ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB344

MACHINE LEARNING MODELS FOR BREAKDOWN PREDICTION IN RF CAVITIES FOR ACCELERATORS

C. Obermair*[1], A. Apollonio, T. Cartier-Michaud, W. L. Millar, F. Pernkopf[1],
W. Wuensch, N. Catalan-Lasheras, L. Felsberger
CERN, Geneva, Switzerland
[1]also at Graz University of Technology, Graz, Austria

Abstract

Radio Frequency (RF) breakdowns are one of the most prevalent limits in RF cavities for particle accelerators. During a breakdown, field enhancement associated with small deformations on the cavity surface results in electrical arcs. Such arcs degrade a passing beam and if they occur frequently, they can cause irreparable damage to the RF cavity surface. In this paper, we propose a machine learning approach to predict the occurrence of breakdowns in CERN's Compact LInear Collider (CLIC) accelerating structures. We discuss state-of-the-art algorithms for data exploration with unsupervised machine learning, breakdown prediction with supervised machine learning, and result validation with Explainable-Artificial Intelligence (Explainable AI). By interpreting the model parameters of various approaches, we go further in addressing opportunities to elucidate the physics of a breakdown and improve accelerator reliability and operation.

INTRODUCTION

The novel RF cavities of CERN's Compact LInear Collider (CLIC) are designed for high gradient operation at ~100 MV/m [1]. Even though RF cavities are operated in vacuum, local field emissions can cause arcs and breakdowns of the electric field in the cavity which have a negative effect on the cavity surface material. The frequency of these arcs, described by the breakdown rate, is the main limitation to increase the electric field in an RF cavity during conditioning and operation. While historically RF structures have been conditioned in a manual way by machine operators, an automated conditioning algorithm is in place at the CLIC test stand to gradually increase the field gradient while maintaining a pre-defined target breakdown rate [2]. These conditioning efforts set the limit to the gradient due to field emission, caused by the geometrical defects of the surfaces and the RF power flow, to reduce the likelihood of a breakdown [3, 4]. Given the limited understanding of the origin and evolution of RF breakdowns, current optimization algorithms aim for a progressive recovery of operating conditions by a temporary limitation of the RF power after a breakdown, but do not avoid breakdowns in the first place. Recently, data-driven machine learning algorithms have been deployed successfully for incorporating sequential dynamics [5, 6] using the large amount of experimental data available. Ongoing efforts already try to predict breakdowns in the RF power source output of CERN's LINAC 4 [7], or to classify superconducting RF faults at Jefferson Laboratory [8].

This paper gives an overview of several data-driven methods for RF breakdown analysis, specifically suited to the properties of the measurement data of the CLIC XBOX-2 test stand at CERN. The paper provides an introduction, comparison, and hands-on experience of existing data-driven modeling approaches to non-machine learning experts. It provides RF physicists and engineers with machine learning based tools, which allow to gain insights in observing abnormal behaviours. Finally, first results of these methods applied to the CLIC XBOX-2 test stand data are presented.

The paper is structured as follows. First, the properties of the CLIC XBOX-2 test stand and its historical data are described. Consecutively, a broad overview of existing machine learning algorithms, suitable for breakdown prediction in the test stand, is given. Finally, their strengths and the limitations are discussed, first results are presented and an outlook is given.

TEST STAND SETUP

The CERN XBOX-2 test stand is part of the CLIC e+e- collider research program for high gradient acceleration in high gradient structures. It is one of three high power test stands at CERN and its primary objective is to study the RF breakdown phenomenon. A low level radio frequency generator creates a 1.5 μs long, 12 GHz phase-modulated pulse. This pulse is amplified by a klystron and a pulse compressor and is then transferred through a copper wave guide to the RF cavity [9]. A diagram of the high-power portion of the test stand layout is provided in Fig. 1. The RF cavity is represented as Device Under Test (DUT). The signals from the upstream and downstream Faraday cups, which measure the dark current in the structure, are symbolized by the blue arrows labelled DC UP and DC DOWN.

METHODOLOGIES

In order to give a hands-on overview, in this paper, the choice of an algorithm is governed by the chronological order of the data processing, i.e. transformation, exploration, modeling, and explanation of RF cavity specific data.

