UNIVERSITÄT ZU LÜBECK

From the Institute for Electical Engineering in Medicine
of the University of Lübeck

Director: Prof. Dr.-Ing. Philipp Rostalski

Probabilistic Model-Based Anomaly Detection for the European X-Ray Free Electron Laser

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

from the Department of Computer Sciences and Technical Engineering

Submitted by
Ayla S. Nawaz, M.Sc.
from Pforzheim

Lübeck 2021

First referee: Prof. Dr.-Ing. Philipp Rostalski

Second referee: Prof. Dr. rer. nat. Thorsten M. Buzug

Date of oral examination: April 20th, 2022

Approved for printing. Lübeck, April 26th, 2022

Dedicated to
Hyder & Margarete Nawaz
and to Mirko Steffen

Abstract

The European X-Ray Free Electron Laser (EuXFEL) is one of the most complex machines on earth. Its large number of components, available system observations, operational parameters, and the interconnectedness of many sub-components motivate the need for automated fault detection and handling.

In order to handle faults, component-specific solutions that meet individual fault detection requirements can be considered. At the same time, root cause analysis for the entire machine should account for existing interdependencies between the components. This goal can be achieved using a versatile and general framework that encompasses various algorithms, can deal with the heterogeneity of the components, and can be used to find solutions in a modular way.

In this thesis, the methodological focus is on the use of factor graphs, which graphically represent probabilistic models, and inference-based algorithms are described by the passing of messages on the graph. Due to their capability of defining a versatile spectrum of models, factor graphs can be used for different control, signal processing, or machine learning tasks. In this thesis, factor graphs are applied to the fault detection of one of the central components in the EuXFEL—the pulsed-mode operated superconducting cavities. In addition, two alternative, deterministic methods, not part of the factor graph framework, were tailored to the cavity model. For one, a nonlinear parity space residual was defined for the cavity model, characterized above all by its computational simplicity. The second deterministic method is a parameter-based approach, which can make use of already existing FPGA implemented components.

The methods are tested on vast amounts of cavity data from several weeks of operation. The evaluation of these cavity measurements provided new insights into the behavior of the cavities. In addition, the methods are tested on the only so-far automatically detected cavity fault known as quench. It was shown that anomalies could be detected before the onset of a quench, which was not possible with previous methods.

The detection results confirm that rigorous and detailed analysis of available component measurements can lead to new insights into system characteristics that can help to avoid unwanted downtime of the machine. The factor graphs of the inference-based algorithms present a first building block for a fault detection scheme that can be expanded to other essential components of the EuXFEL.

Zusammenfassung

Der European X-Ray Free Electron Laser (EuXFEL) ist eine der komplexesten Maschinen der Welt. Seine große Anzahl an Komponenten, die verfügbaren Systembeobachtungen, die Betriebsparameter und die Verflechtung vieler Teilkomponenten machen eine automatisierte Erkennung von Fehlern und deren Behandlung erforderlich.

Um Fehler automatisiert handhaben zu können, sind komponentenspezifische Lösungen notwendig, die den jeweiligen Anforderungen an eine Fehlererkennung gerecht werden. Gleichzeitig sollten bei der Ursachenanalyse für die gesamte Maschine die bestehenden Abhängigkeiten zwischen den Komponenten berücksichtigt werden. Dieses Ziel kann durch ein vielseitiges und allgemeines Framework erreicht werden, das verschiedene Algorithmen umfasst, mit der Heterogenität der Komponenten umgehen kann und mit dem Lösungen modular gefunden werden können.

In dieser Arbeit wird der methodische Schwerpunkt auf die Verwendung von Faktorgraphen gelegt, welche probabilistische Modelle grafisch repräsentieren und auf denen inferenzbasierte Algorithmen durch die Weitergabe von Nachrichten beschrieben werden. Aufgrund ihrer Fähigkeit, ein vielseitiges Spektrum von Modellen zu definieren, können Faktorgraphen für unterschiedliche Aufgaben der Regelungstechnik, der Signalverarbeitung oder des maschinellen Lernens eingesetzt werden. In dieser Arbeit werden Faktorgraphen auf die Fehlerdetektion einer der zentralen Komponenten im EuXFEL, den im Pulsmodus betriebenen supraleitenden Kavitäten, angewendet. Darüber hinaus wurden zwei deterministische Methoden, die nicht Teil des Factor Graph Frameworks sind, auf das Hohlraummodell zugeschnitten. Ein nichtlineares *Parity Space*-Residuum für das Hohlraummodell definiert, welches sich vor allem durch seine Einfachheit auszeichnet. Die zweite deterministische Methode ist ein parameterbasierter Ansatz, der auf bereits vorhandene FPGA-implementierte Komponenten zurückgreifen kann.

Die Methoden werden an einer großen Menge von Kavitätsdaten aus mehreren Wochen Betrieb getestet. Die Auswertung dieser Daten ergab neue Erkenntnisse über das Verhalten der Kavitäten. Darüber hinaus werden die Methoden an dem einzigen bisher automatisch detektierten Kavitätsfehler, dem Quench getestet. Es konnte gezeigt werden, dass Anomalien bereits vor dem Einsetzen der untersuchten Quenche erkannt wurden, was mit bisherigen Methoden nicht möglich war.

Die Ergebnisse der Untersuchung zeigen, dass eine gründliche und detaillierte Analyse der verfügbaren Komponentenmessungen zu neuen Erkenntnissen über die Systemeigenschaften führen kann, die helfen können, ungewollte Stillstandszeiten der Maschine zu vermeiden. Die Faktorgraphen der inferenzbasierten Algorithmen stellen einen ersten Baustein für ein Fehlererkennungssystem dar, welches auf andere wesentliche Komponenten des EuXFEL erweitert werden kann.

Acknowledgments

Sitting in front of my computer, I feel relaxed, excited, and nostalgic while adding a few personal lines.
Three research groups were involved in this thesis: The Institute for Electrical Engineering in Medicine (IME), the German Electron Synchrotron (DESY), and the Hamburg University of Applied Sciences (HAW). Each collaboration has left me with many people to whom I would like to express my gratitude. My supervisors Prof. Dr. Philipp Rostalski and Prof. Dr. Gerwald Lichtenberg let me in on some of the secrets of science and engineering. The thesis evolved the way it did because they both were keen on bringing their backgrounds to the table.
Philipp created a genuinely positive working environment using his clear vision of team spirit and humor. I am very fortunate to have worked with him and the first generation of IME doctoral candidates.
Gerwald used his out-of-the-box thinking to equip me with foresighted hands-on advice that made many steps much easier. I looked forward to every single meeting.
All of the application results stem from working at DESY.
Holger Schlarb allowed me to attend several international conferences and supported the collaborations.
Sven Pfeiffer, above all, always believed in me, keeping his doors open to questions and repeatedly sharing his seemingly limitless knowledge about the machine.
Annika Eichler joined the DESY group in the final years of this project as interest in the community rose to work on fault diagnosis, automation, and machine learning for accelerators. She impressed me on more levels than I care to count, foremost with her powerful leadership approach, including the perfect amount of support, patience, and kindness. I am delighted to see how initial ideas about automation tasks have gained momentum and that she is continuing this work.
Jan Timm put time and effort into helping me get the large amounts of data analyzed and supported my work repeatedly when reruns were necessary.
For most of my thesis, I shared an office with Andrea Bellandi. His skill of clearly explaining complex principles both in physics and programming left me more than once in awe.
Julien Branlard is a master in creating an energetic, positive, and constructive atmosphere in meetings, seminars, conferences, and office building halls, which made collaborative research and development much easier.
There are several more colleagues I will happily think of when I recount the DESY years, such as Christian Schmidt, Martin Hierholzer, Martin Killenberg, Mattias Hoffmann, and all the rest of MSK.
One and a half years into my thesis, the collaboration with the IME intensified, resulting in the work on factor graphs.
Christian Herzog pointed out when I went wrong and when I went right; his ability to find the right words in key moments was uncanny. The feedback

I received reflected his love for detail and helped me increase the quality of my texts and presentations. Although the IME group was still small, consisting exclusively of computer scientists and engineers, his focus on ethics encouraged us all to participate in enriching discussions that took into consideration the social responsibility of our work.
Jan Graßhoff and Eike Petersen, worked with me in a virtual office environment during the corona pandemic. For more than one year, almost every day (including weekends and late nights), we communicated, keeping the feeling of an office alive.
I shared many wonderful moments of focused work with Jan, and I am thankful for his incredible patience, encouragement, and kindness in frustrating situations. I also cherished the discussions about factor graphs, probabilistic models, art, and politics.
When both diligence and efficiency were needed, Eike was the person to turn to, and his openness to try out new things in life made me want to try them, too. Also, it was a lot of fun to think with him about the world as a set of processes that can be optimized with best practices.
I already miss working with Felix Vollmer, Georg Männel, Maryam Nezami, and Julia Sauer with whom I shared meals and seminars, talked and laughed.
Apart from the collaborations, I want to acknowledge the privileges that enabled me to start and finish this thesis. I thank my parents for their love and support, my father in particular for kindling my scientific curiosity, and my mother for teaching me the resilience to see things through, especially when difficult.
My dear partner, Mirko Steffen, for his companionship, support, and interest in my work.
What I have learned in the last years has impacted aspects of my life, my personality, and how I view the world. I feel blessed to take with me the fun, satisfaction, and fantastic memories of those years.

Hamburg, May 13th, 2021

Contents

1

Introduction

The expansion of scientific knowledge is often based on observations made with the help of specially developed instruments. It is crucial that such instruments function reliably and without faults, and their maintenance is therefore of high priority.

With time, the measurement apparatuses used for scientific experiments have evolved significantly. In some cases their setups, scopes and sophistication can only be reached by dedicating decades of research towards their development. These machines have become so complex that large amounts of measuring devices are used to monitor and operate them. The use of the thus acquired large amounts of data can benefit from the highly active fields of research, treating the questions of, e.g., data-based classification, system modeling, and prediction. The fruit of this research can be applied to the detection of anomalous or unwanted behavior to ensure the reliable and safe operation of complex machines.

In this thesis first steps towards automatized fault detection and maintenance is addressed for the European X-Ray Free-Electron Laser (EuXFEL).

1.1 CONTEXT & MOTIVATION

The EuXFEL is a recently commissioned large-scale particle accelerator for the exploration of the nano-world [1]. At the time of first commission in September 2017, it broke several records with its ability to generate high-repetition, high-frequency, coherent light flashes.

[1] Altarelli et al., *The European x-ray free-electron laser technical design report*

To reach the peak performance requirements of the EuXFEL, the components, control systems, and sensors need to fulfill multiple high-end specifications, e.g., handling time scales of femtoseconds.

The increase in computing power and data collections, combined with increasing demands on the machine drive the need for automating the operation of individual components as well as the machine as a whole. Methods for fault detection and handling, tuning of parameters, setting up the machine, and monitoring the performance without the need of operator intervention, all contribute to the overall goal of a fully automated particle accelerator.

Fault detection and handling has gained more interest in order to avoid downtime, which is not only expensive but also compromises ongoing experiments performed by the machine users [2].

[2] Venkatasubramanian et al., "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods"

Over the course of this thesis, work on automatic fault detection and handling has notably increased, i.e., several (other) accelerator facilities have started to direct research capacities towards this goal [3, 4].

[3] Edelen et al., "Opportunities in machine learning for particle accelerators"

[4] Tennant et al., "Superconducting radio-frequency cavity fault classification using machine learning at Jefferson Laboratory"

Early detection of component degradation as well as of sudden faults and the identification of root causes can not only help to automate the handling of such events and reduce downtime but can also reveal issues that should be addressed during planned maintenance or the development of new accelerator specifications.

Building a system capable of monitoring faulty behaviors for all components which can fail or degrade, is challenging. Many of the components of the EuXFEL are operated on high performance requirements using closed-loop control. In addition, the components are interconnected through physical interaction, e.g., mechanical and electromagnetic. They are also interconnected through multiple control systems, detection systems, and monitoring devices. Ten thousands of sensor measurements, operation, and control parameters are available, carrying information about the machine's status and the occurrence of faults. The challenge is to extract that information from the available data. Tackling this task requires a deep understanding of the various sub-components and their individual requirements on fault detection. At the same time it is important to consider that these components are interconnected, and unwanted behavior in one part of the machine may be caused by some other component.

For the development of component-specific solutions, methods developed in, e.g., the fault diagnosis, control, and machine learning communities to detect anomalies or handle faults should be considered. Those communities have worked on vast amounts of methods which can be categorized into methods that are model-based or data-driven, deterministic or probabilistic, and many more.

In this thesis an approach is chosen with which the task of fault diagnosis for this complicated system is addressed modularly. Instead of developing one method which can deal with the heterogeneous measurements and subsystem requirements at once, island-like solutions for individual components should be set up. At the same time, the overall fault diagnosis task can benefit from a framework in which these local solutions can be handled and with which their merging can be achieved.

Probabilistic graphical models such as factor graphs, lend themselves well to this concept. The factor graph framework is an abstraction tool to represent dependencies between variables and with which inference based algorithms can be expressed by the passing of messages on the graph.[1] Factor graphs are used to design probabilistic models and inference algorithms in modular building blocks [6].[2] Probabilistic models describe the distribution over variables. With prior assumptions about the distribution of the involved variables, and data carrying information about the system behavior, they are used to handle decision, classification, or estimation tasks using probabilistic inference. Their mix and match property allows adaptive algorithm design, in which parts of the model can be exchanged when, e.g., new insight was gained, the task specifications have changed, or more observable variables should be added.

First attempts of making use of factor graphs for the fault diagnosis

[1] "A graphical model provides a natural and intuitive medium for displaying dependencies that exist between random variables. In particular, the structure of the graphical model clarifies the conditional independencies in the associated probability models, allowing model assessment and revision", M. Jordan [5].

[6] Murphy, *Machine Learning: A Probabilistic Perspective*

[2] "I basically know of two principles for treating complicated systems in simple ways: the first is the principle of modularity and the second is the principle of abstraction.[...]", M. Jordan [5].

community have been published by [7, 8, 9, 10].

In this work they are used to graphically represent models and perform inference using message passing to detect anomalies in one of the EuXFEL components, the superconducting radio frequency (SRF) cavities.

[7] Shen et al., "Spacecraft fault diagnosis based on empirical mode decomposition and directed factor graph"

[8] Gienger and Sawodny, "Data-based Process Monitoring and Iterative Fault Diagnosis using Factor Graphs"

[9] Escobet et al., *Fault diagnosis of dynamic systems*

[10] Li et al., "A robust graph optimization realization of tightly coupled GNSS/INS integrated navigation system for urban vehicles"

1.2 FOCUS & SCOPE

As a first step towards the overall problem of fault detection, one subcomponent is considered in this thesis for which model-based approaches are used to determine anomalous behavior and detect incipient fault symptoms. This work focuses on the pulsed-mode operated superconducting cavities of the EuXFEL. There are 808 SRF cavities installed in the EuXFEL which are used to generate electromagnetic fields with which electron bunches are accelerated. With a repetition rate of 10 Hz, in total, around 700 million RF-pulses are generated each day.

Monitoring of the cavity currently consists of estimating pulse-wise parameters and thus fault related decisions are only taken on the time scale of the pulses. Since the cavity is connected to several other components, the RF-pulse measurements may also carry information about behavioral changes of other subsystems. A particularly severe cavity-related fault, is the breaking down of the superconductivity. These so-called quenches are the only cavity-related faults that are actively detected.

Until now, no methods are available that may detect anomalous RF-pulse behavior in a more general sense or determine the degree of anomaly *within* the pulse. Anomalies in the RF-pulses are caused for a variety of reasons. This thesis develops methods that can detect currently unspecified anomalies, regardless of their level of severity, root cause, or characteristics. To do so, three model-based residuals are applied to the cavity model, which define a deviation between the expected system behavior of a nominal model and the system's behavior according to observations. The severity of the deviations is assessed using a likelihood-ratio based evaluation.

Furthermore the intra-pulse detection of quenches using a Gaussian mixture model (GMM) is assessed. To better understand what kind of anomalies occur in user-run operation RF-pulses of two stations and over several weeks of data is analyzed. With the detection methods developed, characteristics of anomalous and severely anomalous RF-pulse behavior are determined.

All methods which include probabilistic modeling are represented and described in the factor graph framework, thus contributing the first building blocks to which other components may be included.

1.3 CONTRIBUTIONS

The contributions of this thesis may be formulated as follows:

1. Generation of residuals between model and measurements for the SRF cavities
    - unscented Kalman filter residual generation using two nominal cavity models

- parameter-based residual generation based on the currently used cavity parameter computations
- nonlinear parity space residual generation [11]

2. Evaluation of the residuals for anomaly and fault detection
   - likelihood ratio based evaluation for the detection of anomalies [12]
   - GMM model based evaluation applied to the detection of quenches [13]
3. Extensive data analysis
   - implementation of the methods for the analysis of several weeks of RF-pulses [14]
   - evaluation of the detection results, describing the characteristics of the anomalies
   - observation of data which indicate incipient quench symptoms [15]

[11] Nawaz et al., "Anomaly Detection for the European XFEL using a Nonlinear Parity Space Method"

[12] Nawaz et al., "Fault Detection Method for the SRF Cavities of the European XFEL"

[13] Nawaz et al., "Probabilistic model-based fault diagnosis for the cavities of the European XFEL"

[14] Nawaz et al., "Anomaly Detection for Cavity Signals-Results from the European XFEL"

[15] Nawaz et al., "Self-organzied critical control for the european xfel using black box parameter identification for the quench detection system"

1.4 STRUCTURE OF THE THESIS

The thesis is structured in the following way:

- CHAPTER 2 introduces the main principles and components of the EuXFEL focusing on the SRF cavities for which figure of merits, the available RF-pulse measurements, and a nominal model is described. This chapter also introduced the main probabilistic principles used in this thesis, such as the expectation maximization algorithm, the notation of factor graphs and some fundamental terms.

- CHAPTER 3 introduces the data sets used in this thesis, defines a heuristic with which nominal RF-pulses were selected and introduces two alternative cavity models which are used in later chapters to define the nominal behavior of the cavity.

- CHAPTER 4 describes the fault and anomaly detection methods used in this thesis. It describes three model-based residual generation methods as well as two residual evaluation methods to detect anomalies and to detect quenches. The probabilistic models are represented via message passing on factor graphs. Then, a training of the GMM is given based on the expectation maximization algorithm.

- CHAPTER 5 presents the results of applying the detection methods to all collected RF-pulses, determining the amounts of anomalies detected, analyzing some characteristics of the anomalies and categorizing them into severe and non-severe anomalies. Undetected severe anomalies are described and the proposed method to detect cavity faults is evaluated on the example of quenches.

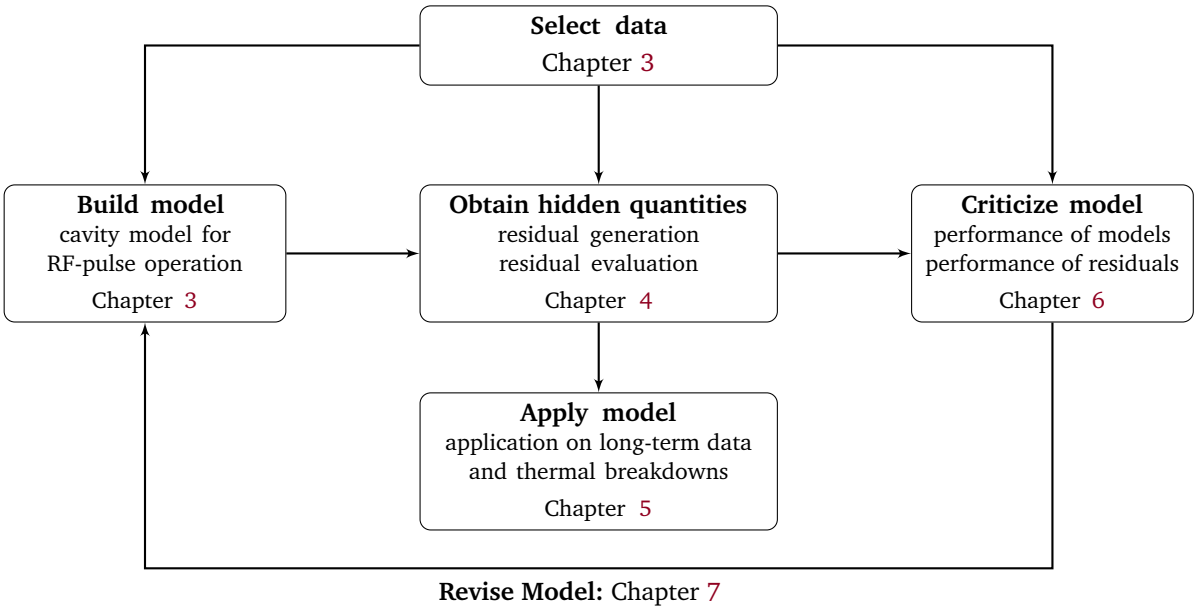


FIGURE 1.1: Structure of the thesis described as one iteration through box's loop adapted from [16], where inference- based algorithm design is described as an iterative process. This thesis thus runs through one iteration this process.

- CHAPTER 6 evaluates the performance of the proposed models and residual generation methods and draws some conclusions about how they can be improved.

- CHAPTER 7 concludes the thesis, points out to some future ideas, and summarizes the main findings of the thesis.

2

Preliminaries

► SYNOPSIS This chapter introduces properties of the X-ray Free-Electron Laser (EuXFEL) facility, as well as the general principles on which the European XFEL is build. It introduces the superconducting cavities as an essential part of the machine and describes their functionality and purpose, as well as the fundamental cavity model, which is used in later chapters a definition of nominal cavity behavior.

In the second part of this chapter, an introduction to probability theory and factor graphs is given. Some key terminology and the used notation are introduced regarding random variables, probability distributions, and factor graphs. Finally, the concept of residual generation and evaluation, which plays an essential role for the remainder of this thesis, is introduced.

2.1 EUROPEAN XFEL

The EuXFEL is a pulsed X-ray laser light source, which generates laser flashes of wavelengths in the X-ray spectrum with unprecedented high brilliance [1]. It is used to study the structure of materials on a nanometer scale and processes with dynamics in the femtosecond range [17]. The EuXFEL is a relatively new facility, first commissioned in 2017. The properties of the light flashes produced are unique worldwide. The exceptional brilliance and the high repetition rate make this light source very attractive for experimental research in biology, chemistry, or material science. Due to the high demand for the facility, the time allotted to each experiment is limited and unpredicted downtime of the machine compromises ongoing experiments. The primary objective of the EuXFEL facility is to reliably and safely produce the required light flashes for the user experiments.

[1] Altarelli et al., *The European x-ray free-electron laser technical design report*

[17] Ackermann et al., "Operation of a free-electron laser from the extreme ultraviolet to the water window"

The following sections introduce the principle behind XFELs, their application as well as some essential technical details.

2.1.1 *Performance and User Experiments*

The EuXFEL is used to study ultra-small objects of a tenth of a nanometer as well as ultrafast processes with dynamics in the femtoseconds. These properties can help to find answers to a broad range of research questions. For example, new insights can be gained from obtaining 3-D images of

FIGURE 2.1: False-colored image of two Klebsiella pneumoniae bacteria (yellow) interacting with a human white blood cell (green). This image was one of the first obtained at the EuXFEL [18].

[19] Tschentscher and Feidenhans'l, "Starting User Operation at the European XFEL"

viruses, biomolecules, or protein crystals. Other experiments aim to study light-matter interaction on a femtosecond time-scale. Figure 2.1 shows the visualization of a bacterial enzyme, which plays a vital role in the resistance towards antibiotics. The data was taken at the EuXFEL.

Typical experimental setups generate diffraction patterns of the object under consideration or use spectroscopy analysis [19]. The properties of the light flashes are key factors when determining which experiments are possible. For example: The lower limit of the object size that can be investigated is determined by the minimal possible wavelength of the light, and the ability to study ultrafast processes is mainly determined by the duration of each single flash. For a higher contrast of the images, the laser flashes' intensity (brilliance) is a key factor.

At the EuXFEL, wavelengths down to 0.05 nm are possible, and the brilliance of the light flashes is considerably higher than other XFEL machines. Another essential property of the EuXFEL facility is the high repetition rate of light flashes per second that can be delivered to the users. This is possible because of the use of *superconducting technology* [20].

[20] Brinkmann et al., "Prospects for CW and LP operation of the European XFEL in hard X-ray regime"

Superconducting radio-frequency technology is referred to as SRF, whereas normal conducting is referred to as NRF. In the next paragraph, a short introduction to the principal setup of the physical process generating the light flashes is given, including details concerning the setup of the EuXFEL.

2.1.2 *General Principles and EuXFEL Setup*

The X-ray flashes are generated using highly energized electron bunches. Each electron bunch results in one X-ray laser flash. The process of generating the X-ray laser flashes can be roughly described in two phases, see Figure 2.2. The first phase consists of generating the electron bunches and accelerating them to their desired maximum energy levels (up to 17.5 GeV at the EuXFEL). The generation and first acceleration of the electron bunches occurs in the *photo-injector* [21]. The electron bunches are extracted from a solid cathode by a laser beam. They are then accelerated to 6 MeV by an interaction with a standing RF electromagnetic wave using an NRF *gun*. A more detailed description of the acceleration process will follow in the next section. After the photo-injector, the bunches are further accelerated in the

[21] Ferrario et al., *Conceptual design of the XFEL photoinjector*

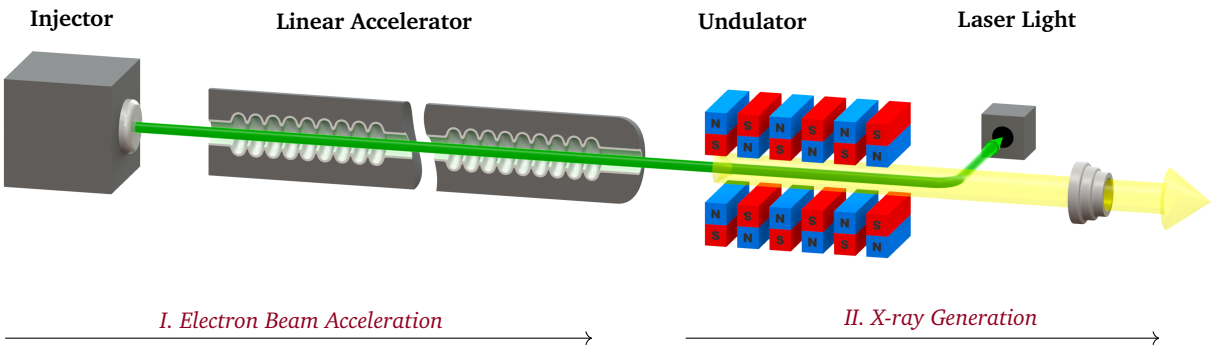


FIGURE 2.2: Main set up of the EuXFEL. First electron bunches generated and accelerated to the desired energy levels. Then these electron bunches are used to generate the desired X-rays.

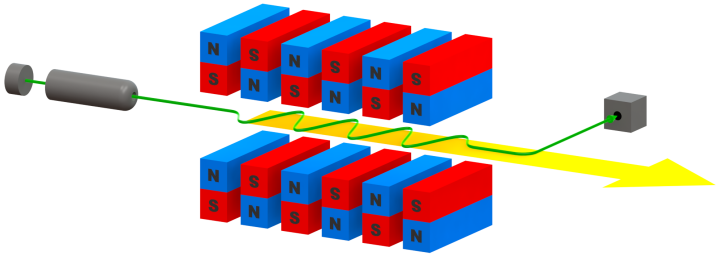


FIGURE 2.3: X-ray laser flashes are generated from electron bunches through SASE. The electrons are forced to follow a sinusoidal path due to the magnetic field polarizations. In following this path, the electron bunch starts emitting light.

first superconducting acceleration module and then enter the main acceleration unit, i.e., the superconducting *linear accelerator* (LINAC) with an energy level of 130 MeV [19]. The superconducting LINAC is divided into three sections that are separated by bunch compressors. The bunch compressors are essential components to shorten the bunch length to about 55 µm [1]. The bunches pass through in total 101 superconducting accelerator modules. Each accelerator module is 12 m long and holds eight superconducting electromagnetic wave resonators, called cavities. At the end of the LINAC, the electron bunches exit with an energy level of up to 17.5 GeV.

[1] Altarelli et al., *The European x-ray free-electron laser technical design report*

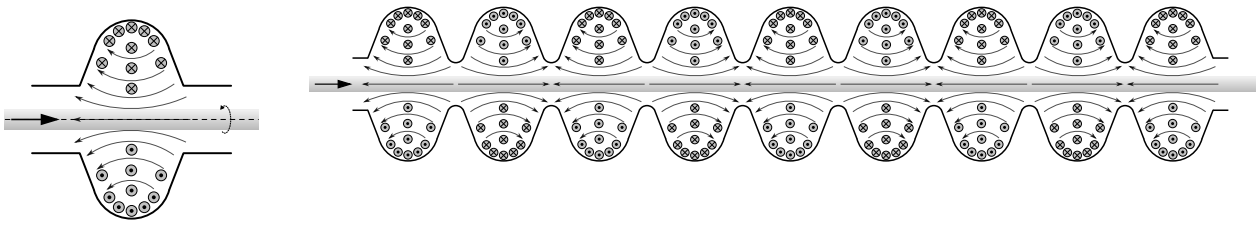
In the second phase, the X-ray laser flashes are generated from the highly energized electron bunches. This is achieved by forcing the electron bunches onto a sinusoidal path using a setup of magnets that produce transverse periodic magnetic fields—the *undulators*. They thereby start emitting photons in a self-amplified spontaneous emission (SASE) process [22]. This process is schematically shown in Figure 2.3. After passing through the undulators, the electron bunches are of no further use and are deflected into the *electron dump*. The X-ray flashes are distributed among five experimental halls in which different research groups can perform their experiments.

[22] Geloni et al., "Coherence properties of the European XFEL"

The EuXFEL facility is mainly built underground with a total length of 3.4 km. The largest part of the facility is the 2.2 km long superconducting LINAC. The focus of this thesis is on on the SRF cavities in the LINAC with which the electrons are accelerated.

2.1.3 *Acceleration with Radio Frequency Resonators*

The electrons are accelerated in SRF cavities by interacting with the electric field gradient of an RF electromagnetic standing wave coupled into the cavity. The electric fields act along the cavity axis. Figure 2.4(a) shows a snapshot of the field distribution in a single-cell cavity. For maximal acceleration, the passing of the bunches through the cavity center has to coincide with the maximum of the RF electric field. This *on-crest* acceleration makes sure that the maximum amount of energy is transferred to the electron bunch, also called the electron beam. To reach energies of 17.5 GeV, the bunches have to pass through such fields several times. It is common to join multiple single-cell cavities together to a multi-cell cavity. In each cell, the bunches are exposed to the same maximal RF-field. The cells are connected via an iris, through which the electron beam can pass. The multi-cell cavities behave like an assembly of weakly coupled single cavities. At the EuXFEL, nine-cell superconducting cavities are used. Nine-cell cavities have nine fundamental



(a) Single cell cavity.

(b) 9-cell cavity. In the π-mode shown here, electric fields in neighboring cells are of opposite direction. [24]

FIGURE 2.4: Electromagnetic field distribution. The electric field is aligned to the beam axis, the magnetic field surrounds it perpendicularly. In on-crest acceleration the electron bunch passes through the center when maximum electric field gradient is reached.

resonance modes. The mode used for the acceleration is the π-mode, for which half an RF wavelength fits into each cell, see Figure 2.4(b). For an on-crest acceleration in each cell, the RF frequency is chosen such, that half of an RF period matches the time it takes for a bunch to pass from one cell to the next. Since the electron bunches behave relativistically, i.e., their velocity is approximately at the speed of light, a fixed frequency for all accelerating cavities can be used [23].

[23] Padamsee et al., *RF superconductivity for accelerators*

The acceleration of the electrons is a high power process, and unnecessary power losses should be avoided. The process of coupling the RF waves into the cavity and transferring their energy to the beam is not lossless and part of the energy is dissipated through the cavity walls in the form of thermal energy. In NRF cavities the heat load is a limiting factor in increasing the RF pulse. Therefore in NRF XFELs a single bunch is accelerated for each RF pulse. Since the SRF cavities' resistance is orders of magnitude lower, the realizable RF pulse length allows to accelerate multiple bunches. Further advantages of using superconducting cavities, as well as the technical challenges in operating these accelerator modules, are given in the next section.

2.1.4 *Superconducting Technology*

Superconductivity is a material property that reduces the ohmic resistance of a material to zero when it is cooled below a critical temperature T_c. Direct currents can flow without losses in superconducting materials. For alternating currents the resistivity is not zero, but far below that of conventional conductors [23]. This effect considerably reduces power loss for the acceleration of particles. Superconducting cavities thus assure a higher conversion efficiency (20%) of the primary energy to the energy of the bunches [25]. It also enables longer bunch trains, i.e., a higher frequency of bunches, at higher electromagnetic field gradients.

All accelerating cavities in the LINAC are superconducting nine-cell TESLA-type cavities, made from solid niobium with a π-mode resonance frequency of f^0 =1.3 GHz. The critical temperature of niobium is $T_c = 9.5$ K, and for their operation, the cavities are bathed in liquid helium, cooling them down to 2 K. The electromagnetic RF fields used for the acceleration in a superconducting cavity induce a flow of alternating currents in a thin surface layer of the cavity walls. The result is power dissipation in the form of heat. This heat load limits the cavity's maximum attainable field

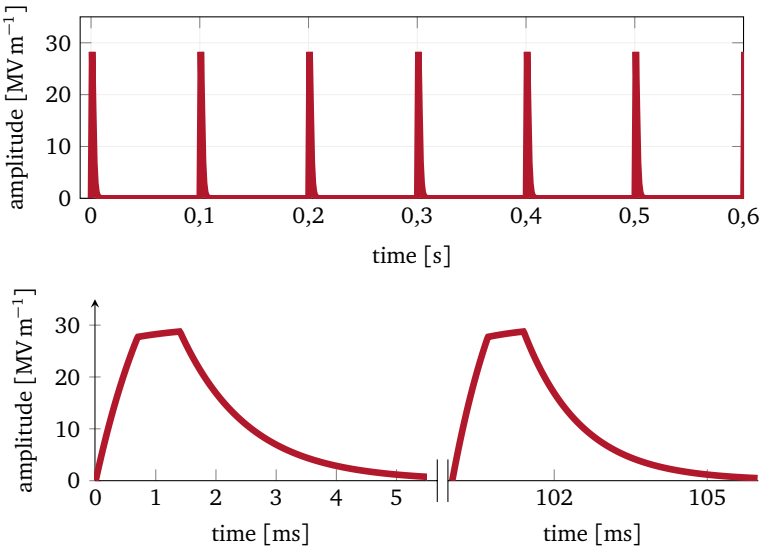


FIGURE 2.5: RF amplitude envelopes of electromagnetic fields in the cavity in pulsed mode operation. The repetition rate of the pulses is 10 Hz. Each pulse duration is 1.3 ms long. Acceleration of the electron bunches is possible during the flattop.

gradient and needs to be considered during the operation. The average design acceleration gradient for each cavity is 23.6 MeV [19].[1]

[19] Tschentscher and Feidenhans'l, "Starting User Operation at the European XFEL"

[1] Tests with TELSA cavities using recent improvements in the production process were able to reach even higher gradients of up to 50 MeV [26].

2.1.5 *Pulsed Mode Operation*

To prevent large heat loads and to increase the maximum achievable electron energy, the cavities are currently operated in *short pulse* (SP) mode. Pulsed mode means, that the RF field is coupled into the cavities in bursts with a constant repetition rate of 10 Hz [20]. Each RF burst is 1.3 ms long. In Figure 2.5 the SP operation mode is depicted by showing the amplitudes of the electromagnetic wave bursts in the cavities. During each RF-pulse, long bunch trains of up to 2700 bunches are accelerated resulting in the repetition rate of 27000 X-ray flashes per second.

[20] Brinkmann et al., "Prospects for CW and LP operation of the European XFEL in hard X-ray regime"

2.1.6 *Subsystems for Cavity Operation*

To meet the conditions for operating SRF cavities, several operations need to be involved, e.g., the cooling of the cavities, the generation of a vacuum inside the cavities as well as the connection to the RF power source.

The set-up for each cavity is organized on three different levels: The single cavity, the cryomodule, and the RF station.

Each single cavity is equipped with a helium tank, a tuning system driven by a stepping motor, a coaxial RF power coupler, a pickup probe (measuring the RF fields), two higher-order mode (HOM) couplers and a piezo-electric tuner. The stepper motor is used to tune the individual cavities for optimal operation [27]. For a more precise tuning of the cavities, feed-forward and iterative learning controlled piezoelectric tuners are installed for each cavity [28, 29]. The higher-order mode couplers are used to assure that the π-mode is the dominant mode in the cavities, and the higher-order modes, generated by the electron beam current, are coupled outside the cavity.

On the second level, a group of eight cavities are cooled down to the necessary temperatures in a cryomodule. In the cryomodule, several pipes

[27] Przygoda et al., "Testing procedures for fast frequency tuners of XFEL cavities"

[28] Sekalsky et al., *Lorentz Force Detuning Compensation System for Accelerating Field Gradients Up To 35 MV/m for Superconducting XFEL and Tesla Nine-Cell Cavities*

[29] Paparella, "A control and systems theory approach to the high gradient cavity detuning compensation"

are used to ensure a constant flow of liquid helium that stabilizes the cavities' temperatures. Thus there is one cryogenic system for the regulation of the temperature of eight SRF cavities.

Finally, four of these cryomodules, i.e., 32 cavities, are powered by one klystron, which is referred to as one RF-station. The klystron amplifies the RF signal of the station, coming from the low-lever RF controller output.

The fields in the cavity are controlled via a feedback controller, as well as learning feed-forward methods implemented on hardware to control the fields online [30, 31]. The controller input is the sum of the electromagnetic field measurements in each cavity and its output is the sum of the non-amplified electromagnetic field inputs to all cavities.

[30] Pfeiffer, "Symmetric grey box identification and distributed beam-based controller design for free-electron lasers"

[31] Schmidt, *RF system modeling and controller design for the European XFEL*

The behavior of the components is dependent on a set of parameters, and they are themselves (open or closed-loop) controlled systems. The behavior of the subcomponents, their parameter values, or operation settings can all have an influence on the behavior of the cavity system.

2.2 SUPERCONDUCTING CAVITIES

Superconducting cavities are electromagnetic field resonators. To effectively obtain a desired field gradient the SRF cavities need to be operated as closely as possible to their resonance frequency. The range of acceptable divergence from the resonance frequency is directly connected to their quality factor. In this chapter, the main figures of merit are introduced, along with the involved signals when operating the SRF cavities. A model describing the input to output signal relationship of an SRF cavity is presented. Finally, sources of disturbances and possible known cavity faults are introduced.

2.2.1 *Figure of Merit*

Some of the parameters introduced in the following sections are assumed to be constant for all considerations in this thesis, whereas others may change over time. Those parameters which can change, e.g., due to a fault, are consistently denoted with superscripts, s.t., the continuous time variable t or discrete time variable k can be used as a subscript. Parameters assumed as static are consistently denoted with subscripts.

A prominent figure of merit for SRF cavities is their *quality factor*. The quality factor relates the amount of dissipated energy in each RF cycle, to the stored energy W inside the cavity, i.e.,

$$Q = 2\pi \frac{f^0 W}{P_{diss}}, \tag{2.1}$$

where P_{diss} is the dissipated power in each RF-cycle, and $f^0 = 1.3\,\text{GHz}$ is the resonance frequency of TESLA SRF cavities used in the EuXFEL [32].

[32] Schilcher, "Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities"

There are two sources of power dissipation that are typically considered, each characterized by its quality factor. The unloaded quality factor Q^0 describes the power dissipated through the walls of the superconducting material. The external quality factor Q_{ext} represents the power dissipation into an external load which occurs during the coupling process of the RF-wave into the cavity. The *loaded* quality factor Q^L combines both dissipation

sources and it is defined by

$$Q^L = \frac{1}{\frac{1}{Q_{ext}} + \frac{1}{Q^0}}, \tag{2.2}$$

The cavities are operated with $Q^0 \gg Q_{ext}$, i.e., power dissipation during the coupling process is significantly higher than through the superconducting walls. The unloaded quality factor lies at $Q^0 \approx 2 \times 10^{10}$, while the external quality factor is in a range of $10^6 < Q_{ext} < 10^7$. Thus a common approximation is that $Q^L \approx Q_{ext}$. The nominal loaded quality factor of the TESLA cavities operated in pulsed-mode is around 4.6×10^6 [1].

Figure 2.6 shows the amplitude and phase dependencies of the RF-waves on the frequency. When the cavity is operated at its resonance frequency f^0, the amplitude of the cavity field peaks, and the phase is zero, which means a maximal efficiency for the acceleration of the electron bunches.

Discrepancies between the driving frequency and the resonance frequency is referred to as the *detuning* of the cavity defined by

$$f^\Delta = f - f^0, \tag{2.3}$$

where f is the constant driving frequency which is provided to all cavities by the master oscillator.

The bandwidth of the cavity determines how sensitive the cavity is towards detuning. The half-bandwidth describes the amount of detuning needed to drop the amplitude by 3 dB. These −3 dB points are marked in Figure 2.6. The bandwidth of the cavity is inversely proportional to the quality factor of the cavity, i.e.,

$$f^\dagger = \frac{f^0}{2Q^L}, \tag{2.4}$$

where $f^\dagger$ is the *half-bandwidth*.

The high quality factor makes the SRF cavity particularly efficient in accelerating the electron bunches, but it makes them also susceptible to detuning. With a half-bandwidth of around 140 Hz, a divergence of 1×10^{-4} % reduces the amplitude by −3 dB.

The following section introduces all relevant RF-signals necessary to represent the input to output cavity behavior which is used extensively in the subsequent chapters.

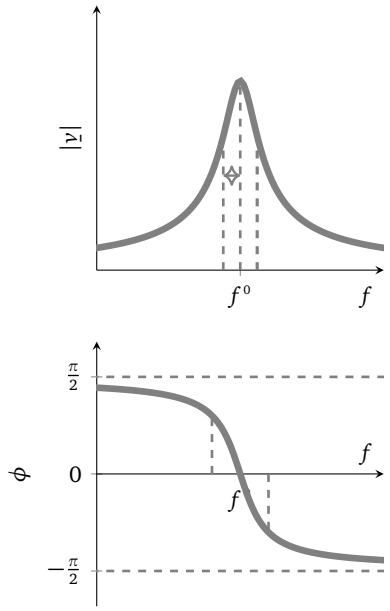


FIGURE 2.6: Frequency dependent amplitude and phase of RF-wave in the cavity. With a bandwidth of 140 Hz the superconducting cavities are very sensitive towards detuning.

[1] Altarelli et al., *The European x-ray free-electron laser technical design report*

Envelope Description

Pick-up antennas and RF waveguide directional couplers at each cavity observe three electromagnetic RF-waves. The observations' sampling frequency is three orders of magnitude smaller than the frequency of a single RF-cycle. The observations thus describe the envelopes of the amplitude and phases of each RF-wave. The driving RF-wave is typically termed the *forward* wave.[2] The standing wave inside the cavity is termed the *probe*. Part of the forward RF-wave is reflected at the coupler, called the *reflected* RF-wave. Each of

[2] It arrives from the master oscillator, is amplified by the klystron, and then distributed through the waveguide system to the individual cavity, into which it is coupled via the input couplers.

these three waves can be represented by a phasor, i.e.,

$$\begin{aligned}\underline{v}_t^F &= |\underline{v}_t^F| \cdot e^{j\phi_t^F} = v_t^{F,I} + jv_t^{F,Q},\\ \underline{v}_t^P &= |\underline{v}_t^P| \cdot e^{j\phi_t^P} = v_t^{P,I} + jv_t^{P,Q},\\ \underline{v}_t^R &= |\underline{v}_t^R| \cdot e^{j\phi_t^R} = v_t^{R,I} + jv_t^{R,Q},\end{aligned} \tag{2.5}$$

where $|\underline{v}_t^F|, |\underline{v}_t^P|, |\underline{v}_t^R| \in \mathbb{C}$ are the amplitudes and $\phi_t^F, \phi_t^P, \phi_t^R \in \mathbb{R}[-180, 180]$ are the phases of the wave and the subscripts F, R, P stand for the forward, reflected and probe signals, respectively. For the nominal cavity model the real *in-phase* (I-) and the imaginary, *quardrature* (Q-) components of the phasor are used. Both amplitude and phase are time-dependent. The amplitude changes throughout a pulse for the forward signals. The phase changes due to the effect of various detuning sources. The relationship between forward and probe signals in the cavity is modeled in the subsequent section with a nonlinear ordinary differential equation.

2.2.2 *Nominal Cavity Model*

Nominal multi-cell SRF cavity models are derived from the assumption that each cell behaves like an LCR circuit, weakly coupled to their neighboring cells.[3] The LCR circuit is driven by an RF current generator, representing the forward RF wave coupled into the cavity. In the following, the total driving voltage $\underline{v}_t^D = \underline{v}_t^F, \in \mathbb{C}$ describes the forward wave to the cavity. From this LCR circuit model, the dynamic behavior of the cavities is derived with the use of Kirchoff's rule, which results in a second-order differential equation for harmonic oscillators, i.e.,

$$\frac{d^2\underline{v}_t^P}{dt^2} + \frac{2\pi f_t^0}{Q^L}\frac{d\underline{v}_t^P}{dt} + (2\pi f_t^0)^2\underline{v}_t^P = \frac{2\pi f_t^0}{Q^L}\frac{d\underline{v}_t^D}{dt}, \tag{2.6}$$

When modeling the RF fields envelopes behavior, the second-order terms can be neglected as they are small compared to the first order terms [32]. By including a linearization around the resonance frequency f_t^0, a first-order cavity model was derived

$$\frac{d\underline{v}_t^P}{dt} + 2\pi\left(f_t^\dagger - jf_t^\Delta\right)\underline{v}_t^P = 2\tilde{f}^\dagger\underline{v}_t^D, \tag{2.7}$$

with the time-varying detuning

$$f_t^\Delta = \left(f - f_t^0\right), \tag{2.8}$$

the constant parameter

$$\tilde{f}^\dagger = \frac{f^0}{2Q_{ext}}, \tag{2.9}$$

and the half-bandwidth $f^\dagger$ given in (2.4), also constant for nominal cavity behavior. Since $Q_{ext} \approx Q^L$ (as described in the previous section), the nominal assumption is that $\tilde{f}^\dagger \approx f^\dagger$. Using the I and Q representations to formulate two coupled first order equations, the nominal cavity behavior may be represented by

$$\begin{aligned}\frac{dv_t^{P,I}}{dt} &= -2\pi f^\dagger v_t^{P,I} - 2\pi f_t^\Delta v_t^{P,Q} + 2f^\dagger v_t^{F,I},\\ \frac{dv_t^{P,Q}}{dt} &= 2\pi f_t^\Delta v_t^{P,I} - 2\pi f^\dagger v_t^{P,Q} + 2f^\dagger v_t^{F,Q}.\end{aligned} \tag{2.10}$$

[3] Details of the full LCR circuit model are omitted here, but can be found in several publications [33, 32, 34, 35].

[32] Schilcher, "Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities"

This model is used later to determine changes in the cavity behavior, reflected in the RF measurements.

2.2.3 *Sources of Detuning*

With the high quality factor of superconducting cavities, detuning poses a considerable challenge for their operation. Small changes in detuning will have a great effect on the fields in the cavities. In the following two sources of detuning are described, Lorentz force detuning and microphonics. Lorentz force detuning is caused by the fields in the cavity. Microphonics are generally caused by external disturbances, e.g., vibrations. Since the longitudinal tuning sensitivity is on the order of hundreds of 100 kHz mm^{-1}, cavity deformations of few hundreds nm results in a detuning variation comparable to the cavity half-bandwidth [36].

