect quenches. It however triggers to other faults as well, such as, numerical errors in the digital detection chain, user-triggered changes of external coupling, or field-emitter induced plasma discharge of the SRF cavity field.

Machine learning approaches can help discriminate quenches against other faults and sort them in predefined categories. To this end, a new algorithm based on parity-space methods has been developed and implemented [2]. This method uses a generalized likelihood ratio (GLR) as anomaly indicator. Compared to the QDS, the GLR approach demonstrated a reduced false positive rate in identifying quenches while keeping low false negative rates [3,4]. Although the GLR approach is currently running offline (i.e. on stored trip snapshots), it provides valuable help to the experts reviewing the trips, supporting them with automated postmortem analysis [5].

The objective of this work is to improve the reliability of the EuXFEL, more specifically of the low-level radio frequency (LLRF) by providing an intelligent anomaly detection and classification system. To fulfill this goal, the GLR approach was first implemented as a software solution. Two external CPUs were installed at the linac station A24, with a direct and dedicated connection to the LLRF CPU. The software approach serves two purposes: direct logging of faulty traces and flexible algorithm development to analyze online data. Although most of the analysis performed to date has been focused on quenches, gaining insight on new anomalies will help identify and categorize other faults, where no fast reaction is required. Part of this approach also includes investigating infrastructure limitations and requirements.

A new firmware block has also been implemented in the LLRF system. This new code was deployed on all RF stations during the Winter 2024 shutdown. One obvious benefit of the Field-Programmable Gate Array (FPGA) implementation is that it allows continuous monitoring of all RF traces (i.e. live data) as opposed to postmortem trip snapshots previously detected as faulty by the QDS. This should allow detecting faults until now left unnoticed.

A second benefit of the firmware implementation is the speed of detection and possible reaction. Anomalies can be detected within the RF pulse, compared to the software QDS which can take two to three pulses to detect and react by stopping the RF. With a FPGA-based anomaly detection, a distributed control concept could be setup to reduce the gradient of the station where a quench occurred while increasing those of the neighboring stations for minimal error in beam energy and maximized accelerator up-time. The concept has been laid out in [6].

METHOD

The analysis is based on the physical SRF cavity model as derived in [7]. An Euler-discretized residual is obtained with the nonlinear parity space method,

$$\begin{aligned} r(k) = &- P_I(k)\big(- P_I(k+1) + P_I(k) \\ &\quad + \omega_{1/2}T\left[-P_I(k) + 2F_I(k) - B_I(k)\right]\big) \\ &- P_Q(k)\big(P_Q(k+1) - P_Q(k) \\ &\quad + \omega_{1/2}T\left[P_Q(k) - 2F_Q(k) + B_Q(k)\right]\big), \end{aligned}$$

* nadeem.shehzad@desy.de

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPS134

where I and Q are the in-phase and quadrature components of the RF signals. $F(k)$ is the forward field coupled into the cavity, $P(k)$ is the probe signal, i.e., the field generated inside the cavity, $B(k)$ is the field induced by the beam, $\omega_{1/2}$ is the half bandwidth, and $T = 1.11 \times 10^{-7}$s the sampling period. The residual is then analyzed with the generalized likelihood method. When the GLR exceeds a predefined threshold, it indicates a high probability of fault occurrence,

$$\mathrm{GLR}(k) = \frac{K}{2}\left(\frac{1}{K}\sum_{i=k-K+1}^{k} r(i)^{\top}\right)\sigma^{-1}\left(\frac{1}{K}\sum_{i=k-K+1}^{k} r(i)\right),$$

with K being the moving averaging window. More details of the derivations can be found in [2, 8]. In order to classify faults and isolate quenches, a third step is necessary. As this is not yet implemented online, we will not go further into detail here. A comparison of different methods can be found in [3]. The link between classification neural networks and firmware requirements is given in [5].