Existing open-source libraries are used instead of handcrafted methods in all processing steps, because the engineering and the maintenance of customized methods is time consuming. For the labeled measurement data from the XBOX-2 test stand, dedicated toolboxes are used for feature

* christoph.obermair@cern.ch

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB344

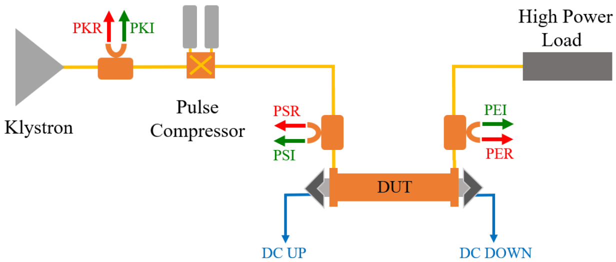

Figure 1: Schematic layout of CERN's Xbox-2 test stand. The red and green arrows show the reflected and forward RF signals, respectively, which are sampled via directional couplers.

calculation [10], time series classification [11], and interpretation of model predictions [12].

Transformation

The sensor data from the XBOX-2 high power tests is divided into so-called *trend* data and *event* data. While the trend data contains single scalar features (e.g. temperatures), the event data contains time series signals, generally sampled with a frequency of 1.6 GHz. In total there are 90 GB of data available from a period of six months of XBOX-2 operation in 2018. These data do not only contain runs in which the operational setting was stable, but also commissioning data with variable operational settings. Thus, it is crucial to initially clean the data, create fast queries, memory efficient storage and file formats with diverse usability. Figure 2 shows the condition summary of the data, where the runs with stable operational settings are highlighted in yellow and the cumulative number of breakdowns is shown in red. The plot further shows the input power in blue, and the pulse width of the input signal in green.

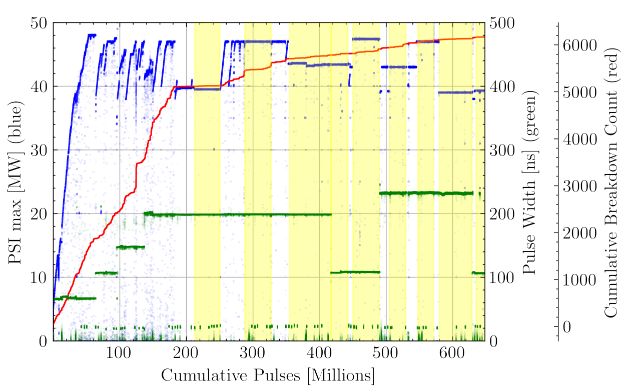

Figure 2: Condition summary of available data. The yellow area represents the runs during which the operational settings were kept stable.

A breakdown results in a burst of current in the cavity, which can be detected by the Faraday cups next to the structure. Therefore, for each event data signal, a label *healthy* ($y = 1$) and *breakdown* ($y = 0$) is assigned by the XBOX-2 experts by setting a threshold on the DC (Faraday cup) signals and the reflected signals. However, as the DC signals are the most reliable filter for structure breakdowns, the RF signals accounting for the reflected power in the structures are not considered for breakdown prediction. Specifically, this means a signal is considered a breakdown, if one of the DC time series signals goes below -0.05 A. In addition, a label is considered a so-called *follow-up breakdown*, if there has already been a breakdown within less than a minute from its occurrence. After filtering out the test stand commissioning data, where most of the breakdowns occurred, 124,448 healthy events and 479 breakdown events, out of which 250 are follow-up breakdowns, remained for further analysis. This class imbalance is tackled by only taking a sub-set of healthy signals and by assigning class weights to the breakdown events during optimization of the algorithm and during computation of the performance measure.

Merging and synchronizing the trend data with the event data is a critical data transformation step. Due to its high sampling frequency, an event data signal with up to 3200 sample points is stored every minute. Exceptions are breakdown events, where the prior two event data signals are stored each time a pulse is injected into the RF cavity. The scalar values of the trend data take up much less space, and are therefore stored every second. During merging of event data and trend data, causality is ensured by always taking the closest information in the past, not in the future.

Exploration

During the exploration phase the goal is to get a quick initial understanding of the data and to validate the transformation step, i.e. if the preceding data cleaning was successful. If there are still outlier signals, which are fundamentally different from the other signals, they have to be understood and, if applicable, neglected. Ideally, a 2D-representation should be found for each event in the high dimensional data, without losing any information due to the dimension reduction. This allows to see correlations and clusters within the representations in one glance. Several unsupervised machine learning methods aim to determine low-dimensional representations from the high dimensional data, including but not limited to principle component analysis [13], stochastic neighbor embeddings [14], and representation learning methods based on neural networks [14–16].

An example is shown in Fig. 3, where the XBOX-2 trend data is transformed into a two dimensional space with 2D-tSNE [14]. 2D-tSNE transforms pairs of data points to joint probabilities, where close points have high probabilities and points which are far apart have low probabilities. Consequently, the Kullback-Leibler divergence within the joint probabilities of the low-dimensional representations and the high-dimensional data is minimized iteratively. While the axes lose their physical meaning during the dimension reduction, one can clearly see clusters of breakdowns and healthy signals in the left plot and the nine different stable runs in the right plot of Fig. 3. Neither the information of the label, nor the information of the runs were given to the algorithm during training.