[36] Pagani et al., *The fast piezo-blade tuner for SCRF resonators*

- LORENTZ FORCES are the primary cause of detuning in pulsed mode operation. They occur due to the interaction between the magnetic fields inside the cavity and RF currents in the cavity wall [23]. The effect of Lorentz forces on TESLA cavities is high, as the walls of the cavities are thin (2.8 mm). When Lorentz forces act upon the cavity walls, they cause a deformation, which in turn results in a shift of the resonance frequency and — depending on the field amplitudes — can cause resonance frequency shifts far beyond the bandwidth. While Lorentz forces are more prominent in magnitude in pulsed operation, it also plays an important role for the continuous mode of operation and both effects have been modeled. In steady state operation (or continuous mode) with on-crest acceleration, Lorentz forces are proportional to the square of the accelerating fields amplitude, i.e.,

[23] Padamsee et al., *RF superconductivity for accelerators*

$$f_t^{\Delta} = -\beta |\underline{v}_t^{p}|^2, \tag{2.11}$$

with

$$|\underline{v}_t^{p}|^2 = (v_t^{pI})^2 + (v_t^{pQ})^2 \tag{2.12}$$

where β is the *Lorentz force detuning* constant [23]. According to (2.11) the detuning is dependent on the squared accelerating fields which is the reason why Lorentz force detuning is particularly dominant in large accelerating fields. For pulsed mode operation, dynamic Lorentz force detuning is described by a second order differential equation which assumes, that the dynamic behavior can be modeled by superimposing the effect of a set of mechanical resonance frequencies, i.e.,

$$\frac{d^2 x_t^m}{dt^2} = -{\alpha_m}^2 x_t^m - \frac{1}{\gamma_m}\frac{dx_t^m}{dt} - 2\pi\beta_m {\alpha_m}^2 |\underline{v}_t^{p}|^2, \; m = 1, .., n_m \tag{2.13}$$

where x_t^m is the detuning caused by the mechanical resonance mode m, γ_m is the respective time-constant and β_m the Lorentz force detuning constant associated to the mode. The detuning over the pulse is then defined as

$$f_t^{\Delta} = \sum_{m=1}^{n_m} x_t^m, \tag{2.14}$$

where n_m is the number of considered modes [37, 34]. The driving sig-

[37] Liepe et al., "Dynamic Lorentz force compensation with a fast piezoelectric tuner"

[34] Neumann, "Compensating microphonics in SRF Cavities to ensure beam stability for future Free-Electron-Lasers"

nal is the square of the accelerating field. To compensate for the detuning during operation, piezo-electric tuners are used in feed-forward compensation, which relies on the assumption, that the Lorentz force detuning characteristics for each cavity are highly repetitive from pulse to pulse [37].

[37] Liepe et al., "Dynamic Lorentz force compensation with a fast piezoelectric tuner"

- MICROPHONICS are mechanical vibrations which affect the cavity. A deformation of the cavity is then not due to the interaction with the RF-field, but due to disturbances acting from outside the cavity. The sources of vibrations, and therefore the frequency spectrum is diverse. Lower frequency sources include heavy machinery, mechanical vacuum pumps, compressors and pumps from the cryogenic system. These kinds of vibrations can be transmitted through the beamline, the ground, the supports and the cryostats [23]. Although microphonics are often unavoidable and are not necessarily due to the malfunction of a component, they do represent unwanted behavior. They have thus been listed as a fault category for the with continuous wave operated superconducting cavities in [4].

[23] Padamsee et al., *RF superconductivity for accelerators*

[4] Tennant et al., "Superconducting radio-frequency cavity fault classification using machine learning at Jefferson Laboratory"

2.2.4 *Determining Detuning and Bandwidth*

The following detuning and half-bandwidth calculations are used to control the cavities' detuning and monitoring.

First, the cavities half-bandwidth of the *pulse* is calculated from the RF-pulse probe data, which can then be used to determine the detuning and half-bandwidth for each *sample* during a pulse. Pulse-wise half-bandwidths of the cavity can be estimated using the decay phase, i.e., the cavity systems' free-response where the forward signal is shut off. The probe amplitude decays according to

$$|\underline{v}_t^P| = e^{\frac{1}{\tau}t}, \tag{2.15}$$

where τ is the systems' time constant which is related to the half-bandwidth by

$$f^\dagger = \frac{f^0}{2Q^L} = \frac{1}{2\pi\tau}. \tag{2.16}$$

The time constant is determined by fitting an exponential function to the probe amplitude samples of the decay. For intra-pulse half-bandwidth and detuning calculations (2.7) is used. Solving the equation for the detuning results into

$$\begin{aligned} f_t^\Delta = &\frac{v_t^{P,I}\left(\frac{dv_t^{P,Q}}{dt} - 4\pi\tilde{f}^\dagger\left(v_t^{F,Q} + v_t^{B,Q}\right)\right)}{2\pi\left(\left(v_t^{P,I}\right)^2 + \left(v_t^{P,Q}\right)^2\right)} \\ &+ \frac{v_t^{P,Q}\left(4\pi\tilde{f}^\dagger\left(v_t^{F,I} + v_t^{B,I}\right) - \frac{dv_t^{P,I}}{dt}\right)}{2\pi\left(\left(v_t^{P,I}\right)^2 + \left(v_t^{P,Q}\right)^2\right)}, \end{aligned} \tag{2.17}$$

whereas the half-bandwidth is given by

$$f_t^{\dagger} = \frac{v_t^{P,I}\left(4\pi\tilde{f}^{\dagger}\left(v_t^{F,Q}+v_t^{B,Q}\right)-\frac{dv_t^{P,I}}{dt}\right)}{2\pi\left(\left(v_t^{P,I}\right)^2+\left(v_t^{P,Q}\right)^2\right)} + \frac{v_t^{P,Q}\left(4\pi\tilde{f}^{\dagger}\left(v_t^{F,I}+v_t^{B,I}\right)-\frac{dv_t^{P,Q}}{dt}\right)}{2\pi\left(\left(v_t^{P,I}\right)^2+\left(v_t^{P,Q}\right)^2\right)}. \tag{2.18}$$

The detuning and half-bandwidth calculations can be carried out by assuming that the half-bandwidth $\tilde{f}^{\dagger}$ can be well approximated with the half-bandwidth $f^{\dagger}$ determined once with (2.16). When calculating the derivative of the discrete probe measurements noise is introduced into the detuning and half-bandwidth values. Filtering with a simple low-pass finite impulse response filter (FIR) decreases the noise. An FPGA implementation of these calculations can be found in [38].

[38] Rybaniec et al., "Real-time estimation of superconducting cavities parameters"

2.2.5 *Typical Cavity Faults*

Typical cavity faults include the thermal breakdown of superconductivity, called *quench*, multipacting and field emissions. Due to improvements in the production process of the cavities, the occurrence of these faults has been considerably lowered in recent years. In this thesis quenches play important role for the results in Chapter 5, which is why this fault is introduced here. Quenches causes the SRF cavity to partly loose their superconductivity. Ideally, the conductivity throughout the cavity material is homogenous. However, regions in the order of mm can show considerable discrepancies of conductivity compared to the rest of the material, [39]. These spots are called defects. Alternating currents' reactive parts flow right through these effects, causing a considerable heat load. If the temperature at the edges of these defects rises above the critical temperature of the superconducting material T_c, the surrounding material becomes normal conducting, which causes a thermal breakdown of the cavity, i.e. a quench.

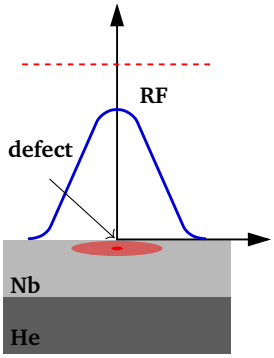
Figure 2.7 depicts this effect schematically. Defects limit the performance of a cavity as they lower the maximal possible peak values of the magnetic fields and thus the maximally safely attainable RF-field gradient.

Cavities can recover from quenches when the RF power is lowered or shut down, thus making it possible for the cavities to cool back down below the critical temperature [15].

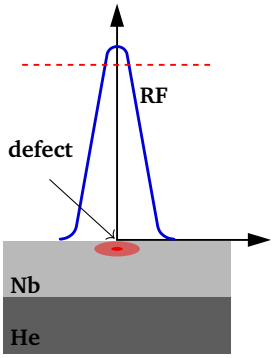
The current method of quench detection sets a fixed threshold on the moving averaged values of the loaded quality factor Q^L, [40]. The quench detection decision rule can best be described in the following way:

$$\mathrm{d}^q(Q^L_{i_p}) = \begin{cases} 1 \text{ if } & Q^L_{i_p} - \frac{1}{K}\sum_{j=i_p-K}^{i_p} Q^L_j > 5\times 10^5, \text{ quench,} \\ 0 \text{ if } & \text{not a quench,} \end{cases} \tag{2.19}$$

where i_p is the index representing the pulse under consideration, and the decision depends on the divergence between the current loaded quality factor values from its moving average over the last $K = 100$ pulses. A detection threshold of 5×10^5 is currently used in operation.



(a) Heating spot caused by defect, temperature below critical limit.



(b) Temperature rises above critical limit, causing a breakdown of superconductivity.

FIGURE 2.7: Local rise of temperature in the superconducting cavity walls due to a defect.

[39] Champion et al., "Quench-limited SRF cavities: failure at the heat-affected zone"

[15] Nawaz et al., "Self-organzied critical control for the european xfel using black box parameter identification for the quench detection system"

[40] Ayvazyan et al., "Superconducting Cavity Quench Detection and Prevention for the European XFEL"

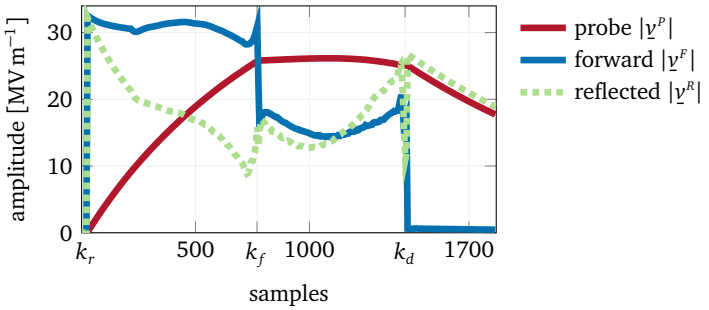


FIGURE 2.8: Uncalibrated RF-pulse measurements of the field amplitudes for the driving wave (forward), the transmitted wave (probe), and the reflected wave.

2.3 ACQUISITION AND CALIBRATION OF DATA

A considerable amount of data was stored and analyzed for the purpose of this thesis. In the following section, the data acquisition system (DAQ) used to access and store the data and the calibration of the data is presented.

2.3.1 *Data Acquisition System*

The DAQ system is designed to collect all front-end data via multicast-enabled push protocol [41]. The DAQ collects the data at its pulse repetition rate, i.e., all samples associated to a pulse are stored after the end of an RF-pulse. Each pulse is stamped with a unique pulse identification number (PID) as well as a time-stamp, facilitating synchronization of different data streams according to the PID.

[41] Wilksen et al., "A Bunch-Synchronized Data Acquisition System for the European XFEL Accelerator"

The RF-data includes measurements of the forward, reflected, and probe envelopes in their amplitude and phase for each RF-pulse. The data was stored with a sampling rate of $f_s = 1\,\text{MHz}$. Instead of storing the entire pulse traces, the RF-measurements are cut off after 1.82 ms, when the pulse has not yet faded away completely, see Figure 2.8.
The discrete RF data of the amplitude and phase is given by $|\underline{v}^P|(kT_s), \phi^P(kT_s)$ for the probe, $|\underline{v}^F|(kT_s), \phi^F(kT_s)$ for the forward and $|\underline{v}^R|(kT_s), \phi^R(kT_s)$ for the reflected, where $k \in \mathbb{Z}^{\geq 0}$ and $T_s = \frac{1}{f_s} = 1\,\mu\text{s}$. The sampling time T_s is omitted from now on. The complex representations for these measurements will be denoted by

$$\begin{aligned}\underline{v}_k^P &= v_k^{P,I} + jv_k^{P,Q},\\ \underline{v}_k^F &= v_k^{F,I} + jv_k^{F,Q},\\ \underline{v}_k^R &= v_k^{R,I} + jv_k^{R,Q},\end{aligned}$$

with $\underline{v}_k^P, \underline{v}_k^F, \underline{v}_k^R \in \mathbb{C}$. The RF-data is stored in units of MV m^{-1} and deg, for the amplitudes and phase traces, respectively.

In normal operation, all pulses have the same general structure, i.e., they consist of three discernible parts. In the first phase, the *filling* phase starting at $k = k_r$, the RF field is ramped up, then the forward power is lowered at the onset of the *flattop* phase at $k = k_f$. In the final phase, the *decay* phase, starting when the RF-pulse is switched off at $k = k_d$, the cavity system is in free-response.

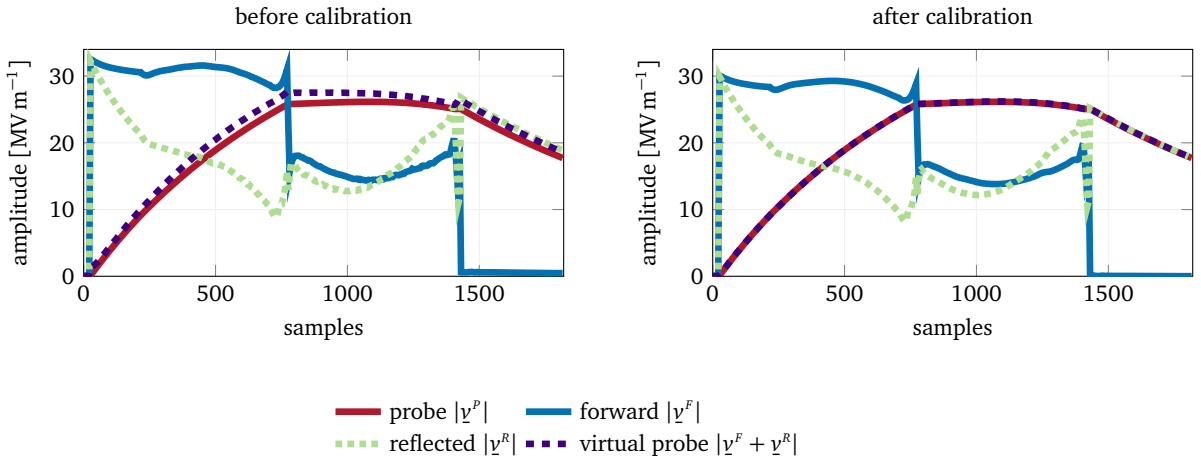


FIGURE 2.9: RF field measurements of the amplitude before and after calibration. A calibration of the signals is a prerequisite for the fault detection methods. When not calibrated, the signals do not represent physical behavior and therefore the cavity model cannot be used.

Pulse Shape Properties

At the beginning of the filling phase, the forward wave is entirely reflected at the input coupler. The acceleration of the electrons only takes place during the flattop, and thus the amplitude and phase-stability requirements are defined for the flattop phase. During the decay, i.e., $\underline{v}_k^F = 0$, the reflected signal is equal to the probe signal inside the cavity. This is because both pick-up antennas detect the same fields, i.e., the probe field fading away. While one pick-up interprets the field as the reflected, the other interprets it as the probe. In principle, the length of the filling, flattop, and decay times can be altered. The RF data considered in this thesis have fixed filling, flattop, and decay times, with the filling starting at sample $k_r = 20$, the flattop at $k_f = 770$ and the decay at sample $k_d = 1420$. The total number of samples in each RF-pulse is $n_s = 1820$. The measurements need to calibrated. The calibration procedure is described in the subsequent section.

2.3.2 *Calibration*

The RF-measurements are calibrated according to [42]. While forward and reflected signals are measured by two independent but not perfect pick-up antennas and a coupling between the two signals is often unavoidable, the probe signal has its own pick-up antenna. The probe signal is thus used as a reference for the calibration. When the signals are well-calibrated, the sum of the complex forward and reflected signals must equal the probe signal, i.e.,

$$\underline{v}_k^P = \underline{v}_k^{FC} + \underline{v}_k^{RC}, \quad k = 1, ..., n_s, \tag{2.20}$$

where $\underline{v}_k^{FC}, \underline{v}_k^{RC}$ are the calibrated forward and reflected signals respectively and

$$\underline{v}_k^{PV} = \underline{v}_k^F + \underline{v}_k^R, \quad k = 1, ..., n_s, \tag{2.21}$$

where $\underline{v}_k^{PV}$ is the 'virtual probe', which is the sum of the uncalibrated forward and probe signals. In case the measurements are not calibrated, the forward signals are not zero during the decay, and the reflected fields are not equal to the probe see Figure 2.8. After calibration, the virtual probe should be approximately equal to the measured probe, see Figure 2.9.

[42] Pfeiffer et al., "Virtual cavity probe generation using calibrated forward and reflected signals"

The calibration procedure is explained in more detail in appendix 7.2.

2.4 PROBABILITY THEORY AND FACTOR GRAPHS

"Probability theory is nothing but common sense reduced to calculation"
—Pierre Laplace, 1812

For notational reasons the following gives a very brief introduction into the well known probability theory relevant for this thesis and introduces factor graphs. Factor graphs are used as a unifying, graphical framework for probabilistic models, and probabilistic inference can be described by the passing of messages on the graph. In this thesis factor graphs are used for the fault and anomaly detection of the pulsed mode operated SRF cavities. The structure of a factor graph is defined by the joint probability distribution which it represents. Nodes and edges are used to represent random variables and probability distributions, respectively. Different notations exist. The rest of this work follows the notation used, e.g., by [43, 44].

[43] Loeliger et al., "The Factor Graph Approach to Model-Based Signal Processing"

[44] Loeliger, "An Introduction to Factor Graphs"

Probability Distributions and Random Variables

A *random variable* is a function that maps the outcomes of a random event to a value. Random variables can either be continuous or discrete. The probability distribution of a discrete random variable (*probability mass function*) is often represented by a table and describes discrete random experiments such as the rolling of a die. In this thesis, a discrete variable is used to represent fault-class categories of data. All other random variables that occur are continuous whose distribution is described by the *probability density function*. Continuous random variables describe states or measurements which have continuous values such as in dynamical systems. To describe the distributions over continuous random variables, the standard notation $p(x)$ is adopted, where x is a variable denoting any of the possible values of X. It is common to use roman capital letters to represent random variables and lower-case letters for their values.

2.4.1 *Joint, Marginal and Conditional Probability*

The *joint* probability is the probability distribution over multiple random variables $X_1, ..., X_n$, i.e., $p(x_1, ..., x_n)$. Often, the distribution over a subset of these variables or even just one is desired, which is called the *marginal* probability distribution. In that case this marginal distribution needs to be inferred from the joint distribution.

The *conditional* probability is the probability, that an event $X = x$ occurs given that another event $Y = y$ has occurred, i.e., $p(x|y)$.

2.4.2 *Independence and Conditional Independence*

Random variables may be *mutually (unconditionally) independent* or *conditionally independent*. Mutual independence is given, when the joint probability of the variables can be expressed as the product of their marginals. For example X and Y are mutually independent if

$$p(x, y) = p(x)p(y). \tag{2.22}$$

Unconditional independence therefore only exists if the two variables do not in any way influence each other. For models of the physical world, this

is often not true. It is more common, that the variables are *conditionally independent*, which means that their influence is not direct, but mediated over some other random variable. For example, Z may be the mediator random variable between X, Y, then

$$p(x, y|z) = p(x|z)p(y|z). \tag{2.23}$$

Conditional independence is an important property for making Bayesian inference more tractable and it is just these independency structures between random variables, which are represented by probabilistic graphical models.

Theorem of Bayes

Many probabilistic algorithms rely on the use of the Bayes Theorem. It is a consequence of the chain rule of probabilities, and describes the conditional probability distribution

$$\underbrace{p(\theta|x)}_{posterior} = \frac{\overbrace{p(x|\theta)}^{likelihood}\overbrace{p(\theta)}^{prior}}{\underbrace{p(x)}_{marginal}}. \tag{2.24}$$

Often, the prior $p(\theta)$ needs to be chosen. It represents the assumptions made concerning the distribution of θ before taking evidence (observations) into account. The likelihood is the distribution over x for a given parameter θ and it describes the measurement model.

2.4.3 *Factor Graphs and Gaussian Message-Passing*

Factor graphs are *bipartite graphs*, that represent existing independency structures of the joint distribution.[4] The graph structure displays the conditional dependencies between the random variables. The remainder of the document adopts the Forney-style factor graph (FFG) notation, in which nodes represent factors whereas *edges or half-edges* represent either *discrete or continuous random variables*, [47, 43].[5] The following constraints characterize FFGs:

1. There is a unique node for every factor.
2. There is a unique edge/half-edge for every variable.
3. The node for factor f connects to the edge for variable $x \iff$ f is a function of x.

Random variables are denoted by roman capital letters whereas observed variables, i.e., instances of the random variable, are denoted by small letters [43].

Example 1. Consider the probability distribution over the random variables X_1, X_2, X_3 and X_4 defined by

$$p(x_1, x_2, x_3, x_4, x_5) = \mathrm{f}_a(x_1, x_2)\mathrm{f}_b(x_2, x_3, x_4)\mathrm{f}_c(x_4, x_5). \tag{2.25}$$

and the factors $\mathrm{f}_a, \mathrm{f}_b, \mathrm{f}_c$ signify the conditional independence of the variables. The factor graph of (2.25) is depicted in Figure 2.10.

[4] Factor graphs are one approach of graphically representing probabilistic models. A concise introduction into probabilistic graphical models can be found in [45, 46].

[47] Forney, "Codes on graphs: Normal realizations"

[5] The notation was first introduced by Forney in [47].

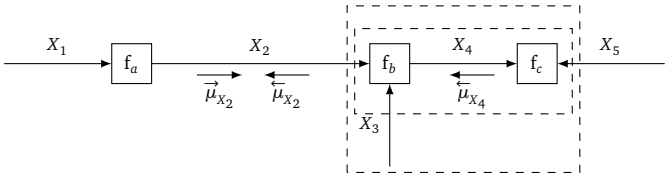


FIGURE 2.10: A Forney-style factor graph representing (2.25). Arrow directions of the edges are added to clarify the direction of the messages. Messages along the arrow direction are referred to as "forward messages", messages against the arrow direction are referred to as "backward messages".

The factor graph structure depends solely on how the joint probability distribution is factorized. *Messages* can then be propagated from one node to the next to determine, e.g., the marginal distribution over one of the variables. The following introduces these messages and their notation in this thesis.

Example 2 (continued)**.** Using the factorization defined in (2.25) the marginal distribution over X_2 is calculated by integrating the joint probability function over all other variables, i.e.,

$$\begin{aligned} p(x_2) &= \int_{x_1,x_3,x_4,x_5} p(x_1,x_2,x_3,x_4,x_5)\,\mathrm{d}x_1\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}x_5 \\ &= \underbrace{\int_{x_1} \mathrm{f}_a(x_1)\mathrm{d}x_1}_{\vec{\mu}_{X_2}} \underbrace{\int_{x_3,x_4} \mathrm{f}_b(x_2,x_3,x_4) \underbrace{\left(\int_{x_5} \mathrm{f}_c(x_4,x_5)\mathrm{d}x_5\right)}_{\overleftarrow{\mu}_{X_4}} \mathrm{d}x_4\mathrm{d}x_3}_{\overleftarrow{\mu}_{X_2}}. \end{aligned} \tag{2.26}$$

where by $\vec{\mu}_{X_2}, \overleftarrow{\mu}_{X_2}$ and $\overleftarrow{\mu}_{X_4}$ are the messages propagated between nodes $\mathrm{f}_a, \mathrm{f}_b$ and f_c as depicted in the factor graph representation in Figure 2.10.

Due to the factorization given in (2.25), solving the integral can be described by a sequential calculation of multiple integrals as exemplified in (2.26). In the factor graph framework, the product-wise computations of (2.26) are interpreted as messages, denoted by $\vec{\mu}_{X_2}, \overleftarrow{\mu}_{X_2}$ and $\overleftarrow{\mu}_{X_4}$ that pass between the individual factor nodes. Thus, message passing captures this step-wise concept of marginalization graphically, and makes use of the computational efficiency gained by taking these factorizations into account. This concept is also known as the sum-product algorithm [48].

[48] Kschischang et al., "Factor Graphs and the Sum-Product Algorithm"

A message may be passed from node f_a to node f_b over the connecting edge X_2 or from node f_b to node f_a. The message is dependent on the node from which it is sent and is a function of the edges's variable. Thus, the direction of message flow needs to be considered in the notation. To do so, arrow directions are added to the edges. Messages propagated with the direction of the arrow are denoted by $\vec{\mu}_X$, and referred to as "forward messages", messages propagated against the direction of the arrow are denoted by $\overleftarrow{\mu}_X$ and referred to as "backward messages".[6]

[6] To sum up: Factor graphs are undirected graphs, but the arrows help to notationally clarify the message passing operations needed for inference. Although the arrows are added to the *edges*, they signify the direction of the *messages* and are more a property of the node than of the edge.

The computation of the marginal for any variable in the factor graph, e.g., $\bar{\mu}_{X_2}$, is then expressed by the multiplication of the forward and backward messages on the edge, i.e., $\bar{\mu}_{X_2} = \vec{\mu}_{X_2}\overleftarrow{\mu}_{X_2}$, where $\vec{\mu}_{X_2}$ and $\overleftarrow{\mu}_{X_2}$ are defined in (2.26). For linear factors and cylce-free graphs exact solutions are obtained. Other inference algorithms like variational Bayes [49], expectation propagation [50], and particle filtering [51] have also been formulated using

[49] Winn and Bishop, "Variational message passing"

[50] Minka, "Expectation propagation for approximate Bayesian inference"

[51] Dauwels et al., "Particle methods as message passing"

message passing.

2.4.4 *Prior Distributions and Observations*

The computation of messages become especially simple when only linear factor nodes and Gaussian priors are considered. All continuous priors considered in this thesis are indeed assumed to be Gaussian. The multivariate, n-dimensional Gaussian distribution is defined by

$$\begin{aligned} p(x) &= \mathcal{N}\left(x; \mathrm{m}_X, \mathrm{V}_X\right) \\ &= \frac{1}{\sqrt{(2\pi)^n |\mathrm{V}_X|}} \exp\left(\frac{1}{2}(x-\mathrm{m}_X)\mathrm{V}_X^{-1}(x-\mathrm{m}_X)^T\right), \end{aligned} \tag{2.27}$$

with $x \in \mathbb{R}^n$, the mean $\mathrm{m}_X \in \mathbb{R}^n$ and the covariance matrix $\mathrm{V}_X \in \mathbb{R}^{n\times n}$. Alternatively the Gaussian distribution can be parameterized in its weighted mean $\xi_X = \mathrm{V}_X^{-1}\mathrm{m}_X$ and information matrix $\mathrm{W}_X = \mathrm{V}_X^{-1}$, which will also be used in this thesis.

Factors connecting only to one edge are prior distributions over that edges' variable, and are typeset slightly smaller compared to other factor nodes. When the variable is observed, its prior is defined as a degenerate Gaussian with zero variance, i.e., $\mathrm{V}_X = 0$ and a mean taking up the value of the observation, e.g., $\mathrm{m}_X = \hat{x}$. This is represented in the factor graph by a small solid black node. When no information about the prior distribution is available, an *uninformative prior* can be used, defined as a degenerate Gaussian with $\mathrm{W}_X = 0$.

Markov Model in Factor Graphs

Markov Models are used to model dynamic systems. The representation of Markov models, e.g., state-space models in factor graphs has extensively been covered, e.g., in [43]. Consider the following linear state-space model,

[43] Loeliger et al., "The Factor Graph Approach to Model-Based Signal Processing"

$$\begin{aligned} X_{k+1} &= \mathbf{A}X_k + \mathbf{B}U_k + I_k, \\ Z_k &= \mathbf{C}X_k + D_k, \end{aligned} \tag{2.28}$$

for $k = 0, \ldots, K$, where the random variables $U_k \in \mathbb{R}^l$, $Z_k \in \mathbb{R}^m$ and $X_k \in \mathbb{R}^n$ are the real-valued input, output and states respectively. The matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are matrices of suitable dimensions. The system is subject to Gaussian process noise I_k and measurement noise D_k. The joint probability distributions up to some instant T of this system can be formulated by

$$p(x_{0:T}) = p(x_0)\prod_{k=1}^{T} p(x_k|x_{k-1}) \quad T \in \{0, \ldots, K\} \tag{2.29}$$

denoting the state transitions where the notation $x_{0:T} = x_0, \ldots, x_T$ is used, and

$$p(z_{0:T}|x_{0:T}) = \prod_{k=0}^{T} p(z_k|x_k), \tag{2.30}$$

denoting the output transitions conditioned on the states.

In factor graphs, these models are represented in slices, where each slice represents the joint probability distribution at time k. The joint probability

function of one slice includes the input, output, measurement and process noise as well as several representations of the state variables.

Example 3. Consider the joint probability distribution of a single slice at instant k:

$$p(x_k, x_k''', x_k'', x_k', x_k, x_k'''', x_{k-1}, d_k, i_k, u_k, z_k), \tag{2.31}$$

and its factorization

$$\begin{aligned} &p(x_k|x_k'''', x_k')p(x_k'|u_k, x_k'')p(x''|i_k, x_k''') \\ &\cdot p(x_k'''|x_{k-1})p(z_k|x_k'''')p(z_k)p(u_k)p(d_k)p(i_k). \end{aligned} \tag{2.32}$$

This factorization is described by the factor graph depicted in Figure 2.11.

The input values U_k and output values Z_k are defined by degenerate Gaussian priors and thus denoted by a solid black node. The priors of the state variable X_k, process noise I_k, and measurement noise D_k are Gaussian distributed and represented by slightly smaller factor nodes in the graph. The state transitions are described by predefined factor node functions. The multiplication with system matrices and the addition of two random variables is achieved with the predefined "multiplication node" f_A and "addition node" f_+ respectively. Branching of a variable is achieved with the "equality node" $\mathrm{f}_=$ definition. For example

$$p(x_k|x_k'''', x_k') = \delta(x_k - x_k'''')\delta(x_k - x_k'), \tag{2.33}$$

$$p(x_k'''|x_{k-1}) = \delta(x_k''' - \mathbf{A}x_{k-1}), \text{ and} \tag{2.34}$$

$$p(x_k|x_k'''', x_k') = \delta(x_k - x_k'''')\delta(x_k - x_k') \tag{2.35}$$

for the addition, multiplication and equality node respectively. These nodes are represented in the graph with symbols such as '=' and '+', which make the factor graph easier to read. For Gaussian messages, these nodes have tabulated updates, which describe the changes of the Gaussian parameters after transitioning though the node. In Table 2.1, their factor function definitions, as well as the tabulated update rules for Gaussian messages, are given.

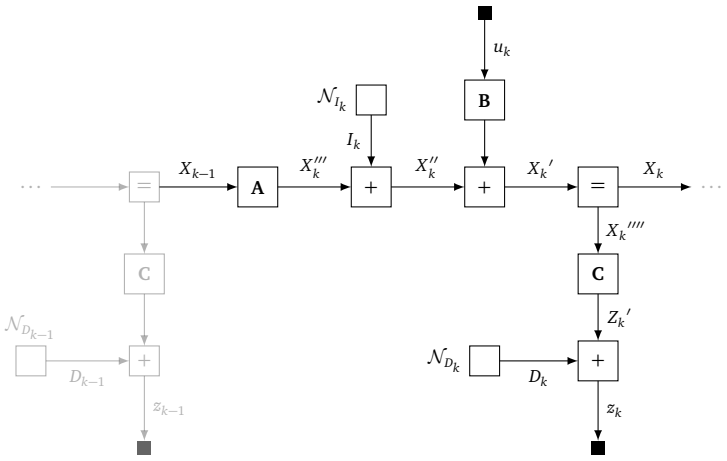


FIGURE 2.11: Representation of the state space model by a chain of identical factor graph slices for each time instant k.

2.4.5 *Gaussian Mixture Models & Expectation-Maximization*

This section briefly recounts the *expectation-maximization* (EM) algorithm which is used in this thesis to train parameters of Gaussian Mixture Models (GMM). A short introduction to the expecation-maximization algorithm, where the training of GMMs is derived can be found, e.g., in [52].

[52] Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models"

A GMM or Gaussian mixture density is the weighted sum over a set of Gaussian distributions [53]. GMMs are able to represent a large class of *sample distributions* as they can approximate arbitrarily shaped densities by a smooth probability density function [52]. The probability density function of a GMM is

[53] Reynolds, "Gaussian Mixture Models."

$$p(x|\theta) = \sum_{i=1}^{n_c} \mathrm{w}_i \; \mathcal{N}\big(x; \mathrm{m}_i, \mathrm{V}_i\big), \quad \sum_{i=1}^{n_c} \mathrm{w}_i = 1, \tag{2.36}$$

where the model parameters $\theta = \big\{\mathrm{m}_i, \mathrm{V}_i, \mathrm{w}_i\big\}_{i=1}^{n_c}$ denote the mean, the covariance matrix and the weights of each of the n_c classes. The classes are represented by a discrete random variable C and the weights w_i of the GMM denote the probability of each class to occur, i.e.,

$$P(C = c_i) = \mathrm{w}_i. \tag{2.37}$$

The GMM parameters denoted by $\theta = \big\{\mathrm{m}_i, \mathrm{V}_i, \mathrm{w}_i\big\}_{i=1}^{n_c}$ are estimated from training data $\boldsymbol{\mathcal{X}}$, using the maximum (log-)likelihood of the training sets, i.e.,[7]

$$\hat{\theta} = \arg\max_{\theta} \; \ln p(\boldsymbol{\mathcal{X}}|\theta), \tag{2.38}$$

[7] Since the logarithmic function is strictly monotonically increasing it can be used to simplify the calculations with Gaussian distributions without changing the results of the optimization problem.

thus the parameters are determined for which the likelihoods of the data set is maximized, i.e., the parameters of a model should be chosen such, that they best fit the data.

When the training data $\boldsymbol{\mathcal{X}}$ is incomplete, which can be represented by a set of latent variables $\boldsymbol{\mathcal{Z}}$, the EM algorithm may be used [54]. Instead of directly maximizing likelihoods (which is often intractable), an iterative procedure is adopted, making use of a convex lower bound to the original likelihood function. The lower bound is defined by the expectation of the log-likelihood

[54] Dempster et al., "Maximum Likelihood from Incomplete Data via the EM Algorithm"

$$\mathcal{Q}\big(\theta|\hat{\theta}^j\big) = E_{p_{\boldsymbol{\mathcal{Z}}}}\big[\ln p(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{X}}|\theta)\big], \tag{2.39}$$

with

$$p_{\boldsymbol{\mathcal{Z}}} = p\big(\boldsymbol{\mathcal{Z}}|\boldsymbol{\mathcal{X}}, \hat{\theta}^j\big), \tag{2.40}$$

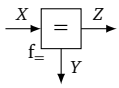
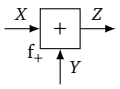
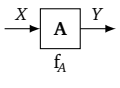
		
$\mathrm{f}_= = \delta(z-x)\delta(z-y)$	$\mathrm{f}_+ = \delta(z-x+y)$	$\mathrm{f}_A = \delta(y-Ax)$
$\vec{\xi}_Z = \vec{\xi}_X + \overleftarrow{\xi}_Y$ $\vec{\mathrm{W}}_Z = \vec{\mathrm{W}}_X + \overleftarrow{\mathrm{W}}_Y$	$\vec{\mathrm{m}}_Z = \vec{\mathrm{m}}_X + \vec{\mathrm{m}}_Y$ $\vec{\mathrm{V}}_Z = \vec{\mathrm{V}}_X + \vec{\mathrm{V}}_Y$	$\vec{\mathrm{m}}_Y = A\,\vec{\mathrm{m}}_X$ $\vec{\mathrm{V}}_Y = A\,\vec{\mathrm{V}}_X\,A^\top$
$\overleftarrow{\xi}_X = \overleftarrow{\xi}_Y + \overleftarrow{\xi}_Z$ $\overleftarrow{\mathrm{W}}_X = \overleftarrow{\mathrm{W}}_Y + \overleftarrow{\mathrm{W}}_Z$	$\overleftarrow{\mathrm{m}}_X = \overleftarrow{\mathrm{m}}_Z - \vec{\mathrm{m}}_Y$ $\overleftarrow{\mathrm{V}}_X = \overleftarrow{\mathrm{V}}_Z + \vec{\mathrm{V}}_Y$	$\overleftarrow{\xi}_X = A^\top\,\overleftarrow{\xi}_Y$ $\overleftarrow{\mathrm{W}}_X = A^\top\,\overleftarrow{\mathrm{W}}_Y\,A$

TABLE 2.1: Tabulated update rules for Gaussian messages of three common nodes, i.e., the equality, the addition and the matrix multiplication node. All three are used to describe the joint probability distributions in this thesis.

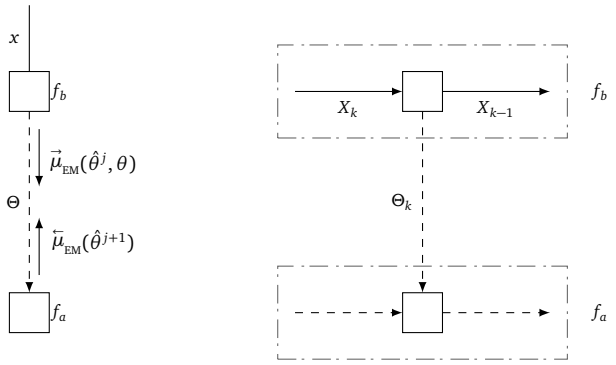


FIGURE 2.12: General FG for EM. The message $\vec{\mu}_{\text{EM}}$ defines the expectation step, whereas $\overleftarrow{\mu}_{\text{EM}}$ carries the updated parameters. This figure was adapted from [57].

where $\hat{\theta}^j$ denote the parameters of iteration step j.

The EM algorithm consist of alternating between

1. the E-step: With $p_{\boldsymbol{\mathcal{Z}}} = p\big(\boldsymbol{\mathcal{Z}}|\boldsymbol{\mathcal{X}}, \hat{\theta}^j\big)$ obtain
$$\mathcal{Q}\big(\theta|\hat{\theta}^j\big) \tag{2.41}$$
and
2. the M-step: Maximize
$$\mathcal{Q}\big(\theta|\hat{\theta}^j\big), \tag{2.42}$$
to obtain the updated parameters, i.e.,
$$\hat{\theta}^{j+1} = \arg\max_{\theta} \ \mathcal{Q}\big(\theta|\hat{\theta}^j\big), \tag{2.43}$$

until the parameters converge, or the allotted time is over [52]. The E-step therefore consists of finding an expression for $p\big(\boldsymbol{\mathcal{Z}}|\boldsymbol{\mathcal{X}}, \hat{\theta}^j\big)$, whereas the M-step results in update rules for the parameters. For the first iteration $j = 0$, initial GMM parameters $\theta^0 = \{\hat{\mathrm{w}}_i^0, \hat{\mathrm{m}}_i^0, \hat{\mathrm{V}}_i^0\}$ should be chosen carefully [55], e.g., using the k-means cluster solutions for the means and making additional guesses for the weights and covariance matrices [56].

[52] Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models"

[55] Arthur and Vassilvitskii, *k-means++: The advantages of careful seeding*

[56] Kerenidis et al., "Quantum expectation-maximization for Gaussian mixture models"

2.4.6 *Expectation-Maximization on Factor Graphs*

The general concept of performing EM on factor graphs was described by Dauwels et. al [57] proposing the general factorization of $f(x, \theta) = f_a(\theta) f_b(x, \theta)$ as shown in Figure 2.12. The EM messages are not an instance of the sum-product algorithm. Rather, they merge well with the factor graph framework, e.g., when Gaussian distributions are used. The open half edge represents the measurements x. Two messages are defined, i.e., a forward EM message, which consists of the expectation step (2.41), represented by

$$\vec{\mu}_{\text{EM}}\left(\hat{\theta}^j, \theta\right) = e^{\eta(\hat{\theta}^j, \theta)} \tag{2.44}$$

where

$$\eta(\hat{\theta}^j, \theta) = \mathrm{E}_{p_b(x|\hat{\theta}^j)}[\ln f_b(x, \theta)], \tag{2.45}$$

and $\mathrm{E}_{p_b(\hat{\theta}^j)}$ is the expectation with respect to the probability distribution $p_b(x|\hat{\theta}^j)$ using the current model parameters $\hat{\theta}^j$, and a backward message

$$\overleftarrow{\mu}_{\text{EM}}\left(\hat{\theta}^{j+1}\right) = \arg\max_{\theta}\left(f_a \ \vec{\mu}_{\text{EM}}\left(\hat{\theta}^j, \theta\right)\right), \tag{2.46}$$

[57] Dauwels et al., "Expectation Maximization as Message Passing"

is defined which carries the updated parameters of the maximization step. In general f_a and f_b may also be factorizable, s.t. more involved distributions can be represented. In particular, it is possible to define "local" expectation messages if (2.45) can be formulated as a product such as

$$e^{\eta(\hat{\theta}^j,\theta)} = e^{\sum_k \eta_k(\hat{\theta}^j,\theta_k)} = \prod_k e^{\eta_k(\hat{\theta}^j,\theta_k)}. \tag{2.47}$$

One iteration step of the EM, assuming local messages are possible, then consists of determining the forward message using the current parameters $\hat{\theta}^j$, performing a forward-backward pass through f_a,[8] leading to the backward message (2.46), which then carry the updates of the parameters $\hat{\theta}^{j+1}$. In this thesis, this relationship is used to define local EM-messages for the estimation of the GMM parameters from noisy training data.

[8] for which f_a must be cycle-free

2.5 MODEL-BASED RESIDUAL GENERATION

Model-based fault detection and diagnosis is a vastly successful approach for automatized fault handling. It relies on the idea, that a nominal system model is available which adequately represent the expected nominal behavior. Then, a signal, i.e., a *residual* is defined which captures the difference between this expected nominal behavior and the observations of the system describing the actual behavior of the system. When the system behaves as expected, and all stochastic processes are neglected, this residual should be equal to zero. If the residual diverges from zero, a fault has occurred. Since the models cannot be perfect and the measurements are always subject to noise, the residuals are commonly assumed to be zero-mean Gaussian distributions and changes in the residuals are detected using change detection approaches. The residuals can also be used to classify the type of fault. The process of interpreting the residual values is termed *residual evaluation*.

This section briefly introduces the two deterministic residual generation concepts which, apart from the probabilistic residual generation method described in factor graphs, were used to analyze the RF-pulses in this thesis. Parity-space and parameter-based are commonly used residual generation approaches for model based fault diagnosis [9]. Other residual generation methods include the use of principal component analysis, neural networks or other black-box model approaches [10].

[9] Surveys covering residual generation methods were, e.g., published by [58, 59, 60]

[10] An overview over the methods can be found, e.g., in [2, 61, 62]

2.5.1 *Parity Space*

Parity space residual generation is based on analytical redundancy [63]. The key concept of analytical redundancy is to define redundant expressions within a nominal system model. In the fault-free, noise-free case, divergences between the redundant expressions should be zero.[11]

[63] Kinnaert, "Fault diagnosis based on analytical models for linear and nonlinear systems-a tutorial"

[11] E. Chow and A. Willsky first proposed parity space approaches for the generation of analytical redundancy in 1984 [64].

The general approach of nonlinear parity space generation is as follows: Consider the nonlinear (deterministic) state-space model

$$\begin{aligned} \frac{dx_t}{dt} &= f(x_t, u_t), \\ z_t &= h(x_t, u_t), \end{aligned} \tag{2.48}$$

where $x_t \in \mathbb{R}^n$ is the state, $u_t \in \mathbb{R}^l$ is the system input and $z_t \in \mathbb{R}^m$ is the system output. It is assumed, that the input and output signals, as well as their derivatives

$$z_t^{(1)}, z_t^{(2)}, z_t^{(i)}, ..., u_t^{(1)}, u_t^{(2)}, u_t^{(i)} \tag{2.49}$$

are known, with (i) indicating the order of the derivative. The states, however, may not be known. The goal is to find redundant expressions, which only depend on the known signals. By multiple differentiations of (2.48), redundant expressions of the unknown states can be formulated. These expressions are then used to eliminate the unknown states and define a set of residuals which are only dependent on the observed signals:

$$\begin{aligned} r_i =& S_i\left(u_t, z_t, u_t^{(1)}, z_t^{(1)}, ..., u_t^{(i)}, z_t^{(i)}\right) = 0, \\ i =& 1, 2, ... \end{aligned} \tag{2.50}$$

where S_i is called the parity space. Depending on the kind of nonlinearity this elimination step can be quite involved [65].

[65] Bokor and Szabó, "Fault detection and isolation in nonlinear systems"

2.5.2 *Parameter-Based*

Unwanted changes in the system behavior can have a significant effect on the system's parameter values. A residual based on those parameters is generated such that the divergence between the system's nominal parameters and those estimated from (current) observations is computed [60]. The nominal parameter values must either be known, e.g., through expert knowledge, or determined from nominal data, e.g., through parameter identification. If the method is supposed to run online, an online estimation of the parameters is needed, i.e., the estimation of the parameter can only depend on current and past observations. If the system model is linear, the parameters are constants, and the residual consists of the divergence of the estimated parameters' value from their nominal [66, 67, 68]. Then, parameter estimation methods such as least-squares based approaches are used to determine the parameters online.

In this thesis, the parameters of the cavity are determined according to the FPGA implemented component, which is capable of computing the parameters online.

[60] Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*

[66] Mulumba et al., "Robust model-based fault diagnosis for air handling units"

[67] Kumar et al., "Development of parameter based fault detection and diagnosis technique for energy efficient building management system"

[68] Hu et al., "Photovoltaic fault detection using a parameter based model"

3

Data Sets & Nominal Cavity Models

SYNOPSIS In this chapter, the data sets are introduced. Then, cavity models are defined which have the objective to represent *nominal* cavity behavior when driven in SP-mode. The use of the model defined in (2.13) proved to be challenging for this task. Thus two alternative models that represent the Lorentz force detuning effect are proposed. A heuristic rule is defined for the selection of nominal RF-pulses, which are needed to identify the model parameters of each cavity. The models' discretized version is given and finally, to account for uncertainties in modeling and measurement noise, uncertainty assumptions are included.

OVERVIEW

Anomaly and fault detection methods determine a measure of deviation from nominal behavior [69]. While anomalies describe not further specified system or measurement behavior, faults are associated to a defined phenomena. The performance of the detection scheme highly depends on how accurately nominal behavior is defined.[1] Thus, assumptions concerning nominal cavity behavior, whether in the form of RF-pulse data or in the form of a physical model (or both), need to be chosen carefully.

The models proposed in this chapter build the foundation from which the difference between analytically obtained cavity outputs and the respective measurements, i.e., residuals, are generated. Ideally the models represent the true cavity behavior perfectly and the residuals obtained from nominal RF-pulse behavior reduces to white noise [60]. The cavity model has so far not been used for the generation of residuals. To fill this gap, two detuning models are proposed with which the cavity output can be simulated.

The cavity-wise model parameters need to be identified from *nominal* RF-pulses. Up until now, there are no methods available with which a preselection of nominal pulses is possible.[2] RF-pulse behavior can be evaluated manually by an experienced operator, but so far no automatic labeling exists. Thus, a heuristic rule is defined to select nominal RF-pulses for the training of the cavity models.