SOFTWARE

The software is designed to provide continuous online detection of anomalies in the GLR. The server utilizes the ChimeraTK [9] software framework to communicate with other servers that provide LLRF signals and timing information. To ensure accurate anomaly detection, the server enforces synchronization of all incoming signals by matching the timing information, specifically, the macro pulse number (MPN) associated with each signal. Due to effects from coupling and stored energy dynamics, signal calibration is required and achieved offline using scripts developed by [10]. For the anomaly detection procedure, a nominal signal and a GLR are calculated for each input signal during every pulse. The rules used to define what constitutes an anomaly are configured in the server by a domain expert, typically after visually inspecting the behavior of the nominal and GLR calculations under normal operating conditions. A live monitoring panel at A24 can be seen in Figure 1 using JDDD [11]. A threshold parameter is set to determine

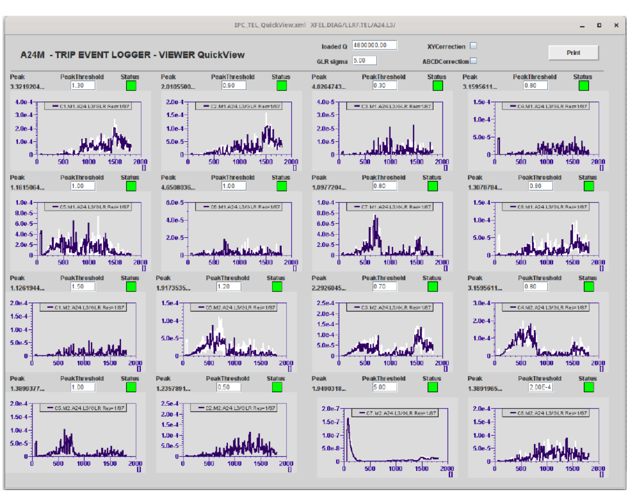

Figure 1: Pulse by pulse anomaly monitoring at EuXFEL.

whether the peak value of the GLR exceeds a predefined limit. All peak values are logged to a history file. To prevent false alarms for example, during the system's ramp-up phase the cavity energy is monitored to ensure it remains within a reasonable range. To further reduce false detections due to noise or under/over-shoots, the starting point for inspecting the GLR peak can be shifted dynamically along the sample points. All detected anomalies are logged to disk for later study. The server maintains a ring buffer that stores nominal data just before an anomaly is detected, which helps in analyzing and understanding the event in greater detail. The in-crate CPUs at EuXFEL host many critical applications and operate near their performance limits. Therefore, two powerful external CPUs have been installed to host this software. However, because the application relies on ZMQ to receive data from other components, observations indicate that during normal operation, outgoing traffic on the interface rises from 20 MB/s to 80 MB/s when anomaly detection is active as shown in Figure 2. This increased traffic load occasionally results in DAQ packet loss, particularly when the available bandwidth is temporarily saturated.

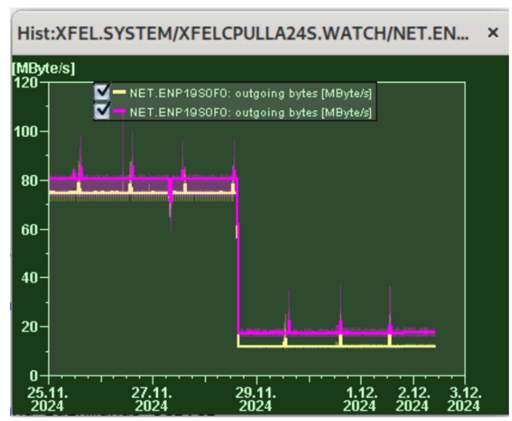

Figure 2: Bandwidth analysis at A24M.

To mitigate this, an additional direct network connection has been established between the external CPU and the crate CPU at the test station.

FIRMWARE

The FPGA implementation process began with developing a double-precision floating-point Python model of the anomaly detection algorithm. This model was fed with actual system data to establish the expected outputs for both nominal operation and anomalous events like SRF cavity quenches. Subsequently, a fixed-point conceptual design was created in Python, specifically tailored for the target FPGA architecture. The fixed-point model was critical for optimizing numerical performance, involving extensive experimentation over the region of interest to analyze dynamic range and precisely adjust resolution and scaling for fixed-point operations in the digital signal processing pipeline as shown in Figure 3. The optimized design was then translated into a VHDL implementation that generates sufficiently matching GLR with respect to floating-point computations. Rigorous

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPS134

verification of the RTL design was performed against machine captured test vectors, with the co-simulated outputs compared against both the fixed-point and floating-point Python models to precisely quantify its numerical performance.