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB344

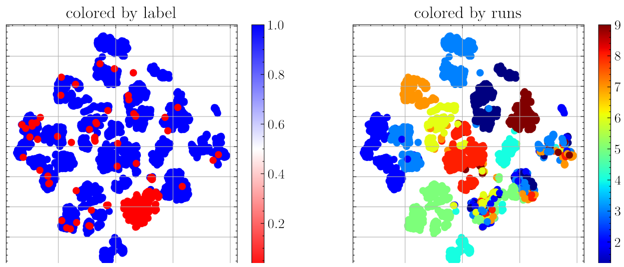

Figure 3: 2D-tSNE of XBOX-2 trend data during stable operation. The algorithm was able to distinguish between healthy and breakdown signals (left) and between stable runs (right). No information about the labels was given to the algorithm.

Modeling

We propose two supervised machine learning stages. First, the behaviour of the trend data over time is investigated. This means that a window, covering a certain time-span of data, is moved over the data-set. For each window a prediction is made if a breakdown will occur in a certain time period. In this step, no shuffling of time series data is allowed due to the sequential dependency. Recurrent neural networks, like long short-term memory networks [17], are especially suited to process this temporally dynamic behavior due to their recurrent neuron connections. The model is trained with several rounds of leave-one-out cross-validation. One round of cross-validation involves splitting the data-set into a training, validation and test set, based on the given runs. This process is repeated until each run was used as training, validation and test set. In a second stage, it is assumed that only the signals before a breakdown are essential to predict the breakdown. Therefore, the signals from the event data are taken and treated independently with their own label *breakdown in the next pulse*. This has the advantage that convolutional neural networks can be used to classify the time-series signals [18]. Additionally, the data-set is shuffled, and the signals are randomly split into training, validation, and test sets.

Explanation

To increase the reliability of a system, understanding why the prediction was made, i.e. looking for a precursor, is often more important than the prediction itself. Especially when designing upgrades of existing systems, a deep understanding of the root cause of the failures can be an invaluable asset. As data-driven models are often black-boxes, explainable-AI does not only help the user to better interpret the behaviours of the models, but it also helps to build trust in the prediction, to validate the results, and to find possible errors within the earlier data processing steps. One can either explore each prediction separately to gain trust in a prediction (instance wise explanation) [19–22], or investigate all predictions to gain trust in a model (population wise explanation) [23]. Both approaches are applicable for explaining predictions of RF cavity breakdowns.

RESULTS & CONCLUSION

Table 1 shows the results of the trained supervised models. The balanced accuracy is used for taking into account the strong class imbalance. It is calculated by averaging the fraction of correctly categorization breakdowns and healthy signals.

Table 1: Balanced Accuracy of Classifying and Predicting Breakdowns With XBOX-2 Data. The Separation Indicates Different Results on Breakdowns / Follow-up Breakdowns

	Classification of Breakdowns	**Prediction of Breakdowns**
Trend Data	100%	91%
Event Data	100%	65% / 98%

The classification step is required for the validation of the algorithms applied on the trend and event data. The balanced accuracy of 100% for trend and event data in the classification shows the successful validation of the algorithms.

The models achieved a balanced accuracy of 91% for predicting breakdowns in the next pulse using trend data. Here, explainable-AI showed that the models made decisions mainly by using the vacuum signals. After further investigation, it was found that a rise in the vacuum pressure mostly occurred just before a breakdown and not only after a breakdown, as generally assumed. This rise in vacuum pressure might be due to small breakdowns happening just before a major breakdown. Further experiments in the test stand are ongoing to validate this result and exclude any artefacts due to signal timing in the experimental setup.

By using the time-series signals of the event data, a balanced accuracy of 65% was achieved for predicting breakdowns, and 98% for predicting follow-up breakdowns. Here, explainable-AI indicates precursors in multiple ways, pointing to the most important part of each measurement, or indicating the three most similar events present in the rest of the data set.

Using this method, an additional precursor has been identified. Faraday cup signals with a small spike, which occurs relatively late in the signal but does not reach the breakdown threshold, often leads to consecutive breakdowns in the next pulse. Following further validations of these results, an operational tool for breakdown reduction based on the described machine learning methods will be developed.