The RF-pulses collected for this thesis are used for the modeling, the residual generation and evaluation, as well as the testing of the proposed anomaly detection methods. RF-pulses over several stations and varying time

[69] Ruff et al., "Deep one-class classification"

[1] The term *nominal* is chosen here (instead of normal) to emphasize that it is not assumed that the cavity behaves in an (however defined) optimal sense, but rather that some selected properties are in an acceptable range.

[60] Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*

[2] The quench detection server is only capable of labeling the observed RF-pulses as 'quench' or 'no quench'

spans were collected. The structure of the data is described and properties of the cavity operation are recounted.

The chapter is structured in the following way: First, the data sets are introduced. Then the selection of nominal RF-pulses from these data sets is described. Next, the proposed nominal cavity models are introduced. The chapter ends with a discretized version of the models and the inclusion of uncertainty for a stochastic cavity model representation.

3.1 DATA SELECTION

The RF-pulse data was collected via the DAQ, see Chapter 2.3.1, at several time-periods between March 2019 and January 2020. The time-periods span between three hours to more than 14 days at a time and include various stations. In this section key properties of the selected time-spans for which the data was stored, and rules with which nominal RF-pulses were selected are described.

3.1.1 *Cavity Operation Properties During Data Collection*

Over 70 TB of RF-data from six different stations was collected and stored. The data sets can be roughly categorized into

1. long-term measurements of nominal operation (several weeks)
2. data sets including thermal breakdowns (quenches)
3. data from "study times".

The long-term measurements are of two arbitrarily picked stations, i.e., A12.L3 and A17.L3.[3] Piezo-electric detuning controllers were installed in all RF-stations of the EuXFEL after December 2019 and thus *most* of the data considered in this thesis stems from a cavity operation *without* piezo-electric detuning control. However, for station A17.L3 (which was used to test the piezo-controller prior to the global installment) two weeks from November 2019 include RF-pulses which were obtained with piezo-electric detuning control.

One data set was stored in which the RF station is not in nominal operation mode, but in "study time". The set-point of the station was altered manually. This data set is used to determine the sensitivity of the detection methods with respect to set-point changes. An overview of the selected data sets is given in Table 3.1.

A total of six instances of quenches were stored. For those, RF-pulses of around three hours prior to the quench were collected for the entire station, i.e., for all 32 cavities. Table 3.2 summarizes the dates and cavities for which the quench was detected.

Ramp-ups and ramp-downs occurring during the selected time-periods were excluded from analysis, as were those cavities which were detuned. Furthermore, pulses affected by a measurement inconsistency, which detects single phase sample measurements as zero, when they are actually around ±180 deg were corrected [14].

[3] The names of the RF-stations result from a numbering of the stations (A1, ..., A25); "L3" indicates that the respective station is located in the main LINAC.

[14] Nawaz et al., "Anomaly Detection for Cavity Signals-Results from the European XFEL"

Station	Beam	Operation	Piezo	Quench	Duration [h]
A17.L3	on	nominal	on/off	yes	792
A12.L3	on	nominal	off	none	480
A16.L3	on	study	off	none	3.5
A11.L3	on	nominal	off	yes	3
A24.L3	on	nominal	off	yes	3
A20.L3	on	nominal	on	yes	3

TABLE 3.1: Data set description of the complete RF-pulse collection analyzed in this thesis.

Date	Station	Cavity	Time	Piezo	Name
22.03.2019	A17.L3	M4.C8	20:02:15	off	I
23.03.2019	A17.L3	M1.C2	23:51:10	off	II
07.10.2019	A11.L3	M1.C3	06:00:15	off	III
07.10.2019	A25.L3	M2.C1	06:02:00	off	IV-1
07.10.2019	A25.L3	M2.C6	06:02:00	off	IV-2
13.02.2020	A20.L3	M4.C4	17:00:15	on	V^{pz}

TABLE 3.2: Overview over six quenches, which happened on five different days and in different cavities. The names of the cavities are defined by the cryogenic module (M1, ..., M4) in which the cavity is installed and the number of the cavity (C1, ..., C8) in that module. To refer directly to singe events, each set was given a name.

3.1.2 *Data Set Descriptions*

The full amount of collected RF-data was divided into different sets. First, the structure of the data set and the notation used in this thesis to describe the data per cavity, pulse and sample is introduced. Then the sets are briefly described. It is important to make a distinction between those RF-pulses collected from cavities with piezo-electric detuning controller and those collected from cavities without. If a data set belongs to the former category, it is marked by "pz", e.g., $\boldsymbol{\mathcal{X}}^{pz}$.

Collected from the DAQ, the *uncalibrated* forward, probe and reflected RF-pulse signals were stored in their amplitude and phase, respectively. Each RF-pulse was then calibrated as described in 2.3.2, for which the reflected signal is needed. After calibration, the reflected signals were no longer needed. The following describes the data structure notation used for the calibrated signals. Each data set contains I and Q measurements of the forward and probe signals of n_p RF-pulses for each of the n_{cav} cavities belonging to the set. A RF-data set is described by

$$\boldsymbol{\mathcal{X}} = \left\{\mathbf{X}^{F,I},\, \mathbf{X}^{P,I},\, \mathbf{X}^{F,Q},\, \mathbf{X}^{P,Q}\right\}, \tag{3.1}$$

where $\mathbf{X}^{F,I}, \mathbf{X}^{F,Q}, \mathbf{X}^{P,Q}, \mathbf{X}^{P,I}, \mathbf{X}^{R,Q} \in \mathrm{R}^{n_{cav} \times n_p \times n_s}$ are tensors of the calibrated forward I and Q, as well as the probe I and Q signals, and n_s denotes the total number of samples in each RF-pulse. When only parts of the full sets of each cavity are needed, the following indexing notation is used: Choosing the sample k for one cavity i_c and all pulses i_p is denoted as $\boldsymbol{\mathcal{X}}_{i_c,:,k}$, i.e.,

$$\boldsymbol{\mathcal{X}}_{i_c,:,k} = \left\{\mathbf{X}^{F,I}_{i_c,:,k},\, \mathbf{X}^{P,I}_{i_c,:,k},\, \mathbf{X}^{F,Q}_{i_c,:,k},\, \mathbf{X}^{P,Q}_{i_c,:,k}\right\}. \tag{3.2}$$

TABLE 3.3: Overview of different data sets. RF-pulses collected when the piezo-electric detuning controller was running, are denoted with "^{pz}".

Data	Description	Stations/ Cavities
$\mathcal{M}, \mathcal{M}^{pz}$	selected nominal RF-pulses for training	all
$\tilde{\mathcal{M}}, \tilde{\mathcal{M}}^{pz}$	selected nominal RF-pulses for validation	all
$\mathcal{Q}, \mathcal{Q}^{pz}$	quenching RF-pulses	stations listed in Table 3.2
$\mathcal{S}$	RF-pulses of C6.M1 of L17.L3	C6.M1 of L17.L3
$\mathcal{T}, \mathcal{T}^{pz}$	non-severe data	A12.L3, A17.L3

This data structure is used for all data sets which are now introduced by their names, their associated variable and their purpose in this thesis.

The nominal data sets $\mathcal{M}, \mathcal{M}^{pz}$ are used to determine nominal cavity parameter values as well as nominal characteristics of the anomaly/fault detection methods. They are selected from the beginning of each collection period for every cavity and are chosen such that they conform to specific rules described in the next section. Each cavity has its own set of nominal RF-pulses.

The validation data sets $\tilde{\mathcal{M}}, \tilde{\mathcal{M}}^{pz}$ are chosen according to the same criteria as the nominal data sets, however, they are not used to determine the cavity parameters but were set aside for model validation purposes.

RF-pulses associated to a quench event are denoted by $\mathcal{Q}, \mathcal{Q}^{pz}$. The quench data sets are the only ones with labels (through the current quench detection system), i.e., each RF-pulse is either labeled as "quench" or as "no quench". In this thesis a quench event is associated to multiple quenching RF-pulses, which often have a cascading property. The severity of a quench often gets worse from pulse to pulse and thus the first quenching RF-pulse usually shows the smallest effect of the quench. After the detection of the quench, the RF-signal of the station is switched off. Between the detection of the first quench and the RF turn-off the station generated at least three RF-pulses. This lead to the fact that each quench event contained at least three quenching RF-pulse.

Table 3.2 shows six instances where a quench was detected. The first two instances occurred in two different cavities of station A17.L3 with more than 24 hours apart. Those quench events are referred to as events I and II, respectively. Another quench occurred in A11.L3. This event is referred to as event III. Only two minutes later, in a different station, i.e., A25.L3 two cavities quenched with one pulse apart. Although they can be seen as one quench data set, in order to be able to refer to each cavity individually, the two quench events are named IV-1 and IV-2. Finally one quench event is considered, which occurred in January 2020. This quench event is the only one analyzed in this thesis for which piezo-electric detuning control was running. This event will be referred to as V^{pz}.

Anomalous behavior was manually observed (but not automatically detected) in parts of the RF-pulses collected for cavity C6.M1 of station A17.L3 when it was run without piezo-electric detuning control. To further

analyze this behavior, the RF-pulses of that cavity were separated from the rest of the station and collected in the data set $\mathcal{S}$.

The non-severe data set $\mathcal{T}$, $\mathcal{T}^{pz}$ is defined by those RF-pulses which were recorded for the long-term data and for which no quenches were detected and data set $\mathcal{S}$ was excluded. Those data sets are referred to as non-severe, since any anomalies detected in that time-span did not immediately lead to a shut-down of the station, and were not deemed severe by an expert.

Except for nominal training and nominal validation sets, the data sets all entail consecutive RF-pulses and thus define a time period between the fist and final time of recording the pulses.

3.1.3 *Defining Nominal RF-Pulse Data*

Each cavity has its own individual properties. These include the unavoidable differences resulting from their manufacturing to the way they are tuned, and at what set-point they are operated. Thus, to determine each cavities' nominal parameter values, nominal RF-pulses of that cavity were needed. While an expert may be able to manually select nominal RF-pulses, the amounts of RF-pulses produced (600 each second for each cavity), and the amount of data stored and analyzed for this thesis make such a manual selection tedious. The following describes the heuristic rules that aim at selecting RF-pulses such that the possibility of picking anomalous RF-pulses is reduced.

For the RF-pulses to be nominal, they need to stem from nominal cavity operation, i.e., those of a user-run. This entails that the gradient set-points are fixed and do not exceed the limits of the stations. The cavities are tuned close to resonance using the stepper motors, and the RF-stations are on-beam, i.e., accelerating electron bunches for the generation of X-ray flashes. In nominal operation, all feedback loops controlling the RF fields in the cavities are closed, including the feed-forward, learning feed-forward, and the MIMO controller. Since active resonance control using piezo-electric tuners became available only after January 2020, for the data collected before that date, operation without piezo controllers was considered nominal.

Furthermore, the RF-pulses are defined as nominal when they were not labeled as or lead up to a quench or other RF-station trip, and when the *static detuning* $\bar{f}^{\Delta}$ stayed within an acceptable range of ±25 Hz, [32]. The static detuning is defined as the mean detuning over the flattop of the pulse, i.e.,

[32] Schilcher, "Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities"

$$\bar{f}^{\Delta} = \frac{1}{n_{flt}} \sum_{k=k_f}^{k_d-1} f_k^{\Delta}, \tag{3.3}$$

where n_{flt} are the number of samples in the flattop and f_k^{Δ} is determined from (2.17) using numerical differentiations. For cavities tuned near resonance, the mean static detuning is around 0 Hz. In the data sets taken in 2019, although in user-run, individual cavities were *not* tuned around 0 Hz. Instead of excluding those cavities from the evaluation, the nominal detuning is determined relative to the mean static detuning over several RF-pulses. The threshold of acceptable divergence from the mean is again set at 25 Hz. In Figure 3.1 static detuning values over three hours for a cavity without piezo-electric control in the flattop are depicted. These variations are an

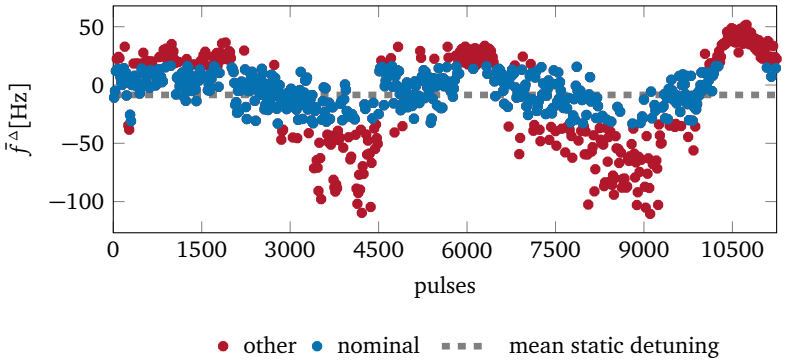


FIGURE 3.1: Static detuning values over the period of three hours. It is assumed that nominal RF-pulses deviate from the mean static detuning by ±25 Hz

extreme example and this behavior is *not* assumed to be nominal even for cavity operation without piezo-electric detuning control [70, 71].

[70] Branlard et al., "LLRF testing of superconducting cryomodules for the European XFEL"

[71] Branlard et al., "LLRF tests of XFEL cryomodules at AMTF: first experimental results"

In an online operation, nominal RF-pulses would be selected at the beginning of the detection period for each cavity. The parameters derived from those RF-pulses would then be used for the evaluation of the anomalies of any *future* pulses.

For the estimation of the model parameters described in the subsequent section, $n_{\mathrm{n}} = 1000$ pulses are used, selected randomly from the set of pulses fulfilling the above defined requirements. Choosing $n_{\mathrm{n}} = 1000$ corresponds to a confidence of 90%, that the estimation of the distribution parameters is close to their actual values [72]. The following section introduces two models for the Lorentz force detuning effects on the cavities.

[72] Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability"

3.2 NOMINAL CAVITY MODELS

The effect of Lorentz forces on the cavity were described in Chapter 2.2.3, and a second order differential equation was introduced with which their effects on the RF-fields can be modeled. However, identifying the necessary parameters with the given RF-pulses turns out to be challenging and thus, two alternative cavity models are proposed.

3.2.1 *Detuning Model*

In Chapter 2.2.3, a second-order system was used to describe superimposed mechanical oscillations affecting the cavity due to field-dependent Lorentz forces. When using this model structure for the cavities, the mechanical resonance frequencies α_m, the time constants γ_m, and the Lorentz force detuning constants β_m accompanying the most dominant resonant modes n_m need to be estimated.[4] Furthermore, the initial values of each detuning mode, i.e., the initial state vector, also needs to be determined. Determining the parameters for each cavity should not interfere with a user-run. Thus, measuring the cavities' mechanical resonances, e.g., by exciting the modes mechanically, is not practicable. Instead, the unknown parameters may be determined from the RF-pulse measurements. Mechanical resonance frequencies for TESLA SRF cavities were determined for the first three modes in [73], and for the first four modes in [74], see Table 3.4. The mechanical frequencies are slow compared to the length of a single RF-pulse

[4] In continuous-wave operation, 20 mechanical modes were identified [34].

[73] Czarski et al., "Cavity parameters identification for TESLA control system development"

[74] Jugo et al., "Control algorithm tests using a virtual CW SRF cavity"

	Mode 1	Mode 2	Mode 3	Mode 4	Unit
Frequency [74]	41.7	93.7	171.7	193.5	Hz
Frequency [73]	235	290	450		Hz
β_m [73]	0.4	0.3	0.2	–	$Hz/(MV)^2$

TABLE 3.4: Frequency and Lorentz force detuning parameters.

and the "missing" information about the system behavior between the pulses is comparatively large.

Figure 3.2 depicts the first four frequencies as determined by [74], assuming a zero phase at time = 0. The dashed line at t_f indicates the end of a single RF-pulse measurement, its duration only barely surpassing the first quarter of a full oscillation period for the highest frequencies. A proper estimation of these frequencies from the pulses alone would result in a high uncertainty of the estimated parameters, and first attempts to identify them have been unsuccessful.

The following alternative detuning models make use of the fact that in nominal user-run operation, the pulse to pulse variation of the detuning is low. Both models are thus restricted to fixed vector-sum set-points of the forward fields, which is assumed for nominal RF-station operation. Only then are detuning traces reproduced accurately from pulse to pulse and are only slightly (up to ±10 Hz) modulated by microphonics [37].

3.2.2 *First Order Detuning Model*

A first-order system is assumed which is similar to the polynomial fittings proposed in [73] where piece-wise polynomials for the filling, flattop and decay were used to determine the detuning estimations online. The first-order system as used in this thesis, was previously proposed in [32]. It is given by

$$\frac{dx_t^m}{dt} = -\frac{1}{\gamma_m} x_t^m - 2\pi\beta_m |\underline{v}_t^p|^2, \tag{3.4}$$

where γ_m, β_m are the characteristic detuning time and Lorentz force detuning constants, respectively, and the total detuning is obtained by superimposing the individual (first-order) detuning components $x_t^m \in \mathbb{R}$, i.e.,

$$f_t^\Delta = \sum_{m=1}^{n_m} x_t^m, \tag{3.5}$$

[32] Schilcher, "Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities"

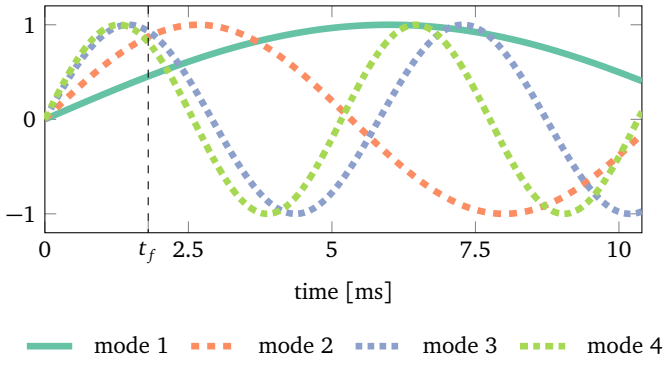


FIGURE 3.2: Mechanical modes in comparison to the length of a single RF-pulse. The mechanical oscillations are slow in comparison to the single RF-pulse, ending at t_f. In addition two successive pulses are in relation to those oscillations far apart.

where n_m is the number of mechanical modes considered for the model. The detuning parameters are estimated using the linear state space model for the detuning described by (3.4) and (3.5). This is possible, because the model defined by (3.4) and (3.5) can be decoupled from (3.6), when using the detuning estimation f_t^{Δ} as the system output and the squared probe signals $|\underline{v}_t^P|^2$ as its input.
The estimation of the model parameters is described in more detail in a later part of this chapter. First, an alternative model is introduced which uses the nominal detuning estimations as a fixed additional input to the cavity model (3.6).

3.2.3 *Model with Nominal Detuning Trace*

The following model uses the nominal detuning over the RF-pulse as an additional (and fixed) input to the cavity model.

This has, e.g., the practical advantage of restricting the number of states to two such that the cavity model of (3.6) can then be formulated as

$$\begin{aligned}\frac{dv_t^{P,I}}{dt} &= -2\pi f^{\dagger} v_t^{P,I} - 2\pi u_t^{III} v_t^{P,Q} + 2\tilde{f}^{\dagger}\left(v_t^{F,I} + v_t^{B,I}\right),\\ \frac{dv_t^{P,Q}}{dt} &= 2\pi u_t^{III} v_t^{P,I} - 2\pi f^{\dagger} v_t^{P,Q} + 2\tilde{f}^{\dagger}\left(v_t^{F,Q} + v_t^{B,Q}\right),\end{aligned} \tag{3.6}$$

where u_t^{III} is the nominal detuning value estimated from nominal RF-pulses. Additionally for a fixed set-point operation of the cavity, the pulse to pulse variations of the detuning should be negligible and the thus obtained nominal model for the cavities may be very accurate. Instead of having to estimate the mechanical parameters for the n_m modes, the nominal detuning values for the entire RF-pulse are used. Before describing the estimation procedures for both models, their discretized versions are given.

3.2.4 *Discretization*

With the comparatively high sampling rate $f_s = 1\,\text{MHz}$, the Euler forward method is chosen for the discretization of the model. The discrete-time model may then be summarized to a set of difference equations describing the electromagnetic properties of the cavity by

$$\begin{aligned}x_{k+1}^{I} &= -a_0 x_k^{I} - 2\pi T_s f_k^{\Delta} x_k^{II} + b_0 u_k^{I},\\ x_{k+1}^{II} &= 2\pi T_s f_k^{\Delta} x_k^{I} - a_0 x_k^{II} + b_0 u_k^{II}\end{aligned} \tag{3.7}$$

where $T_s = \frac{1}{f_s}$ and

$$a_0 = -1 + f^{\dagger} T_s, \quad b_0 = 2\pi f^{\dagger} T_s. \tag{3.8}$$

are the model parameters after discretization. The model outputs are then defined by

$$\begin{aligned}y_k^{I} &= x_k^{I},\\ y_k^{II} &= x_k^{II},\end{aligned} \tag{3.9}$$

where

$$u_k = \begin{bmatrix} u_k^I \\ u_k^{II} \end{bmatrix} = \begin{bmatrix} v_k^{F,I} \\ v_k^{F,Q} \end{bmatrix}, \quad y_k = \begin{bmatrix} y_k^I \\ y_k^{II} \end{bmatrix} = \begin{bmatrix} v_k^{P,I} \\ v_k^{P,Q} \end{bmatrix}, \tag{3.10}$$

are the input and output measurements.

When using the first-order models for the mechanical properties of the cavity, the discretized states

$$\begin{aligned} x_{k+1}^m &= -a_m x_k^m + b_m\left({x_k^I}^2 + {x_k^{II}}^2\right), \\ f_k^\Delta &= \sum_{m=3}^{m=n_m-3} x_k^m, \end{aligned} \tag{3.11}$$

are added, where

$$a_m = -1 + \frac{1}{\gamma_m} T_s, \quad b_m = \beta_m T_s, \tag{3.12}$$

are the discretized model parameters.

When alternatively using the nominal detuning values as an additional input to the system, then

$$f_k^\Delta = u_k^{III}, \tag{3.13}$$

where u_k^{III} represents the additional discretized input.

Identifying the nominal parameters of the first-order and fixed-trace models is described in the subsequent section.

3.3 DETERMINING NOMINAL CAVITY PARAMETERS

This section describes the identification of nominal parameters. For each cavity i_c, the parameters are determined from the calibrated nominal RF-pulses $\boldsymbol{\mathcal{M}}_{i_c,:,:}$, if i_c refers to a cavity operated without piezo-electric control and $\boldsymbol{\mathcal{M}}^{pz}_{i_c,:,:}$, if for that cavity piezo-electric control was on. For the cavity model defined by (3.7) the time-varying detuning of the cavity f_k^Δ, and the pulse-wise half-bandwidths $f^\dagger, \tilde{f}^\dagger$ with $f^\dagger \approx \tilde{f}^\dagger$ need to be determined for each cavity.[5]

[5] As defined in Chapter 2: The parameter $\tilde{f}^\dagger$ describes the bandwidth dominated by the external quality factor and is assumed constant over the RF-pulse, $f^\dagger$ on the other hand can change during an RF-pulse.

First the half-bandwidth is determined from the decay samples of the amplitude probe measurements as described in (2.16) for each of the RF-pulses in $\boldsymbol{\mathcal{M}}_{i_c,:,:}$, or $\boldsymbol{\mathcal{M}}^{pz}_{i_c,:,:}$. The distribution of the half-bandwidth is assumed to be Gaussian, i.e., for each cavity i_c, it is described by the random variable $F^\dagger$, with

$$F^\dagger \sim p(f^\dagger | \boldsymbol{\mathcal{M}}_{i_c,:,k_d:n_s}) = \mathcal{N}\left(f^\dagger; \mathrm{m}_{F^\dagger}, \mathrm{V}_{F^\dagger}\right), \tag{3.14}$$

where $k_d : n_s$ denote the samples of the decay of the RF-pulse and $\mathrm{m}_{F^\dagger}, \mathrm{V}_{F^\dagger} \in \mathbb{R}$ are the estimated mean and variance, respectively. It is thus assumed that the best (maximum likelihood) estimation of the nominal half-bandwidth parameter is given by $f^\dagger = \mathrm{m}_{F^\dagger} \approx \tilde{f}^\dagger$.

The nominal bandwidth $\mathrm{m}_{F^\dagger}$ is then used to determine the cavity-dependent detuning, computed with the RF-pulses of $\boldsymbol{\mathcal{M}}_{i_c}, \boldsymbol{\mathcal{M}}^{pz}_{i_c}$ with the intra-pulse parameter calculation described in (2.17), which is used for determining the nominal detuning values for each cavity.

Each sample k of the detuning is assumed to be Gaussian distributed over the set of nominal RF-pulses belonging to each cavity, i.e.,

$$F_k^\Delta \sim p(f_k^\Delta | \boldsymbol{\mathcal{M}}_{i_c,:,k}) = \mathcal{N}\left(f_k^\Delta; \mathrm{m}_{F_k^\Delta}, \mathrm{V}_{F_k^\Delta}\right), \tag{3.15}$$

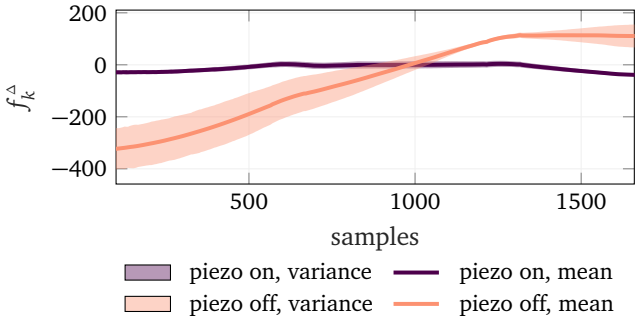


FIGURE 3.3: Nominal detuning traces of a cavity operated with and without piezo-electric detuning control. The piezo-electric controller assures that the detuning values during the flattop of the pulse are around zero.

where $m_{F_k^{\Delta}}, V_{F_k^{\Delta}} \in \mathbb{R}$ is the mean and variance of each detuning sample in a pulse. Figure 3.3 shows nominal detuning traces $m_{F_k^{\Delta}}$ including their variance bounds of $V_{F_k^{\Delta}}$ for the RF-pulses collected during a cavity operation with ($\mathcal{M}^{pz}$) and without ($\mathcal{M}$) piezo-electric control.

The nominal detuning values $m_{F_k^{\Delta}}$ are used as the additional input to each cavity as described in (3.13), i.e., $u_k^{III} = m_{F_k^{\Delta}}$. It is also used to determine the first-order mechanical detuning parameters of (3.11).

3.3.1 *Detuning Parameter Identification*

The parameter values are identified using the grey-box model structure in the form of a linear state space model given by a set of differential equations, i.e.,

$$\frac{dx_t}{dt} = \begin{bmatrix} -\frac{1}{\gamma_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -\frac{1}{\gamma_{n_m}} \end{bmatrix} x_t + 2\pi \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{n_m} \end{bmatrix} u_t^*, \tag{3.16}$$

$$y_t^* = \begin{bmatrix} 1, \ldots, 1 \end{bmatrix} x_t$$

where $u_t^* = |\underline{y}_t^p|^2 \in \mathbb{R}$ and $y_t^* = f_t^{\Delta} \in \mathbb{R}$ are virtual, i.e., estimated input and outputs and

$$x_t = [x_t^1, \ldots, x_t^{n_m}]^T \in \mathbb{R}^{n_m}. \tag{3.17}$$

The inputs u_t^* are the estimated nominal detuning values $m_{F_k^{\Delta}}$ of 3.15. The identification of the parameters consists of determining the initial state vector $x_0 \in \mathbb{R}^{n_m}$, the order of the model n_m and identifying the time constants γ_m and the Lorentz force detuning constants β_m for each mode.
The algorithm is based on the least-squares method and was performed using the grey-box estimation algorithm implemented in the System Identification toolbox of MATLAB [75].
Best identification performances were obtained when choosing a model order of $n_m = 4$ for each cavity.

The quality of the identification is assessed by comparing the measured probe signals to the forward simulation of the model using the forward signal measurements. For comparison the normalized mean squared error is determined, with a maximum fit of one. The initial states and parameters used for the identification procedure are different for each cavity and are determined using a random search with 60 iterations, see e.g. [76, 77], over a range of possible initial parameters and then selecting those initial parameters resulting in the highest overall fits. Both v_k^{PI} and v_k^{PQ} measurements

[75], *MATLAB version 9.3.0.713579 (R2017b)*

[76] Bergstra and Bengio, "Random search for hyper-parameter optimization"

[77] He et al., "AutoML: A Survey of the State-of-the-Art"

could be predicted with normalized mean squared error fits between 0.70 and 0.98.

3.4 STOCHASTIC CAVITY MODEL

So far, uncertainties inherent to all physical processes are not accounted for in the cavity model. To do so, the cavity state, input and output vectors are now represented as random variables. Additional random variables are introduced, modeling the process and measurement noise of the cavity. Thus the following extended, stochastic cavity model may be considered:

$$\begin{aligned} X^{I}_{k+1} &= -a_0 X^{I}_{k} \quad -2\pi T_s F^{\Delta}_{k} X^{II}_{k} + b_0 U^{I}_{k} + I^{I}_{k}, \\ X^{II}_{k+1} &= 2\pi T_s F^{\Delta}_{k} X^{I}_{k} - \quad a_0 X^{II}_{k} + \quad b_0 U^{II}_{k} + I^{II}_{k}, \\ Z^{I}_{k} &= X^{I}_{k} + D^{I}_{k}, \\ Z^{II}_{k} &= X^{II}_{k} + D^{II}_{k}, \end{aligned} \tag{3.18}$$

as well as the mechanical properties described by four detuning modes, i.e.,

$$\begin{aligned} X^{m}_{k+1} &= -a_m X^{m}_{k} + b_m \left(X^{I\,2}_{k} + X^{II\,2}_{k} \right) + I^{m}_{k}, \\ F^{\Delta}_{k} &= \sum_{m=3}^{m=n_m-3} X^{m}_{k}, \end{aligned} \tag{3.19}$$

where D^{I}_{k} and D^{II}_{k} are Gaussian distributed random variables representing the measurement noise of the system and I^{I}_{t} and I^{II}_{t} are Gaussian distributed random variables representing the process noise/ modeling uncertainty of the electromagnetic system, whereas I^{m}_{k} is the random variable representing the noise in the detuning for each mode m.

4

Residual Generation & Evaluation

"A graphical model can be thought of as a probabilistic database, a machine that can answer "queries" regarding the values of the sets of random variables. We build up the database in pieces, using probability theory to ensure that the pieces have a consistent overall interpretation."
—M. Jordan [46]

SYNOPSIS In this chapter, the nominal cavity assumptions introduced in the previous chapter are used to define three model-based residuals, each of which describe a measure of divergence between the expected nominal cavity behavior and the calibrated electromagnetic field measurements. Then, two residual evaluation methods are introduced. The first method assumes that only the nominal distribution is known and can be represented as a Gaussian. The second method assumes that data pertaining information about the fault distribution is available and different fault classes can be modeled by a GMM. Both models are introduced in factor graphs and the evaluations are described via Gaussian message passing. To obtain the GMM, parameters need to be estimated. Finding the parameters from unclassified data (training) is also described as message passing. The chapter ends exploring the possibilities of combining the residual generation methods with the evaluation methods in the factor graph framework.

OVERVIEW

All methods considered in this thesis rely on the principle of analytical redundancy,[1] which mainly involves the use of signal processing techniques such as state and parameter estimation. Observer-based estimators or Kalman filters are used for linear state estimations whereas least squared methods have successfully been applied for the estimation of parameters in the context of analytical redundancy. Only the parity space approach does *not* use signal processing techniques, rather, it rearranges the input/output equations to obtain redundant expressions of the same variable.[2] This chapter describes three residuals for the cavity system. The first is represented in a factor graph using Gaussian message passing, extending the stochastic nominal cavity model given in (3.18) with the assumption that a fault/anomaly has an *additive* effect on the outputs. Other assumptions are possible, e.g., considering multiplicative and additive effects on all the states or the parameters as described for other systems in [78, 79]. That the faults have a direct impact on the parameters of the cavity system is considered in the deterministic parameter-based method.

For the factor graph based residual, the marginal distribution of the residual is inferred via message passing, for which a linear approximation

[1] as opposed to hardware redundancy, common for analogue fault detection

[2] A thorough overview over the different approaches is given, e.g., in [2, 61, 62]

[78] Ding, *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*

[79] Gertler, "Structured residuals for fault isolation, disturbance decoupling and modelling error robustness"

of the nonlinear state transition is chosen. Using linear approximations has the advantage that the resulting messages are again Gaussian, making subsequent calculations easier. The approximations are chosen such, that when used to infer the states, they are the same as obtained by the unscented Kalman filter which was first introduced in [80]. The residuals thus generated are also known as the *innovation* process of the unscented Kalman filter, which is a well-established residual generator for nonlinear system models, first introduced in [81], and used, e.g., in [82, 83].

Although Bayesian networks have long been recognized as a modeling tool for probabilistic fault diagnosis schemes, see [84], factor graphs have not been applied as extensively.[3] Using factor graphs in the context of fault diagnosis has also been explicitly proposed by [8], whereas [85] has described the task of fault diagnosis as an inference problem at its core.

The second residual proposed makes use of the computations of the intra-pulse cavity parameters described by (2.17) and (2.18). Taking advantage of the fact that these computations are already implemented as an FPGA component, a straight-forward parameter-based method is defined.

Finally, a nonlinear, deterministic parity space residual for the electromagnetic part of the cavity model is proposed.

The aim of residual evaluation is to detect changes in the residual and classify them either within a *binary* classification into nominal, anomalous or into *multiple* fault classes. There is a diverse range of different residual evaluation methods, such as statistical tests, geometric classifiers, support vector machines, change-point detection algorithms or neural networks, to name just a few. For an overview see, e.g., [60], whereas [86] gives a thorough introduction to change point detection.

Two residual evaluation methods that have been applied to different systems in the past for the detection of anomalies or faults, are here defined as Gaussian message passing algorithms in factor graphs. Both methods assume Gaussian distributions for both nominal and faulty residuals.

The first assumes that only the nominal (Gaussian) distribution of the residual is known, from which a statistically significant divergence can be detected using *likelihood ratios*. Both model selection in general and likelihood ratios in particular have already been mentioned in the factor graph context in [87, 88]. Likelihood ratios are used extensively for change detection, model selection and hypothesis testing and play an important role for the detection of changes in data-streams [89, 90, 86].

The second evaluation method assumes that the distribution over the residuals for both nominal and anomalous may be represented as a *Gaussian mixture model* (GMM), see Chapter 2.4.5. Different Gaussian components may refer to different classes of anomalies, making a multi-class evaluation possible. Using GMMs for the purpose of anomaly and fault detection has been studied before, see, e.g., [91, 92]. Different from the likelihood ratio based method, the classification results are probabilistic, i.e., the evaluation result gives the probability that the current measurements describe one of the considered fault classes or the nominal [93, 94]. Both the evaluation of residuals and the training of the GMM parameters are described in factor graphs including both noisy and deterministic residuals. GMM parameters can be obtained, e.g., using the expectation-maximization (EM) algorithm.

[80] Julier and Uhlmann, "New extension of the Kalman filter to nonlinear systems"

[81] Mehra and Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems"

[82] Nag et al., "Model based fault diagnosis of low earth orbiting (leo) satellite using spherical unscented kalman filter"

[83] Wei et al., "Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques"

[84] Cai et al., "Bayesian networks in fault diagnosis"

[3] Fault diagnosis is used as an umbrella term and refers to detecting and locating the affected component and identifying the type of fault that occurred.

[8] Gienger and Sawodny, "Data-based Process Monitoring and Iterative Fault Diagnosis using Factor Graphs"

[85] Verbert, "Fault diagnosis and maintenance optimization for interconnected systems: With applications to railway and climate control systems"

[60] Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*

[86] Basseville and Nikiforov, *Detection of abrupt changes: theory and application*

[87] Zarrin and Lim, "Belief propagation on factor graphs for cooperative spectrum sensing in cognitive radio"

[88] Laar and Vries, "A probabilistic modeling approach to hearing loss compensation"

[89] Lu et al., "Change detection techniques"

[90] Chandola et al., "Anomaly detection: A survey"

[91] Heydarzadeh and Nourani, "A two-stage fault detection and isolation platform for industrial systems using residual evaluation"

[92] Yu, "Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models"

[93] Schwall et al., "A Probabilistic Vehicle Diagnostic System Using Multiple Models."

[94] Zhang et al., "A probabilistic fault detection approach: Application to bearing fault detection"

Training GMMs with noisy data has been described, e.g., in [95]. In one of the following sections, the updates of the parameters defined in [95] are obtained through message passing. Outlier detection has previously been proposed to be handled in the factor graph framework using for example a normal prior with unknown variance, for the estimation of impulsive signals [96].

Finally the residual generation with the unscented Kalman filter, the evaluation using GMMs and the training of the GMM can all be represented in one factor graph, describing one possibility for an overall method to classify the RF-pulse samples.

The residual generation, evaluation and GMM parameter training is published in [13]. The parity space residual generation is published in [11].

[95] Ozerov et al., "Uncertainty-Based Learning of Gaussian Mixture Models from Noisy Data"

[96] Loeliger et al., "On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing"

[13] Nawaz et al., "Probabilistic model-based fault diagnosis for the cavities of the European XFEL"

[11] Nawaz et al., "Anomaly Detection for the European XFEL using a Nonlinear Parity Space Method"

4.1 PROBABILISTIC RESIDUAL GENERATION

The stochastic cavity model described with (3.18) can be generalized to a hidden Markov model with l inputs, m outputs and n states, i.e.,

$$\begin{aligned} X_{k+1} &= \mathfrak{a}(X_k) + \mathbf{B}U_k + I_k \\ Z_k &= \mathbf{C}X_k + D_k, \end{aligned} \tag{4.1}$$

where $X_k \in \mathbb{R}^n, U_k \in \mathbb{R}^l, Z_k \in \mathbb{R}^m$ are the state, input and output random variables and $I_k \in \mathbb{R}^n, D_k \in \mathbb{R}^m$ denote the random variables for process (modeling) and measurement uncertainty, respectively. The real-valued matrices $\mathbf{B} \in \mathbb{R}^{n\times l}$ and $\mathbf{C} \in \mathbb{R}^{n\times m}$ and the nonlinear, deterministic function $\mathfrak{a} : \mathbb{R}^n \to \mathbb{R}^n$ define the dependency between the random variables for each discrete time-step k.

Figure 4.1 depicts the factor graph of this model. The graph structure is the same as that of the linear state-space model introduced in Chapter 2. The nonlinearity of the cavity system dynamic is represented by a nonlinear factor node, defined as

$$f_{\mathfrak{a}} = \delta(x_k'' - \mathfrak{a}(x_{k-1})), \tag{4.2}$$

and labeled with the nonlinear function "$\mathfrak{a}$" in the factor graph. For this node, the message passing updates for a linear system matrix (Chapter 2.4.4,

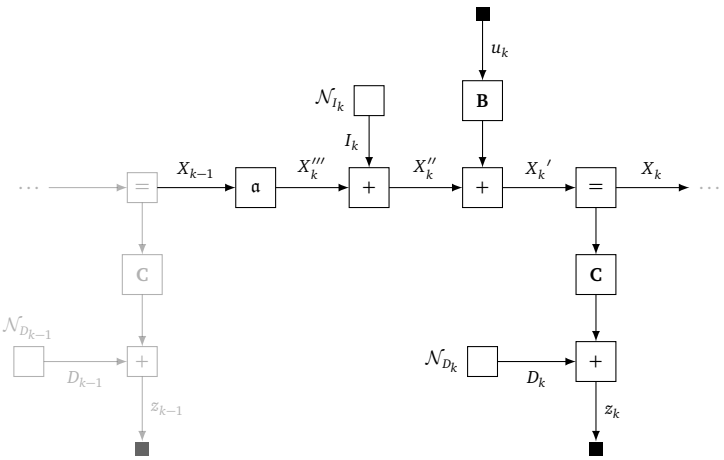


FIGURE 4.1: State-space factor graph with nonlinear factor node $\mathfrak{a}$. The structure of the graph is the same as in Chapter 2.4.4.

Table 2.1) do not apply, instead message passing definitions for a nonlinear transformation is used. When inferring the distribution over any variable in the graph, the Gaussian messages that pass through this node are subjected to a nonlinear transformation. A nonlinear transformation of a Gaussian distribution is, in general, not Gaussian. To be able to handle nonlinear transformations while still retaining the computational simplicity of Gaussian message passing everywhere else in the graph (but at the nonlinear factor node), an approximation (linearization) of the nonlinear function $\mathfrak{a}$ is desirable. Various approximations that result in a Gaussian with features of the exact distribution, e.g., the mode, or the mean and covariance can be found in the literature. Message passing computations to perform such approximations in the factor graph framework were proposed by [97]. One such approximation was chosen, which, when used for the estimation of the state variable X_k is equal to the unscented Kalman filter.

[97] Herzog et al., "Iterative Approximate Nonlinear Inference via Gaussian Message Passing on Factor Graphs"

4.1.1 *Residual Generation with Factor Graphs*

So far, the state-space model in (4.1) does not account for faulty system behavior. A fault can have a multiplicative or an additive influence on the states, the output, the parameters of the model or a combination of all.

In the following it is assumed that anomalous behavior of the system can be detected via a continuous random variable R_k which enters the state-space equations through the output equation, i.e.,

$$\begin{aligned} X_{k+1} &= \mathfrak{a}(X_k) + \mathbf{B}U_k + I_k \\ Z_k &= \mathbf{C}X_k + D_k + R_k. \end{aligned} \tag{4.3}$$

In Figure 4.2 the residual R_k is added to the factor graph. The residual reveals information about a divergence between the estimated outputs using the nominal model and the measurements which may or may not reflect that nominal behavior. Assuming (for now) that nothing is known about possible cavity faults, and the residual should *not* be used when estimating the distribution over any other random variable in the factor graph, the connection with the factor graph is denoted by a dashed line.[4]

[4] Another possibility is to choose a non-informative prior for the residual, which should have no effect on the inferred distributions of any other variable in the graph. However, a non-informative Gaussian prior, is defined as a Gaussian with zero mean and infinite variance. Assuming this kind of prior for the residual is not an option as that would mean that this variance would be added to the variance of the measurements. This, in turn, would mean that the measurements would not influence the estimation of the state variables, i.e., after the transition through the equality node, see Table 2.1.

Determining the marginal likelihood of R_k,i.e.,

$$p(r_k | u_{0:k}, z_{0:k}, i_{0:k}, d_{0:k}, x_0) \tag{4.4}$$

the residual distribution can be inferred. In the factor graph this is equivalent to the backward message $\overleftarrow{\mu}_{R_k}$. Assuming the observations of Z_k and U_k are known and deterministic, the forward message $\vec{\mu}_{U_k} = \mathcal{N}(u_k; \vec{\mathrm{m}}_{U_k}, 0)$ describing the inputs and the backward message $\overleftarrow{\mu}_{Z_k} = \mathcal{N}(z_k; \overleftarrow{\mathrm{m}}_{Z_k}, 0)$ describing the outputs are degenerate Gaussian distributions and the associated prior nodes in the factor graph are therefore denoted by a solid black node. Assuming further that both measurement and process noise are zero-mean distributions, i.e., $\vec{\mathrm{m}}_{D_k} = \vec{\mathrm{m}}_{I_k} = 0$ the mean and covariance of the residual distribution

$$\overleftarrow{\mu}_{R_k} = \mathcal{N}\left(r_k; \overleftarrow{\mathrm{m}}_{R_k}, \overleftarrow{\mathrm{V}}_{R_k}\right), \tag{4.5}$$

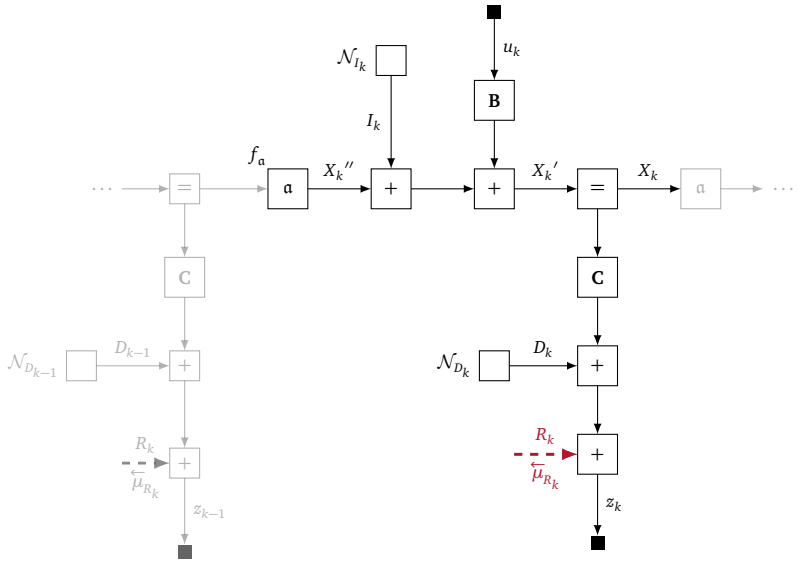


FIGURE 4.2: State-space model is extended with the continuous random variable R_k. The marginal likelihood of that variable is its backward message $\overleftarrow{\mu}_{R_k}$ which is used to define the residual.

are given by

$$\begin{aligned}\overleftarrow{\mathrm{m}}_{R_k} &= \overleftarrow{\mathrm{m}}_{Z_k} - \mathbf{C}\left(\vec{\mathrm{m}}_{X_k''} + \mathbf{B}\,\vec{\mathrm{m}}_{U_k}\right),\\ \overleftarrow{\mathrm{V}}_{R_k} &= \mathbf{C}\left(\vec{\mathrm{V}}_{X_k''} + \vec{\mathrm{V}}_{I_k}\right)\mathbf{C}^T + \vec{\mathrm{V}}_{D_k},\end{aligned} \tag{4.6}$$

where X_k'' denotes the state variable to the right of the nonlinear factor node, and the forward message on that edge

$$\vec{\mu}_{X_k''} = \mathcal{N}\left(x_k''; \vec{\mathrm{m}}_{X_k''}, \vec{\mathrm{V}}_{X_k''}\right) \tag{4.7}$$

is the approximated Gaussian distribution which results from the nonlinear factor node definitions.

Measurement and modeling noise priors need to be selected to tune the estimations. They are selected such that the residuals for nominal data are zero-mean Gaussian, e.g., see [98].

[98] Åkesson et al., "A generalized autocovariance least-squares method for Kalman filter tuning"

The generated residual is a Gaussian distributed continuous random variable with an estimated mean and covariance. When evaluating the residual, the covariance estimation may or may not be taken into consideration. Both possibilities are explored in later sections of this chapter.

Applied to the Cavity Model

Since the cavity system equations provide two outputs, the inferred nominal residuals are two-dimensional Gaussian distributions. Given that two detuning models were proposed for the cavity system, two versions of the unscented Kalman filter residuals are considered in this thesis. The first residual is the one generated using the first-order detuning model, i.e.,

$$R_k^{\mathrm{fo}} \sim \mathcal{N}\left(r_k^{\mathrm{fo}}; \mathrm{m}_{r_k^{\mathrm{fo}}}, \mathrm{V}_{r_k^{\mathrm{fo}}}\right) \tag{4.8}$$

where $\mathrm{m}_{r_k^{\mathrm{fo}}} \in \mathbb{R}^{2,1}$ and $\mathrm{V}_{r_k^{\mathrm{fo}}} \in \mathbb{R}^{2,2}$ are the estimated mean vector and covariance matrix, respectively generated for each sample k in the RF-pulse and

$$r_k^{\mathrm{fo}} = \begin{bmatrix} r_k^{I,\mathrm{fo}} \\ r_k^{Q,\mathrm{fo}} \end{bmatrix} \tag{4.9}$$

is the two-dimensional residual for I and Q components of the RF-signals, respectively. Equivalently, the residuals generated using the fixed-detuning model is denoted by

$$R_k^{\text{ft}} \sim \mathcal{N}\left(r_k^{\text{ft}}; \mathrm{m}_{r_k^{\text{ft}}}, \mathrm{V}_{r_k^{\text{ft}}}\right) \tag{4.10}$$

where again $\mathrm{m}_{r_k^{\text{ft}}} \in \mathbb{R}^{2,1}$ and $\mathrm{V}_{r_k^{\text{ft}}} \in \mathbb{R}^{2,2}$ are the mean vector and covariance matrix estimations generated and

$$\text{and } r_k^{\text{ft}} = \begin{bmatrix} r_k^{I,\text{ft}} \\ r_k^{Q,\text{ft}} \end{bmatrix} \tag{4.11}$$

are the residuals of I and Q components of the RF-signals when using the fixed-detuning model.