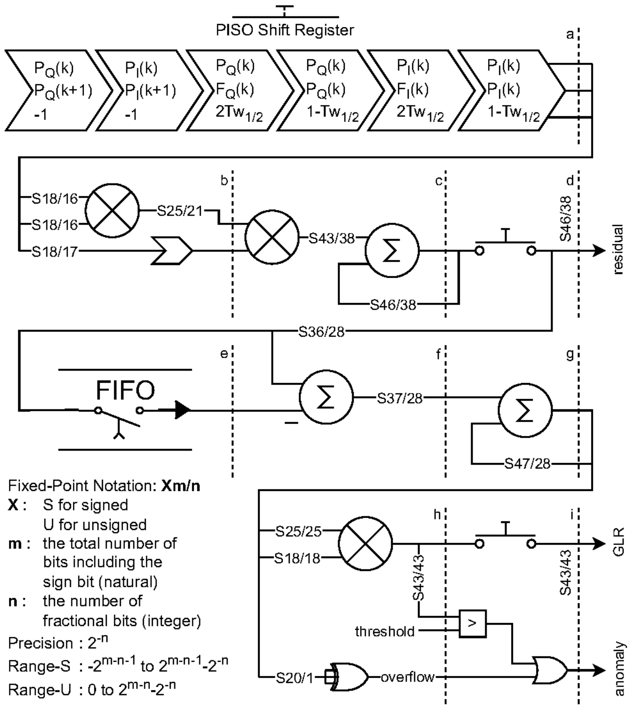

Figure 3: Block diagram of FPGA implementation.

Beyond numerical accuracy, the FPGA implementation was also carefully optimized for its performance and resource use (Table 1). The design achieves a throughput of 1/7 clock cycles. This optimization led to a pipeline latency of 9 clock cycles for the residual calculation and 12 clock cycles for the GLR computation. However, the effective latency for the GLR is further influenced by its embedded moving average. For a typical window size of $K = 450$ samples and a sample aperture of 9 clock cycles, this results in an effective GLR latency of 4062 clock cycles. At the application clock rate of 81.25 MHz, this translates to approximately 50 µs.

Table 1: FPGA Resource Utilization

Resource Type	Total Available	LLRF Control	Anomaly Cavity / System
DSP	1680	270	3 / 48
BRAM	835	422.5	1 / 16
LUT	260600	102579	371 / 5936
FF	521200	145471	519 / 8304

CONCLUSION

Numerous validation campaigns have consistently demonstrated the system's robust capabilities. An experiment involving a controlled cavity quench at C8.M2.A24 highlights its effectiveness. Both the software and firmware implementations identified anomalous behavior six pulses before the RF was turned off. While direct GLR waveforms were unavailable from the firmware during the experiment, the captured machine data was subjected to simulation employing a bit-accurate firmware model. This simulation precisely mirrored the anomaly flag pattern observed during the experiment, thereby substantiating the fidelity of the simulated firmware GLR. Figure 4 presents the comparison of these simulated firmware GLR plots with those computed by production software, alongside the amplitude and phase of the probe signals captured from the station over eight successive pulses.

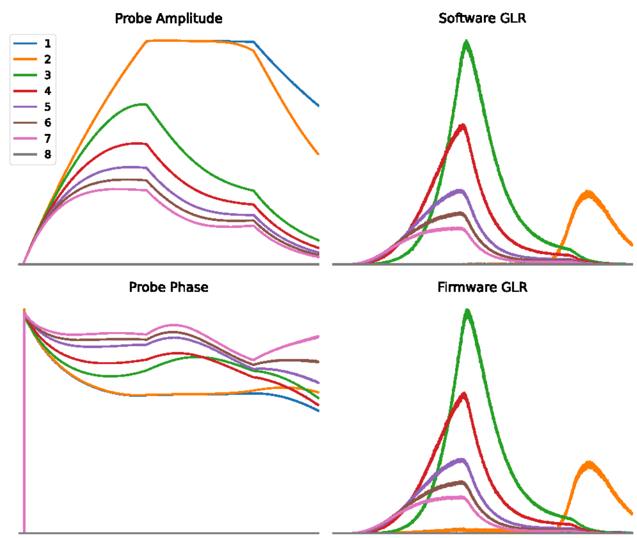

Figure 4: Quench at C8.M2.A24 over eight successive pulses.