REFERENCES

[1] E. Sicking and R. Ström, "From precision physics to the energy frontier with the Compact Linear Collider", *Nature Physics*, vol. 16, pp. 386–392, 2020. doi:10.1038/s41567-020-0834-8

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB344

[2] W. Wuensch *et al.*, "Experience Operating an X-band High-Power Test Stand at CERN", in *Proc. 5th Int. Particle Accelerator Conf. (IPAC'14)*, Dresden, Germany, Jun. 2014, paper WEPME016, pp. 2288–2290.

[3] A. Grudiev and W. Wuensch, "A New Local Field Quantity Describing the High Gradient Limit of Accelerating Structures", in *Proc. 24th Linear Accelerator Conf. (LINAC'08)*, Victoria, Canada, Sep.-Oct. 2008, paper THP063, pp. 936–938.

[4] A. Descoeudres, Y. Levinsen, S. Calatroni, M. Taborelli, and W. Wuensch, "Investigation of the dc vacuum breakdown mechanism", *Physical Review Special Topics - Accelerators and Beams*, vol. 12, p. 092001, 2009. doi:10.1103/PhysRevSTAB.12.092001

[5] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling", in *Proc. of the Annual Conf. of the Int. Speech Communication Association(PACISCA'14)*, USA, 2014, pp. 338-342.

[6] N. Catalan-Lasheras *et al.*, "Commissioning of XBox-3: A Very High Capacity X-band Test Stand", in *Proc. 28th Linear Accelerator Conf. (LINAC'16)*, East Lansing, MI, USA, Sep. 2016, paper TUPLR047, pp. 568–571.

[7] Y. Donon *et al.*, "Extended anomaly detection and breakdown prediction in LINAC 4's RF power source output", in *Proc. of Int. Conf. on Information Technology and Nanotechnology (ITNT'20)*, Samara, Russia, May 2020, pp. 1-7. doi:10.1109/ITNT49337.2020.9253296

[8] C. Tennant, A. Carpenter, T. Powers, A. S. Solopova, L. Vidyaratne, and K. Iftekharuddin, "Superconducting radiofrequency cavity fault classification using machine learning at Jefferson Laboratory", *Physical Review Accelerators and Beams*, vol. 23, p. 114601, 2020. doi:10.1103/PhysRevAccelBeams.23.114601

[9] B. J. Woolley, "High Power X-band RF Test Stand Development and High Power Testing of the CLIC Crab Cavity", Ph.D. dissertation, Lancaster University, UK, 2015.

[10] Tsfresh, https://tsfresh.readthedocs.io

[11] Github, https://github.com/hfawaz/dl-4-tsc

[12] Shap, http://shap.readthedocs.io

[13] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis", *Chemometrics and Intelligent Laboratory Systems*, vol. 2, pp. 37-52, 1987. doi:10.1016/0169-7439(87)80084-9

[14] G. Hinton and R. Sam, "Stochastic Neighbor Embedding", in *Proc. of the 15th Int. Conf. on Neural Information Processing Systems (NIPS'02)*, Cambridge, MA, USA, Jan. 2002, pp. 857–864.

[15] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1798–1828, 2013. doi:10.1109/TPAMI.2013.50

[16] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised Scalable Representation Learning for Multivariate Time Series", in *Proc. 33rd Conf. on Neural Information Processing Systems (NeurIPS'19)*, Vancouver, Canada, Dec. 2019, pp. 4650–4661.

[17] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, vol. 9, pp. 1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735

[18] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification", *Data Mining and Knowledge Discovery*, vol. 33, pp. 917–963, 2019. doi:10.1007/s10618-019-00619-1

[19] J. V. Jeyakumar, J. Noor, Y.-H. Cheng, L. Garcia, and M. Srivastava, "How Can I Explain This to You? An Empirical Study of Deep Neural Network Explanation Methods", in *34th Conf. on Neural Information Processing Systems (NeurIPS'20)*, Vancouver, Canada, Dec. 2020, pp. 4211–4222.

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions", in *Proc. 31st Conf. on Neural Information Processing Systems (NIPS'17)*, Long Beach, CA, USA, Dec. 2017.

[21] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier", 2016. arXiv:1602.04938

[22] S. Tonekaboni, S. Joshi, K. Campbell, D. K. Duvenaud, and A. Goldenberg, "What went wrong and when? Instance-wise feature importance for time-series black-box models", in *Proc. 34th Conf. on Neural Information Processing Systems (NeurIPS'20)*, Vancouver, Canada, Dec. 2020, pp. 262–269.

[23] M. A. Wojtas and K. Chen, "Feature Importance Ranking for Deep Learning", in *Proc. 34th Conf. on Neural Information Processing Systems (NeurIPS'20)*, Vancouver, Canada, Dec. 2020, pp. 5105–5114.