4.2 DETERMINISTIC RESIDUAL GENERATION

In classical model-based fault detection, three methods have been used extensively: The parity space, the observer-based and the parameter-based approach.

In Chapter 2, these methods were introduced and the following thus only defines their generation as specified for the cavity system. The parameter-based residual is motivated by the fact that the values of the cavity parameters over the duration of the pulse is already part of a FPGA component as described in [38]. Deviations of the parameters can thus be obtained, when the online estimations are compared to nominal expected parameter values of each cavity. Furthermore the parameter space, consisting of detuning f^{Δ} and half-bandwidth $f^{\dagger}$ may be more sensitive towards anomalies or faults affecting the cavity parameter properties, than both parity space and unscented Kalman filter residuals.

[38] Rybaniec et al., "Real-time estimation of superconducting cavities parameters"

This section presents two methods that define residuals based on the cavity model described in Chapter 2.2.4. The methods are deterministic since uncertainty in measurement and modeling is neglected.

4.2.1 *Parameter-Based Residual*

The SRF cavity is described by the nonlinear system model 2.2.4, which includes the RF input-dependent detuning f_k^{Δ} and the half-bandwidth $f_k^{\dagger}$. Since the cavities are currently operated in pulsed mode, the detuning is time-varying over the pulse.[5] The variations over the pulse for both parameters should be highly repetitive for nominal behaviors.

[5] When the cavities are operated in CW, detuning and half-bandwidth are assumed constant, and an online estimation of the parameters can be compared to constant values.

Nominal parameter changes over the RF-pulses for each cavity are determined from nominal data using the detuning (2.17) and half-bandwidth (2.18) computation. In nominal operation, where the same inputs are used from pulse to pulse, the half-bandwidth values and the detuning at sample k should—apart from unavoidable uncertainties—be constant. For each cavity i_c, data stream vectors containing the values of all n_s samples in an RF-pulse for the nominal detuning $\mathbf{f}^{\Delta} = \{f_{k=1}^{\Delta}, ..., f_{k=n_s}^{\Delta}\}$ and nominal half-bandwidth $\mathbf{f}^{\dagger} = \{f_{k=1}^{\dagger}, ..., f_{k=n_s}^{\dagger}\}$ for each cavity with data from the nominal set $\boldsymbol{\mathcal{M}}$ or $\boldsymbol{\mathcal{M}}^{pz}$.

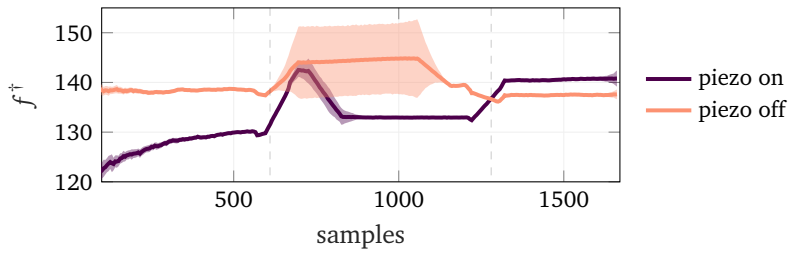


FIGURE 4.3: Half-bandwidth over an RF-pulse with and without piezo-electric detuning control. In the flattop the beam-loading changes the half-bandwidth. Without a beam, the half-bandwidth should not change over the pulse.

The sample distribution of each element in $\mathbf{f}^{\Delta}$ or $\mathbf{f}^{\dagger}$ can be described by a Gaussian distributed random variable for each cavity i_c and sample k, i.e.,

$$F_k^{\dagger} \sim p(f_k^{\dagger}|\boldsymbol{\mathcal{M}}_{i_c,:,k}) = \quad \mathcal{N}\left(f_k^{\dagger}; \mathrm{m}_{F_k^{\dagger}}, \mathrm{V}_{F_k^{\dagger}}\right), \tag{4.12}$$

$$F_k^{\Delta} \sim p(f_k^{\Delta}|\boldsymbol{\mathcal{M}}_{i_c,:,k}) = \quad \mathcal{N}\left(f_k^{\Delta}; \mathrm{m}_{F_k^{\Delta}}, \mathrm{V}_{F_k^{\Delta}}\right), \tag{4.13}$$

where $\mathrm{m}_{F_k^{\dagger}}, \mathrm{V}_{F_k^{\dagger}} \in \mathbb{R}$ and $\mathrm{m}_{F_k^{\Delta}}, \mathrm{V}_{F_k^{\Delta}} \in \mathbb{R}$ are the mean and variance of the half-bandwidth and detuning estimations, respectively. An exemplary evolution of the half-bandwidth over the RF-pulse of a cavity is depicted in Figure 4.3, showing mean and standard deviation for each sample. Detuning over the pulse was already shown in Figure 3.3.

Using the mean values of the nominal distributions, the following zero-mean residuals for nominal RF-pulse behavior can be defined

$$r_k^{\Delta} = \hat{f}_k^{\Delta} - \mathrm{m}_{F_k^{\Delta}} \tag{4.14}$$

$$r_k^{\dagger} = \hat{f}_k^{\dagger} - \mathrm{m}_{F_k^{\dagger}}, \tag{4.15}$$

where $\hat{f}_k^{\Delta}$, $\hat{f}_k^{\dagger}$ are the detuning and half-bandwidth values of the current RF-pulse at sample k.
Both (4.14) and (4.15) are used to define the parameter-based residual as a two-dimensional vector in each sample, i.e.,

$$r_k^{\mathrm{PB}} = [r_k^{\Delta}, r_k^{\dagger}]^T, \tag{4.16}$$

where $r_k^{\mathrm{PB}} \in \mathbb{R}^{1\times 2}$ for each evaluated sample at instant k.

4.2.2 *Nonlinear Parity Space Method*

The parity space method was introduced in Chapter 2. For the cavity system, consider the discretized electromagnetic cavity model, first introduced in (3.7) and repeated here for convenience:

$$x_{k+1}^{I} = -a_0 x_k^{I} - 2\pi T_s f_k^{\Delta} x_k^{II} + b_0 u_k^{I} \tag{4.17}$$

$$x_{k+1}^{II} = \ 2\pi T_s f_k^{\Delta} x_k^{I} - a_0 x_k^{II} + b_0 u_k^{II} \tag{4.18}$$

$$\begin{aligned} y_k^{I} &= x_k^{I}, \\ y_k^{II} &= x_k^{II}, \end{aligned} \tag{4.19}$$

where a_0, b_0 are constant parameters related to the nominal half-bandwidth, see (3.7) and u_k^{I}, u_k^{II} and y_k^{I}, y_k^{II} are the I and Q input and output measurements respectively.

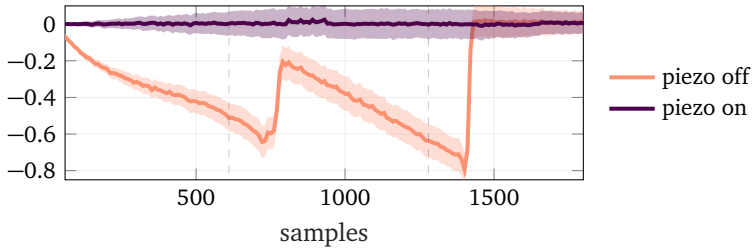


FIGURE 4.4: Residuals over RF-pulse generated from the set of nominal RF-pulses $\mathcal{M}, \mathcal{M}^{pz}$ as defined in (4.22). The mean values are deducted to obtain zero-mean residuals over the pulse.

Two (redundant) expressions of the detuning f_k^{Δ} can be defined, by solving both equations (4.17) and (4.18) for the detuning, which yields

$$\left(f_k^{\Delta}\right)^{*} = \frac{-y_{k+1}^{I} - a_0 y_k^{I} + b_0 u_k^{I}}{2\pi T_s y_k^{II}}, \tag{4.20}$$

$$\left(f_k^{\Delta}\right)^{**} = \frac{y_{k+1}^{II} + a_0 y_k^{II} - b_0 u_k^{II}}{2\pi T_s y_k^{I}}, \tag{4.21}$$

where $\left(f_k^{\Delta}\right)^{*}$ is the detuning derived from (4.17) and $\left(f_k^{\Delta}\right)^{**}$ is the detuning derived from (4.18). The nonlinear parity space residual is then defined as $r_k = \left(f_k^{\Delta}\right)^{*} - \left(f_k^{\Delta}\right)^{**}$. Reformulating with a common denominator, leads to

$$r_k^{ps'} = \left(f_k^{\Delta}\right)^{*} y_k^{I} - \left(f_k^{\Delta}\right)^{**} y_k^{II}. \tag{4.22}$$

For the fault-free, noise-free case, i.e., only when model uncertainty is non-existent and the measurements are noise-free $r_k^{ps'} = 0$.

In reality, the parity space residuals generated from the nominal RF-pulses $\mathcal{M}, \mathcal{M}^{pz}$ have some deterministic divergences from zero, where the model does not accurately represent the system behavior. Figure 4.4 depicts the deterministic divergences from zero, showing both the mean and the variance bounds over the RF-pulse. To account for the inevitable uncertainties in both measurement and model, the residual generated can be expressed by

$$r_k^{ps'} = r_k^{ps} + \mathcal{N}\left(r_k^{ps}; \mathrm{m}_{r_k^{ps}}, \mathrm{V}_{r_k^{ps}}\right), \tag{4.23}$$

$$r_k^{ps} = r_k^{ps'} - \mathcal{N}\left(r_k^{ps}; \mathrm{m}_{r_k^{ps}}, \mathrm{V}_{r_k^{ps}}\right), \tag{4.24}$$

where $\mathrm{m}_{r_k^{ps}} \in \mathbb{R}$ is determined from the set of nominal RF-pulses $\mathcal{M}, \mathcal{M}^{pz}$ for each cavity i_c. When neglecting the variance and assuming $\mathrm{V}_{r_k^{ps}} = 0$, i.e., the parity space residual proposed may be formulated as

$$r_k^{ps} = r_k^{ps'} - \mathrm{m}_{r_k^{ps}}, \tag{4.25}$$

for each sample k.

Figure 4.5 summarizes the residual generation methods, with respect to their modeling assumptions.

In the following section a likelihood ratio based evaluation method for both the deterministic and probabilistic residuals is described in the factor graph framework. To emphasize the mix and match property of factor graphs, the following evaluation approaches are described independently from the residual generation methods. Instead, some general data stream is considered for which the evaluation methods are used to detect anomalies.

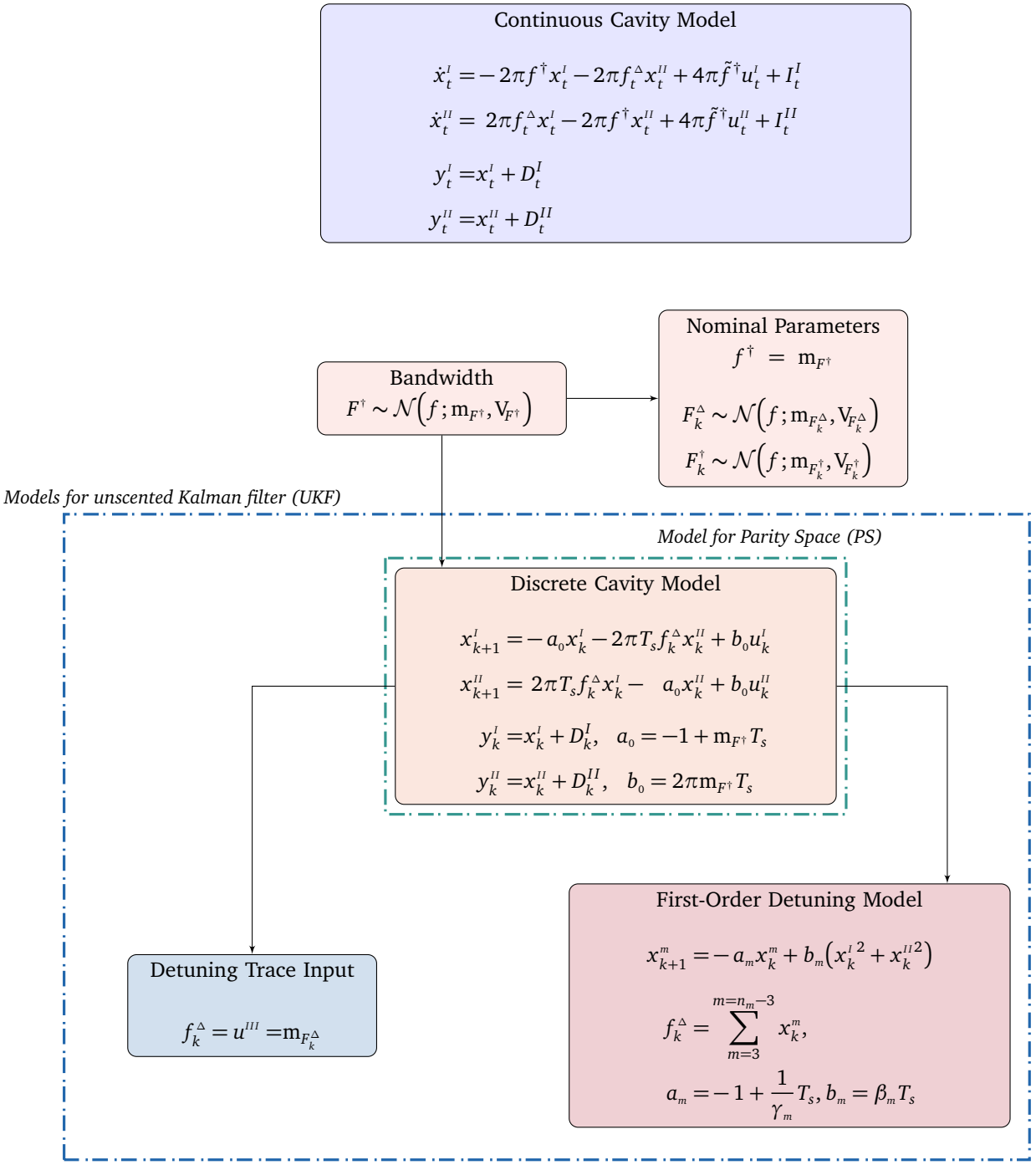


FIGURE 4.5: Nominal cavity models used in this thesis and an overview over the methods that are associated to the respective cavity model aspects.

4.3 EVALUATION WITH MODEL SELECTION/ HYPOTHESIS TESTING

The following model selection method assumes that the latest K, independent observations from a data stream $\mathbf{r} = \{\hat{r}_{j-K}, ..., \hat{r}_j\}$ can best be modeled by one of n_M models $\{h_1, ..., h_{n_M}\}$. The models may be represented by a discrete random variable H, where $P(H = h_i)$ is the probability that the model h_i is the generative model of the data stream. In this context, the terms model selection and hypothesis testing can be used interchangeably. The likelihood of the observed data for one of the models $H = h_i$ is then defined as $p(\mathbf{r}|H = h_i)$. The best model is determined by computing the posterior probabilities over the models, i.e., using Bayes' theorem (2.24)

$$P(H = h_i|\mathbf{r}) \propto p(\mathbf{r}|H = h_i)P(H = h_i), \tag{4.26}$$

where $P(H = h_i|\mathbf{r})$ is the posterior probability that the model $H = h_i$ generated the observed data. Then, the best model is chosen by selecting the model with the highest $P(H = h_i|\mathbf{r})$.
In practice, the prior probability $P(H = h_i)$ can be determined from prior knowledge but is often chosen subjectively or is assumed to be uniform [6]. When a uniform prior is assumed, the likelihoods $p(\mathbf{r}|H = h_i)$ determine the selection result.

[6] Murphy, *Machine Learning: A Probabilistic Perspective*

Latent Model Parameters

Each of the considered models is defined by n_Θ latent model parameters $\theta_i = \{\theta_i^1, ..., \theta_i^{n_\Theta}\}$. When the parameters of each model are unknown, it is common to model them via a continuous random variable Θ with model-dependent distributions $p(\theta_i|H = h_i)$. Then

$$p(\mathbf{r}|H = h_i) = \int_{\theta_i} p(\mathbf{r}|\theta_i, H = h_i)p(\theta_i|H = h_i)\mathrm{d}\theta_i, \tag{4.27}$$

is the marginal likelihood, also called the *model evidence*, obtained by marginalizing over all parameters for each model $H = h_i$. The full posterior distribution is then given by

$$P(H = h_i|\mathbf{r}) \propto \int_{\theta_i} \underbrace{p(\mathbf{r}|\theta_i, H = h_i)p(\theta_i|H = h_i)P(H = h_i)}_{p(\mathbf{r},\theta_i,H\mathrm{d}\theta_i=h_i)}\mathrm{d}\theta_i, \tag{4.28}$$

where each model $H = h_i$ is constituted by their respective parameters θ_i and thus

$$p(\mathbf{r}|\theta_i, H = h_i) = p(\mathbf{r}|\theta_i). \tag{4.29}$$

This straightforward approach to model selection/hypothesis testing can easily be translated into a factor graph, and the model selection can be performed using message passing.

4.3.1 *Factor Graph Representation of Model Selection*

In Figure 4.6 the factor graph of the joint distribution $p(\mathbf{r}, \theta_i, H = h_i)$, defined by the right hand side of (4.28) is given. The graph represents the joint distribution of each generative model for the observations $\mathbf{r}$.

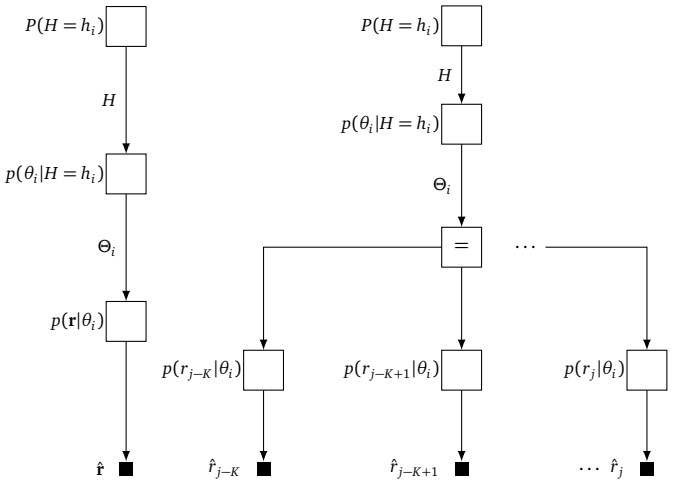


FIGURE 4.6: Factor graph representation of the model selection/hypothesis testing approach. Left: FFG model selection/hypothesis testing distribution. A fully Bayesian approach uses the marginal distribution over H. Right: FFG of likelihood-based change detection algorithms using the maximum likelihood of the marginal on Θ.

The model selection of (4.26) consist of determining the marginal on the edge representing H, i.e.,

$$\overline{\mu}_{H=h_i} = P(H=h_i|\hat{\mathbf{r}}) = P(H=h_i) \underbrace{\int_{\theta_i} p(\theta_i|H=h_i)p(\hat{\mathbf{r}}|\theta_i)\mathrm{d}\theta_i}_{\overleftarrow{\mu}_{H=h_i}}. \tag{4.30}$$

For anomaly detection only two competing models, i.e., a zero and an alternative hypothesis are considered. The zero hypothesis h_0 represents the nominal, whereas the alternative hypothesis h_1 represents the anomalous distribution.

Bayes Factor

When no prior probability distribution of two competing models is available, a uniform prior $P(H=h_0) = P(H=h_1)$ is assumed. In that case the *Bayes factor* is often used [6]. For a binary classification the Bayes factor is defined by

$$\mathrm{BF}_{0,1} = \frac{\overline{\mu}_{H=h_1}}{\overline{\mu}_{H=h_0}} = \frac{\overleftarrow{\mu}_{H=h_1}}{\overleftarrow{\mu}_{H=h_0}}, \tag{4.31}$$

This means that the Bayes factor is determined from the ratio of the messages representing one hypothesis, each. The model selection is thus equivalent to picking the model with the higher marginal likelihood. To include a sense of confidence over the selection, after all one model could be only slightly better than the other, a scale of evidence for the Bayes factor is commonly used [6].

Integrating over the parameters in (4.30) may, however, involve cumbersome calculations. One reason for using likelihoods, which are described in detail in the next paragraph is to circumvent these cumbersome calculations.

Likelihood-Based Tests

The following introduces likelihood-based hypothesis testing in factor graphs, focusing on the likelihood ratio test. Here, this test is treated and understood again through the factor graph model in Figure 4.6. Instead of marginalizing

over the parameter distributions of the models, likelihoods are used and the maximum likelihood of the parameters are determined.

The likelihoods are provided by message passing via the backward message on Θ, i.e.,

$$\overleftarrow{\mu}_{\Theta} = \int_{\mathbf{r}} p(\mathbf{r}|\theta_i)p(\mathbf{r})\mathrm{d}\mathbf{r} = p(\hat{\mathbf{r}}|\theta_i), \tag{4.32}$$

and the maximum likelihood estimates of the parameters are used to compare the models via likelihood ratios

$$\mathrm{LR}_{0,1} = \frac{\max\limits_{\theta_1 \in \Theta_1} \overleftarrow{\mu}_{\Theta}(\theta_i = \theta_1)}{\max\limits_{\theta_0 \in \Theta_0} \overleftarrow{\mu}_{\Theta}(\theta_i = \theta_0)} = \frac{\max\limits_{\theta_1 \in \Theta_1} p(\hat{\mathbf{r}}|\theta_1)}{\max\limits_{\theta_0 \in \Theta_0} p(\hat{\mathbf{r}}|\theta_0)} = \frac{p(\hat{\mathbf{r}}|\hat{\theta}_1)}{p(\hat{\mathbf{r}}|\hat{\theta}_0)}, \tag{4.33}$$

[86] Basseville and Nikiforov, *Detection of abrupt changes: theory and application*

where $\Theta = \Theta_0 \cup \Theta_1$ [86]. Equivalently the log-likelihood ratio which may simplify the calculation is

$$\Lambda = \ln \frac{\max\limits_{\theta_1 \in \Theta_1} \overleftarrow{\mu}_{\Theta}(\theta_i = \theta_1)}{\max\limits_{\theta_0 \in \Theta_0} \overleftarrow{\mu}_{\Theta}(\theta_i = \theta_0)} = \ln \frac{p(\hat{\mathbf{r}}|\hat{\theta}_1)}{p(\hat{\mathbf{r}}|\hat{\theta}_0)}, \tag{4.34}$$

where Λ is the log-likelihood ratio of the set of observations taken from a data stream at $k = j$. The null hypothesis is rejected, if the ratio exceeds a certain threshold ψ_{th} which is chosen according to an application-dependent acceptable probability of false alarms ϵ, as is, e.g., described in [78]. This test is referred to as generalized likelihood ratio test. In the following the generalized likelihood ratio for Gaussian distributed hypotheses is described.

[78] Ding, *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*

4.3.2 *Generalized Likelihood Ratio for Gaussian Distributions*

Again, the data is assumed to be i.i.d., i.e.,

$$p(\mathbf{r}|\theta_i) = \prod_{k=j-K}^{j} p(r_k|\theta_i), \tag{4.35}$$

represented by the factor graph in Figure 4.6, where the K individual measurements of the data stream are connected through equality nodes.

In addition, the data is assumed to be zero-mean Gaussian distributed when nominal, i.e. $\mathrm{m}_R = \theta_0 = 0$, and a Gaussian distribution with mean other than zero in case of an anomaly, i.e., $\mathrm{m}_R = \theta_1 \neq 0$. It is thus assumed, that changes in the residual only have an effect on the mean of the distribution [78], i.e.,

$$p(r_k|\theta_i) = \mathcal{N}\left(r_k; \mathrm{m}_R = \theta_i, \mathrm{V}_R\right), \tag{4.36}$$

where V_R is the covariance of the zero hypothesis.

With known samples $\hat{\mathbf{r}}$, the backward message on Θ is the product of all local likelihoods of each observation branch, i.e.,

$$\overleftarrow{\mu}_{\Theta}(\theta_i) = \prod_{k=j-K}^{j} p(\hat{r}_k|\theta_i). \tag{4.37}$$

The mean value (and therefore the maximum value) for the zero-hypothesis distribution is at $\theta_0 = 0$ whereas the change in mean θ_1 is unknown but

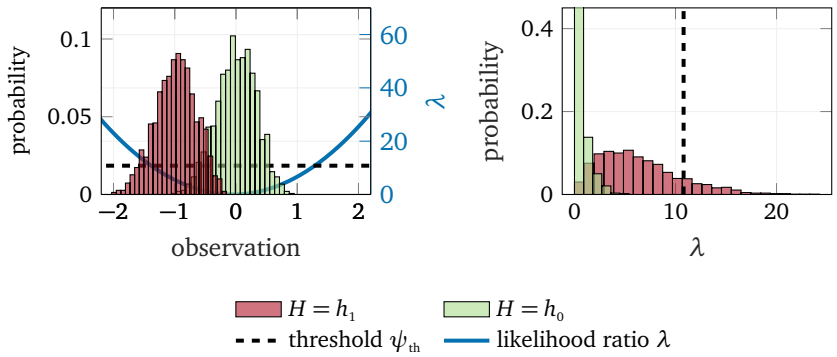


FIGURE 4.7: The general likelihood ratios transform the *moving average* of a Gaussian distributed variable into a χ^2 distributed variable according to the quadratic function (left). Values further away from the Gaussian mean are thus quadratically amplified while values close to the mean are reduced to zero.

can be estimated using the maximum likelihoods. According to (4.34), the log-likelihood ratio is then

$$\lambda_k = \ln \frac{\max\limits_{\theta_1 \in \Theta_1} \overleftarrow{\mu}_\Theta(\theta_i = \theta_1)}{\max\limits_{\theta_0 \in \Theta_0} \overleftarrow{\mu}_\Theta(\theta_i = \theta_0)} = \ln\big(\overleftarrow{\mu}_\Theta(\theta_i = \hat{\theta}_1)\big) - \ln\big(\overleftarrow{\mu}_\Theta(\theta_i = \theta_0)\big), \quad (4.38)$$

where

$$\hat{\theta}_1 = \arg\max_{\theta_1 \in \Theta_1} \ln\big(\overleftarrow{\mu}_\Theta(\theta_i = \theta_1)\big), \quad (4.39)$$

for each evaluation of the latest K observations.

The maximum log-likelihood for θ_1, leads to the log-likelihood ratio

$$\lambda_k = \frac{K}{2}\Big(\frac{1}{K}\sum_{j=k-K}^{k} \hat{r}_j^T\Big) \mathrm{V}_R^{-1} \Big(\frac{1}{K}\sum_{j=k-K}^{k} \hat{r}_j\Big), \quad (4.40)$$

where at each evaluation k, $\lambda_k \in \mathrm{R}_+$ has positive values. The derivation of this result can, e.g., be found in [78]. The likelihood ratios are thus a quadratic function of the last K observations' mean value.

In Figure 4.7 the distributions of the *moving average* over two Gaussian distributions are depicted. The Gaussian with zero-mean represents the zero hypothesis. The quadratic function which determines the likelihood ratios and given by (4.40) is depicted on the left. When the average of the observations $\hat{\mathbf{r}}$ for a given window K are close to zero, the log-likelihood ratios also take up values close to zero. With increased deviation from zero, the log-likelihood ratios increase quadratically. Depending on what probability of false alarm ϵ is chosen to be acceptable, the likelihood ratio values may still be associated to the zero hypothesis. The transformation of the Gaussian distribution results in χ^2 distributions, depicted on the right.

Observations can then be evaluated according to a decision function given by

$$\mathrm{d}_{\mathrm{GLR}}(\hat{\mathbf{r}}) = \begin{cases} 0 \text{ if } \ln \overleftarrow{\mu}_\Theta(\theta_i = \theta_0) > \psi_{\mathrm{th}} \ln \overleftarrow{\mu}_\Theta(\theta_i = \hat{\theta}_1); & h_0 \text{ is chosen,} \\ \\ 1 \text{ if } \ln \overleftarrow{\mu}_\Theta(\theta_i = \theta_0) \leq \psi_{\mathrm{th}} \ln \overleftarrow{\mu}_\Theta(\theta_i = \hat{\theta}_1); & h_1 \text{ is chosen,} \end{cases} \quad (4.41)$$

where ψ_{th} is set according to an acceptable probability of false alarms ϵ. The decision rule in terms of the log-likelihood ratio is thus

$$\mathrm{d}_{\mathrm{GLR}}(\hat{\mathbf{r}}) = \begin{cases} 0 \text{ if } \lambda_k \leq \psi_{\mathrm{th}}; & h_0 \text{ is chosen,} \\ \\ 1 \text{ if } \lambda_k > \psi_{\mathrm{th}}; & h_1 \text{ is chosen.} \end{cases} \quad (4.42)$$

Using either the marginal or maximum likelihoods of (4.34), several other standard model selection criteria, such as the Akaike information criterion can be derived, this relationship is, e.g., described in [6]. A reconciliation between the Bayesian approach of Bayes factors, and the likelihood ratios is the Bayesian information criterion (BIC), which was derived by Efron [99].

[6] Murphy, *Machine Learning: A Probabilistic Perspective*

[99] Efron et al., "Scales of evidence for model selection: Fisher versus Jeffreys"

4.4 EVALUATION WITH GAUSSIAN MIXTURE MODEL

The model selection method described in the previous section assumed little knowledge about the distribution of the residuals in case of anomalous system behavior. In this section, the residual space is assumed to be represented by several Gaussians, i.e., it can be modeled with a GMM. GMMs were introduced in Chapter 2.4.5.

The GMM may be used to classify any number of discernible residual groups, which, in turn, may also be represented by a GMM. For example, when used for a binary classification (nominal/anomalous), one subset of the Gaussian mixtures may represent the residuals' anomalous behavior, whereas all other Gaussian mixtures define the nominal behavior. To simplify the notation, however, the following assumes that each Gaussian mixture component represents one class of fault.

Evaluating data samples with a GMM consists of determining with what probability that sample belongs to each class. Both deterministic and noisy data can be evaluated.

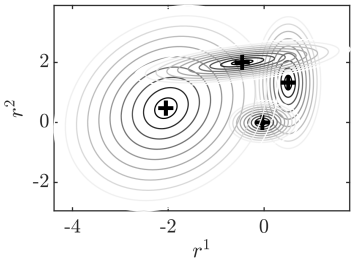


FIGURE 4.8: Example of a two-dimensional Gaussian mixture with four components.

4.4.1 *A Factor Graph Representation for GMM*

It is assumed, that each class of fault is modeled by a Gaussian distribution, i.e.,

$$p(r_k|C=c_i) = \mathcal{N}\Big(r_k; \mathrm{m}_i, \mathrm{V}_i\Big). \tag{4.43}$$

In Figure 4.8, as an example, such a GMM is depicted, where the underlying data is two-dimensional and in total four Gaussian components were included.

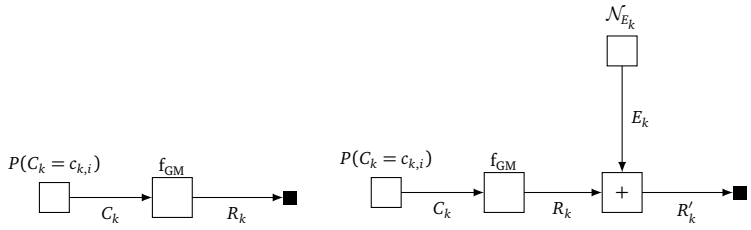
Classifying data with GMMs can be formulated as message passing in factor graphs. To that end, a Gaussian mixture factor node is defined, which represents the joint distribution of the time-dependent classes $C_k = c_{k,i}$, and the data. This node can be used for the evaluation of deterministic data or noisy, Gaussian distributed data with known covariance. A prior on the class probabilities $P(C_k = c_{k,i})$ for each time-step k is included into the factor graph model, denoted by $P(C_k = c_{k,i})$. This prior may, e.g., represent additional information about the class probabilities deduced from other observations.

The Gaussian mixture node describes the joint distribution of the data and the class $p(r_k, C_k = c_{k,i}|\theta)$, and is defined by a list, where each item describes one class, i.e.,

$$\mathrm{f}_{\mathrm{GM}}(r_k, C_k = c_{k,i}) = \mathrm{w}_i \, \mathcal{N}\Big(r_k; \mathrm{m}_i, \mathrm{V}_i\Big), \tag{4.44}$$

with $i \in 1, ..., n_c$. The discrete distribution prior $P(C_k = c_{k,i})$ is defined by weights $\mathrm{w}'_{k,i}$, i.e.,

$$P(C_k = c_{k,i}) = \mathrm{w}'_{k,i} \quad \text{with } i \in 1, ..., n_c. \tag{4.45}$$



(a) Deterministic sample residual. (b) Gaussian residual.

FIGURE 4.9: Residual evaluation with GMM in factor graphs.

see Figure 4.9. Then, the forward message on $\vec{\mu}_{R_k}$ is a Gaussian mixture message defined as

$$\vec{\mu}_{R_k} = \sum_{i=1}^{n_c} \mathrm{w}'_{k,i} \mathrm{w}_i \mathcal{N}\left(r_k; \mathrm{m}_i, \mathrm{V}_i\right). \tag{4.46}$$

The evaluation of data from sample time k with a GMM consists of determining the marginal distribution over the classes $C_k = c_{k,i}$, which is obtained via message passing with $\bar{\mu}_{C_k} = \vec{\mu}_{C_k} \overleftarrow{\mu}_{C_k}$. This marginal distribution depends on whether noisy or deterministic data samples are evaluated. The next two paragraphs distinguish between noisy and deterministic data samples.

- FOR NOISY DATA, R_k can be modeled as

$$R'_k = E_k + R_k, \quad \text{where } E_k \sim \mathcal{N}\left(e_k; 0, \vec{\mathrm{V}}_{E_k}\right), \tag{4.47}$$

the marginal over the classes C_k is

$$\begin{aligned} \overleftarrow{\mu}_{C_k}(C_k = c_{k,i}) &= \int_{r_k} \mathrm{f}_{\mathrm{GM}}(r_k, C_k = c_{k,i})\, \overleftarrow{\mu}_{R_k}(r_k) \mathrm{d}r_k \\ &= \mathrm{w}_i \mathcal{N}\left(\overleftarrow{\mathrm{m}}_{R_k}; \mathrm{m}_i, \mathrm{V}_i + \overleftarrow{\mathrm{V}}_{R_k}\right) \end{aligned} \tag{4.48}$$

with

$$\overleftarrow{\mu}_{R_k}(r_k) \propto p(r_k | r'_k) = \mathcal{N}\left(r_k; r'_k, \vec{\mathrm{V}}_{E_k}\right), \tag{4.49}$$

and $\overleftarrow{\mathrm{m}}_{R_k} = r'_k$.

Note that (4.48) is no longer a Gaussian distribution but the scalar value of the non-normalized posterior class probability for class $c_{k,i}$. The forward message $\vec{\mu}_{C_k}$ is the discrete distribution prior, i.e., $\vec{\mu}_{C_k} = P(C_k = c_{k,i})$ of (4.45). The product of the forward and backward messages on C_k is then given by

$$\rho_{k,i} \propto \vec{\mu}_{C_k} \overleftarrow{\mu}_{C_k} = \mathrm{w}_i \mathrm{w}'_{k,i} \mathcal{N}\left(\mathrm{m}_{R_k}; \mathrm{m}_i, \mathrm{V}_i + \vec{\mathrm{V}}_{E_k}\right), \tag{4.50}$$

which is consistent with the (non-normalized) *responsibilities* $\rho_{k,i}$ used for the classification of noisy data as derived for example in [100].

[100] Kolossa et al., "Independent Component Analysis and Time-Frequency Masking for Speech Recognition in Multitalker Conditions"

- FOR DETERMINISTIC DATA, the variance is $\mathrm{V}_{E_k} = 0$, i.e.,

$$\rho_{k,i} \propto \vec{\mu}_{C_k} \overleftarrow{\mu}_{C_k} = \mathrm{w}_i \mathrm{w}'_{k,i} \mathcal{N}\left(\overleftarrow{\mathrm{m}}_{R_k}; \mathrm{m}_i, \mathrm{V}_i\right), \tag{4.51}$$

which corresponds to the results for deterministic classifications as derived, e.g., in [101].

[101] Bishop, *Pattern Recognition and Machine Learning*

Example 4. Consider the 2D-GMM build of $n_c = 4$ Gaussian components, with the prior class probabilities $w_i = 0.25$, shown in Figure 4.10. The evaluation of an observation sample, indicated in red at $r_k = [-1, 1]^T$ has the following responsibilities:

$$\rho_{k,1} = 0.0001, \quad \rho_{k,3} = 0.0006,$$
$$\rho_{k,2} = 0.9817, \quad \rho_{k,4} = 0.0177.$$

If instead of the deterministic observation sample, the uncertainty of the observation is known and described by a Gaussian, see Figure 4.11 then the evaluation yields

$$\rho_{k,1} = 0.001, \quad \rho_{k,3} = 0.5397,$$
$$\rho_{k,2} = 0.0218, \quad \rho_{k,4} = 0.4372.$$

The deterministic evaluation thus assigns the sample to the Gaussian component 2, whereas the noisy data sample with a mean equal to the deterministic sample is assigned to the Gaussian component 3.

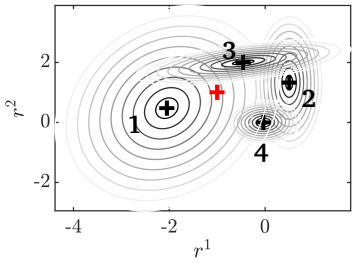


FIGURE 4.10: Deterministic sample evaluation.

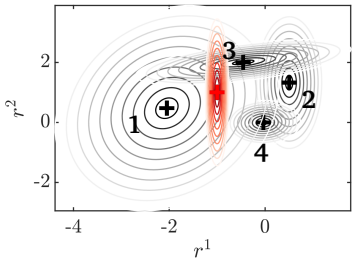


FIGURE 4.11: Noisy sample evaluation.

GMMs are commonly determined through a training procedure from unclassified data. Both training and evaluation of data using a GMM are described in the following sections via message passing on factor graphs. In particular, it is shown how noisy observations can be incorporated without changing the factor graph definitions.

4.5 PARAMETER TRAINING FOR GAUSSIAN MIXTURE MODELS

The GMM model parameters are estimated in a training process from sets of unlabeled data. One common approach to determine the parameters of a GMM is to use the EM algorithm. Defining appropriate message passing schemes in factor graphs to express parameter estimation using the EM algorithm has already been described by [57]. Chapter 2.4.5 gave a short introduction to the EM algorithm and its representation within the factor graph framework. Now, the message passing details are presented for learning a GMM from noisy measurements, i.e., when information about the samples' uncertainty is available for the training process. This training has already been described independently from the factor graph framework, e.g., by [95].

This section is structured in the following way: First, the messages proposed by [57] are determined for the noisy GMM training. The results of the updates, which were derived independently from the factor graph framework, are shortly recounted. Finally, these updates are recovered using Gaussian message passing.

[57] Dauwels et al., "Expectation Maximization as Message Passing"

[95] Ozerov et al., "Uncertainty-Based Learning of Gaussian Mixture Models from Noisy Data"

Expectation-Maximization with Noisy Residuals

In Chapter 2.4.5 the EM algorithm was described in two steps. The first step was to define the expectation concerning the probability density function over the latent variables given the current GMM parameter update $\hat{\theta}^j$. In the second step—the maximization step—this expectation function is maximized

with respect to the unknown GMM parameters, which then result in the updated parameters $\hat{\theta}^{j+1}$. The following describes these steps and defines the messages with which the EM algorithm can be performed in the factor graph framework.

For the case that the data set consist of noisy observations, both the class-memberships of the data *and* the underlying, noise-free observations are unknown. Then, the training data set $\boldsymbol{\mathcal{X}}$ consists of K noisy observations r'_k and an estimation of the noise E_k, as defined in (4.47), i.e.,

$$\boldsymbol{\mathcal{X}} = \left\{r'_k, E_k\right\}_{k=0}^{n_K}. \tag{4.52}$$

Since neither the noise-free data R_k, nor the class memberships C_k can be measured, both are considered latent variables, i.e.,

$$\boldsymbol{\mathcal{Z}} = \{C_k, R_k\}_{k=0}^{n_K}. \tag{4.53}$$

The EM message (2.44) representing the expectation step described in Chapter 2.4.6 is then defined using

$$p_b = p(\mathbf{C} = \mathbf{c}, \mathbf{r}, |\mathbf{r}', \hat{\theta}^j), \tag{4.54}$$

which is the joint distribution of the complete training sets which includes the latent class, i.e., $\mathbf{c} = \{c_{1,i_1}, ..., c_{n_K,i_{n_K}}\}$, and $\mathbf{r} = \{r_1, ..., r_{n_K}\}$, and the observations $\mathbf{r}' = \{r'_1, ..., r'_{n_K}\}$. Then,

$$f_b = p(\mathbf{C} = \mathbf{c}, \mathbf{r}, \mathbf{r}'|\theta), \tag{4.55}$$

represents the joint distribution over both latent and observed data. The forward message in (2.44) was defined as

$$\vec{\mu}_{\text{EM}}\left(\hat{\theta}^j, \theta\right) = e^{\eta(\hat{\theta}^j, \theta)} \tag{4.56}$$

and here leads to

$$\eta(\hat{\theta}^j, \theta) = \mathcal{Q}\big(\theta|\hat{\theta}^j\big) = \mathbf{E}_{p(\mathbf{C}=\mathbf{c},\mathbf{r},|\mathbf{r}',\hat{\theta}^j)}\left[\ \ln p(\mathbf{C} = \mathbf{c}, \mathbf{r}, \mathbf{r}'|\theta)\right]. \tag{4.57}$$

When the joint distribution can be described by the product of "local" distributions, local messages can be used which contribute to the E-step. Assuming that

$$\begin{aligned}\ln\big(p(\mathbf{C} = \mathbf{c}, \mathbf{r}, \mathbf{r}'|\theta)\big) &= \ln\Big(\prod_{k=1}^{n_K} p(C_k = c_{k,i}, r_k, r'_k|\theta)\Big) \\ &= \sum_{k=1}^{n_K} \ln\big(p(C_k = c_{k,i}, r_k, r'_k|\theta)\big),\end{aligned} \tag{4.58}$$

and the fact that the summation over the GM components can be pulled infront of the expectation, see [52], the function $\mathcal{Q}\big(\theta|\hat{\theta}^j\big)$ necessary for the E-step can indeed be represented by the summation of local messages, i.e.,

$$\mathcal{Q}\big(\theta|\hat{\theta}^j\big) = \sum_{k=1}^{n_K} \underbrace{\sum_{i=1}^{n_c} \mathbf{E}_{p(C_k=c_{k,i},r_k|r'_k,\hat{\theta}^j)}\left[\ \ln p(C_k = c_{k,i}, r_k, r'_k|\theta_i)\right]}_{\eta_k}. \tag{4.59}$$

The factor graph in Figure 4.12 describes the local distributions and their connections through the node f_a. The factors of f_a connect the local

[52] Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models"

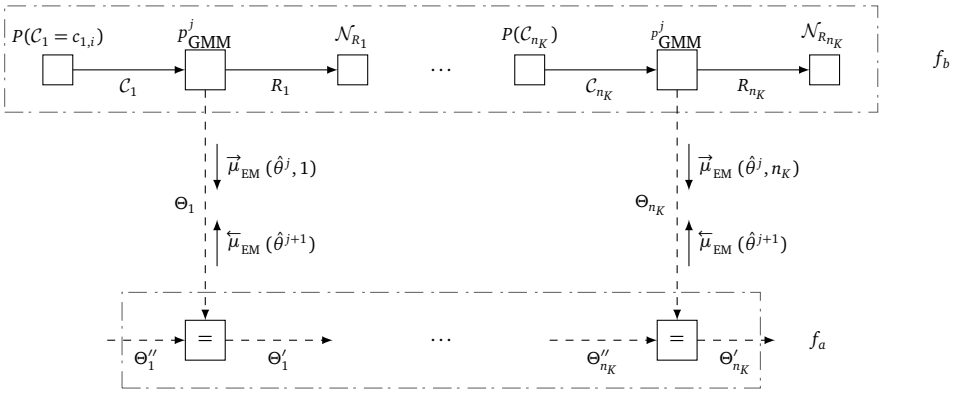


FIGURE 4.12: Factor graph for the training of the GMM. In this thesis the priors $\mathcal{N}_{R_k}$ stem from the unscented Kalman filter residual generation.

messages with equality nodes. In order to use the same Gaussian message passing rules, i.e., the tabulated definitions of the equality node, the forward EM message (2.44) needs to be reformulated. Then, the updates of the noisy GMM training are obtained by the following message passing schedule. A forward, backward pass through the equality nodes is done using the local messages. For the backward pass, an uninformative prior is used at the final equality node Θ'_{n_K}. The message containing the updated GMM parameters $\hat{\theta}^{j+1}$ is defined by the marginal on any of the parameter branches, i.e., $\overleftarrow{\mu}_{\text{EM}}(\hat{\theta}^{j+1}) = \bar{\mu}_{\hat{\theta}^{j+1}}$. The backward messages $\overleftarrow{\mu}_{\text{EM}}$ is reformulated such, that a forward/backward pass through the equality nodes using the rules of Table 2.1 results in the updates needed to obtain $\hat{\theta}^{j+1}$.

4.5.1 *GMM Training with Noisy Data Using EM*

The EM parameter estimation for GMM models with noisy data has previously been described independently from the factor graph framework, e.g., by Ozerov et.al. [95]. The maximization of $\mathcal{Q}\big(\theta|\hat{\theta}^j\big)$ with respect to θ, results in following parameter updates:

[95] Ozerov et al., "Uncertainty-Based Learning of Gaussian Mixture Models from Noisy Data"

$$
\begin{aligned}
\hat{\mathrm{w}}_i^{j+1} &= \frac{1}{n_K}\sum_{k=1}^{n_K}\rho_{k,i}^j, \\
\hat{\mathrm{m}}_i^{j+1} &= \frac{1}{\sum_{k=1}^{n_K}\rho_{k,i}^j}\sum_{k=1}^{n_K}\rho_{k,i}^j\bar{\mathrm{m}}_{R_{k,i}}^j \\
\hat{\mathrm{V}}_i^{j+1} &= \frac{1}{\sum_{k=1}^{n_K}\rho_{k,i}^j}\sum_{k=1}^{n_K}\rho_{k,i}^j\,\Omega_{k,i}^j - \hat{\mathrm{m}}_i^{j+1}\hat{\mathrm{m}}_i^{j+1^T} \\
&\text{with } \Omega_{k,i}^j = \bar{\mathrm{V}}_{R_{k,i}}^j + \bar{\mathrm{m}}_{R_{k,i}}^j\bar{\mathrm{m}}_{R_{k,i}}^{j^T},
\end{aligned}
\tag{4.60}
$$

where $\hat{\mathrm{w}}_i^{j+1}, \hat{\mathrm{m}}_i^{j+1}$ and $\hat{\mathrm{V}}_i^{j+1}$ are the updates of the weights, the mean and the covariance matrix for each $i = 1, ..., n_c$. They are dependent on $\bar{\mathrm{m}}_{R_{k,i}}^j, \bar{\mathrm{V}}_{R_{k,i}}^j$, which are the mean and covariance of the unobserved noise-free marginal distribution, respectively. A derivation of the "classical" approach without using factor graphs is given in Appendix 7.2.