Currently, the firmware implementation of the GLR approach is operational throughout all EuXFEL stations for diagnostics. The software implementation, however, is restricted to station A24, as it necessitates dedicated server hardware for its intricate computations. This software setup proves instrumental for further algorithm development and comprehensive analysis. Notably, it presently encompasses computations for additional calibration designed to eliminate coupling between forward and reflected signals, and it furthermore provides compensation for the beam's influence on the probe signal.

This study represents a critical step toward reducing accelerator downtime by implementing anomaly response mechanisms at the firmware level, with continuous software development used for method validation and verification.

ACKNOWLEDGEMENTS

This work was funded in the context of the R&D program of the EuXFEL. The authors acknowledge support from DESY (Hamburg, Germany), a member of the Helmholtz Association HGF. They also thank Vladimir Rybnikov for his input.

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPS134

REFERENCES

[1] J. Branlard, V. Ayvazyan, O. Hensler, H. Schlarb, C. Schmidt, and W. Cichalewski, "Superconducting Cavity Quench Detection and Prevention for the European XFEL", in *Proc. ICALEPCS'13*, San Francisco, CA, USA, Oct. 2013, paper THPPC072, pp. 1239–1241. https://jacow.org/ICALEPCS2013/papers/THPPC072.pdf

[2] A. Eichler, J. Branlard, and J. H. K. Timm, "Anomaly detection at the European X-ray Free Electron Laser using a Parity-Space-Based Method", *Phys. Rev. Accel. Beams*, vol. 26, no. 1, p. 012801, 2023. doi:10.1103/PhysRevAccelBeams.26.012801

[3] L. Boukela, A. Eichler, J. Branlard, and NZ. Jomhari, "A Two-Stage Machine Learning-Aided Approach for Quench Identification at the European XFEL", *IFAC-PapersOnLine*, vol. 58, no. 4, pp. 402–407, 2024. doi:10.1016/j.ifacol.2024.07.251

[4] J. Branlard, A. Eichler, J. H. K. Timm, and N. Walker, "Machine Learning Assisted Cavity Quench Identification at the European XFEL", in *Proc. LINAC'22*, Liverpool, UK, 2022, pp. 798–801. doi:10.18429/JACoW-LINAC2022-THPOPA26

[5] L. Boukela, J. Branlard, and A. Eichler, "Exploring NAS for Anomaly Detection in Superconducting Cavities of Particle Accelerators", *Front. Phys.*, vol. 13, p. 1553993, 2025. doi:10.3389/fphy.2025.1553993

[6] S. Pfeiffer, H. Werner, C. Schmidt, and H. Schlarb, "Distributed Controller Design For a Free-Electron Laser", in *2013 American Control Conference*, Washington, DC, USA, Jun. 2013, pp. 6553–6558. doi:10.1109/ACC.2013.6580867

[7] T. Schilcher and A. Gamp, "Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities", Ph.D. thesis, University of Hamburg, 1998. https://bib-pubdb1.desy.de/record/416390

[8] A. Nawaz, C. H. né Hoffmann, J. Graßhoff, S. Pfeiffer, G. Lichtenberg, and P. Rostalski, "Probabilistic Model-Based Fault Diagnosis for the Cavities of the European XFEL", *at - Automatisierungstechnik*, vol. 69, no. 6, pp. 538–549, 2021. doi:10.1515/auto-2020-0064

[9] ChimeraTK: Control system and Hardware Interface with Mapped and Extensible Register-based device Abstraction Tool Kit. https://github.com/ChimeraTK/

[10] A. Bellandi, J. Branlard, M. Diomede, M. Herrmann, S. Pfeiffer, and C. Schmidt, "Calibration of Superconducting Radio-Frequency Cavity Forward and Reflected Channels based on Stored Energy Dynamics", *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 1069, p. 169825, 2024. doi:10.1016/j.nima.2024.169825

[11] DOOCS Collaboration, The Distributed Object Oriented Control System (DOOCS). http://doocs.desy.de/