In the following section, these updates are recovered using a message passing schedule for which the forward message $\vec{\mu}_{\text{EM}}$ is reformulated as a Gaussian.

4.5.2 *EM Messages for GMM Training in Factor Graphs*

The forward message $\vec{\mu}_{\text{EM}}$ is reformulated as a Gaussian, making it possible to use the tabulated updates when passing through the equality nodes of f_a. It is shown that such a Gaussian—after forward and backward propagation through equality nodes—carries results with which the updates of (4.60) can be obtained. Rewriting the expectation function as a Gaussian distribution is a common approach for message passing based EM. It allows the use of the tabulated rules, which after all are defined only for Gaussian distributed messages. Up until now, the forward message was defined by the expectation given in (4.57), which can be reformulated into

$$\mathcal{Q}\big(\theta|\theta^j\big) = \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\rho_{k,i}\Big(\ \ln \mathrm{w}_i + \ln 2\pi|\mathrm{V}_i| + \frac{1}{2}\mathrm{trace}\Big(\mathrm{V}_i^{-1}(\bar{\mathrm{V}}_{R_{k,i}} + \bar{\mathrm{m}}_{R_{k,i}}\bar{\mathrm{m}}_{R_{k,i}}^T) - 2\mathrm{V}_i^{-1}\bar{\mathrm{m}}_{R_{k,i}}\mathrm{m}_i^T + \mathrm{V}_i^{-1}\mathrm{m}_i\mathrm{m}_i^T\Big)\Big). \tag{4.61}$$

This formulation shows that the expectation function is dependent on the responsibilities $\rho_{k,i}$, the marginal mean $\bar{\mathrm{m}}_{R_{k,i}}$ and the covariance $\bar{\mathrm{V}}_{R_{k,i}}$ of R_k. A derivation for this can be found in Appendix 7.2. For each sample of training data $k = 1, ..., n_K$ the marginal distributions of $C_k = c_{k,i}$ and R_k, can be determined using message passing, i.e.,

$$\bar{\mu}_{C_k} = \vec{\mu}_{C_k}\overleftarrow{\mu}_{C_k} = \rho_{k,i} \propto \mathrm{w}_i, \mathcal{N}\Big(\overleftarrow{\mathrm{m}}_{R_k}; \mathrm{m}_i, \overleftarrow{\mathrm{V}}_{R_k} + \mathrm{V}_i\Big), \tag{4.62}$$

$$\bar{\mu}_{R_k} = \vec{\mu}_{R_k}\overleftarrow{\mu}_{R_k} = p(r_k|r_k', \hat{\theta}^j) = \mathcal{N}\Big(r_k; \bar{\mathrm{m}}_{R_{k,i}}, \bar{\mathrm{V}}_{R_{k,i}}\Big), \tag{4.63}$$

where $\overleftarrow{\mathrm{V}}_{R_k} = \vec{\mathrm{V}}_{E_k}$, $\overleftarrow{\mathrm{m}}_{R_k} = r_k'$ and the priors $P(C_k = c_{k,i})$ are for now assumed to be uniform but can be used to incorporate additional information from other subsystems. To compute the marginals mean and covariance for $\bar{\mu}_{R_k}$ the tabulated definitions of the equality node can be directly applied because computing the marginal of two Gaussian messages is equal to combining them through the equality node. The mean and covariance marginals are thus defined as

$$\bar{\mathrm{m}}_{R_{k,i}} = (\vec{\mathrm{V}}_i^{-1} + \overleftarrow{\mathrm{V}}_{R_k}^{-1})^{-1}(\vec{\mathrm{V}}_i^{-1}\vec{\mathrm{m}}_i + \overleftarrow{\mathrm{V}}_{R_k}^{-1}\overleftarrow{\mathrm{m}}_{R_k}) \tag{4.64}$$

and

$$\bar{\mathrm{V}}_{R_{k,i}} = (\vec{\mathrm{V}}_i^{-1} + \overleftarrow{\mathrm{V}}_{R_k}^{-1})^{-1}. \tag{4.65}$$

With the marginals computed, the local forward message is reformulated as a Gaussian distribution with

$$\vec{\mu}_{\text{EM}}(\hat{\theta}_k^j) = \mathcal{N}\Big(\hat{\theta}; \mathrm{m}_{\hat{\theta}^j}, \mathrm{V}_{\hat{\theta}^j}\Big). \tag{4.66}$$

Its mean $\mathrm{m}_{\hat{\theta}^j}$ is composed of the marginal mean and covariance over R_k, see (4.63), s.t.,

$$\mathrm{m}_{\hat{\theta}_k^j} = \Big[\bar{\mathrm{m}}_{R_{k,1}}^j, \mathrm{rvect}(\Omega_{k,1}^j), \ldots, \bar{\mathrm{m}}_{R_{k,n_c}}^j, \mathrm{rvect}(\Omega_{k,n_c}^j)\Big]^T$$
$$\text{with} \tag{4.67}$$
$$\Omega_{k,i}^j = \bar{\mathrm{V}}_{R_{k,i}}^j + \bar{\mathrm{m}}_{R_{k,i}}^j \bar{\mathrm{m}}_{R_{k,i}}^{j^T}$$

where $\mathrm{m}_{\hat{\theta}_k^j} \in \mathbb{R}^{(n \cdot n_c \cdot (1+n))}$, n is the dimension of the data sample, and rvect rewrites the entries of a matrix as a row vector, e.g., considering some matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \text{ then } \quad \mathrm{rvect}(A) = [a_{11}, a_{12}, a_{21}, a_{22}]. \tag{4.68}$$

The covariance matrix $\mathrm{V}_{\hat{\theta}_k^j}$ is a diagonal matrix and its entries are the responsibilities $\rho_{k,i}^j$, i.e.,

$$\mathrm{V}_{\hat{\theta}_k^j} = \begin{bmatrix} \rho_{k,1}^j & 0 & \dots & 0 \\ 0 & \rho_{k,2}^j & & \\ \vdots & & \ddots & \\ 0 & & & \rho_{k,n_c}^j \end{bmatrix}^{-1} \otimes \mathbf{I}, \tag{4.69}$$

where $\mathbf{I} \in \mathbb{R}^{n(1+n)}$, and $\otimes$ denotes the Kroneker product.

Updates for the GMM parameters (M-step) are obtained by connecting the local expectation messages of (4.66) on the factor graph via equality nodes.

With a forward propagation followed by a backward propagation through these nodes, the message

$$\overleftarrow{\mu}_{\mathrm{EM}}(\hat{\theta}^{j+1}) = \mathcal{N}(\hat{\theta}; \mathrm{m}_{\hat{\theta}^{j+1}}, \mathrm{V}_{\hat{\theta}^{j+1}}) \tag{4.70}$$

with

$$\mathrm{m}_{\hat{\theta}^{j+1}} = \left[\tilde{\hat{\mathrm{m}}}_1^{j+1}, \dots, \tilde{\hat{\mathrm{m}}}_{n_c}^{j+1}, \ \mathrm{rvect}(\tilde{\hat{\mathrm{V}}}_1^{j+1}), \dots, \mathrm{rvect}(\tilde{\hat{\mathrm{V}}}_{n_c}^{j+1})\right]^T \tag{4.71}$$

and

$$\mathrm{V}_{\hat{\theta}^{j+1}} = \begin{bmatrix} \frac{1}{\sum_{k=1}^{n_K} \rho_{k,1}^j} & 0 & \dots & 0 \\ 0 & \frac{1}{\sum_{k=1}^{n_K} \rho_{k,2}^j} & & \\ \vdots & & \ddots & \\ 0 & & & \frac{1}{\sum_{k=1}^{n_K} \rho_{k,n_c}^j} \end{bmatrix} \otimes \mathbf{I}, \tag{4.72}$$

is obtained, which contain the updates of (4.60), i.e.,

$$\begin{aligned} \tilde{\hat{\mathrm{w}}}_i^{j+1} &= \sum_{k=1}^{n_K} \rho_{k,i}^j, \\ \tilde{\hat{\mathrm{m}}}_i^{j+1} &= \frac{1}{\sum_{k=1}^{n_K} \rho_{k,i}^j} \sum_{k=1}^{n_K} \rho_{k,i}^j \bar{\mathrm{m}}_{\mathrm{R}_{k,i}}^j, \\ \tilde{\hat{\mathrm{V}}}_i^{j+1} = \frac{1}{\sum_{k=1}^{n_K} \rho_{k,i}^j} \sum_{k=1}^{n_K} \rho_{k,i}^j \, \Omega_{k,i}^j &= \frac{1}{\sum_{k=1}^{n_K} \rho_{k,i}^j} \sum_{k=1}^{n_K} \rho_{k,i}^j (\bar{\mathrm{V}}_{\mathrm{R}_{k,i}}^j + \bar{\mathrm{m}}_{\mathrm{R}_{k,i}}^j \bar{\mathrm{m}}_{\mathrm{R}_{k,i}}^{j^T}). \end{aligned} \tag{4.73}$$

Thus, the second order non-central moment $\tilde{\hat{\mathrm{V}}}_i^{j+1}$ is obtained instead of the covariance $\hat{\mathrm{V}}_i^{j+1}$ as it was defined in (4.60). From these the updates given in (4.60) are obtained by normalizing $\tilde{\hat{\mathrm{w}}}_i^{j+1}$, such that $\sum_{i=1}^{n_c} \tilde{\hat{\mathrm{w}}}_i^{j+1} = 1$ and by computing $\hat{\mathrm{V}}_i^{j+1} = \tilde{\hat{\mathrm{V}}}_i^{j+1} - \tilde{\hat{\mathrm{m}}}_i^{j+1} \tilde{\hat{\mathrm{m}}}_i^{j+1^T}$. The straight forward derivation of this result uses the tabulated update rules of the equality node recounted in Table 2.1 and are given in detail in the Appendix 7.2.

4.6 COMBINING RESIDUAL GENERATION AND EVALUATION

The probabilistic residual generation was represented in factor graphs in Figure 4.2. A complete factor graph including residual generation and evaluation can be determined, for which some considerations regarding the message passing are discussed in the following paragraphs. For the deterministic residual generation methods, the residuals can be understood as deterministic observations, represented by the solid black nodes in Figure 4.6 and Figure 4.9(a).

4.6.1 *Likelihood Ratio*

As the likelihood ratio approaches described in 4.3.1 do not take into consideration that the variance of the observations are available, and only the deterministic residual mean values are used for the evaluation, a connection between the two factor graphs is not necessary. If, however, it is assumed that the variance is also used for the determination of the hypothesis test, then a connection between the two factor graphs will automatically lead to some loops between past unscented Kalman filter updates and the likelihood ratio evaluations. In such a case convergence is not guaranteed, and loopy belief propagation would have to be considered.

4.6.2 *Gaussian Mixture Model*

In Figure 4.13 the state-space model is merged with the GMM evaluation in factor graphs, s.t., the forward message $\vec{\mu}_{R_k}$ is a Gaussian mixture *message*.

Gaussian mixture distributions in Kalman filtering lead to an exponential increase in the summands contributing to the mixture. Pruning and merging methods are necessary to reduce the Gaussian mixture summands to a single Gaussian [102]. To avoid this, the most likely mixture component is passed as forward message $\vec{\mu}_{R_k}$, i.e., the Gaussian mixture message is pruned to a single Gaussian. The forward message is thus fixed to the most probable class, which is denoted as g, i.e., $i = g \in n_c$ of the GMM. Then, it is advantageous to include the measurement noise D_k into the GMM, since according to (4.6) the forward message is otherwise a degenerate Gaussian and the marginal would always result into $\bar{\mu}_{R_k} = \vec{\mu}_{R_k}$. Then

$$\vec{\mu}_{\tilde{R}_k}(\tilde{r}_k) = \mathcal{N}\left(\tilde{r}_k; \vec{\mathrm{m}}_g + \vec{\mathrm{m}}_{D_k}, \vec{\mathrm{V}}_g + \vec{\mathrm{V}}_{D_k}\right) \forall\, g.$$

Including the measurement noise into the GMM definition does not change the residual evaluation results.

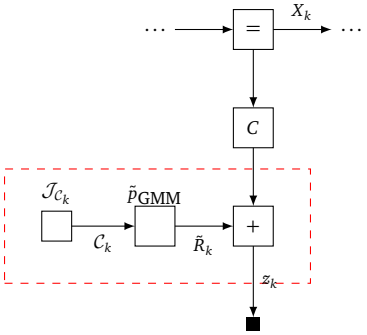


FIGURE 4.13: Merging the GMM factor node to the state space model via the connection of the residual variable.

[102] Runnalls, "Kullback-Leibler Approach to Gaussian Mixture Reduction"

4.7 DISCUSSION & SUMMARY

This chapter developed a full factor graph based method for the generation of model based residuals and the evaluation for anomaly detection as well as for fault classification. With the use of the factor graph framework multiple aspects of the overall fault detection scheme for the SRF cavities were uniformly represented. The stochastic nominal cavity model was extended and the resulting residual was inferred using Gaussian message passing. The

versatility of the approach was confirmed, e.g., both the Bayes factor and the likelihood ratio test can be represented by the same factor graph. Although the factor graphs can handle the propagation of uncertainty, it was shown that they can also be used for the evaluation of deterministic data streams. Training using the expectation-maximization algorithm to obtain the parameters needed for the GMM based fault classification approach were also derived via Gaussian message passing on the same factor graph. It should be noted, that the graph structure may represent other time-varying systems in the EuXFEL. The message passing rules and the resulting inferences may thus also be used to detect and classify faults in other systems. The model selection representation could, e.g., also be directly used on steady state systems or constant but noisy observables of any of the data streams in the EuXFEL. Two deterministic, alternative approaches were also considered. The first is based on the computations of the cavity-wise parameters, which is therefore already implemented in the FPGA framework. In addition the parity space algorithm was defined for the cavity model.

5

Detection Results

"Anomaly is when you don't fit the expected norm. Like "wait what is this?! It doesn't belong here." It's what the system never planned for but now has to adapt to. [...] "
—LeCrae

► SYNOPSIS In this chapter the results of the data analysis is presented. Results pertaining to the long-term data in which no critical events occurred and to the quench data sets which recorded instances of thermal breakdown in six different cavities are recounted. The results distinguish between the different methods of residual generations introduced in the previous section. Applying both detuning models, two unscented Kalman filter residuals are distinguished. The one which generated the residuals using the first-order model is referred to as UKF^{fo}, whereas the other which generated residuals using the fixed detuning traces is referred to as UKF^{ft}. The parameter-based residual is referred to as PB and the parity space method is referred to as PS. Furthermore a distinction is made between those results generated from RF-pulses *with*, and those generated *without* piezo-electric detuning control. The likelihood ratio based residual evaluation is considered for the detection of anomalies in general, whereas the GMM residual evaluation is used specifically to detect quenches.

OVERVIEW

Monitoring of the EuXFEL cavities has so-far relied on the inspection of pulse-wise cavity parameter values, i.e., the static detuning $\bar{f}^{\Delta}$ and the half-bandwidth $f^{\dagger}$.[1] In the time-range of micro-seconds, the cavities' behavior can "spontaneously," i.e., without intentional interference, deteriorate due to, e.g., a quench, multipacting, or field emissions, which makes intra-pulse detection desirable. However, slow-going effects or so-far undetected phenomena may *also* deteriorate the quality of the RF-pulses.

Model-based approaches were proposed in the previous chapter with which anomalous RF-pulses can be determined. The extent to which detected anomalies correspond to meaningful RF-pulse deviations must be evaluated.[2] Usually, the detection capability of a newly proposed method is evaluated by comparing its results to those of an already established method. Then, e.g., the number of *anomalies* detected by both methods (true positive), the number of *nominals* detected by both methods (true negative), as well as those when the methods are not in agreement (false negatives, false positives) are often used as a metric. Since the methods described in this thesis are the first to tackle this task and no ground truth is available, such

[1] In practice, instead of the half-bandwidth $f^{\dagger}$, often the loaded quality factor Q^{L} (which is inversely proportional to $f^{\dagger}$) is inspected. Here, the half-bandwidth values are preferred, but the quality factor values are also specified when necessary.

[2] "Defining a normal region that encompasses every possible normal behavior is very difficult. In addition, the boundary between normal and anomalous behavior is often not precise. Thus an anomalous observation that lies close to the boundary can actually be normal, and vice versa."—V. Chandola, [90].

an evaluation is not possible. Instead, the validity of a detected anomaly is evaluated by comparing the severity of the anomalies to the pulse-wise static detuning and half-bandwidth values.

Descriptive statistics are introduced to determine the severity of an anomaly. The times of occurrence of the *strongest anomalies* per cavity was determined and "hourly events" were identified, i.e., an increase of number of anomalies in the first ten minutes of the hour.

Though these results confirm that so-far unnoticed effects can be found when analysing the RF-pulses, none of those anomalies severely impaired the RF-pulse operation. Thus, thresholds on the descriptive statistics of a pulse are introduced, with the aim of detecting *severe anomalies*. With these thresholds, severe anomalous pulses were detected in the data sets $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{Q}}$, $\boldsymbol{\mathcal{Q}}^{pz}$,[3] see Table 3.3 for a description of these data sets.

[3] *pz* indicates that when this data was collected, the piezo-electric detuning control was on.

To determine the detection capability of quenches, the number of quenches classified as severe anomalies were investigated. It is assessed at what point *during* the pulse the *first* detected quench of each quench event can be classified as severely anomalous. Then, the available quench examples together with a set of nominal and anomalous RF-pulses were used to train GMMs as proposed in Chapter 4.5. To train the GMM, both unscented Kalman filter residuals UKF^{fo} and UKF^{ft} were used. The performance of the models was evaluated by determining the number of false positives and false negatives from the classifications of the quench detection system.

5.1 DESCRIPTIVE STATISTICS FOR RF-PULSE EVALUATION

All RF-pulses were evaluated based on the likelihood ratios described in Chapter 4.3.2, i.e., λ was determined for each sample of all residuals. Each sample in an RF-pulse is evaluated with

$$\mathrm{d}_k(\lambda_k) = \begin{cases} 0 \text{ if } & \lambda_k < \psi_{\text{th}}, \text{ sample is nominal,} \\ 1 \text{ if } & \lambda_k \geq \psi_{\text{th}}, \text{ sample is anomalous,} \end{cases} \tag{5.1}$$

where d_k is the decision function that classifies the samples in an RF-pulse. The threshold ψ_{th} was chosen according to a significance level of 0.01%, i.e., a probability of false alarm $\epsilon = 0.01$ was chosen. The threshold values were determined from the table of the $\chi^2(\text{df})$ distribution and depend on the degrees of freedom df that the residuals have [103].

[103] Kokoska and Nevison, "Critical Values For The Chi-Square Distribution"

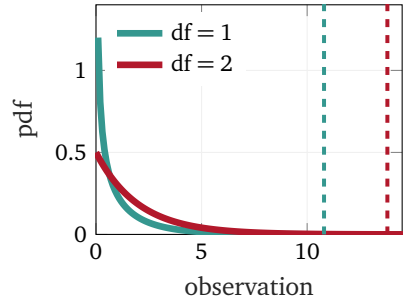


FIGURE 5.1: χ^2 probability density function with one degree of freedom df = 1 and two degrees of freedom df = 2, and the chosen respective thresholds.

The parameter-based and unscented Kalman filter residuals have two dimensions, therefore df = 2. The parity-space residual has only one, therefore df = 1. For one degree of freedom the threshold is $\psi_{\text{th}} = 10.8$, whereas for two degrees of freedom it is $\psi_{\text{th}} = 13.8$. Figure 5.1 shows the probability density function of a $\chi^2(\text{df})$ distribution for df = 1 and df = 2, i.e., nominal values of PS residuals should have lower λ values than those of UKF^{fo},UKF^{ft}or PB.

To be able to handle the huge amounts of RF-pulse evaluation results, pulse-wise descriptive statistics were stored instead of sample results. They are used in the following chapters to gain insight into the severity of an anomaly. Figure 5.2 depicts an example of the likelihood ratio values for

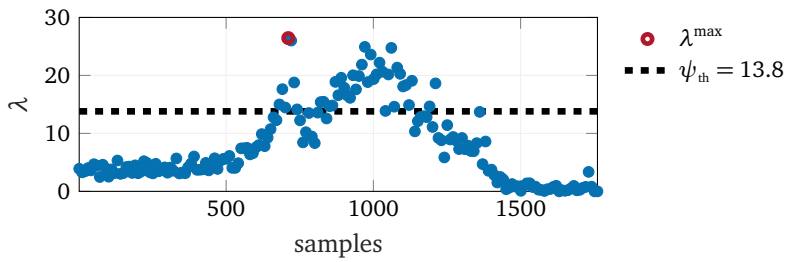


FIGURE 5.2: Likelihood ratio values λ for each sample in a pulse of the unscented Kalman filter residual $\mathrm{UKF}^{\mathrm{fo}}$. When $\lambda \geq \psi_{\mathrm{th}}$, the sample is classified as anomalous. A pulse that contains an anomalous sample is classified as anomalous. To store information about the severity of the anomaly, the mean and the maximum values of all λ in a pulse are determined and stored.

the samples over a single pulse. The pulse-wise descriptive statistics are the following:

- THE NUMBER of anomalous samples per pulse i_p, i.e.,

$$n^{\mathrm{a}}_{i_p} = \sum_{k=1}^{n_s} \mathrm{d}_k(\lambda_k), \tag{5.2}$$

where $\mathrm{d}_k(\lambda_k)$ is defined in (5.1) and $n_s = 1820$ is the number of samples in each pulse. It thus determines how much of the evaluated RF-pulse was detected as anomalous.

- THE MEAN of all λ_k per pulse, i.e.,

$$\mathrm{m}^{\lambda}_{i_p} = \frac{1}{n_s} \sum_{k=1}^{n_s} \lambda_k, \tag{5.3}$$

which determines an average pulse-wise divergence of the samples. Since the likelihood ratio values can only be positive, i.e., $\lambda_k \in \mathrm{R}_+$, the mean over the pulse $\mathrm{m}^{\lambda}_{i_p}$ indicates on how strongly the residual diverges from its nominal behavior, independent of the direction of divergence. The lager this value is, the stronger the average divergence of the pulse.

- THE LARGEST value of λ_k in a pulse

$$\lambda^{\max}_{i_p} = \max_{k=1,\ldots,n_s} \{\lambda_k\}, \tag{5.4}$$

which gives insight into the largest divergence within a pulse.

Other statistics as the mode or the variance of λ over the RF-pulse were also considered, but did not prove to give additional information relevant for the subsequent result descriptions of the detection results.

5.2 ANOMALY DETECTION

For each residual generation method, the percentage of anomalies detected in $\mathcal{T}$ and $\mathcal{T}^{pz}$ is presented. The descriptive statistics introduced in the previous section were used to determine the detected anomalies' severity for each cavity and the correlation of the mean divergence m^{λ} to the pulse-wise static detuning $\bar{f}^{\,\Delta}$ and half-bandwidth $f^{\,\dagger}$ values was analyzed. The time of occurrence for the strongest anomalies per cavity were analyzed for all stations.

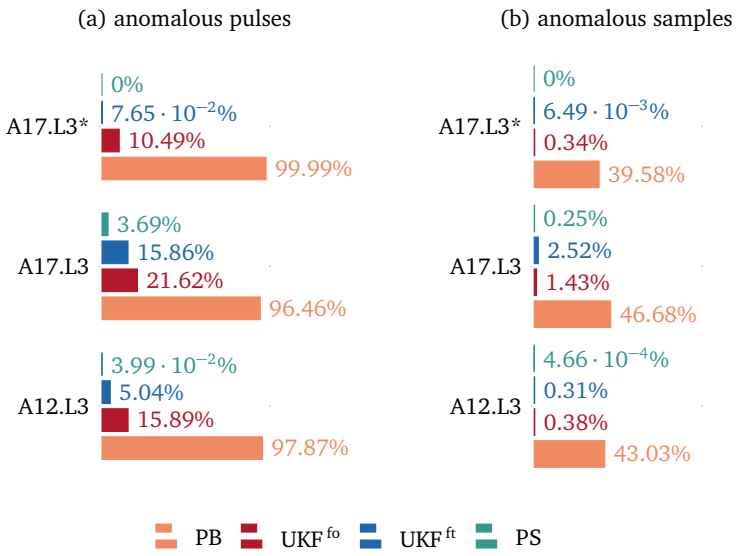


FIGURE 5.3: Percentage of detected anomalous pulses and samples for the evaluated long-term data obtained from stations A12.L3 and A17.L3 without piezo-electric detuning control, and long-term data with running piezo-electric detuning control A17.L3^{pz}. (a) Percentage of anomalous pulses detected, where a pulse is classified as anomalous when the number of detected anomalous samples in that pulse is larger than zero, i.e., $n^{\mathrm{a}} > 0$. (b) Percentage of anomalous samples, where a sample is classified as anomalous when the divergence measure exceeds a defined threshold $\lambda > \psi_{\mathrm{th}}$.

5.2.1 *Method-Dependent Detection Results*

The residuals generated with each method using the samples in $\mathcal{T}$ and $\mathcal{T}^{pz}$ were classified as either nominal or anomalous according to (5.1). Pulses were classified as anomalous if at least one of the samples was classified as anomalous, i.e., a pulse is anomalous if $n^{\mathrm{a}} > 0$. The percentage of anomalous pulses and the percentage of anomalous samples is shown in Figure 5.3. The results are grouped according to the involved stations, A12.L3, A17.L3 and A17.L3^{pz}, where A17.L3^{pz} stands for the evaluation results of $\mathcal{T}^{pz}$, i.e., the data where piezo-electric detuning control was running.

The parameter-based approach PB classified more than 95% of the pulses and between 39% and 43% of the samples as anomalous. The least amount of anomalies were detected using the parity-space PS method, where 0%, 0.039% and 3.7% of the pulses, and 0%, 0.25% and 0.0004% of the samples were classified as anomalous for A17.L3^{pz}, A12.L3 and A17.L3 respectively. The amount of anomalies detected with either one of the unscented Kalman filter residuals lie in between these two methods. When using the first-order detuning model for the residual, i.e., UKF $^{\mathrm{fo}}$, 10.55%, 21.62% and 15% anomalous pulses, and 0.34%, 1.43% and 0.38% anomalous samples for A17.L3^{pz}, A12.L3 and A17.L3 respectively were detected. With the fixed detuning model UKF $^{\mathrm{ft}}$ 0.59%, 15.86% and 5.04% anomalous pulses and 0.006%, 2.52% and 0.31% anomalous samples were detected.

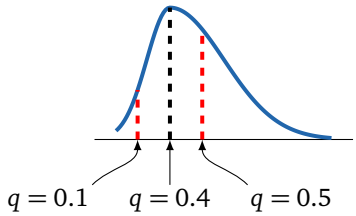


FIGURE 5.4: Quantiles divide observations into groups of occurrences. For example, the quantile $q = 0.1$ is defined by the value for which 10% of the observations have lower values.

5.2.2 *Anomalies per Cavity*

Quantiles q, see Figure 5.4, were determined over all pulses per cavity to represent the results. The lowest value represents $q = 0.1$ the highest represents the $q = 0.9$ and $q = 0.5$ is represented by the highlighted marker in the middle of the two values.

The first row of Figure 5.5 and Figure 5.6 shows the percentage of RF-pulses classified as anomalous by the different residual generations for each cavity of station A17.L3 and A17.L3^{pz}, respectively. The number of anomalies is different for each cavity and depends on the detection method.

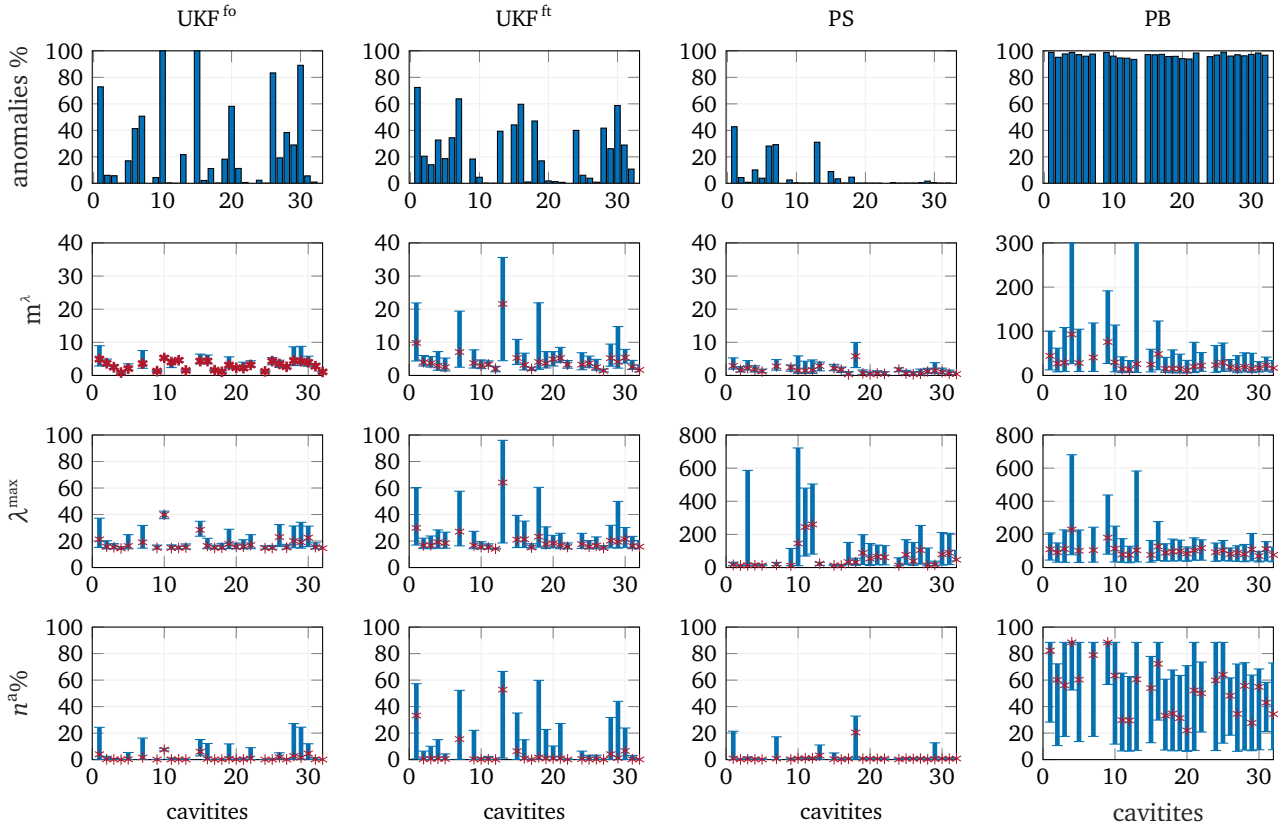


FIGURE 5.5: Cavity-wise results of A17.L3. The station was running without piezo-electric detuning control.

While for some cavities no RF-pulse was classified as anomalous, for others up to 100% of the RF-pulses were classified as anomalous.

Further characteristics of these anomalies described by m^{λ}, λ^{max} and n^{a} are depicted in the second, third and final row of Figure 5.5.

Without Piezo Detuning Control

In Figure 5.5, it can be seen that none of the n^{a} values of $q = 0.9$ for UKF fo values for exceeded 35%, and $q = 0.5$ did not exceed 15%, which means that for a majority of the detected anomalies, less than half of the RF-pulse samples exceeded the threshold at $\psi_{th} = 13.8$. The maximum sample divergence, except for cavity 10 and 14, was less than twice the threshold value and the mean divergence did not exceed ten. This shows that for a majority of the detected anomalies the evaluated samples only barely exceeded their threshold.

The total percentages of anomalies for UKF ft is smaller in comparison and the values of $q = 0.9$ for m^{λ}, λ^{max} as well as n^{a} are higher than for UKF fo. Thus, anomalies detected with UKF ft are fewer, but show more severe divergence from nominal.

A much smaller percentage of anomalies was detected with PS, which can mostly be characterized as RF-pulses for which less than 10% of the pulse is anomalous, and the mean divergence m^{λ} is low.

However for three cavities, for which less than 10% of the pulses were classified as anomalous, the maximum divergence λ^{max} was considerably higher, reaching values above 700. This suggests, that those anomalous

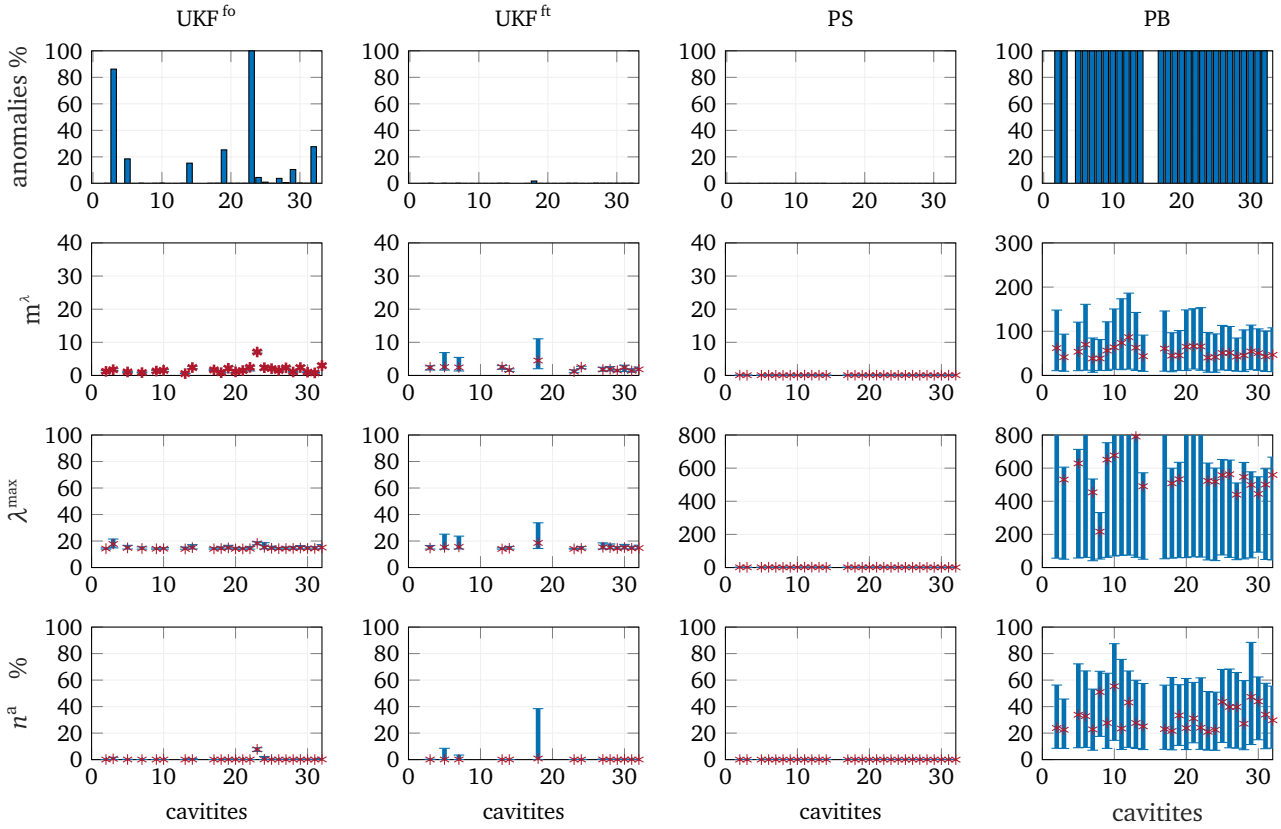


FIGURE 5.6: Cavity-wise results of A17.L3^{pz}. The station was running with piezo-electric detuning control.

RF-pulses had a few, strongly diverging samples. The PB method detected almost 100% of the RF-pulses as anomalous. Other than for the methods UKF^{fo}, UKF^{ft}, or PS, the mean and maximum divergences were comparatively high and for more than 50% of the samples exceeded the detection threshold.

With Piezo Detuning Control

Figure 5.6 represents the results for A17.L3^{pz}. Using UKF^{fo}, UKF^{ft} or PS less anomalies were detected. For UKF^{ft} the amounts of detected anomalies decreased considerably and for PS none of the pulses classified as anomalous. The majority of anomalies detected with UKF^{ft} were associated to low numbers of anomalous samples (less than 10% for $q = 0.5$), which only barely exceeded the thresholds. The anomalies detected with UKF^{fo} also had low numbers of anomalous samples, although for two cavities almost 100% of the RF-pulses had at least one sample which exceeded the threshold. PB is the only approach for which similar amounts of anomalies were detected even in piezo-on mode.

The differences in the detection results are addressed in the subsequent chapter, in which the model assumptions for the different methods are compared and evaluated.

The subsequent sections analyze the sensitivity of the different methods toward changes in the pulse-wise cavity parameters, i.e., the static detuning $\bar{f}^{\Delta}$ and the half-bandwidth $f^{\dagger}$.

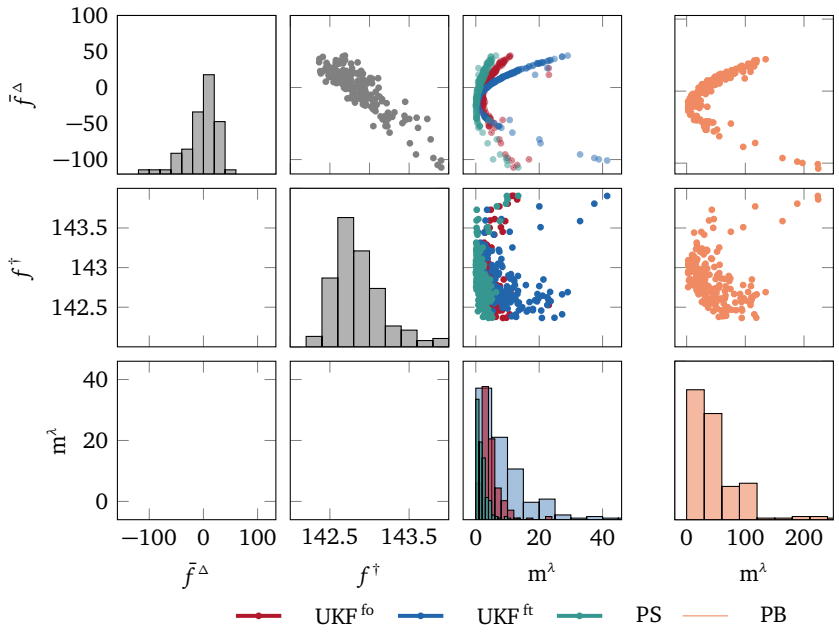


FIGURE 5.7: Correlations between pulse-wise static detuning $\bar{f}^{\Delta}$ and half-bandwidth $f^{\dagger}$ values to the mean likelihood ratio m^{λ} of RF-pulses from one cavity without running piezo-electric detuning control. It can be seen that the methods have a different sensitivity towards changes in static detuning and half-bandwidth. Half-bandwidth and detuning seem to have a linear correlation, whereas the nonlinear correlation of detuning and m^{λ} describes a quadratic relation.

5.2.3 *Correlation to Pulse-Wise Cavity Parameters*

Further insight into what significance the detected anomalies have can be gained when comparing them to changes of the pulse-wise static detuning $\bar{f}^{\Delta}$, defined by (3.3) and half-bandwidth $f^{\dagger}$ defined by (2.16). Figure 5.7 shows a "plot matrix", in which each tile depicts the pairwise correlations between the static detuning, half-bandwidth and the mean divergence m^{λ} for each RF-pulse. The different residual generation methods are distinguished by color, and the parameter-based values are depicted in separate tiles, since the ranges of maximally and minimally attained values is around ten times higher and an elegant depiction in the same tile was not possible. Instead of showing the auto-correlations on the diagonal tiles, histograms of the respective parameters are depicted.

The plot matrix in Figure 5.7 was generated using the values of cavity C1.M1.A17.L3 as an example, but qualitatively similar results were also obtained for other cavities. The static detuning values are between −100 Hz and 50 Hz, whereas the half-bandwidth values are between 142.5 Hz and 144 Hz. The second tile in the first row suggests a linear correlation between the half-bandwidth values $f^{\dagger}$ and the static detuning.

Centered around 0 Hz, the values m^{λ} increase quadratically with the divergence of the static detuning in both positive and negative directions. How much the m^{λ} values change with the static detuning is different for each method. The parity space residuals change the least, whereas the parameter-based residuals change the most. Between the two unscented Kalman filter methods, UKF^{fo} is less sensitive towards detuning changes than UKF^{ft}. A change in static detuning is thus, more or less severely, also reflected in the residuals. The same correlation matrix plot was defined for cavity C1.M1.A17.L3, i.e., operated *with* piezo-electric detuning control. The results can be seen in Figure 5.8, where the static detuning values lie between ±10 Hz. The quadratic correlations between $\bar{f}^{\Delta}$ and m^{λ} are still

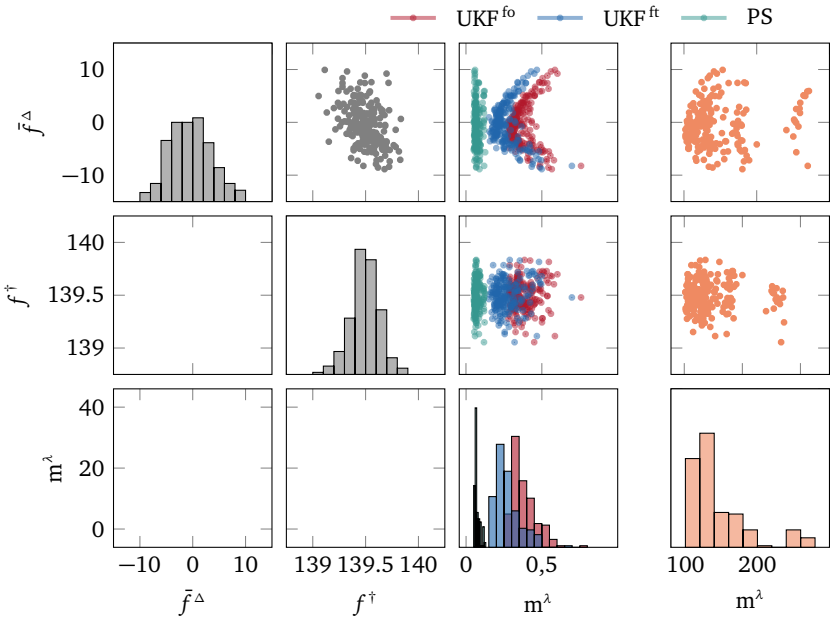


FIGURE 5.8: Correlations between pulse-wise static detuning $\bar{f}^{\Delta}$ and half-bandwidth $f^{\dagger}$ values to the mean likelihood ratio m^{λ} of RF-pulses from one cavity with running piezo-electric detuning control. The correlations between static detuning and bandwidth have disappeared. The quadratic correlation between the mean likelihood ratios and the static detuning is still visible for PB, UKF fo and UKF ft.

visible for UKF fo,UKF ft and PB but seem to have no effect on PS. It should be noted that *none* of the RF-pulses of UKF fo, UKF ft or PS displayed here was detected as anomalous, i.e., changes in detuning of that magnitude did not lead to surpass the detection thresholds. The evaluations with PB, on the other hand were *all* detected as anomalous. A smaller group of PB samples can be seen to diverge further than the rest from zero. Explanations for this are given in the subsequent chapter.

5.2.4 *Time of Occurrence of Strong Anomalies*

In this section the detected anomalies are further characterized with respect to their times of occurrence. As the results showed so far, many anomalous RF-pulses are associated to benign deviations from nominal, with low number of samples only barely exceeding the thresholds. At the same time, the severity of divergence differs from cavity to cavity. The following results show the times in which the strongest anomalies selected from each cavity occured.

Instead of choosing a threshold for each cavity, the 2000 strongest anomalies of each cavity according to their m^{λ} values obtained from UKF ft were selected. The choice of UKF ft is arbitrary and analyzing the strongest anomalies as defined by any of the other methods (UKF fo, PS, or PB) lead to qualitatively similar results. Since a station consists of 32 cavities, in total 62000 pulses were selected for A17.L3, A17.L3^{pz} and A12.L3. Then, the time of occurrence of these RF-pulses was determined via their individual RF-pulse time-stamps. Figure 5.9 shows the amount of occurrences for each hour (0 to 24) of the day as well as each minute of the hour (0 to 59).

According to the results shown in Figure 5.9, when run without piezo-electric detuning control, between 48% and 52% occurred between the 4th and 9th hour and another 10% occurred at the 20th hour. When run with piezo-electric detuning, an increase of the number of anomalies between the 11th and the 13th hour was detected. For all stations, with and without

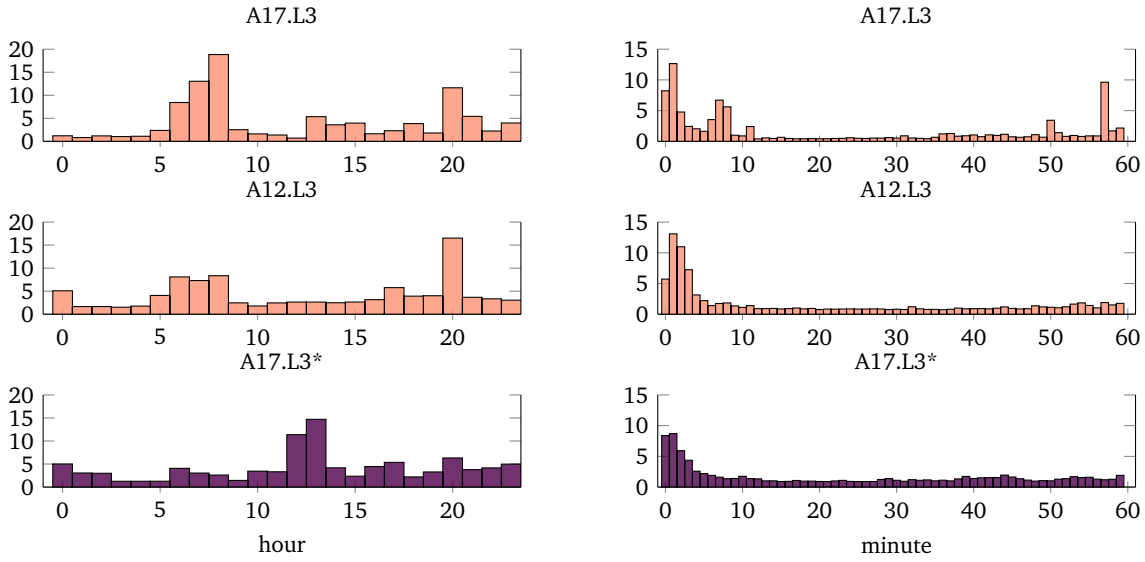


FIGURE 5.9: Distribution of strong anomalies over time for each hour in the day and each minute of the hour.

piezo-electric detuning control, more than 50% of anomalies happened in the first ten minutes of each hour. This behavior is noteworthy and additional information pertaining to what kind of anomalies were detected in the first ten minutes are given in the subsequent section.

5.2.5 *Characteristics of Hourly Events*

The previous section has shown, that around half of the strongest anomalies in each cavity occurred in the first ten minutes of the hour. In addition, large values of m^{λ} have been shown to be associated with large changes in static detuning, see Chapter 5.2.3. In this section the "hourly events" are characterized in more detail. First the pulse and sample-wise change of the likelihood ratio λ values for such hourly events were studied. One representative example is depicted in Figure 5.10 using the likelihood ratios of UKF^{fo}. The darker the color, the higher the value of λ. In total 2000 consecutive pulses are shown which is equivalent to almost 3.3 min.

The largest values of λ_k in that time interval occurred around the 200th pulse and between the 1000th and 1500th sample. This means that the largest anomalous behavior occurred during the end of the flattop and beginning of the decay. It can be noted, that the divergence of the residual, as evaluated by the likelihood ratios, did not immediately go back to zero but slowly faded out during the course of around 2 min. It can further be noted, that just before the main divergence, the likelihood ratio values were particularly small, which is represented in the figure by the almost white coloring. The hourly events were further studied by comparing the distributions of the pulse-wise cavity parameters according to the classifications of nominal, anomalous and strongly anomalous. The result of one cavity of A17.L3 (running without piezo-control) is presented by a histogram in Figure 5.11. The values were normalized for each group, s.t. they are represented as probabilities, i.e., summing over the values of each group is equal to one. From the distribution of the m^{λ} values it can be seen, that anomalous and

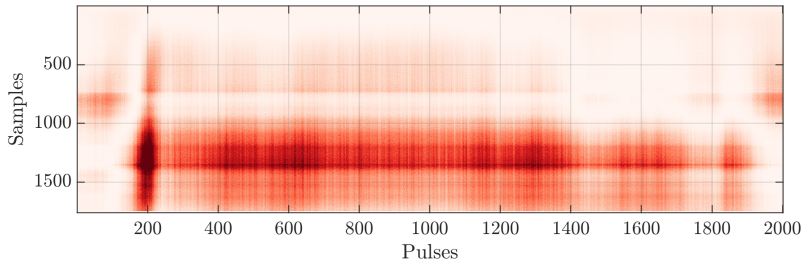


FIGURE 5.10: This is the hourly event represented as a heat graph. The darker the color, the higher the values of λ. It represents 2000 consecutive pulses starting with one of the hourly events resulting in the strong anomalies.

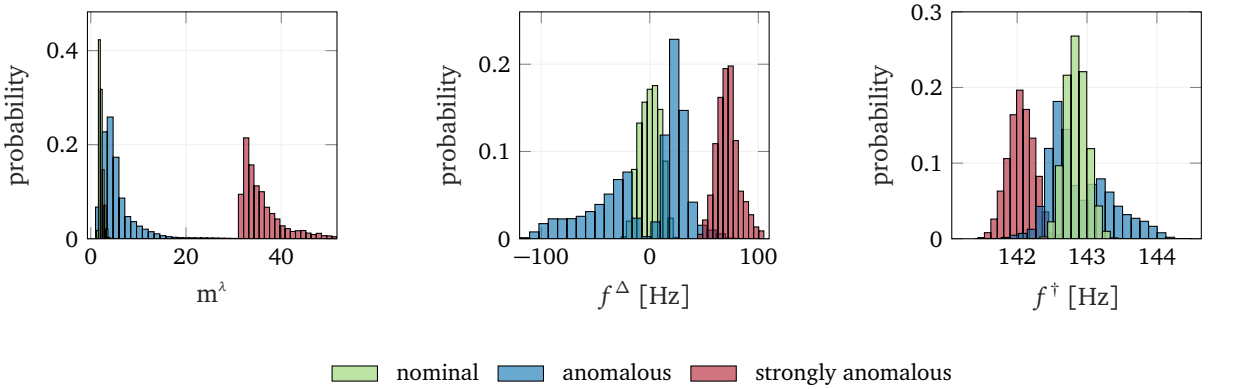


FIGURE 5.11: Histogram normalized to probabilities of the hourly events compared to the rest of the data. The events happening every hours correspond to static detuning increase and half-bandwidth decrease.

strongly anomalous pulses differed in their severity of mean divergence from nominal. It can also be seen that in contrast to other anomalies, both bandwidth and static detuning values changed for the strongest anomalies, which in this case were all associated to an hourly event. For the hourly event, the bandwidth *decreased*. In this example the detuning increased, but in general both an increase or a decrease was observed. Independent of the direction, for strong anomalies the detuning change compared to nominal was in the range of 50 Hz to 150 Hz.

The following section describes severely anomalous pulses, which were defined as those anomalies which diverged more strongly from nominal than the so far analyzed strong anomalies.

5.3 DETECTION OF SEVERE ANOMALIES

While the anomalies detected in the previous section indicated sub-optimal and undesired behavior, none of them could be associated to an unwanted shut-down of the station. This section describes two detection rules to distinguish the so-far encountered anomalies from "severe anomalies".[4] An example of so-far undetected, severely anomalous behavior is presented using the data set $\mathcal{S}$ and the quench data sets $\mathcal{Q}, \mathcal{Q}^{pz}$.

To explore anomalies with more severe divergence from nominal, two decision rules $(d^{s,I}, d^{s,II})$ are formulated, which classify a pulse as severely anomalous using the mean likelihood ratio divergences m^{λ} and maximum likelihood values $\lambda^{\max}$, respectively and thresholds for each method PB, PS, UKF^{fo}, and UKF^{ft} are defined. RF-pulses with larger mean divergence m^{λ} or with larger maximum sample divergence $\lambda^{\max}$ in the data sets $\mathcal{T}$ and $\mathcal{T}^{pz}$, (a total amount of around 17.5 million RF-pulses) were defined as severe

[4] Classification labels in order of severity:

- nominal
- anomalous
- strongly anomalous
- severely anomalous

	PB	PS	UKF $^{\text{fo}}$	UKF $^{\text{ft}}$
piezo off				
α_{th}	14075	38.5	72.2	229
β_{th}	21144	1118	221	662
piezo on				
α_{th}	211	0.15	7.9	18.6
β_{th}	1158	10.8	25.1	51.5

TABLE 5.1: Detection thresholds for each method determined by the maximum attained values in the non-severe data set. A differentiation was only made between piezo-on and piezo-off data sets.

anomalies.

The decision rules are formulated as follows:

- THE FIRST DECISION RULE classifies anomalous RF-pulses, i.e., pulses with $n^{\text{a}} > 0$ as severely anomalous with

$$\mathrm{d}^{s,I}(\mathrm{m}^{\lambda}) = \begin{cases} 0 \text{ if } & \mathrm{m}^{\lambda} \leq \kappa_I \alpha_{th}, \text{ anomalous pulse,} \\ 1 \text{ if } & \mathrm{m}^{\lambda} > \kappa_I \alpha_{th}, \text{ severely anomalous pulse,} \end{cases} \tag{5.5}$$

where α_{th} is the threshold on the mean likelihood divergences m^{λ} determined by

$$\alpha_{th} = \max_{\substack{i_c=1,\dots,n_{cav} \\ i_p=1,\dots,n_p}} \left\{ \mathrm{m}^{\lambda}_{i_c,i_p} \right\}, \tag{5.6}$$

where $\mathrm{m}^{\lambda}_{i_c,i_p}$ are all values encountered in the evaluation of the non-severe data sets $\mathcal{T}$ or $\mathcal{T}^{pz}$.

- THE SECOND DECISION RULE classifies anomalous RF-pulses, $n^{\text{a}} > 0$ as severely anomalous according to

$$\mathrm{d}^{s,II}(\lambda^{\max}) = \begin{cases} 0 \text{ if } & \lambda^{\max} \leq \kappa_{II} \beta_{th}, \text{ anomalous sample,} \\ 1 \text{ if } & \lambda^{\max} > \kappa_{II} \beta_{th}, \text{ severely anomalous sample,} \end{cases} \tag{5.7}$$

where β_{th} is the threshold on the maximum divergence $\lambda^{\max}$, which is determined by

$$\beta_{th} = \max_{\substack{i_c=1,\dots,n_{cav} \\ i_p=1,\dots,n_p}} \left\{ \lambda^{\max}_{i_c,i_p} \right\}, \tag{5.8}$$

which, in turn, is the largest sample divergence $\lambda^{\max}$ encountered in pulses of all cavities in $\mathcal{T}$ or $\mathcal{T}^{pz}$.

The parameters κ_I, κ_{II} are introduced to tune the divergence between the severe and not severe anomalies. For both decision rules, the tuning parameters were set to $\kappa_I = \kappa_{II} = 1.05$, setting the detection threshold at a 5% larger value than the highest values encountered in $\mathcal{T}$ or $\mathcal{T}^{pz}$.

The threshold α_{th} and β_{th} are determined for each residual generation method PB, PS, UKF $^{\text{fo}}$, UKF $^{\text{ft}}$ once for an operation without piezo tuning and once for an operation with piezo tuning, see Table 5.1. The determined thresholds α_{th}, β_{th} are smaller for the piezo-on operation. A large difference can also be noted between the different residual generation methods. The

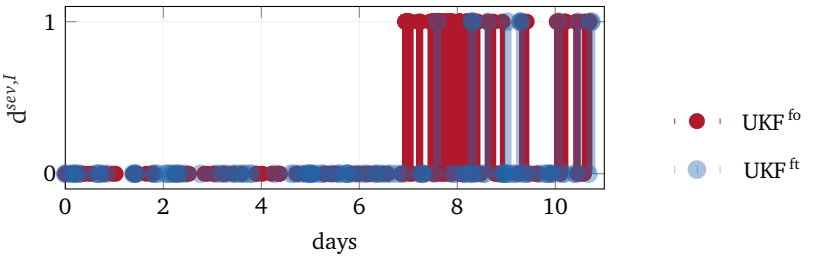


FIGURE 5.12: The cavity C6.M1 of station A17.L3 showed large amounts of severe anomalies for over three days. Using the detection thresholds defined in (5.5), large amounts of severe anomalies were detected using the UKF fo or UKF ft residual generation methods.

parameter-based maximum value of m^{λ} is much higher than the ones for PS, UKF fo, or UKF ft. The lowest values were obtained for the PS residuals. Since no anomalies were detected in $\mathcal{T}^{pz}$ using the PS residuals, the threshold on the mean divergence was set to $\alpha_{th} = 0.26$ and the threshold on the maximum divergence was set on the sample threshold for PS, i.e., $\beta_{th} = \psi_{\text{th}} = 10.8$.

Both decision rules are used to identify severely anomalous RF-pulses in the data sets $\mathcal{Q}$, $\mathcal{Q}^{pz}$ and $\mathcal{S}$. While $d^{s,I}$ is only able to decide about the severity on the level of pulses, $d^{s,II}$ can decide about the severity both on the level of pulses and on the level of samples.

5.3.1 *The Strange Behavior of C6.M1*

The data set $\mathcal{S}$ stems from 11 days of operating cavity C6.M1.A17.L3 in March 2019. The cavity stood out, because the pulse divergence for both m^{λ} and $\lambda^{\max}$ showed a change after around six days of operation. The detection thresholds defined in Table 5.1 were used to determine severely anomalous pulses for the cavity C6.M1.A17.L3. Using the UKF fo and UKF ft methods of residual generation, severely anomalous behavior was detected over several days. Evaluating the severity of the divergence with the detection thresholds in Table 5.1, resulted in the detection of severe anomalies.

Figure 5.12 shows the detection results, for each pulse over the course of 12 days. Just after six days, the first severely anomalous pulses were detected by UKF fo, followed by UKF ft. After the first severe anomalous pulse, both anomalous and even nominal RF-pulses were detected. Although the PS residuals as well as the PB residuals did not exceed the thresholds to be classified as severely anomalous, their values also increased notably over that period of time. During that period, the vector-sum forward signal in the station showed a steady increase. This points to the fact that the MIMO controller compensated for the energy loss from the detuning and half-bandwidth changes. After several days of severely anomalous RF-pulses a different cavity of that same station quenched. Due to the quench, the station was restarted, but the severely anomalous behavior of C6.M1 continued. Two days later, another cavity in that station quenched which ended the data collection period.

To gain additional insight, the histograms of the static detuning $\bar{f}^{\Delta}$ and the half-bandwidth $f^{\dagger}$ for nominal, anomalous and severely anomalous RF-pulses was analyzed. In Figure 5.15 the static detuning and half-bandwidth values for nominal, anomalous and severely anomalous RF-pulses according to the detection results of UKF fo can be seen. The histograms were normalized such that each group represents the distribution in probabilities, i.e.,

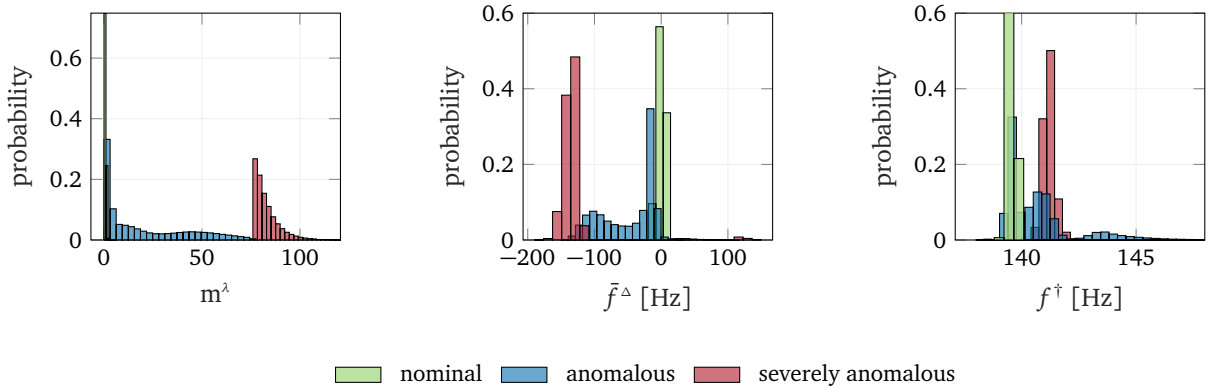


FIGURE 5.13: Distribution of the static detuning $\bar{f}^{\Delta}$ and half-bandwidth $f^{\dagger}$ of pulses exceeding the detection threshold for severe anomalous behavior in C6.M1.L3 compared to those which did not exceed the threshold.

the summarized values for each group equals one. The distribution of the m^{λ} values for the different groups show the thresholds set between nominal, anomalous and severely anomalous.

The static detuning values vary between the three groups. Whereas the nominal static detuning values stayed between $\bar{f}^{\Delta} = 0 \pm 20$ Hz, the anomalous RF-pulses' static detuning fell in a range of $\bar{f}^{\Delta} = -110$ Hz, and $\bar{f}^{\Delta} = 10$ Hz, and the severely anomalous RF-pulses showed static detuning values of above 100 Hz and below −122 Hz.

The distribution of the half-bandwidth also differs for each group. For the nominal RF-pulses, the half-bandwidth takes up values $f^{\dagger} = 139.5 \pm 0.8$ Hz, the half-bandwidth values corresponding to anomalous RF-pulses are more spread out between 138.5 Hz and 155.2 Hz. The majority of the half-bandwidth increased from around $f^{\dagger} = 139$ Hz to $f^{\dagger} = 141$ Hz for the severe anomalies. The lowest and highest half-bandwidth values corresponded to a maximum loaded quality factor of $Q^L = 4.7 \times 10^6$ and a minimum of $Q^L = 4.2 \times 10^6$, respectively. This nearly corresponds to a change of loaded quality factor detected as a quench, see 2.2.5,[5] which indicates that the severely anomalous pulses corresponded to unwanted and yet undetected behavior at the time.

[5] The current quench detection system evaluates entire pulses on the basis of changes in the loaded quality factor Q^L. When the moving average of Q^L drops by more than $0.5 \cdot 10^6$, a quench is detected

5.3.2 *Severe Anomalies Incipient to Quench*

Severe anomalies were also detected in the quench data sets. This section summarizes the results obtained from analyzing the quench data sets with respect to anomalous RF-pulses. Severely anomalous behavior was detected in three of the five quench events, detected *before* the quench occurred. In all three cases they were exclusively detected in at least one cavity which was not the one that eventually quenched.

How many RF-pulses prior to the quench detection were classified as severely anomalous depends on the residual generation method. In all cases the earliest detection times and most amounts of severely anomalous pulses were detected using the UKF^{fo} residuals. The detection results for $d^{sev,I}$ are presented in Figure 5.14.

For quench event I, severe anomalies were detected ≈1200 RF-pulses (≈2 min) prior to a quench according to the first decision rule (5.5) whereas no severe anomalies were detected according the second decision rule (5.7). For quench events II and III, up to ≈ 400 RF-pulses (≈20 s) before the

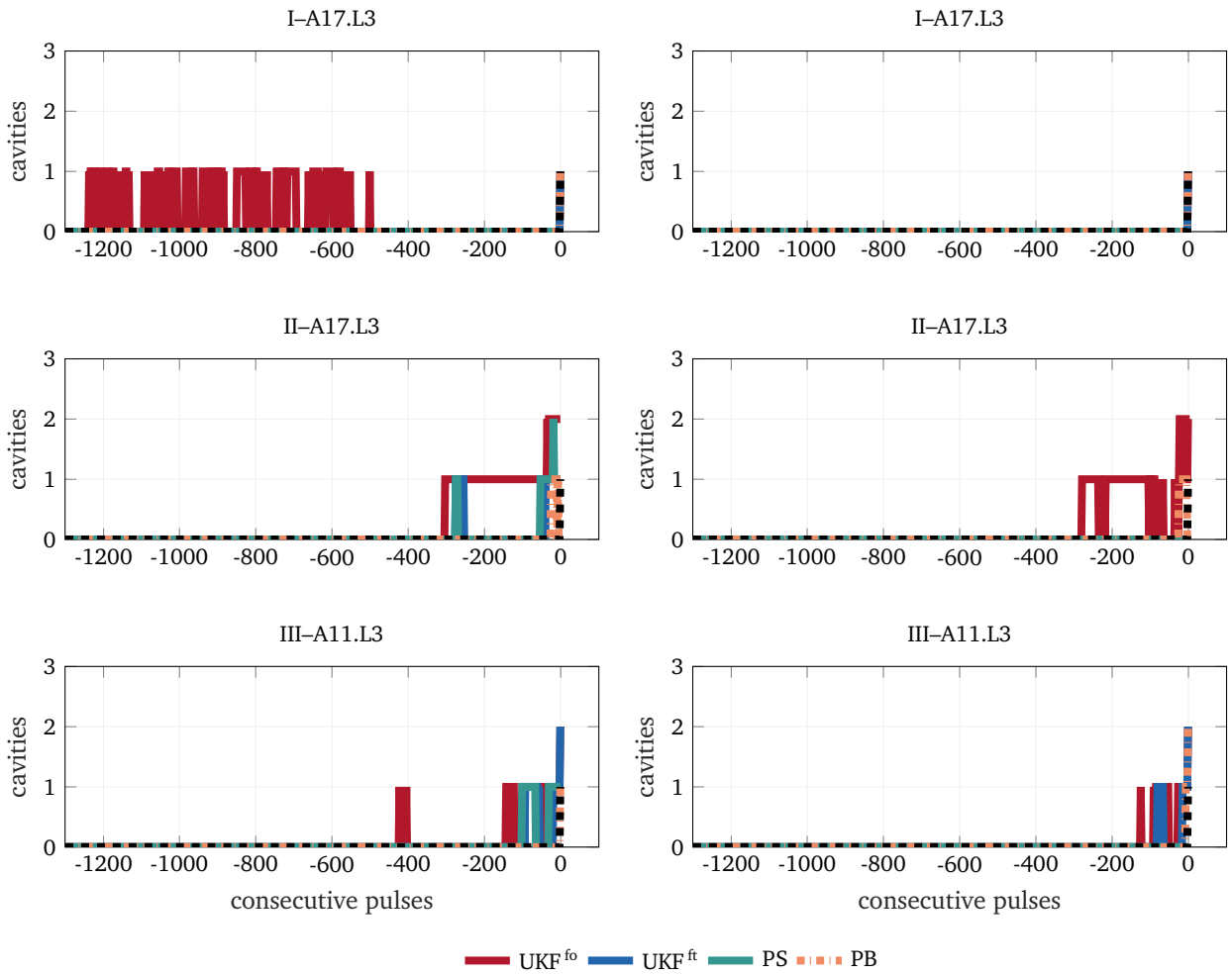


FIGURE 5.14: Detection of severe anomalies prior to the detection server time here denoted by $t = 0$. For three of five analyzed quench events severely anomalous behavior was detectable for at least one of the proposed methods. The most sensitive and earliest detection results come from UKF ^{fo}. In some cases severe anomalies are detected in more than one cavity.

quench were classified as severely anomalous for both decision rules in one cavity at first and then in two cavities simultaneously.

To further investigate characteristics of the severely anomalous pulses detected prior to the quench, the static detuning and half-bandwidth distributions for the groups nominal, anomalous, and severely anomalous (pre-quench) were considered. In Figure. 5.15 an example of severely anomalous pulses is shown where the UKF ^{fo} residuals were used for the evaluation.

The static detuning values for nominal RF-pulses are distributed around zero, i.e., taking up values of $\bar{f}^{\Delta} = 0 \pm 20$ Hz, and the half-bandwidth values are between $f^{\dagger} = 141.7$ Hz and $f^{\dagger} = 143.0$ Hz. For the anomalous pulses, the static detuning changes to $\bar{f}^{\Delta} = -60 \pm 8$ Hz and the half-bandwidth increases to values between $f^{\dagger} = 142.8$ Hz and $f^{\dagger} = 143.5$ Hz. Severely anomalous pulses show static detuning values between -150 Hz and -200 Hz and the half-bandwidth values increase even further to values between -143.5 Hz and 145.1 Hz. This corresponds to a decrease of the loaded quality factor from 4.59×10^6 to 4.48×10^6, i.e., the loaded quality factor decreased by 1.1×10^5. This is around one fifth of what is necessary for the quench detection system to set the alarm. For the time period in which the severely anomalous pulses were detected prior to the quenches, not only the pulse-wise descriptive statistics were stored but the likelihood ratios λ for each

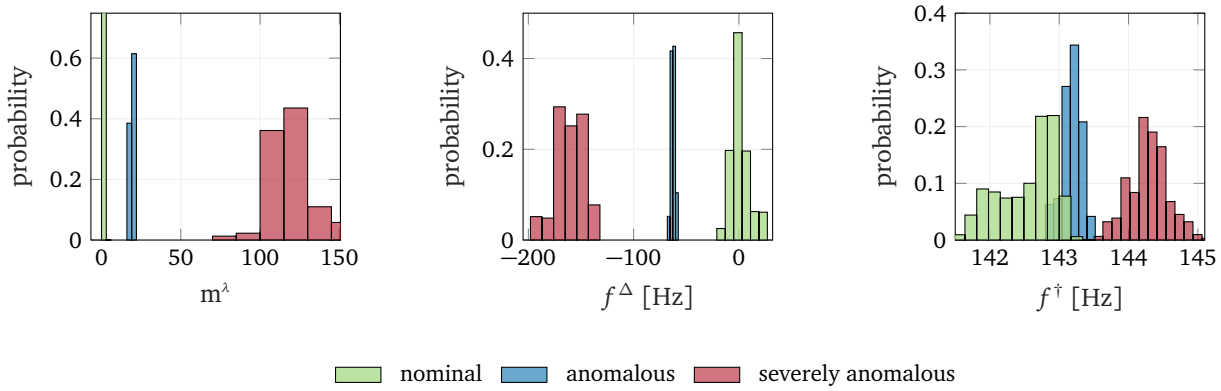


FIGURE 5.15: Distribution of the static detuning $\bar{f}^{\Delta}$ and bandwidth $f^{\dagger}$ of pulses exceeding the detection threshold for severe anomalous behavior before a quench was detected in not this but another cavity of the station.

sample of every pulse. The results can be seen in Figure 5.16, where peculiar behaving cavities were hand-picked and the magnitude of λ in each pulse, i.e., the divergence from nominal is indicated for each sample (y-axis) and each pulse (x-axis) through the intensity of the red color. The darker the color, the stronger the divergence from nominal behavior as determined by UKF^{fo}. Qualitatively similar results were obtained using the values of UKF^{ft}, PS or PB.

The severe anomalies have a distinct signature development over the pulses and samples. Particularly M4.C8 and M2.C7 of quench event I, M1.C5, M2.C7 of quench event II and M1.C8, M3.C8 and M1.M8 of quench event III suggest some kind of oscillating pattern. The strength of the anomalies build up over several pulses, especially noticeable in flattop and decay of the pulse. Even in quench event IV—for which no severe anomalies were detected—a similar pattern arises albeit at a lower level of the likelihood values λ.

During the strange behavior of the cavities, the vector sum forward field increased, driven by the closed-loop controller. This steady increase may have facilitated the quenching of the cavity.

In the subsequent section the time of quenches are investigated.

5.3.3 *Times of Quenches*

To determine the times the quenches occurred, additional quench events, detected after the data collection period of this thesis, were taken under consideration. The detection times and cavities were obtained from the automatically logged records between June and December of 2020. The quench events were separated into those that were detected when piezoelectric detuning control was on and those for which it was off. The results can be seen in Figure 5.17, where the times of occurrence was tracked for each hour in the day and minute of the hour.

The quenches were detected on different days and in six different stations. In total, 23 quench events were considered, of which 18 were with piezoelectric detuning control. Nine of the piezo-on quench events happened in one week, where the EuXFEL was operated at its full gradient, making it more likely for the cavities to quench. Especially cavity M1.C3 of station A7 proved to be prone to quenching; eight of the quench events were detected

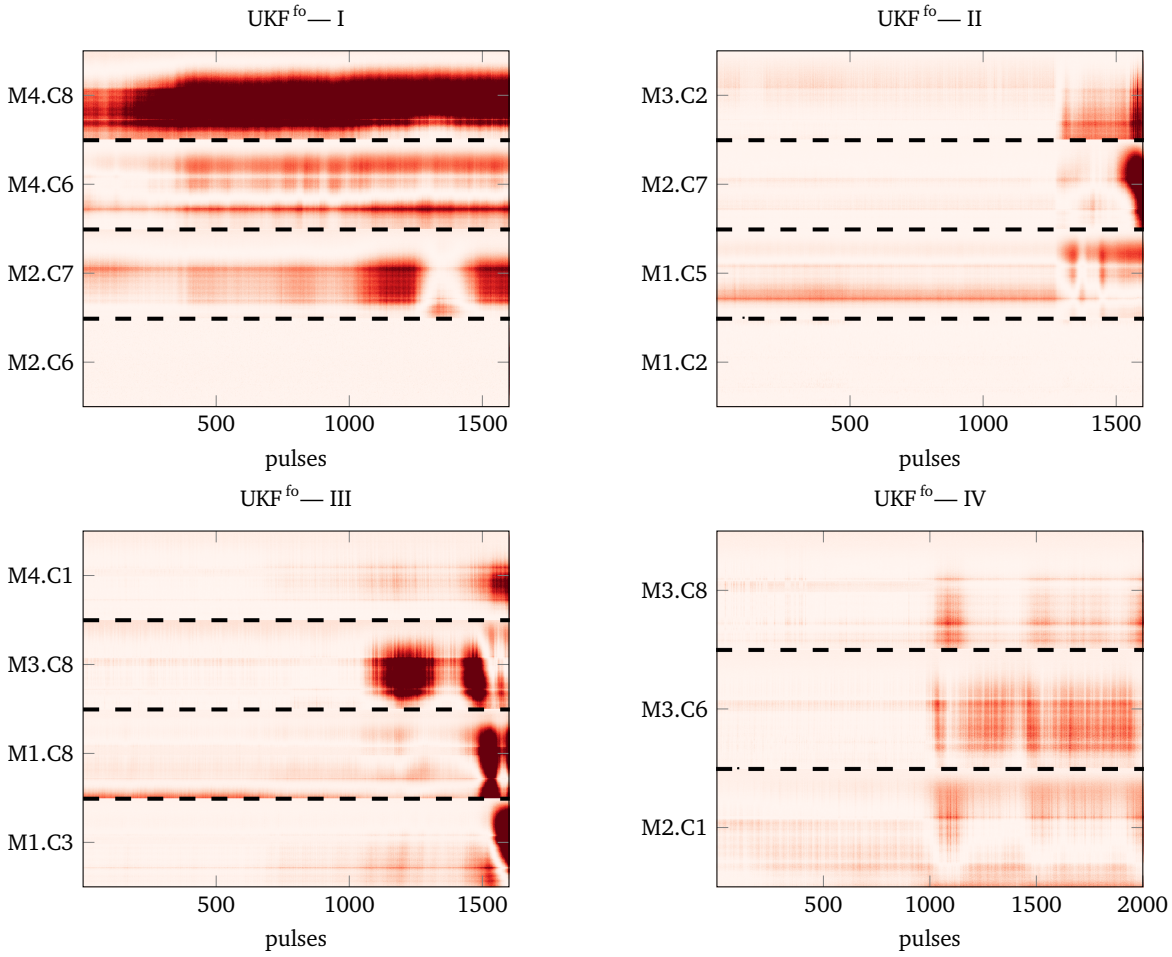


FIGURE 5.16: Selected cavities for four of the five quench events, showing severe anomalous behavior before a quench is detected. For each cavity the y-axis depicts the samples in the pulse, and the x-axis depicts the number of pulses. In quench event I and II, the quenching cavity shows no prior severely anomalous behavior.

in that cavity. For the quenches with piezo-electric detuning control 37% occurred in the first ten minutes of the hour. In the piezo-off cases 75%, i.e., three of four quenches occurred in the first ten minutes of the hour. In addition, almost one-fifth of the detected quenches happened between 6.00 and 6.03am, albeit on different days and cavities. The station of the quenches analyzed in the previous chapter showed that a severe deviation of nominal happened before the quenches occurred, leading to a vector-sum increase of the forward power, which in turn made it more likely for the weakest cavity to quench. The piezo-electric detuning control stabilizing the cavity performance should prevent the station from increasing the vector-sum gradient. Nevertheless, the detected quenches occurred more often in the first ten minutes, as expected if the timing of the quench was random.

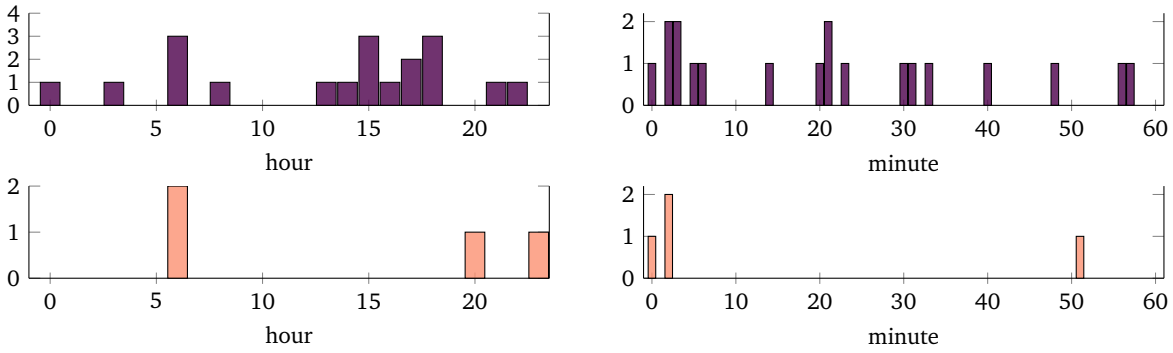


FIGURE 5.17: Distribution of detected quenches for each minute of the hour and hour of the day. Seven out of 18 quench events with running piezo-control were detected in the first ten minutes of the hour, which corresponds to almost 39%.

5.4 QUENCH DETECTION

In the previous section, characteristics of the severely anomalous RF-pulses were analyzed by comparing the distribution of their static detuning and half-bandwidth to those of nominal and anomalous RF-pulses. In this section, the focus is on the detection of quenches rather than on incipient quench behavior. The quenching data sets $\mathcal{Q}$, $\mathcal{Q}^{pz}$ are used to determine how many quenches are classified as severely anomalous by either one of the proposed decisions rules $d^{s,I}, d^{s,II}$. Furthermore, the different methods are compared concerning their ability to detect the first quenching RF-pulse of a quench event.

In the second part of this section, the data from the quenching cavities was used to train GMMs, one for each cavity. For the training of the GMMs, both UKF^{fo} and UKF^{ft} residual generation methods were considered. The detectability of quenches using the GMM is evaluated, determining false positives and false negatives.

5.4.1 *Likelihood Ratio Evaluation*

In Table 5.7 the number of quenches not classified as severely anomalous are given for each decision rule. The number was determined for all piezo-off RF-pulses, i.e., quench events I to IV, as well as for the one piezo-on quench event V^{pz}. In total 18 RF-pulses were classified as quench by the quench detection system for the quench events I to IV whereas for V^{pz} in total 36 RF-pulses were classified as quench. Thus, there are more RF-pulses classified as a quench from piezo-on operation as from piezo-off operation. However, the quenches of quench event V^{pz} all stem from the same cavity.

For both decision rules, the least amount of quenches not classified as severely anomalous were determined using PB. Both times, the quenches came from quench event II. None of the other methods classified these two

	PB		PS		UKF^{fo}		UKF^{ft}	
	$d^{s,I}$	$d^{s,II}$	$d^{s,I}$	$d^{s,II}$	$d^{s,I}$	$d^{s,II}$	$d^{s,I}$	$d^{s,II}$
piezo off	2	2	3	17	2	3	4	3
piezo on	0	0	0	0	1	0	1	0

TABLE 5.2: Number of quenches *not* detected as severe anomalies for each method.

	PB		PS		UKF^{fo}		UKF^{ft}	
	κ_I	κ_{II}	κ_I	κ_{II}	κ_I	κ_{II}	κ_I	κ_{II}
I	27	35	0.5	0.08	2.7	3.4	2.4	3.0
II	0.07	0.3	0.1	0.2	0.08	0.1	0.1	0.07
III	32	86	0.5	0.1	4.1	6.8	1.1	1.6
IV-1	12	59	0.9	0.2	3.4	6.4	1.1	1.7
IV-2	17	42	1.4	0.3	3.0	5.5	0.96	1.3
V^{pz}	21	111	13.4	2.6	0.7	1.6	0.4	1.3

TABLE 5.3: Factors by which the evaluated residuals of the first quench in a quench event exceeded the thresholds of severely anomalous pulses. When the factor is below 1.05, the pulse was not classified as severely anomalous. This is indicated by the gray coloring.

as severely anomalous. The most amounts of quenches not classified as severely anomalous stem from the PS approach. For the piezo-on operation, on the other hand, both PB and PS were equally sensitive to the quenches, classifying all of them as severely anomalous. For quenches detected in piezo-on mode the methods UKF^{fo} and UKF^{ft} show the same performance, whereas in piezo-off mode UKF^{fo} is more sensitive in classifying a quench as severely anomalous than UKF^{ft}. The quenches not classified as severe anomalies for the quench event V^{pz} are the first quenches of the set. In the following, the detectability of the first quench of each set is considered.

Quench events usually have a cascading nature, in which the first quench is the hardest to detect since the quenching of the SRF cavity just started. For each first quench, the factors κ_I and κ_{II} were determined. If $\kappa_I > 1.05$ or $\kappa_{II} > 1.05$, then the quench was classified as severely anomalous. The results can be seen in Table 5.3. Those quenches not classified as severely anomalous are highlighted in the table in gray. The PB method detects most of the first quenches as severely anomalous for both detection rules, with the only exception of quench event II, for which none of the methods detected the first quench as severely anomalous. The factors κ_I and κ_{II} for PB are considerably higher than any of the other methods. The largest factors are obtained by the PB method, reaching values for the examples without piezo-electric detuning control between 12 and 32 for κ_I as well as 35 to 86 for κ_{II}.

The PS approach detects almost none of the first quenches as severely anomalous for the piezo-off quench events I to IV-2. The piezo-on event V^{pz}, however, paints quite a different picture. Here, both detection thresholds are not only surpassed but also the factors by which they are surpassed are higher than those of UKF^{fo} or UKF^{ft}.

The UKF^{fo} and UKF^{ft} approaches both reach the same results, detecting most of the first quenches, except for quench event II with neither decision rule and quench event V^{pz} with the first decision rule $d^{s,I}$. The factors κ_I, κ_{II} for UKF^{fo} are up to three times higher than the ones for UKF^{ft}.

The second decision rule $d^{s,II}$ can also be used to determine at which *point* in the RF-pulse a sample is detected as severely anomalous. Thus the earliest detection of severely anomalous behavior for the first quench was determined. The results can be seen in Table 5.4, where the point

	PB	PS	UKF^{fo}	UKF^{ft}
I	23%	-	30%	29%
II	-	-	-	-
III	56%	-	58%	63%
IV-1	68%	-	70%	73%
IV-2	57%	-	58%	61%
V^{pz}	37%	66%	74%	77%

TABLE 5.4: Percentage of the first quenching RF-pulse after which the severely anomalous detection thresholds β_{th} was surpassed. The earliest a quench was classified as severely anomalous was during the filling of the pulse (quench event I). Most of the quenches classified as severely anomalous were detectable after around 60% of the pulse was over.

of first detection is given in terms of after what percentage of the pulse the severity levels were surpassed. The first quenches were classified as severely anomalous after 23% of the pulse was over for quench event I, which corresponds to the end of the filling. However, most of the detections were possible after 56% and 77% of the pulse was over.

The earliest detection was obtained using the PB method, particularly obvious for V^{pz}.

The subsequent section explores the possibility of detecting quenches via a GMM as proposed in Chatper 4.5.1

5.4.2 *Gaussian Mixture Model Evaluation*

For each of the quench events listed in Table 3.2, four GMMs were considered, see Table 5.5. The first and second GMMs were trained using the UKF^{fo} residuals, and the third and fourth GMMs were trained using the UKF^{ft} residuals.

TABLE 5.5: For each quench event, four GMMs were trained.

	determ.	noisy
UKF^{fo}	1st GMM	2nd GMM
UKF^{ft}	3rd GMM	4th GMM

In addition, the training of the first and third GMMs, included the noise of the residuals, i.e., the estimated covariance matrix of each residual sample (noisy training). For the training of the second and fourth GMMs, only the residuals' mean estimation was considered (deterministic training).

The training of the GMM entails the estimation of the Gaussian mixture parameters $\theta = \{m_i, V_i, w_i\}_{i=1}^{n_c}$, i.e., of the mean m_i, the covariance matrix V_i and the weight w_i for each of the n_c Gaussian components, see Chapter 4.5.1. The EM-based training of the GMMs was initialized with the following choice of parameters:

- The number of Gaussian mixture components n_c was selected from a range between $n_c = 2, ..., 10$. GMMs were trained for each number of classes, and the lowest number with the highest performance with respect to detecting quenches was chosen. For all cavities, the best performing number of classes was either three or four. The performance measure is explained in detail in the subsequent paragraph.

- The initial Gaussian mixture weights were selected to be uniform, i.e., $w_i = \frac{1}{n_c}$.

- The initial mean values were determined using the k-means clustering algorithm, which determines n_c number of clusters with a maximal divergence between the clusters' centers. Initializing the GMM means

with the centers obtained from the k-means algorithm is a common approach [56].

[56] Kerenidis et al., "Quantum expectation-maximization for Gaussian mixture models"

- The initial covariance matrices for each Gaussian mixture component were assumed to be a diagonal matrix in the same order of magnitude as the covariance matrices of the nominal residuals.

The training of the GMMs was performed using randomly picked samples from non-quenching and quenching RF-pulses. The size of the training data set was determined by the number of available quenching RF-pulses, which in turn depended on the quench set data. The smallest amount of quenching RF-pulses was three, i.e., a total number of $3 \cdot 1820$ samples. The largest amount of quenching RF-pulses for one cavity was 36, i.e. $36 \cdot 1820$ samples. As has been shown in the previous section, not all samples of a quenching RF-pulse describe anomalous behavior and thus up to 70% of an RF-pulse was not classified as severely anomalous. To make sure quenching samples are well represented, the ratio between non-quenching and quenching RF-pulses was chosen to be one to two. From the total number of available samples, 70% were used for the training and 20% for the validation for each GMM. The rest of the *quenching* samples was set aside for testing the models.

The sample-based evaluation is defined by (4.50) when the covariance estimations of the residuals are taken into account (noisy evaluation), or by (4.51) when they are *not* taken into account (deterministic evaluation). The evaluation in both cases consists of determining the responsibilities $\rho_{k,i}$ for each Gaussian component i in each sample k. Some of the GMs represent non-quenching and nominal behavior, whereas the rest represent quenches. Since only a binary decision is needed for a quench detection, the Gaussian mixture components were interpreted in the following way: The Gaussian mixture with smallest Euclidean distance from the origin to its mean was interpreted as the "not quenching or nominal" component with $i = i_{nq} \in 1, ..., n_c$. All other GMs with $i = 1, ..., n_c$ except $i = i_{nq}$ were summarized to represent quenches. The quench class is thus represented by a *mixture* of Gaussians, whereas the not quenching class is represented by a *single* Gaussian. Thus the probability of a sample belonging to a quench is determined by

$$\rho_k^{i_q} = \sum_{i=1,...,n_c \setminus i_{nq}} \rho_{k,i}, \tag{5.9}$$

whereas the probability to not belong to a quench is given by

$$\rho_k^{i_{nq}} = \rho_{k,i}, \quad \text{for } i = i_{nq}, \tag{5.10}$$

and

$$\rho_k^{i_q} + \rho_k^{i_{nq}} = 1. \tag{5.11}$$

The classification of the samples is based on a comparison between those two probabilities, formalized by

$$\mathrm{d}^q(\rho_k^{i_q}, \rho_k^{i_{nq}}) = \begin{cases} 1 & \text{if} \quad \sum_{j=k-K}^{k} \rho_j^{i_{nq}} \leq \sum_{j=k-K}^{k} \rho_j^{i_q}, \text{ quench,} \\ 0 & \text{else} \quad \text{not a quench, possibly nominal,} \end{cases} \tag{5.12}$$

	noisy		deterministic	
	UKF $^{\text{fo}}$	UKF $^{\text{ft}}$	UKF $^{\text{fo}}$	UKF $^{\text{ft}}$
FP	0	1	180	543
FN	2	3	2	3
FP*	1	0	0	3
FN*	1	1	0	0

TABLE 5.6: Amounts of false positves (FP) and false negatives (FN) when comparing GMM-based detection results with the classification of the quench detection system.

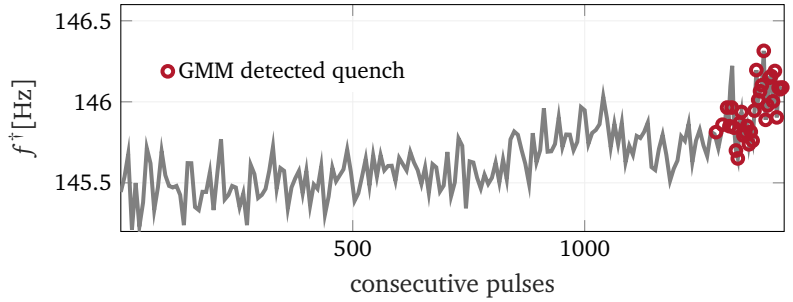


FIGURE 5.18: The evolution of the pulse-wise half-bandwidth values, and the detection of quenches according the GMM evaluation just before a quench was detected.

i.e., when the moving average with a window size of $K = 50$ of the responsibilities $\rho_k^{i_q}$ exceeds the moving average of $\rho_k^{i_{nq}}$, the sample is classified as quenching.

The performance of the GMMs was assessed by the number of false positives and false negatives. The smaller the number of false positives and false negatives, the better the model's performance. Since the quench detection system can only classify entire pulses (intra-pulse detection is not possible), a decision rule was used to derive the sample-level decision onto a pulse-level decision. The pulse was classified as a quench when it contained at least one sample classified as a quench.

The number of false positives and false negatives was determined from 3000 RF-pulses which were randomly picked over the *complete* time-span available for that particular cavity and the entire set of quenching RF-pulses, including the samples set aside for testing. The number of false positives and false negatives are presented in Table 5.6. The total number of quenching RF-pulses available from the data sets I to IV-2 are 18, whereas 36 quenching RF-pulses are included in the quenching data set V^{pz}. At first glance, the noisy approach's performance appears to be better, i.e., fewer false positives and false negatives than the deterministic approach. In particular, the number of false positives for both UKF $^{\text{fo}}$ and UKF $^{\text{ft}}$ residuals is higher for the deterministic approach. However, all detected false positives turn out to be detected *just before* the first quench. Figure 5.18 shows the pulses in which false positives were detected concerning the pulse-wise half-bandwidth values from those RF-pulses. Using the deterministic GMM evaluation approach, it was thus possible to detect quenches several pulses before they occurred.

The false negatives mainly stem from the first quench. While the GMMs obtained from noisy training can detect three of six first quenches, the deterministic approach is capable of detecting four out of six. For the piezo-off examples I to IV-2, both training methods (noisy/deterministic) produce

	noisy		deterministic	
	UKF ^{fo}	UKF ^{ft}	UKF ^{fo}	UKF ^{ft}
A	34%	29%	20%	10%
B	-	-	-	-
C	59%	60%	29%	25%
D1	-	-	-	-
D2	58%	58%	56 %	55%
E*	-	-	79%	74 %

TABLE 5.7: Percentage of pulse after which the detection classified the samples as a quench.

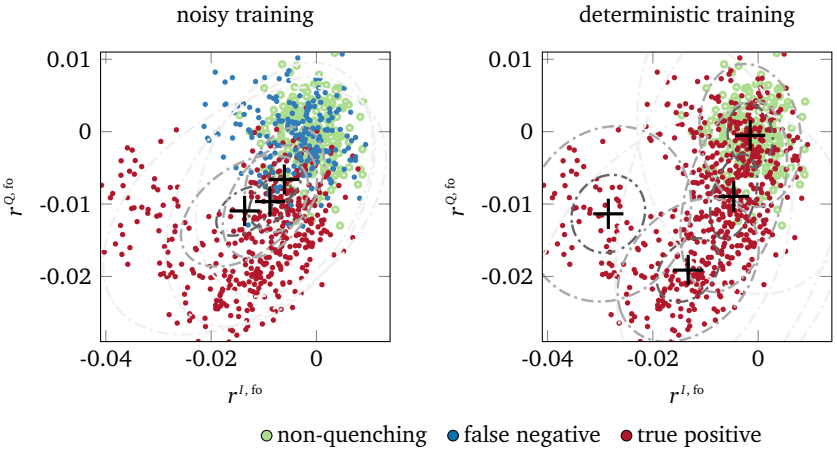


FIGURE 5.19: UKF ^{fo} residual data of quenches and non-quenching behavior. The Gaussian mixture components are represented by their means and variance and were trained using the noisy data (left) and the deterministic data (right). False negatives describe those quenches which were not detected by the GMM.

the same amounts of undetected quenches.

The detectability of the first quenches is further examined, determining after which percentage of the RF-pulse the quench was detected. The results can be seen in Table 5.7.

In Figure 5.19 the data of non-quenching and quenching RF-pulses is shown, as well as the mean and covariance of the trained GMMs using the noisy training (left) and the deterministic training (right). The best performance was obtained using three Gaussian mixture components for the noisy training and four components for the deterministic training. The green samples represent non-quenching RF-pulse behavior. The blue samples represent a false negative, i.e., a pulse that was wrongly classified as non-quenching. While the noisy trained GMM does not entail a Gaussian mixture component that represents the green (non-quenching) RF-pulse adequately, the deterministically trained does. The GMM trained with noisy data consideration was not able to detect the first quench, whereas the determinstically trained GMM detected it correctly. In Figure 5.20 the data samples over each RF-pulse are smoothed using a moving average of $K = 50$ samples. The data represented is thus directly generated from the same residuals of Figure 5.19. The difference is that through smoothing the residual data over the pulse, it becomes clear that multiple of the samples belonging to the false-negatives clearly differentiate themselves from the non-quenching behavior.

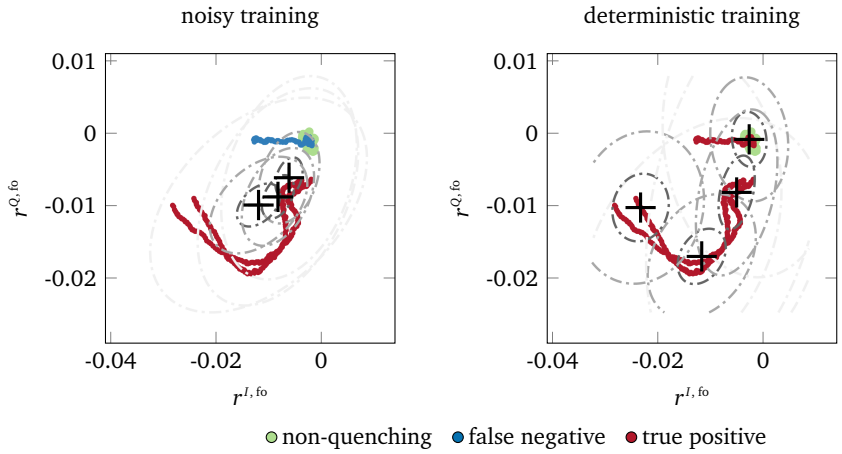


FIGURE 5.20: The moving average of the UKF$^{\mathrm{fo}}$ residuals are presented for the same pulses as depicted in Figure 5.19. All quenches have distinct traces in the residual space and can be distinguished from the non-quenching residual data, close to the origin.

The differences between the resulting GMMs of noisy and deterministic training can be explained as follows: The marginal mean (4.64) which is needed for the GMM training, can be reformulated into

$$\tilde{\mathrm{m}}_{R_{k,i}} = \vec{\mathrm{V}}_i \left(\vec{\mathrm{V}}_i + \vec{\mathrm{V}}_{R_k}\right)^{-1} \overleftarrow{\mathrm{m}}_{R_k} + \vec{\mathrm{V}}_{R_k} \left(\vec{\mathrm{V}}_i + \vec{\mathrm{V}}_{R_k}\right)^{-1} \overleftarrow{\mathrm{m}}_i \,. \tag{5.13}$$

The updated parameter for the GMM is determined by the weighted average over these marginal means. The higher the uncertainty of the (observation) sample $\vec{\mathrm{V}}_{R_k}$, the more "trust" is put onto the Gaussian mixture mean $\overleftarrow{\mathrm{m}}_i$ in each iteration step.

The covariance $\vec{\mathrm{V}}_{R_k}$ is the unscented Kalman filter residual covariance estimate. The covariance estimated by a linear Kalman filter depends entirely on the process noise covariance and measurement covariance and always decreases with time [104]. The estimated state covariance using an unscented Kalman filter depends on the nonlinear function. For the nominal RF-pulses, the covariance showed a highly repetitive behavior, while the covariance during a quenching pulse either increased or decreased relative to the nominal covariance. In this particular example, the covariance of the residual decreased for many of the anomalous pulses, thus putting a greater weight onto the sample means, with the result that the GMM means drifted towards those samples with low covariance.

[104] Simon, "Optimal state estimation: Kalman, H, and nonlinear approaches. Hoboken"

5.5 DISCUSSION & SUMMARY

The following summarizes the main findings of this chapter and discusses aspects of the results.

Discussion

The classification thresholds ψ_{th} were determined from the χ^2 distribution according to the acceptable false alarm rate of $\epsilon = 0.01$. Thus, by design, even when only nominal pulses were observed, 0.01% would be falsely classified as anomalous. However, the percentages of detected anomalies in the non-severe data sets $\mathcal{T}, \mathcal{T}^{pz}$ were higher than $\epsilon = 0.01$, which means that the methods detected statistically significant changes in the RF-pulses.

To determine in what sense those detected divergences do not conform to the expected nominal behavior and whether they carry valuable information about RF-pulses and/or cavity behavior, measures of divergence were analyzed.[6]

[6] "Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior. [...] Anomalies might be induced in the data for a variety of reasons, [...] but all of the reasons have the common characteristic that they are interesting to the analyst."—V. Chandola et al. [90]

It was shown that the mean divergence m^{λ} of the detected anomaly changes quadratically with the static detuning values $\bar{f}^{\Delta}$, i.e., stronger anomalies can be associated to stronger changes of static detuning. The anomalies detected in $\mathcal{T}, \mathcal{T}^{pz}$ can thus be associated to microphonics, i.e., with disturbances acting on the cavities' detuning, see 2.2.3. Microphonics are part of any cavity operation and are treated as unavoidable disturbances for which, e.g., the piezo-controller was designed. For the selection of nominal RF-pulses in Chapter 3 and in line with expert opinion [32], it was assumed that microphonics leading to a change of detuning of ±25 Hz was acceptable. However, the analysis showed that the cavities were subject to much higher microphonics associated with changes in the static detuning of up to ±100 Hz.

[32] Schilcher, "Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities"

The quadratic relationship between static detuning and mean likelihood ratios reflects the property of the likelihood ratio function, see Figure 4.7. This shows that a change of static detuning corresponds to a divergence of the residual distributions from their nominal, zero-mean distribution, as assumed by the zero hypotheses described in Chapter 4.3.2.

The amounts of anomalies detected in nominal operation were determined for both a station operated with piezo-electric detuning control and one without piezo-electric detuning control, which lead to considerably different detection results. When no piezo-electric detuning control was running, more anomalies were detected than when operated with piezo-electric detuning control. Since the piezo-controllers stabilize detuning during the flattop, the cavities are more robust against microphonics. It should be noted, however, that the anomaly detection methods evaluated the *entire* RF-pulse (filling, flattop, and decay), whereas the detuning control only acts on the flattop. Microphonics acting on filling and decay may thus also be detected, even if, due to the piezo control, the flattop would not show anomalous values.

Looking at the distribution of mean divergence, and maximum divergence for all detected anomalies shows that around 50% of the anomalies, independent of the methods are benign, only barely exceeding the thresholds between nominal and anomalous. More interesting insight was gained by focusing on the strongest anomalies per cavity.

It was also shown that around 50% of the anomalies with strongest mean divergence stem from the first ten minutes of the hour caused by some so-far unclassified "hourly event". These events occurred (more or less severely) in all cavities of all considered stations and were shown to be associated with a deviation of both half-bandwidth and static detuning. While the static detuning either increased or decreased, the half-bandwidth decreased. A decrease of the half-bandwidth $f^{\dagger}$ is the same as an increase of the loaded quality factor Q^{L}, depending on the loaded and external quality factor, i.e.,

$$Q^{L} = \frac{1}{\frac{1}{Q^{0}} + \frac{1}{Q_{ext}}}. \tag{5.14}$$

An increase of the loaded quality factor Q^L is most likely caused by an increase of the external quality factor Q_{ext}. The external quality factor can change when the coupling of the RF signal into the cavity changes. For example, it increases when a larger portion of the coupling antenna reaches into the cavity. At this point, the cause for the changes is subject to speculation. For example, it is possible that a mechanical or fluid-mechanical phenomenon had occurred each hour, which causes the cavities to be subject to additional pressure.

Detecting the hourly events was not vital to the safe operation of the cavities since, so far, it could not be shown that they caused downtime. To separate anomalies of similar and lower severity than the hourly events, two detection thresholds were introduced. The thresholds were chosen such that *none* of the anomalies so-far encountered would classify as severely anomalous. This includes RF-pulses of 64 cavities (two stations) and several *weeks* of operation. It turned out that severely anomalous (and so-far undetected) RF-pulse behavior could be identified before three of the six quench events and in the data sets containing RF-pulses of cavity C6.M1.A17.L3.

It was shown that the severe anomalies detected for A17.M1.C6 are also associated with considerable changes in the static detuning and half-bandwidth parameters. The static detuning of the pulses changed both in the positive and in the negative direction obtaining values above 100 Hz. At the same time, the values for the half-bandwidth increased. Other than for the hourly events, this suggests that it was not the external quality factor Q_{ext} causing the change of half-bandwidth, but the unloaded cavity factor Q^0. The amount of change was so high that it nearly resulted in the detection of a quench. This cavity might have suffered from irregular soft quenches over several days without any currently implemented system being able to detect it. It can be assumed that as a result of this the RF-station's forward set-points increased slowly over time, compensating for the energy loss due to this abnormal behavior of the cavity. The increase of the set-point, on the other hand, also meant that the RF station was slowly pushed further towards its limit, making quenches more likely to happen. Indeed, the two quench events I and II happened while C6.M1.A17.L3 behaved in an unexpected manner. After the onset of anomalous RF-pulses in C6.M1.A17.L3, the vector-sum drive slowly increased. Then quench event I happened. After restarting the station, C6.M1.A17.L3 still behaved in an unexpected manner, again causing a steady increase of the forward drive, and after another day, quench event II happened. If the behavior of C6.M1.A17.L3 had been detected at the time, an increase of forward power that compensates the losses would not have been necessary, and both quench events could have possibly been prevented.

Severe anomalies detected before a quench had similar consequences on the vector sum. The severely anomalous RF-pulses could be associated with a considerable increase of static detuning and an increase of the half-bandwidth in at least one cavity of the affected station. As a result, the vector sum control increased the forward power, making a quench more likely. What brought on this kind of severe detuning change of half-bandwidth should be investigated. Due to the stabilizing effect that the piezo-control has on the system, such severe detuning should no longer lead to an increase of

forward power. Assuming that the increase of RF power makes the quenches more probable, they should thus become less likely when operated with piezo-electric detuning control. However, 38% of the quenches from the piezo-on operation were detected in the first ten minutes of the hour. This may suggest that the identified hourly events have an influence on the likelihood of quenches, which are still a limiting factor to the maximum gradient operation of the EuXFEL and novel insights into the mechanisms behind "spontaneous quenches" can help to prevent them in the first place.

The detectability of quenches was further analyzed, determining the amounts of quenches classified as severely anomalous. The time of detection during a pulse was determined for the first quench of each quench event. It was also shown that the majority of first quenches analyzed were detectable after 70% of the RF-pulse was over, which is approximately 200 µs before the end of the flattop. These results suggest that an intrapulse detection followed by an adequately quick reaction, i.e., the lowering of the station's set-point, may be able to prevent the station from quenching.

In the final part of the chapter, GMMs were trained to detect quenches. For each quench event GMM models were trained using the $\mathrm{UKF}^{\mathrm{fo}}$ and $\mathrm{UKF}^{\mathrm{ft}}$ residuals once assuming them to be noisy and once assuming them to be noise-free. The detection results showed very promising behavior in detecting quenches, and no false positives were obtained. Using GMMs, quenches were detected earlier in the pulse than with the severely anomalous thresholds using the same residuals and—in some cases—even several pulses prior to the first detection of a quench.[7]

[7] Understanding the mechanics behind quenches and developing detection methods have been studied for many years [105].

The chapter also gave insight into the detection capabilities of the proposed residual generation methods. While the parity space approach showed the least sensitivity towards changes in the RF-pulse behavior correlated to static detuning changes, the parameter-based methods turned out to be very sensitive towards changes, detecting over 50% of all analyzed samples as anomalous. However, for the detection of quenches, the parameter-based residuals showed the most promising results, with the least amounts of undetected quenches and the earliest detection during the first quench. The residuals' properties are described and analyzed in greater detail in the subsequent chapter, where all model assumptions on the residuals are evaluated and critiqued.

Summary

In this chapter the detection results of analyzing several weeks of RF-pulses was described. Metrics with which the severity of the detected anomalies could be evaluated were defined as well as detection thresholds on those metrics with which severe anomalies could be separated from the rest. It was shown that the severity of the anomaly increased quadratically with increasing changes of static detuning values and around 50% of the strongest anomalies in each cavity occurred in the first ten minutes of the hour. Severe anomalies were detected before some of the analyzed quench events, making a severe detection before the onset of the quench possible. Severe anomalies were also detected in a set of RF-pulses for which no fault was detected. The detection of quenches was further analyzed showing that the

parameter-based residual generation is the most sensitive towards quenches, detecting most of the quenches as severely anomalous and having the earliest time in the pulse for the first quench in which the samples exceed the severely anomalous thresholds. The evaluation using GMMs showed promising potential, being able to detect quenches early on in the pulse. Using the timings of quench detection, it was shown that they, too have a disproportionately large number in the first ten minutes of the hour. Further investigations are necessary to find out the reason for the "hourly events" as well as the reason why the quenches occur more often in the same time slots.

6

Model Performance & Criticism

"All models are wrong"
—Box, 1976

► SYNOPSIS In this chapter both cavity models and the proposed residuals are criticized regarding their performance in *nominal* conditions. First the two detuning models described by (3.7) and (3.13) are evaluated with respect to their simulation capability, i.e., how well they determine the outputs of the cavity for given inputs. Then, characteristics of the residuals generated from nominal RF-pulses are determined. The parameter-based method (PB), the parity space method (PS), and the two unscented Kalman filter residuals from the first-order detuning model (UKF^{fo}), and the fixed detuning trace input (UKF^{ft}) are analyzed regarding their nominal properties. Finally the methods' sensitivity towards set-point changes and the effect of concept drift is analyzed.

OVERVIEW

Model criticism seeks to determine whether a model is adequate for its task,[1] to identify its shortcomings, and determine where the model needs to be improved [106, 16]. The models introduced in Chapter 3, formalize assumptions about nominal cavity behavior. Using these models, residuals were defined, whereas the fact that the RF-pulses have a repetitive nominal behavior played an important role. These assumptions are cross-validated with the set of validation data $\tilde{\mathcal{M}}, \tilde{\mathcal{M}}^{pz}$ introduced in Chapter 3 using a normalized mean squared error between the simulated and measured system outputs. The previous chapter has shown, that the generated residuals are capable of revealing interesting and so-far undetected cavity behavior. However, differences in the residual generation methods were also noticeable leading to different detection results between the methods. Reasons for these differences are analyzed in this chapter by determining the validity of assumptions used to generate the residuals. A "good" residual is zero-mean Gaussian distributed when generated from nominal observations. Shortcomings of the residual model are pointed out, i.e., where they fail to produce a zero-mean Gaussian. In the final part of this chapter, the residual generation methods' sensitivity towards set-point changes is tested as well as the influence of concept drift on their detection capabilities.

All analyses are performed using the nominal RF-pulses of station A17.L3 without and with piezo-electric detuning control.

[1] A measure of how well a model performs is dependent on what task the model was built for. If, for example, the model is used to obtain predictions on a certain parameter, then its ability to correctly make predictions of that parameter determines the model quality.

[106] Guttman, "The use of the concept of a future observation in goodness-of-fit problems"

[16] Blei, "Build, compute, critique, repeat: Data analysis with latent variable models"

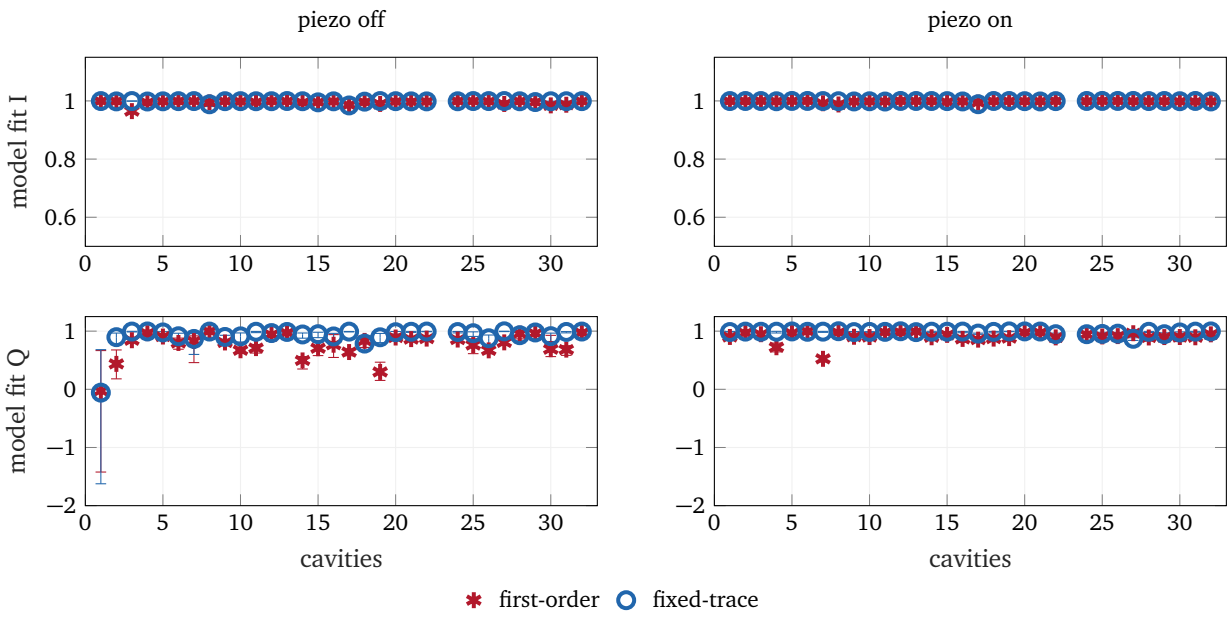


FIGURE 6.1: Model fits for cavity forward simulations using the 'first-order' as defined in (3.7, 3.11) and the 'fixed-trace' detuning mode as defined in (3.13). Cavity 23 (M3.C7) was not taken into consideration as it was not operated for beam acceleration.

6.1 DETUNING MODELS

In Chapter 3.2, two models for the RF-input dependent dynamic detuning were proposed. The first uses a set of first-order differential equations to emulate the detuning of the cavity over the course of an RF-pulse, and the latter uses the nominal detuning estimations as an additional input to the system. To determine the model performance, the inputs to the cavities $M^{F,I}, M^{F,Q}$ of $\tilde{\boldsymbol{\mathcal{M}}}, \tilde{\boldsymbol{\mathcal{M}}}^{pz}$ are used to simulate outputs, which are then compared to the respective outputs $M^{P,I}, M^{P,Q}$. A comparison between the simulation outcomes and the output data for all RF-pulses of the validation sets are used to determine cavity-wise goodness of fits as well as the effect of model uncertainties over the RF-pulses. The cavity model has two inputs, the I and Q component of the forward RF wave, and two outputs, i.e., the I and Q components of the RF probe wave. The model fits were determined for each output individually.

6.1.1 *Cavity-Dependent Model Performance*

The model performance for each cavity model is evaluated using the normalized mean squared error of every RF-pulse in $\tilde{\boldsymbol{\mathcal{M}}}, \tilde{\boldsymbol{\mathcal{M}}}^{pz}$, which normalizes between minus infinity and one, with one being the optimal fit, i.e.,

$$\text{model fit} = 1 - \frac{(y_k - \hat{y}_k)^2}{(y_k - \frac{1}{n_s}\sum_{k=1}^{k=n_s} y_{:})^2}, \tag{6.1}$$

where $\hat{y}_k$ is the simulated model output and y_k is the measurement. From the cavity-wise model fits obtained from each RF-pulse, the quantiles $q = 0.1, 0.5$ and $q = 0.9$ are determined. The results can be seen in Figure 6.1, where the fits for the I and Q components are shown with the lowest boundary representing $q = 0.1$, the middle marker representing $q = 0.5$ and the highest boundary representing $q = 0.9$.

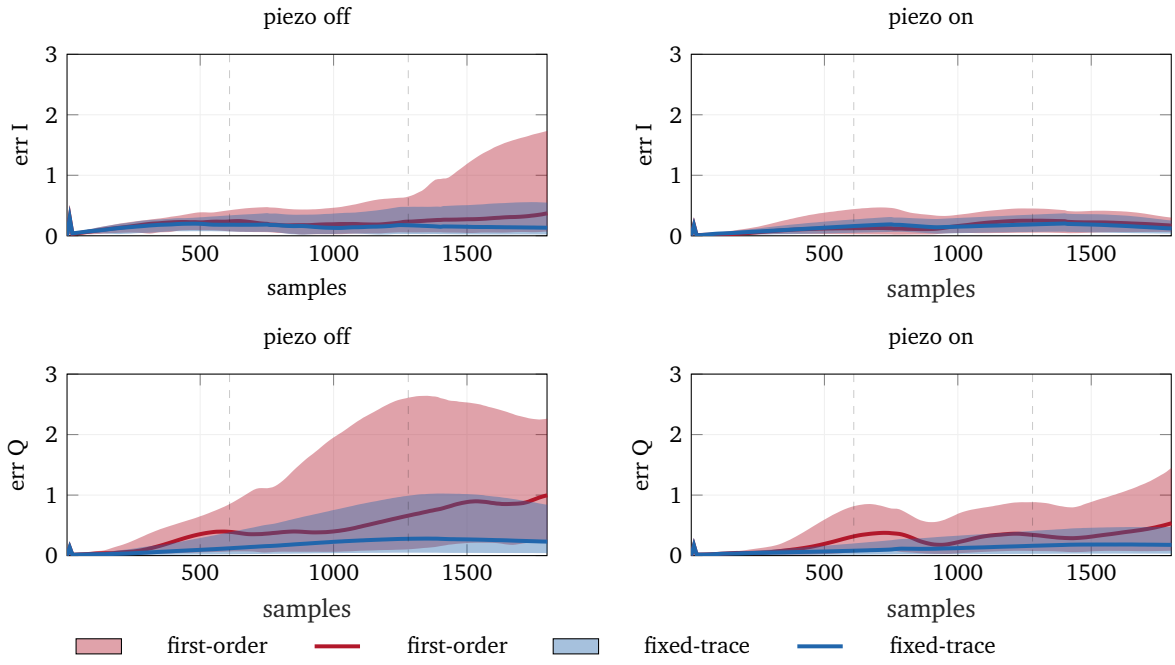


FIGURE 6.2: Absolute forward simulation errors represented with the values of $q = 0.1$, for the first-order detuning model and the fixed-trace model in both I and Q components. The lower and upper boundaries of the shaded areas denote $q = 0.5$ and $q = 0.9$ over all RF-pulses for each sample.

The I component can be simulated well with either first-order or fixed-trace models, both resulting in an accuracy for all cavities above 0.96, and small pulse to pulse variations (high precision). The Q component, however, is more accurately modeled by the fixed-trace model, with fits above 0.7 as opposed to fits above 0.3 for the first-order model. An exception to this performance is the first cavity, which shows considerably lower accuracy and precision for both models. For the piezo-on mode of operation, the models for both I and Q gain both accuracy and precision for both first-order and fixed-trace detuning model.

6.1.2 *Modeling Errors Over the RF-pulse*

In the above section, a goodness of fit for the overall RF-pulses was determined. For this paragraph, the absolute simulation error of the I and Q components for each pulse and cavity was determined. Then the quantiles of $q = 0.1, 0.5, 0.9$ were determined over all cavities i_c and pulses i_p for each sample. The results can be seen in Figure 6.2, where the sample-wise $q = 0.5$ values are depicted by a solid line and $q = 0.1$ and $q = 0.9$ are represented by the lower and upper bounds of the shaded areas. For a piezo-off mode of operation, the first-order and the fixed-trace model have similar $q = 0.5$ values for the I component (err I), whereas the values of $q = 0.9$ of the first-order models increase over the decay phase of the pulse, surpassing those of the fixed-trace model and reaching their highest values at the end of the pulse. For the Q component (err Q) the first-oder model has larger absolute values which increase from filling over flattop to the decay, reaching higher values than err I. In piezo-on operation, both models perform similarly in the I component, for the Q component, the first-order model resulted in higher absolute simulation errors. As in the previous section, model errors are lower for piezo-on operation mode.

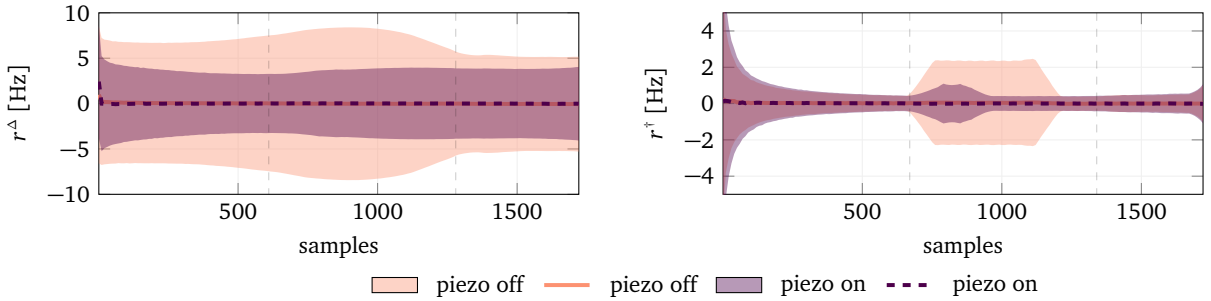


FIGURE 6.3: Mean and standard deviation over the RF-pulse of the parameter based residuals $r^{\dagger}$ and r^{Δ} for all cavities in $\tilde{\mathcal{M}}, \tilde{\mathcal{M}}^{pz}$. Both the divergence of the bandwidth and the detuning are frequencies and their unit is thus in Hz.

6.2 RESIDUAL CHARACTERISTICS OF NOMINAL RF-PULSES

In this section characteristics of the proposed residuals (PB, PS, UKF$^{\text{fo}}$, UKF$^{\text{ft}}$) generated from nominal RF-pulses are analyzed. Residuals generated from $\tilde{\mathcal{M}}$ are referred to as piezo-off residuals and residuals from $\tilde{\mathcal{M}}^{pz}$ are referred to as piezo-on residuals. The resulting residuals of all cavities are analyzed with respect to their mean and standard deviation over the pulse.

6.2.1 *Parameter-Based Residual*

The parameter-based residuals $r_k^{\text{PB}} = [r^{\dagger}, r^{\Delta}]$ were determined as described in (4.16), with $r^{\dagger}$ and r^{Δ} representing the bandwidth and detuning divergences over the pulse from their nominal mean values, respectively. Figure 6.3 shows the mean and standard deviation for each sample in the pulse of all considered cavities. Both residuals describe the divergence of parameters which represent frequencies, and their units are thus in Hz. On average (over the cavities), both piezo-on and piezo-off residuals based on the detuning estimation have mean values close to zero for each sample in the RF-pulse. The standard deviation of the piezo-off residuals is lower than for the piezo-on residuals, the former reaching a maximum of 8 Hz.

The residuals based on the bandwidth estimations also have mean values close to zero over the RF-pulse for both piezo-on and piezo-off operation. The standard deviation is higher at the beginning of the pulse and shows considerable increase during the flattop phase of the pulse at those time intervals in which the beam was switched on. It should be noted that the intervals of beam for the piezo-on data differed from the piezo-off data. Hence the different times and lengths of divergence.

Figure 6.4 shows the results of analyzing to what degree each parameter based residual is Gaussian distributed for a single cavity. Again a distinction between piezo-on and piezo-off residuals was considered. The sample distribution is compared to the theoretical, Gaussian distribution. A reference line is drawn between the first quartile $q = 0.25$ and the third quartile $q = 0.75$ of the data, and extended to encompass all data values. When the residuals are indeed Gaussian distributed, they align with the reference line describing the Gaussian distribution. Although the figure represents only the pulses of one cavity, qualitatively similar results were found for all cavities with this method.

The residuals based on the detuning values r^{Δ}, are close to a Gaussian distribution when operated with piezos. When the piezo-electric detuning

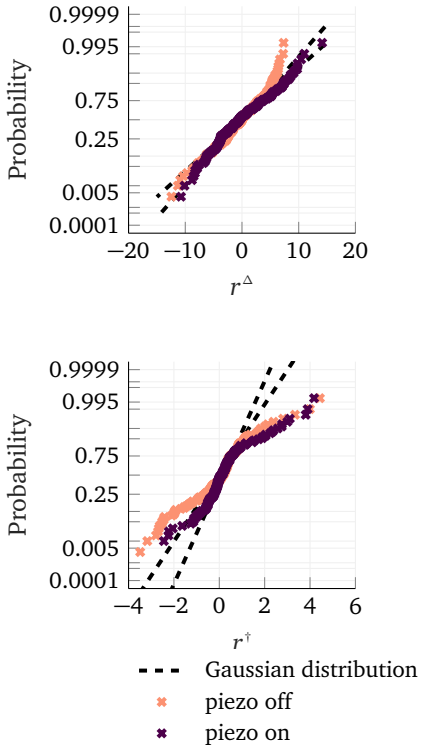
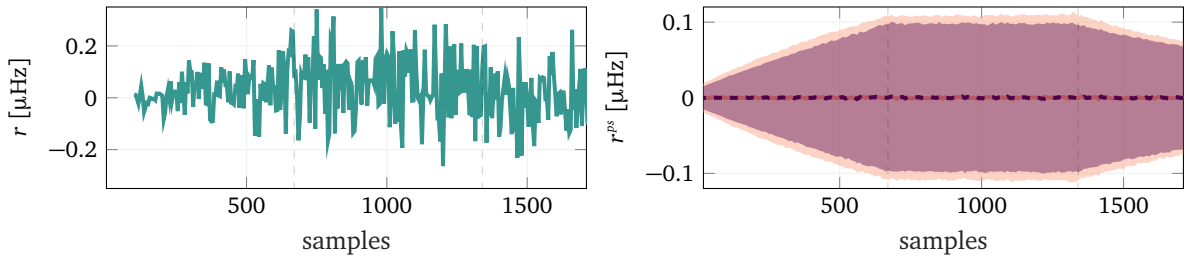


FIGURE 6.4: Comparison of the parameter-based residuals to a Gaussian distribution.



(a) Parity space residual over a single pulse.

(b) Mean and standard deviation for all nominal pulses.

FIGURE 6.5: Parity space residuals r^{ps} generated from nominal RF-pulses $\tilde{\mathcal{M}}$ and $\tilde{\mathcal{M}}^{pz}$.

control is off, their distribution diverges from a Gaussian. The bandwidth based residuals $r^{\dagger}$ both with and without piezo-electric control are not Gaussian distributed.

6.2.2 *Parity Space Residual*

The parity space residual r^{ps} determines the difference between two analytically redundant expressions of the detuning for each sample in the pulse, see (4.22) which resulted in non-Gaussian distributed nominal traces. The average of these nominal sample-wise values are subtracted to (ideally) obtain a zero mean residual, see (4.24). The result of a single RF-pulse parity space residual is depicted in Figure 6.5(a).

The values of the residual are between ± 0.25 µHz. The mean and standard deviation over all nominal piezo-on and piezo-off parity space residuals are depicted in Figure 6.5(b). While the residual mean is close to zero over the pulse, the standard deviation increases during the filling, stays constant during the flattop and decays during the decay phase of the RF-pulse.

Other than for the parameter-based residuals the difference between the piezo-on and piezo-off residuals are small, with the piezo-off residuals showing slightly lower standard deviation values.

Both piezo-on and piezo-off residuals can not be perfectly represented by a Gaussian distribution as can be seen in Figure 6.6, where the lowest and highest values diverge from the optimal Gaussian line. The reasons being two fold. For one, the values of the residuals are restricted by the quantization steps of the analogue to digital converter (ADC) and are thus not continuous. Secondly, the standard deviation in filling and decay are lower than in the flattop.

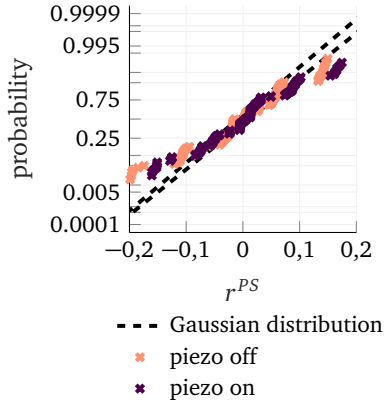


FIGURE 6.6: Gaussian distribution compared to distribution of the parity space residual.

6.2.3 *Unscented Kalman Filter Residuals*

In this section quantiles of the generated residuals over all RF-pulses in the nominal datasets are described, including the differences noted in the ones with and without piezo-electric detuning.

Figure 6.9 shows the I and Q residuals for both detuning models of all considered cavities. The averaged residual of the I component r^{I} of UKF$^{\text{fo}}$ and UKF$^{\text{ft}}$, are similar both in mean and in standard deviation over the pulse. For both piezo-on and piezo-off residuals, the mean values are close to zero with a divergence burst at the end of the flattop and during the intervals

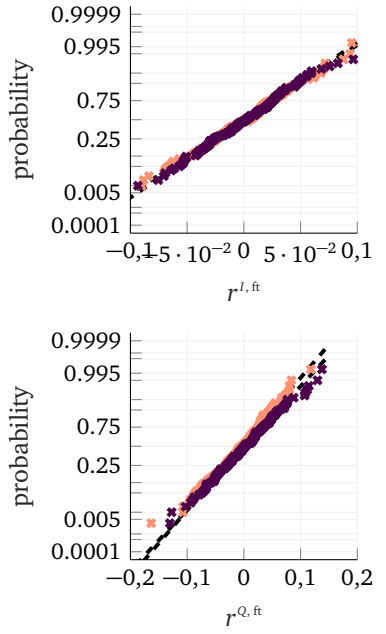


FIGURE 6.7: Distribution of UKF$^{\text{ft}}$ residuals compared to a Gaussian distribution. The residuals of UKF$^{\text{fo}}$ generate qualitatively the same results.

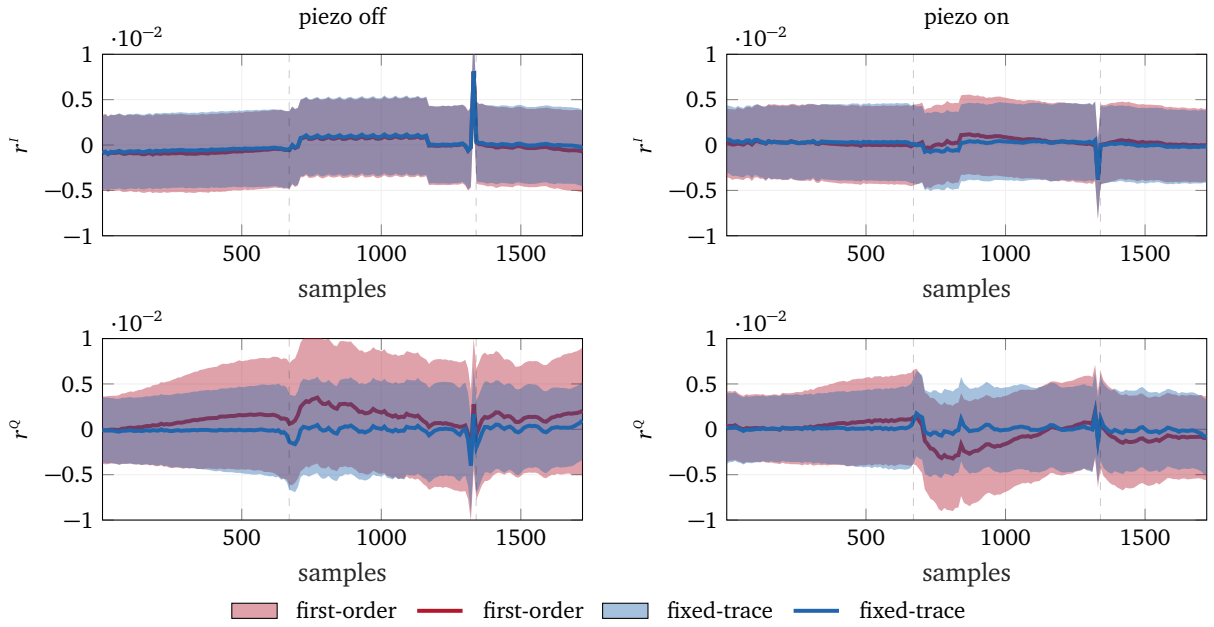


FIGURE 6.8: Piezo-electric detuning control off

FIGURE 6.9: Mean and standard deviation of the residuals generated from the datasets $\tilde{\mathcal{M}}, \tilde{\mathcal{M}}^{pz}$ for each sample using UKF^{fo} and UKF^{ft}.

of beam loading during the flattop. The residuals r^Q for the Q components show a difference between UKF^{fo} and UKF^{ft}. While UKF^{fo} diverges from a zero-mean for both piezo-on and piezo-off residuals, the divergences for UKF^{ft} are benign. Again, the transition from flattop to decay, but also from filling to flattop is noticeable.

Additional divergences during beam-loading are noticeable during the flattop, for which the mean residuals either slightly increase (piezo-off) or decrease (piezo-on). Both for I and Q the residuals have values in the region between ±0.005.

Residuals generated from the unscented Kalman filter are Gaussian distributed. This is confirmed for UKF^{fo} and UKF^{ft} in Figure 6.7, in which the nominal residuals of one cavity are shown. However, for those cavities for which the model was less accurate, the Gaussianity is not as good.

6.3 GLR VALUES OF NOMINAL RESIDUALS

In this section the nominal residuals are evaluated with the likelihood ratios described in Chapter 4.3.

The likelihood ratios of all considered cavities are analyzed over the RF-pulse for each method. A distinction between piezo-on and piezo-off is made and the first, fifth and ninth deciles are used to characterize the evaluated data.

6.3.1 *Cavity-Wise Characteristics*

The cavity-wise likelihood ratios were determined from the nominal data sets $\tilde{\mathcal{M}}, \tilde{\mathcal{M}}^{pz}$ and their distributions were analyzed. The results can be seen in Figure 6.10. Each bar represents the $q = 0.1$ as its lowest value, and $q = 0.9$ as its highest value. The fifth decile value is represented by the

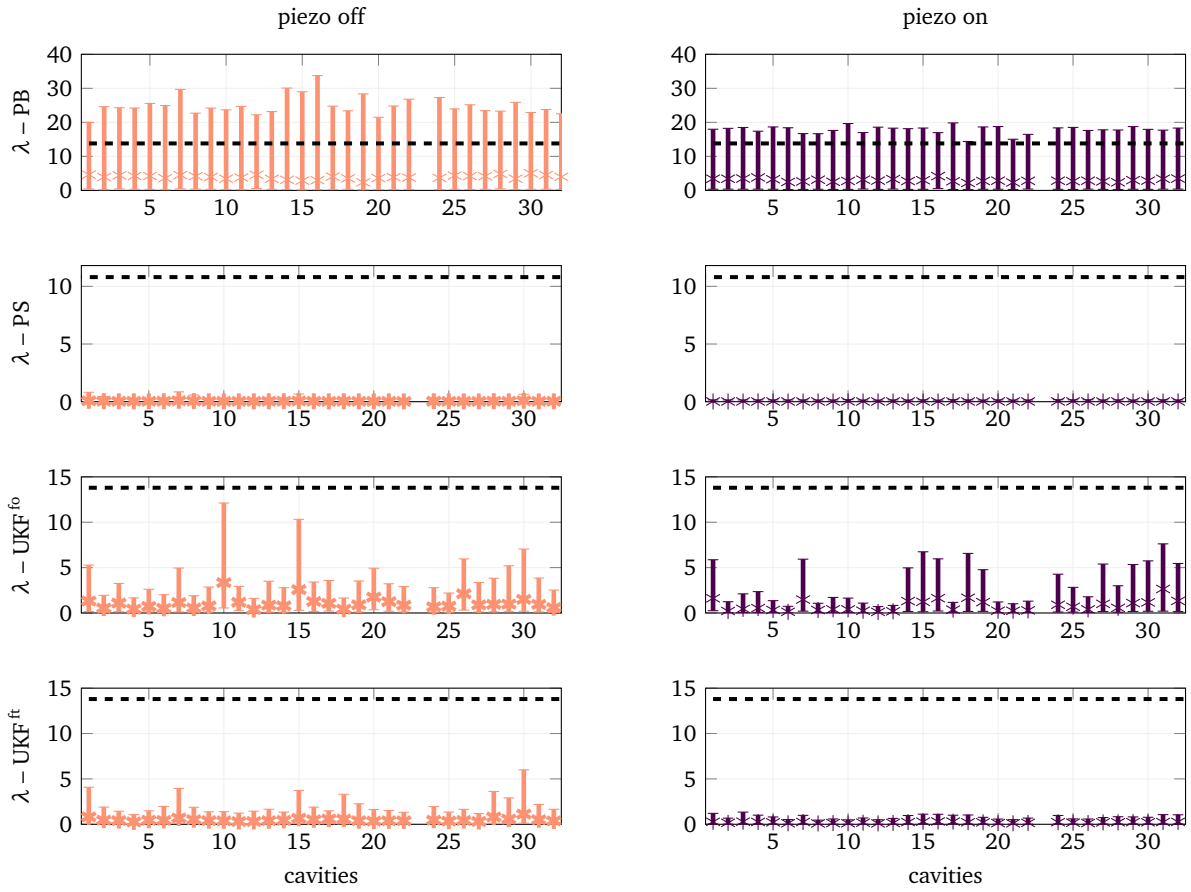


FIGURE 6.10: Likelihood ratio values λ for each of the 32 cavities represented by their $q = 0.1, 0.5, 0.9$ values. On the left, operated without piezoelectric detuning control. On the right, operated with piezoelectric detuning control.

middle marker. The corresponding thresholds ψ_{th} were added to the graphs. If all residuals do indeed stem from a zero-mean Gaussian distribution, the values for $q = 0.5$ should be at 2.7 and 4.7 for one and two degrees of freedom, respectively.

The parameter-based method PB values of all cavities exceed the detection threshold, and little variance between the cavities was observed when comparing $q = 0.5$ or $q = 0.9$. This means that the thresholds are exceeded for all cavities when evaluating nominal residuals. The PS values are far below the detection threshold. Both unscented Kalman filter likelihood ratios vary from cavity to cavity, whereas UKF^{fo} generated the larger likelihood ratio values, e.g., for cavity 10 or 15.

The piezo-on residuals show lower variations in general. The variations from cavity to cavity are lower as are the values of $q = 0.9$. For the PB method, for example, the $q = 0.5$ values are similar for both piezo-on and piezo-off residuals, but the $q = 0.9$ values are notably smaller across all cavities.

Pulse-Wise Characteristics

Taking the complete set of likelihood ratios of all cavities in A17.L3, the quantiles $q = 0.1$, $q = 0.5$ and $q = 0.9$ for each sample in a pulse was determined for every method. The results can be seen in Figure 6.11, where a dashed line is included, indicating the detection thresholds. The fifth

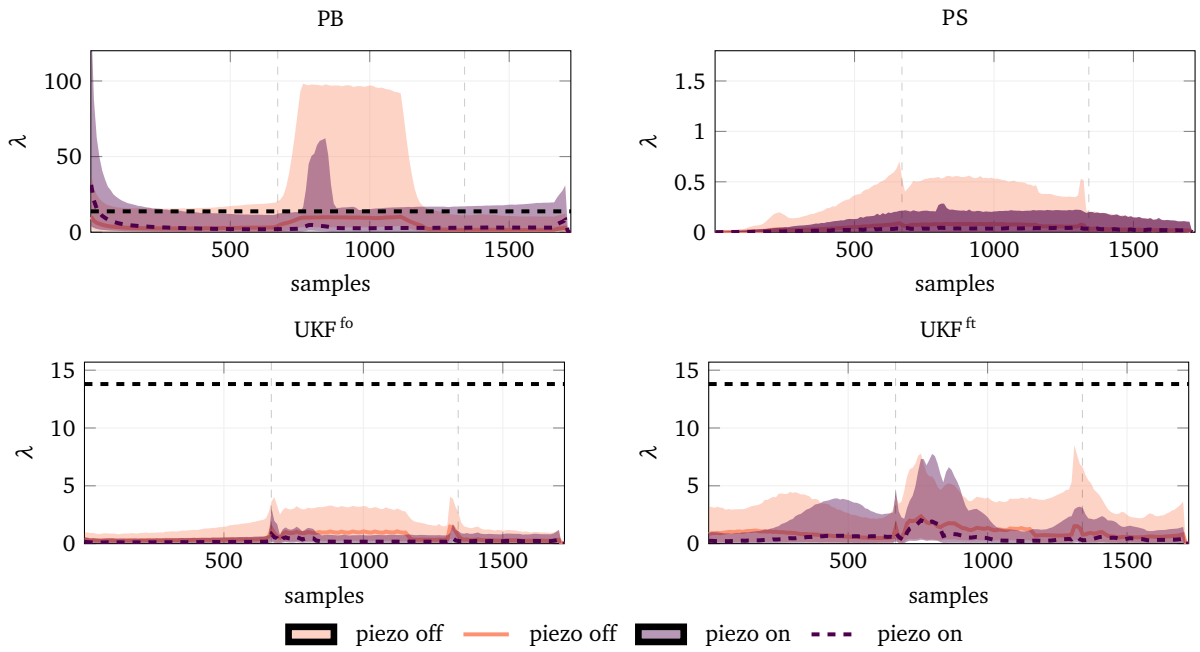


FIGURE 6.11: Likelihood ratios of the different residual generation methods. The values of $q = 0.1$ and $q = 0.9$ for each sample are depicted as the upper and lower bounds of the shaded area. They were determined from the data of all cavities over the RF-pulse. Values that exceed the detection threshold are counted as anomalous samples.

decile is represented by a solid and a dashed line for piezo-on and piezo-off operation respectively, whereas the quantiles $q = 0.1$ and $q = 0.9$ are defined as the lower and upper boundaries of the shaded areas.

- PARAMETER-BASED For both piezo-on and piezo-off mode of operation, the evaluated parameter-based residuals stay below the detection threshold for $q = 0.5$ of each sample. The values of $q = 0.9$ all exceed the detection threshold, particularly in the beginning of the pulse and during the beam loading phases.

- PARITY SPACE The values of $q = 0.9$ for all samples in the pulse are well below the detection threshold. When the piezo-electric detuning control is on, the ninth decile values are slightly smaller than when it is off. The transitions from filling to flattop and from flattop to decay show discernible spikes. In addition the beam loading phases are visible, i.e., during those times the ninth decile values slightly increase.

- UNSCENTED KALMAN FILTER Values for $q = 0.9$ of the evaluated residuals of UKF ^{fo} and UKF ^{ft} stay below the detection threshold. In the filling and in the flattop, the model of UKF ^{fo} show slight divergences from zero. In addition, the influence of the beam can be seen between sample 700 and 800 of the pulse. Not considering the beam in the model does however not result in likelihood ratios exceeding the thresholds, which means that the effect of the beam is not statistically significant as defined by the selected threshold.

Up until now, this chapter described the properties related to the different models and residuals concerning a set of nominal RF-pulses. Model uncertainty due to a change of operation setting or due to drifts can also have an effect on the detection properties. In the following section, large

amounts of not classified data is analyzed to determine the relevance of such effects on the proposed methods.

6.4 RF SET-POINT DEPENDENCY AND CONCEPT DRIFTS

The first sections of this chapter were concerned with the characterization of the evaluated RF-pulse nominal data, and determining model mismatches over the RF-pulse. In the following section the focus shifts towards characterizing the model quality over longer periods of time. Up until now it was assumed that the forward fields are approximately constant and that the behavior of the cavities does not change over time. However, since the cavities are operated in closed-loop, the forward set-points may change over time, which in turn can affect the detection results. It was also assumed that the cavity system does not change over time, i.e., that the model parameters are not subject to drifting. In this section both the effect of RF-field set-point changes and possible cavity parameter drifts on the detection results are analyzed.

6.4.1 *Model-Dependent Concept Drift*

Up until now, it was assumed that the nominal cavity parameters could be determined once with nominal RF-pulses from the *beginning* of a detection period. The anomaly detection is thus dependent on how well this data reflects future nominal behavior of the system. When the distribution of the RF-pulse data is non-stationary, the methods will suffer from *concept drift* [107, 108, 109]. This can lead to an increase of misclassification of the data over time. Concept drifts are common in data streams, as the real world is not as static as the determined models and especially slow drifts only become noticeable over longer periods of time.

To analyze the drifting behavior, the magnitude and the average rate of the drifts as defined by Webber et al. [107] are used. The drift magnitude determines the difference between the distribution of some random variable in two points in time, e.g., between time t_0 and some later time $t \neq t_0$, using a distribution distance function returning a non-negative value.

The average drift rate was determined by calculating the path length between the two times. Figure 6.12 shows an example of a changing parameter, e.g., the mean of a two-dimensional Gaussian distribution over time. While the magnitude simply describes the difference between the two times, the path lengths characterizes the cummulated changes of the parameters in between the two time-points. The average drift rate is then defined by

$$average\ rate_{t_0,t} = \frac{path\ length}{t - t_0}. \tag{6.2}$$

To analyze concept drifts for the cavities, changes in the distributions of the average likelihood ratio m^{λ} over the pulse were compared to changes of the pulse-wise static detuning $\bar{f}^{\Delta}$ and half-bandwidth $f^{\dagger}$ parameters, as well as to changes of the average forward $|\underline{\bar{v}}^{F}|$ and probe $|\underline{\bar{v}}^{P}|$ gradients, which

[107] Webb et al., "Characterizing concept drift"

[108] Elwell and Polikar, "Incremental learning of concept drift in nonstationary environments"

[109] Ditzler et al., "Learning in nonstationary environments: A survey"

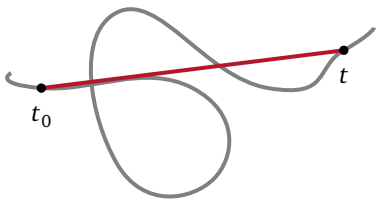


FIGURE 6.12: Scetch of a two-dimensional drifting parameter in time. The magnitude is shown by the red line as the distance between the two points. The path length is the path taken by the variable between the two times. Both characterize the drifts.

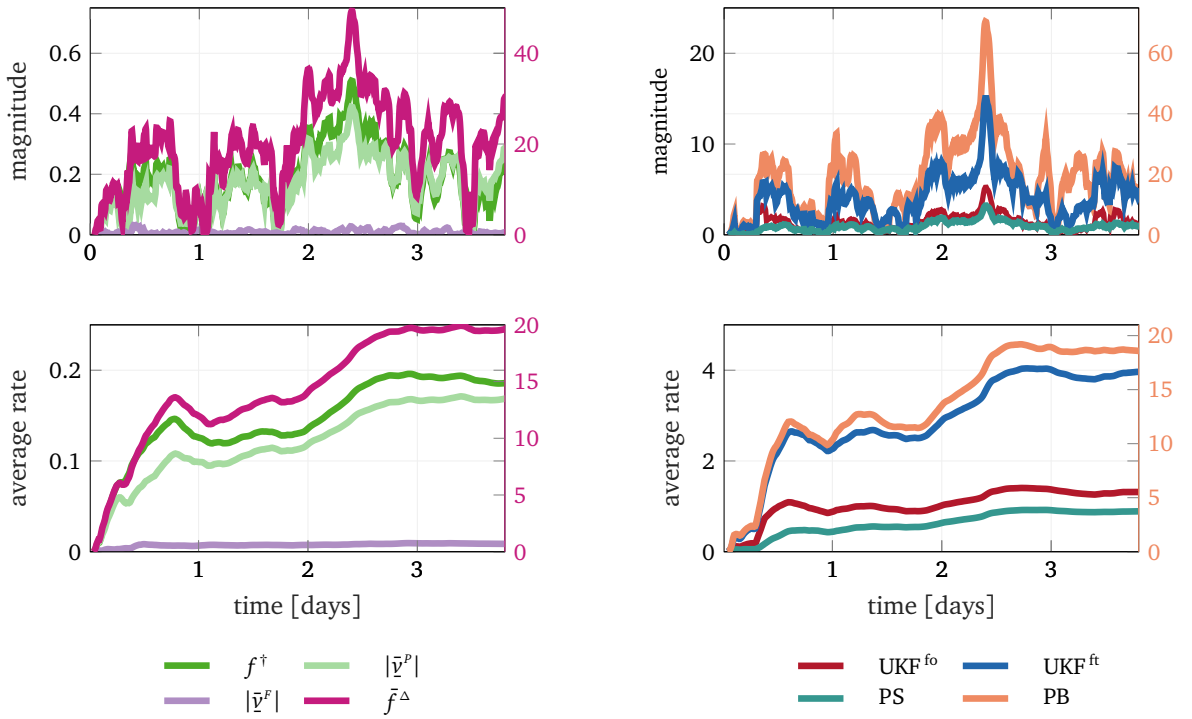


FIGURE 6.13: Magnitude and average drift rate for the pulse-wise cavity parameters over several days of operation in piezo-off mode.

are defined by

$$|\underline{\bar{\nu}}^{F}| = \frac{1}{n_{flt}} \sum_{k=k_f}^{k_d} |\underline{\nu}^{F}|, \quad |\underline{\bar{\nu}}^{P}| = \frac{1}{n_{flt}} \sum_{k=k_f}^{k_d} |\underline{\nu}^{P}|, \tag{6.3}$$

which means they describe the average forward and probe amplitude value during the flattop. The analysis of the drifts should reveal long-term changes of the cavity behavior, which is why the moving averages of the above mentioned parameters were determined with a window size of 72000, which corresponds to a two hour window.

The results of magnitude and average drift rate can be seen in Figure 6.13 and Figure 6.14.

All plots have two y-labels, which was necessary to reasonably depict all four methods and all four cavity parameters in the same plot. The cavity parameters depicted in the plot have different units. While the detuning f^{Δ} and half-bandwidth $f^{\dagger}$ changes are in Hz, the averaged forward and probe measurements are in $\mathrm{MV\,m^{-1}}$. The methods on the other hand are represented by the values of m^{λ}, i.e., the values in the plot inform about the change in magnitude for a given time t. The results can be summarized as follows

Piezo-electric detuning control off

Figure 6.13 shows that the averaged forward fields $|\underline{\bar{\nu}}^{F}|$ do not drift over the course of the depicted three days. Half-bandwidth $f^{\dagger}$, static detuning $\bar{f}^{\Delta}$ and the probe fields $|\underline{\bar{\nu}}^{P}|$ do show drifts. The static detuning values, drifted with magnitudes of up to 40 Hz. Half-bandwidth $f^{\dagger}$ and averaged probe $|\underline{\bar{\nu}}^{P}|$ show changes of maximally 0.4 Hz and $0.4\,\mathrm{MV\,m^{-1}}$, respectively, whereas the forward parameter showed comparatively little drifts. The average

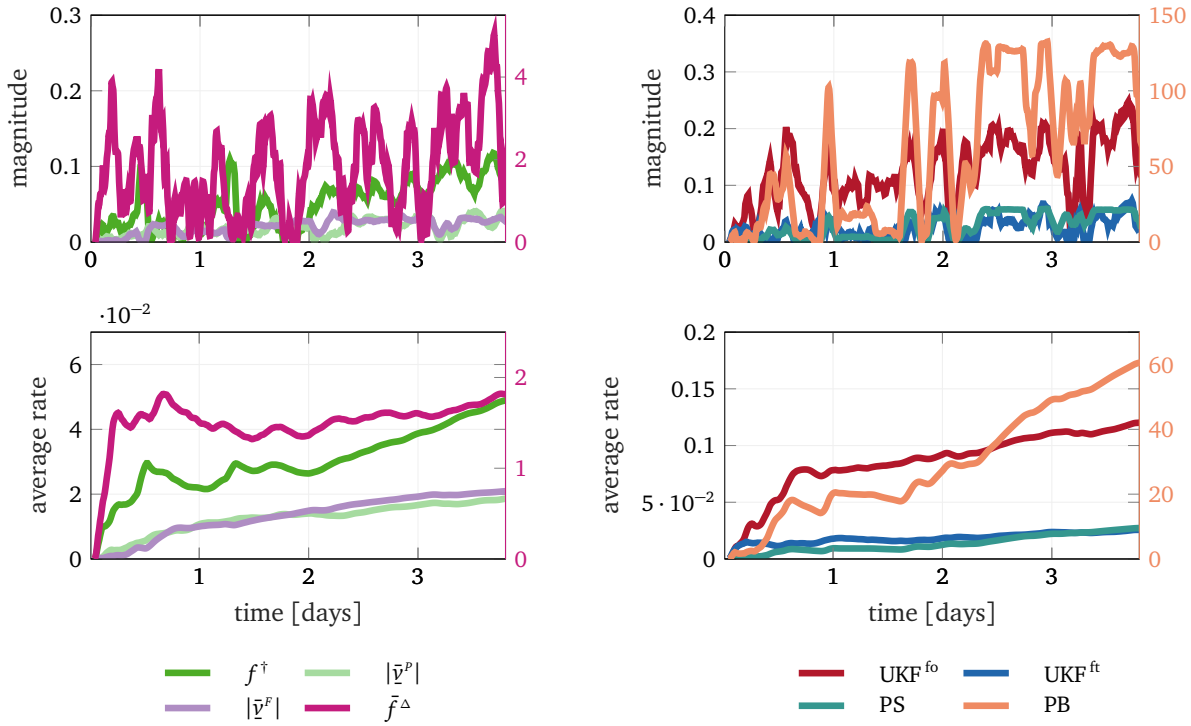


FIGURE 6.14: Magnitude and average drift rate for the pulse-wise cavity parameters over several days of operation in piezo-off mode.

likelihood ratio values m^{λ} drift with different magnitude depending on the method. The values generated with PB have the largest drift magnitudes of up to 60, whereas the lowest magnitudes are obtained from PS. The shapes of the magnitude and average rates are very similar throughout the different methods and can also be found in the pattern of the pulse-wise cavity parameters. In particular, the drifts in the average probe and static detuning $\bar{f}^{\Delta}$ show a qualitative resemblance to the drifting of the m^{λ} values.

Piezo-electric detuning control on

When the piezo-electric detuning control is on, the magnitude of the drifts decreased considerably for all methods. The average rates of drifting in the cavity parameter values decreased by a factor of ten and the same holds for the methods UKF fo,UKF ft and PS. However, the effect of the drift on the parameter-based method PB on the contrary increased by a factor of three. The drifting of the static detuning $\bar{f}^{\Delta}$ is considerably lower. Average forward and probe values drifted at a similar average rate. It can be noted, that the UKF ft drifting shape is similar to the one of the static detuning.

6.4.2 *Detection Result Dependence on Set-Point Changes*

To determine the dependency of the methods UKF fo, UKF ft, PB, and PS on set-point changes, RF-pulses were analyzed stemming from a set-point sweep. The nominal model was trained on one set-point level, and, with the model parameters obtained for that set, the RF-pulses were analyzed. The results can be seen in Figure 6.15, where the m^{λ} values are plotted for each consecutive pulse in the set-point sweep. For the examined cavity, the average forward values $|\underline{\bar{\nu}}^{F}|$ was altered from 12.5 to 11 MV m^{-1}. Due to the change of set-point, the static detuning value rose from around −50 Hz to

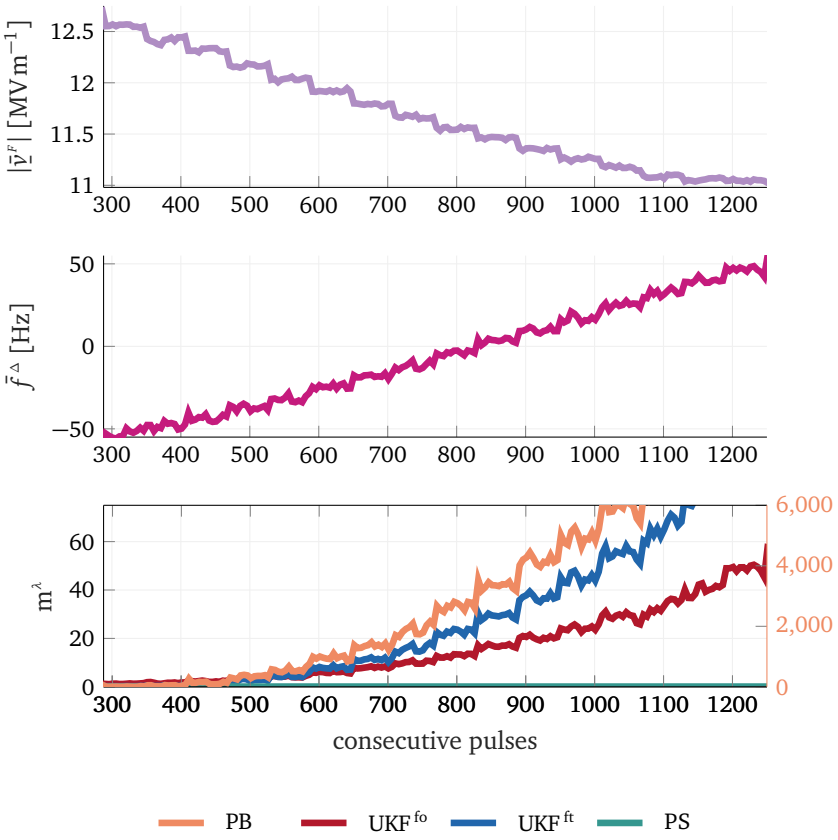


FIGURE 6.15: The nominal parameters were determined at a set-point level of 12.6 MV m^{-1}. Then the set-point was lowered sucessively until a value of 11.1 MV m^{-1} was reached. The changes in the glrt values for the different methods were determined. The parameter-based method shows the strongest sensitivity to set-point changes, while the parity space approach shows the least sensitivity.

around 50 Hz. The changes of m^{λ} values are visible from the first step-wise change of the set-point. The PS residuals show the least sensitivity towards the set-point changes, showing no changes for the range considered. The sensitivity of PB is the highest. Between the two unscented Kalman filter approaches, UKF$^{\mathrm{fo}}$ is less sensitive towards set-point changes than UKF$^{\mathrm{ft}}$. The PB changes from $\mathrm{m}^{\lambda} = 0$ to $\mathrm{m}^{\lambda} = 7000$, which is around a factor 100 higher than the other methods. It can be noted that while the detuning and the set-points show a linear change over the set-point sweep, the changes in the m^{λ} values seem to be better approximated by a quadratic behavior. The further away detuning and set-point are to the training conditions of the model, the larger the rate of divergence change.

6.5 DISCUSSION & SUMMARY

With the results of the previous chapter showing detection differences between the methods and the analysis of this chapter, the strengths and weaknesses of all methods are discussed. Suggestions towards future improvement are given in the final chapter of this thesis.

▸ DETUNING MODELS The fixed detuning trace model was able to generate good simulation results in both accuracy and precision. Differences between the cavities and the pulse to pulse variation were small. Simulations using the first order detuning model, were also capable of generating good simulation results in both accuracy and precision. However, the performance was not consistent for all cavities. The first-order model is dependent on a

least-square based parameter identification step to determine the model parameters, whereas the fixed-trace models are not. It appears that not all parameters estimated through the identification method show the same performance for different cavities.

- UNSCENTED KALMAN FILTER The results and performance of the unscented Kalman filter approach depended on which model was used for the residual generation. The UKF^{fo} based residuals detected more pulses as anomalous, but the majority of these pulses only slightly exceeding the threshold (benign anomalies). For some of the cavities, almost 100% of the pulses had a few samples that exceeded the detection threshold. These can be accounted for model uncertainty of that cavity. The influence of the beam was not significant and did not lead to nominal RF-pulses being classified as anomalous. Concept drift and changes of set-point had a greater effect on the detection result when using the fixed detuning model than when using the first order model. This can be explained in two ways: For one, the first-order model describes the detuning as a set of dynamic equations, dependent on the squared probe amplitude as its inputs. Though the parameters of the model were determined only for one set-point, smaller changes may still be representable by those parameters. Another explanation is that for the first order model larger variance was obtained for the nominal residuals and thus the sensitivity towards changes is lower.

  Some of the cavities did not produce Gaussian distributed residuals for nominal RF-pulses. This, in turn lead to detected anomalies in nominal RF-pulses. It also explains why for some of the cavities, 100% of the pulses were anomalous in Chapter 5.

- PARAMETER-BASED METHOD Using the calculation of the intra-pulse detuning and half-bandwidth as currently implemented for the generation of a residual showed undesirable properties for the detection of anomalies in general, classifying almost all pulses as anomalous. For the detection of quenches, on the other hand this approach surpassed all other methods with its sensitivity, reliably detecting quenches earlier in the pulse than the other methods. In comparison to the other methods, the influence of the beam is very noticable. Since the beam has not been included, most of the detected anomalies can be classified as false positives. In addition, noise especially in the beginning of the half-bandwidth calculations is significant. The strong drifting property can be accounted to its sensitivity towards detuning changes, which is primarily a challenge when operated without piezo-electric detuning control. Since January of 2020 the operation of the cavities was using piezos control, stabilizing the detuning in the flattop, i.e., stabilizing the static detuning values. Drifting due to detuning should considerably have lowered in piezo-on operation. However, the drifting was also strong in piezo-on operation , but could not be directly associated to changes of detuning. It is more likely that the drifts were caused by the multiple changes of beam times in the pulse.

- PARITY SPACE METHOD Residuals generated with the parity space approach were not Gaussian distributed for nominal RF-pulses. The reason here was

	PB	PS	UKF ^{fo}	UKF ^{ft}
Beam loading dependence	high	low	low	low
Set-point dependence	high	low	low to medium	medium
Drift	high	low	low	low
Gaussian nominal residuals	not fulfilled	not fulfilled	largely fulfilled	fulfilled

TABLE 6.1: Comparing model uncertainty for residual approaches.

that the variance over the pulse changed and thus the values could not be represented by a single Gaussian distribution. For the evaluation of the residuals, however, this did not lead to false alarms. Instead, the resulting likelihood ratios showed smaller values than expected. Since the parity space method is not very sensitive towards changes in detuning, it also shows no dependency on set-point changes or concept drifts. In piezo-on mode of operation it was able to detect quenches as severely anomalous, which was not the case for piezo-off operation. The parity space residuals approach is the easiest in terms of implementation. To improve its performance, the beam should be included into the model. While the two-dimensional residuals could be used to distinguish faults according to direction *and* magnitude, the one-dimensional parity space residual is only capable of distinguishing according to magnitude. Since the concept drifts were mainly driven by changes of static detuning, the parity space method showed the least amount of concept drift sensitivity.

This chapter closes with a short summary of the obtained results, whereas the focus lies on a direct comparison between the different residual models and detection characteristics when using the generalized likelihood ratios. The following sources of model uncertainty can be distinguished

1. Beam loading
2. Set-point changes
3. Concept drifts
4. Gaussianity of nominal residuals

Table 6.1 gives a qualitative overview over the performance of the methods with respect to these sources of model uncertainty. The results are described in relative terms, i.e., comparing the performance to all other residual models.

Summary

This chapter determined the performance of the detuning models and described some differences between the residuals. Both detuning models have a good performance, however the fixed trace model showed slightly better simulation fits and less cavity wise variations in the model performance than the first order method. Assumption of Gaussianity was best confirmed by the unscented Kalman filter-based approach. The residuals methods were further analyzed with respect to their sensitivity towards set-point changes as well as concept drifts. The analysis showed that the parity space method

was least affected by set-point changes or concept drifts, whereas the largest sensitivity towards both set-point changes or concept drifts was determined for the parameter-based method. The results also showed, that all methods may benefit from an inclusion of the beam loading into the system model. While the exclusion of the beam for parity space and unscented Kalman filter residuals turned out to be non-critical, the performance of the parameter-based residual showed obvious downsides.

7

Conclusion & Outlook

"And now you do what they told ya, now you're under control"
—Rage Against the Machine

In this thesis probabilistic model-based methods were applied to the fault and anomaly detection task of the pulsed mode operated SRF cavities of the EuXFEL.

7.1 CONCLUSION

The probabilistic methods were conveyed in the factor graph framework, which was able to lend itself to multiple tasks in the detection process. Residual generation, residual evaluation, and parameter training were all encompassed in the same factor graph. In addition to probabilistic methods, two deterministic methods were considered.

The usefulness of the developed approaches as well as the benefits of anomaly detection for the SRF cavities was shown by evaluating considerable amounts of RF-pulses, analyzing long-term changes in those pulses, and describing the main characteristics of the detected anomalies. Unwanted and so far undetected behavior of the EuXFEL's cavities was discovered that may lead to new insights pertaining to the operation of the cavities in the following way:

Firstly, it was shown that all considered cavities in the EuXFEL were affected by an "hourly event". The characteristics of these events point to a repetitive, strong microphonics phenomenon. The reliability with which this phenomenon occurs in the same time period suggests a machine-related cause. If the reason is understood, the effect on the cavities could be reduced, or it could be prevented altogether.

Secondly, severe anomalies were detected directly before a quench, pointing thus to incipient system behavior resulting in a quench. These results can help to understand why quenches appear to occur spontaneously, although the gradient with which the cavity is operated lies below its quenching limit. If the detection of the incipient behavior can be achieved reliably, it can also result in the avoidance of quenches. It was also shown that an intrapulse detection of quenches is possible. The times at which the quenches were detected, suggest a relationship to the "hourly events", however, further investigation is needed. Quenches occur more often when operated at maximum gradients. These results are therefore particularly interesting for the high gradient operation of the EuXFEL.

Finally, severe anomalies were detected for several days in one cavity without being detected by the current system. It was shown that these anomalies were related to considerable changes in static detuning and half-bandwidth. The high detuning values in turn lead to an increase of forward power as determined by the vector sum controller. When operated with piezo-electric detuning, such substantial changes of static detuning would no longer occur. However, a cavity affected by strong vibrations or soft quenching over several days of operation should be detected as this could lead to unnecessary wear of the involved components. Thus, this thesis confirmed the need and benefit of a rigorous fault and anomaly detection scheme for the SRF cavities. In particular, it was shown that relevant information could be gained when monitoring the entire RF-pulses. The causes of the detected anomalies have not been further investigated, but their characteristics suggest that other sub-components may be involved.

The described detection results were obtained using three different residual generation approaches:

- THE UNSCENTED KALMAN FILTER residuals describe deviations between the estimated and measured system outputs. They are thus capable of detecting measurement faults directly. However, it was shown that they could also detect changes that affect the cavity parameters, i.e., quenches or microphonics. The unscented Kalman filter methods showed the best performance for detecting severely anomalous pulses. They were successfully used to detect quenches, in some cases several pulses earlier than the current detection system. The unscented Kalman filter approach includes measurement and process noise assumptions and it obtained zero-mean Gaussian distributed residuals for all anomaly-free cavities. The approach can be advanced to operate during ramp-ups or ramp-downs by extending the cavity model such that it represents the nominal behavior also during set-point changes. For an online operation, it is necessary that this approach can be programmed into an FPGA component.

- THE PARITY SPACE residual determines the difference between two expressions describing the detuning in the cavity model. It is the least sensitive towards detuning changes and was not able to distinguish quenches from anomalies associated with microphonics when the cavity was operated without a piezo-electric detuning controller. When operated with piezo-electric detuning control, the parity space residuals showed promising capabilities in detecting quenches. Furthermore, it was the least dependent on set-point changes. Its simplicity makes for easy online implementation. In fact, while writing this thesis, the parity space approach has been implemented for online operation and is currently tested on one to four cavities of the EuXFEL.

- THE PARAMETER-BASED approach showed to be most sensitive towards changes in parameters of the model, which is of advantage for the detection of both quenches and severe microphonics. Furthermore, this approach can rely on already implemented FPGA components, which would take the

least effort towards online implementation. For this method the effects of the beam should be included in the model and the noise inherent to the parameter computation method should be taken into consideration.

7.2 OUTLOOK

Future work may be summarized in the following aspects:

- MODEL REVISION The influence of the beam should be included into the cavity model. In particular when working with the parameter-based method, including the beam is unavoidable for a good performance. To do so, one can make use of the recently developed FPGA component for detuning and half-bandwidth estimation which also includes the beam, see [110].

  Unexpected behavior can also happen during ramp up or ramp down phases of the machine. To be able to handle changes in set-points, the cavity model needs to be extended.

  It can be assumed that after including the beam into the model and when the cavities are exclusively operated with piezo-electric detuning controllers, issues of concept drift will decrease. Any residual slow changes in the model may be considered for example by periodically updating the model [108].

  To reduce computation time, the cavity model should be linearized in which case the unscented Kalman filter could be exchanged by a linear Kalman filter. Efforts towards FPGA component design for detuning control may then be reused or adapted for the residual generation purpose [111].

  As an alternative to using the parameter computations, the cavity parameters could also be estimated using a Kalman filter, thereby filtering the noise.

- HOURLY EVENT AND QUENCHING The temporal overlap of quenches and the hourly events, as well as the incipient behavior before a quench should be further investigated. If indeed an increase of pressure on the cavities, occurring each hour makes quenches more likely, the RF field set-points could be lowered in those minutes when operated in maximum gradient to avoid the quenches.

- INTRAPULSE QUENCH DETECTION The use of GMMs for the detection of quenches could be examined further. Instead of using the unscented Kalman filter residuals, the parameter-based residuals could be used to train the GMM. Then, the residual space spanned by half-bandwidth and detuning has a direct physical interpretation which could be helpful to distinguish between quenches and other faults. With more faulty data available, the GMMs could be periodically updated and thus, over time, obtain a cavity-specific model which may distinguish between nominal and different classes of faults [112].

- FAULT DIAGNOSIS FOR THE EUXFEL In order to be able to determine the root causes of faults for the cavities, information from other subsystems should be included. Helium pressure, klystron signals or piezo-electric controller

[110] Bellandi et al., "Online Detuning Computation and Quench Detection for Superconducting Resonators"

[108] Elwell and Polikar, "Incremental learning of concept drift in nonstationary environments"

[111] Ushakov et al., "Developing Kalman filter based detuning control with a digital SRF CW cavity simulator"

[112] Yan et al., "Gaussian mixture model using semisupervised learning for probabilistic fault diagnosis under new data categories"

signals can all be included into the detection system. At the same time, fault and anomalies should also be detected in those components.

- FACTOR GRAPHS FOR OTHER SYSTEMS For the detection of faults in other systems, new factor graphs may be developed, tailored to that components' specific need. Since the factor graphs defined in this thesis describe general models, they may also be used for other components. Both the GMM and the state space model factor graph could be applied to components that are best described by a dynamic model. Observations from other subsystems, or assumptions about their behavior can be added to the existing factor graph.

Appendix

CALIBRATION

Taking into consideration, that cross-couplings between the forward and reflected signals are possible, both in their real and imaginary parts, the following relationship holds

$$\underline{v}_k^P = \underbrace{\underline{a}\ \underline{v}_k^F + \underline{b}\ \underline{v}_k^R}_{\underline{v}_k^{Fc}} + \underbrace{\underline{c}\ \underline{v}_k^F + \underline{d}\ \underline{v}_k^R}_{\underline{v}_k^{Rc}}, \tag{7.1}$$

where $\underline{a}, \underline{b}, \underline{c}, \underline{d} \in \mathbb{C}$ are complex calibration parameters which need to be determined using pulse measurements. These parameters are determined by the method of least squares. Since the measurement matrix of complex forward and reflected is only two, i.e., the rank of the measurement matrix two, the posed calibration problem is thus under-determined. To lower the number of free parameters additional boundary conditions are introduced. The first boundary condition poses the constraint that the calibrated forward signal during the decay of the pulse should be equal to zero, i.e.,

$$\underline{v}_k^{Fc} = 0, k = k_d, ..., n_s, \tag{7.2}$$

and the second boundary condition states that the probe and reflected measurements during the decay should be equal, i.e.,

$$\underline{v}_k^P = \underline{v}_k^{Rc}, k = k_d, ..., n_s. \tag{7.3}$$

In addition, the calibration coefficients $\underline{b}$ and $\underline{c}$, which determine the cross couplings between the forward and reflected signals need to be restricted. To do so, some weighting factors are introduced, which bind the cross couplings.

GMM PARAMETER TRAINING WITH EM

To obtain the maximum likelihood solution for a GMM with noisy measurements and latent class memberships, the EM algorithm is used to obtain the parameters in an iterative way. The function for the maximization is given by

$$\theta^{j+1} = \underset{\theta}{\operatorname{argmax}}\ \mathcal{Q}\Big(\theta|\theta^j\Big). \tag{7.4}$$

Taking into consideration, that

$$
\begin{aligned}
\mathcal{Q}\big(\theta|\theta^j\big) &= \mathbf{E}_{p(\mathbf{C}=\mathbf{c},\mathbf{r},|\mathbf{r}',\theta^j)}\left[\ln p(\mathbf{r},\mathbf{r}',\mathbf{C}=\mathbf{c}|\theta)\right] \\
&= \sum_{i=1}^{n_c}\int_{\mathbf{r}}\prod_k p(C_k=c_{k,i},r_k,|r_k',\theta^j)\Big(\sum_k \ln p(r_k,r_k',C=c_{k,i}|\theta)\Big)\,d\mathbf{r} \\
&= \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\int_{r_k} P(C_k=c_{k,i}|r_k{}',\theta^j)p(r_k|\theta^j)\ln p(r_k,r_k',C_k=c_{k,i}|\theta)\,dr_k \\
&= \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\underbrace{P(C_k=c_{k,i}|r_k',\theta^j)}_{\rho_{k,i}}\int_{r_k} p(r_k|\theta^j)\ln p(r_k,r_k',C_k=c_{k,i}|\theta)\,dr_k \\
&= \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\rho_{k,i}\,\mathbf{E}_{p(r_k|\theta^j)}\left[\ln p(r_k,r_k',C_k=c_{k,i}|\theta)\right] \\
&= \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\rho_{k,i}\,\mathbf{E}_{p(r_k|\theta^j)}\left[\ln p(r_k|r_k')+\ln p(r_k,C_k=c_{k,i}|\theta)+\ln p(r_k')\right]
\end{aligned}
\tag{7.5}
$$

for the maximization with regard to θ, the pdf of $p(r_k|r_k')$ and $p(r_k')$ are independent of the model parameters θ and therefore play no role when determining the maximum with respect to θ. A simplified expression is used to determine the updated parameters

$$
\begin{aligned}
\tilde{\mathcal{Q}}\big(\theta|\theta^j\big) &= \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\rho_{k,i}\,\mathbf{E}_{p(r_k|\theta^j)}\ln p(r_k,C_k=c_{k,i}|\theta) \\
&= \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\rho_{k,i}\,\mathbf{E}_{p(r_k|\theta^j)}\ln \mathrm{w}_i\,\mathcal{N}\Big(r_k;\mathrm{m}_i,\mathrm{V}_i\Big)
\end{aligned}
\tag{7.6}
$$

with $p(r_k,C_k=c_{k,i}|\theta)=\mathrm{w}_i\,\mathcal{N}\Big(r_k;\mathrm{m}_i,\mathrm{V}_i\Big)$, the function reduces to

$$
\begin{aligned}
\tilde{\mathcal{Q}}\big(\theta|\theta^j\big) &= \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\rho_{k,i}\,\mathbf{E}_{p(r_k|\theta^j)}\Big[\ln \mathrm{w}_i+\ln 2\pi|\mathrm{V}_i|+\frac{1}{2}(r_k-\mathrm{m}_i)^T\mathrm{V}_i^{-1}(r_k-\mathrm{m}_i)\Big] \\
&= \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\rho_{k,i}\,\mathbf{E}_{p(r_k|\theta^j)}\Big[\ln \mathrm{w}_i+\ln 2\pi|\mathrm{V}_i|+ \\
&\quad \frac{1}{2}\mathrm{trace}\Big(\mathrm{V}_i^{-1}(r_k{r_k}^T-r_k\mathrm{m}_i^T-\mathrm{m}_i r_k^T+\mathrm{m}_i\mathrm{m}_i^T)\Big)\Big]
\end{aligned}
\tag{7.7}
$$

taking into consideration, that the trace operator is commutable, and that

$$
\begin{aligned}
\mathbf{E}_{p(r_k|\theta^j)}\big[r_k{r_k}^T\big] &= \bar{\mathrm{V}}_{R_{k,i}}\bar{\mathrm{m}}_{R_{k,i}}\bar{\mathrm{m}}_{R_{k,i}}^T \\
\mathbf{E}_{p(r_k|\theta^j)}\big[r_k\big] &= \bar{\mathrm{m}}_{R_{k,i}},
\end{aligned}
\tag{7.8}
$$

see [113], the expression results into

$$
\begin{aligned}
\tilde{\mathcal{Q}}\big(\theta|\theta^j\big) = \sum_{k=1}^{n_K}\sum_{i=1}^{n_c}\rho_{k,i}\Big(&\ln \mathrm{w}_i+\ln 2\pi|\mathrm{V}_i|+ \\
&\frac{1}{2}\mathrm{trace}\Big(\mathrm{V}_i^{-1}(\bar{\mathrm{V}}_{R_{k,i}}+\bar{\mathrm{m}}_{R_{k,i}}\bar{\mathrm{m}}_{R_{k,i}}^T)-2\mathrm{V}_i^{-1}\bar{\mathrm{m}}_{R_{k,i}}\mathrm{m}_i^T+\mathrm{V}_i^{-1}\mathrm{m}_i\mathrm{m}_i^T\Big)\Big)
\end{aligned}
\tag{7.9}
$$

Note, that the $\ln 2\pi$ can be neglected. In order to maximize with regard to θ, the above expression has to be differentiated with respect to $\mathrm{w}_i,\mathrm{m}_i,\mathrm{V}_i$

1. Differentiation with respect to w_i and adding the constraint that $\sum_{i=1}^{n_c} \mathrm{w}_i = 1$ with the Lagrange multiplier λ

$$\frac{\partial}{\partial \mathrm{w}_i} \sum_{k=1}^{n_K} \sum_{i=1}^{n_c} \ln \mathrm{w}_i + \lambda(\mathrm{w}_i - 1) \; = 0 \tag{7.10}$$

$$\hat{\mathrm{w}}_i = \frac{1}{n_K} \sum_{k=1}^{n_K} \rho_{k,i} \tag{7.11}$$

2. Differentiation with respect to m_i

$$\frac{\partial}{\partial \mathrm{m}_i} \mathcal{Q}\big(\theta|\theta^j\big) \; = 0$$

$$\mathrm{trace}\Big(\sum_k \rho_{k,i} \big(\frac{1}{2} 2\mathrm{V}_i^{-1} \bar{\mathrm{m}}_{R_{k,i}} + \frac{1}{2} 2\mathrm{V}_i^{-1} \mathrm{m}_i\big)\Big) \; = 0$$

$$\mathrm{V}_i^{-1} \sum_k \rho_{k,i} \, \bar{\mathrm{m}}_{R_{k,i}} = \mathrm{V}_i^{-1} \mathrm{m}_i \sum_k \rho_{k,i}$$

$$\mathrm{m}_i^{j+1} = \frac{1}{\sum_k \rho_{k,i}} \sum_k \rho_{k,i} \, \bar{\mathrm{m}}_{R_{k,i}}$$

3. Differentiation with respect to V_i

$$\frac{\partial}{\partial \mathrm{V}_i} \mathcal{Q}\big(\theta|\theta^j\big) \; = 0$$

$$\sum_k \rho_{k,i} \Big(\frac{1}{2} V_i^{-1} - \frac{1}{2} \mathrm{V}_i^{-2} \bar{\mathrm{V}}_{R_{k,i}} - \frac{1}{2} \mathrm{V}^{-2} \bar{\mathrm{m}}_{R_{k,i}} \bar{\mathrm{m}}_{R_{k,i}}^T + \frac{1}{2} 2\mathrm{V}_i^{-2} \bar{\mathrm{m}}_{R_{k,i}} \mathrm{m}_i^T - \frac{1}{2} \mathrm{V}_i^{-2} \mathrm{m}_i \mathrm{m}_i^T\Big) \; = 0$$

$$\sum_k \rho_{k,i} \Big(1 - \mathrm{V}_i^{-1} \bar{\mathrm{V}}_{R_{k,i}} - \mathrm{V}_i^{-1} \bar{\mathrm{m}}_{R_{k,i}} \bar{\mathrm{m}}_{R_{k,i}}^T + 2\mathrm{V}_i^{-1} \bar{\mathrm{m}}_{R_{k,i}} \mathrm{m}_i^T - \mathrm{V}_i^{-1} \mathrm{m}_i \mathrm{m}_i^T\Big) \; = 0$$

$$\sum_k \rho_{k,i} \; = \mathrm{V}_i^{-1} \sum_k \rho_{k,i} \, \bar{\mathrm{V}}_{R_{k,i}} + \mathrm{V}_i^{-1} \sum_k \rho_{k,i} \, \bar{\mathrm{m}}_{R_{k,i}} \bar{\mathrm{m}}_{R_{k,i}}^T - 2\mathrm{V}_i^{-1} \sum_k \rho_{k,i} \, \bar{\mathrm{m}}_{R_{k,i}} \mathrm{m}_i^T + \mathrm{V}_i^{-1} \sum_k \rho_{k,i} \, \mathrm{m}_i \mathrm{m}_i^T$$

$$\mathrm{V}_i = \frac{1}{\sum_k \rho_{k,i}} \Big(\sum_k \rho_{k,i} \, \bar{\mathrm{V}}_{R_{k,i}} + \sum_k \rho_{k,i} \, \bar{\mathrm{m}}_{R_{k,i}} \bar{\mathrm{m}}_{R_{k,i}}^T - 2 \sum_k \rho_{k,i} \, \bar{\mathrm{m}}_{R_{k,i}} \mathrm{m}_i^T + \sum_k \rho_{k,i} \, \mathrm{m}_i \mathrm{m}_i^T\Big)$$

$$\mathrm{V}_i^{j+1} = \frac{1}{\sum_k \rho_{k,i}} \Big(\sum_k \rho_{k,i} \, \bar{\mathrm{V}}_{R_{k,i}} + \sum_k \rho_{k,i} \, (\bar{\mathrm{m}}_{R_{k,i}} - \mathrm{m}_i)(\bar{\mathrm{m}}_{R_{k,i}} - \mathrm{m}_i)^T\Big)$$

□

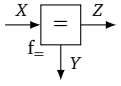
GMM UPDATES FOR EM ALGORITHM

The updates of the EM algorithm for the GMM training were described by a forward pass through the via equality nodes connected local messages of each time step k. This means that the updates, using the described mean and variance definitions of the Gaussian EM message, can be described using the tabulated update rules of the equality node.

The equality node update is recounted for convenience in Table 1.

The full forward pass through all equality nodes connected to the local messages of each time step k results in an update of the information matrix

TABLE 1: Update equations for the equality node.



$$\mathrm{f}_= = \delta(z-x)\delta(z-y)$$

$$\vec{\xi}_Z = \vec{\xi}_X + \overleftarrow{\xi}_Y$$
$$\vec{\mathrm{V}}_Z = \vec{\mathrm{V}}_X + \vec{\mathrm{V}}_Y$$

given by

$$\begin{aligned}
\mathrm{W}_{\hat{\theta}^{j+1}} &= \mathrm{W}_{\hat{\theta}_1^j} + \mathrm{W}_{\hat{\theta}_2^j} + \ldots + \mathrm{W}_{\hat{\theta}_{n_K}^j} \\
&= \mathrm{V}_{\hat{\theta}_1^j}^{-1} + \mathrm{V}_{\hat{\theta}_2^j}^{-1} + \ldots + \mathrm{V}_{\hat{\theta}_{n_K}^j}^{-1} \\
&= \begin{bmatrix} \sum_{k=1}^{n_K} \rho_{k,1}^j & 0 & \ldots & 0 \\ 0 & \sum_{k=1}^{n_K} \rho_{k,2}^j & & \\ \vdots & & \ddots & \\ 0 & & & \sum_{k=1}^{n_K} \rho_{k,n_c}^j \end{bmatrix} \otimes \mathbf{I},
\end{aligned} \tag{7.12}$$

and thus the updated variance is defined by

$$\mathrm{V}_{\hat{\theta}^{j+1}} = \begin{bmatrix} \frac{1}{\sum_{k=1}^{n_K} \rho_{k,1}^j} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sum_{k=1}^{n_K} \rho_{k,2}^j} & & \\ \vdots & & \ddots & \\ 0 & & & \frac{1}{\sum_{k=1}^{n_K} \rho_{k,n_c}^j} \end{bmatrix} \otimes \mathbf{I}. \tag{7.13}$$

The weighted mean, on the other hand, can also be directly defined by the update rules of the equality node, i.e.,

$$\begin{aligned}
\xi_{\hat{\theta}^{j+1}} =& \xi_{\hat{\theta}_1^j} + \xi_{\hat{\theta}_2^j} \ldots + \xi_{\hat{\theta}_{n_K}^j} \\
=& \mathrm{V}_{\hat{\theta}_1^j}^{-1} \mathrm{m}_{\hat{\theta}_1^j} + \mathrm{V}_{\hat{\theta}_2^j}^{-1} \mathrm{m}_{\hat{\theta}_2^j} + \ldots + \mathrm{V}_{\hat{\theta}_{n_K}^j}^{-1} \mathrm{m}_{\hat{\theta}_{n_K}^j} \\
=& \Bigg[\sum_{k=1}^{n_K} \rho_{k,i}^j \bar{\mathrm{m}}_{R_{k,1}}^j, \sum_{k=1}^{n_K} \Big(\mathrm{rvect}(\Omega_{k,1}^j) \Big)_{1,1}, \ldots, \sum_{k=1}^{n_K} \Big(\mathrm{rvect}(\Omega_{k,1}^j) \Big)_{2,2}, \ldots, \\
& \sum_{k=1}^{n_K} \rho_{k,n_c}^j \bar{\mathrm{m}}_{R_{k,n_c}}^j \sum_{k=1}^{n_K}, \Big(\mathrm{rvect}(\Omega_{k,n_c}^j) \Big)_{1,1}, \ldots, \sum_{k=1}^{n_K} \Big(\mathrm{rvect}(\Omega_{k,n_c}^j) \Big)_{2,2} \Bigg]^T
\end{aligned} \tag{7.14}$$

and since $\mathrm{m}_{\hat{\theta}^{j+1}} = \mathrm{V}_{\hat{\theta}^{j+1}} \xi_{\hat{\theta}^{j+1}}$, the updates for each EM iteration defined in (4.73) is obtained. The resulting messages carry all necessary components of the final update solutions. Note also that all messages carrying the updates of the parameters are the same for all time edges.

List of Figures

List of Tables

Notation

Cavity Figures Of Merit	
f	driving frequency
f^0, f_t^0	resonance frequency
$f^\Delta, f_t^\Delta, f_k^\Delta$	detuning
$f^\dagger, f_t^\dagger, f_k^\dagger$	half-bandwidth related to loaded quality factor
$\tilde{f}^\dagger$	half-bandwidth related to external quality factor
$\bar{f}^\Delta, \bar{f}_{i_p}^\Delta$	static detuning
Q^L	loaded quality factor
Q_{ext}	external quality factor
Q^0	unloaded quality factor

RF signals	
$\underline{v}^F, \underline{v}_t^F$	forward phasor
$\lvert\underline{v}^F\rvert, \lvert\underline{v}_t^F\rvert$	forward amplitude
$\lvert\underline{v}^F\rvert$	on
ϕ^F, ϕ_t^F	forward phase
$v^{F,I}, v_t^{F,I}, v_k^{F,I}$	in-phase forward (I-component)
$v^{F,Q}, v_t^{F,Q}, v_k^{F,Q}$	quadrature forward (Q-component)

Detuning Model	
x^m	resonance mode m
β_m	Lorentz force detuning constant
γ_m	time constant of mode m
α_m	mechanical frequency of mode m

Evaluation Parameters	
m^λ	mean likelihood ratio of pulse
$\lambda^{\max}$	maximum likelihood ratio of pulse
ψ_{th}	threshold on likelihood ratio
λ, λ_k	general likelihood ratio

Data Sets

$\mathcal{T}, \mathcal{T}^{pz}$	all RF-pulsescollected
$\mathcal{M}, \mathcal{M}^{pz}$	selected nominal RF-pulses
$\tilde{\mathcal{M}}, \tilde{\mathcal{M}}^{pz}$	selected nominal RF-pulsesfor validation
$\mathcal{Q}, \mathcal{Q}^{pz}$	RF-pulsesof quench event
$\mathcal{Z}, \mathcal{Z}^{pz}$	incipient/severe RF-pulses
$\mathcal{S}$	strange behavior in C6.M1.A17.L3
$\mathcal{A}, \mathcal{A}^{pz}$	all RF-pulseswhich are not severe nor belong to quench event

Number of ...

n_{cav}	... cavities
n_p	... pulses
n_s	... samples
n_{flt}	... samples during the flattop
n^{a}	... anomalies
n_c	... classes
n_{n}	... nominal RF-pulses
n_Θ	... parameters
l	... inputs
m	... outputs
n_M	... hypothesis
n_K	... training samples
n_m	... mechanical modes

Residuals

r^{ps}, r_k^{ps}	parity space residual
$r^{\mathrm{fo}}, r_k^{\mathrm{fo}}$	unscented Kalman filter residual using first order detuning ($\mathrm{UKF}^{\mathrm{fo}}$) model
$r^{\mathrm{ft}}, r_k^{\mathrm{ft}}$	unscented Kalman filter residual using fixed detuning input ($\mathrm{UKF}^{\mathrm{ft}}$)
$r^{I,\mathrm{fo}}, r^{Q,\mathrm{fo}}$	I and Q component of $\mathrm{UKF}^{\mathrm{fo}}$ residuals
$r^{I,\mathrm{ft}}, r^{Q,\mathrm{ft}}$	I and Q component of $\mathrm{UKF}^{\mathrm{ft}}$ residuals
r^{Δ}, r_k^{Δ}	detuning based residual
$r^{\dagger}, r_k^{\dagger}$	half-bandwidth based residual

Bibliography

[1] M. Altarelli, R. Brinkmann, M. Chergui, W. Decking, B. Dobson, S. Düsterer, G. Grübel, W. Graeff, H. Graafsma, J. Hajdu, et al. *The European x-ray free-electron laser technical design report*. 2006 (cit. on pp. 3, 9, 11, 15).

[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods". In: *Computers & chemical engineering* 27.3 (2003), pp. 293–311 (cit. on pp. 3, 29, 45).

[3] A. Edelen, C. Mayes, D. Bowring, D. Ratner, A. Adelmann, R. Ischebeck, J. Snuverink, I. Agapov, R. Kammering, J. Edelen, et al. "Opportunities in machine learning for particle accelerators". In: *arXiv preprint arXiv:1811.03172* (2018) (cit. on p. 4).

[4] C. Tennant, A. Carpenter, T. Powers, A. Shabalina Solopova, L. Vidyaratne, and K. Iftekharuddin. "Superconducting radio-frequency cavity fault classification using machine learning at Jefferson Laboratory". In: *Phys. Rev. Accel. Beams* 23 (11 2020), p. 114601 (cit. on pp. 4, 18).

[5] M. I. Jordan, T. J. Sejnowski, T. A. Poggio, et al. *Graphical models: Foundations of neural computation*. 2001 (cit. on p. 4).

[6] K. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning series. MIT Press, 2012. ISBN: 9780262018029 (cit. on pp. 4, 54, 55, 58).

[7] Y. Shen, X. Zhang, and Z. Wang. "Spacecraft fault diagnosis based on empirical mode decomposition and directed factor graph". In: *Harbin Gongye Daxue Xuebao(Journal of Harbin Institute of Technology)* 45.1 (2013), pp. 19–24 (cit. on p. 5).

[8] A. Gienger and O. Sawodny. "Data-based Process Monitoring and Iterative Fault Diagnosis using Factor Graphs". In: *2020 IEEE International Conference on Industrial Technology (ICIT)*. IEEE. 2020, pp. 35–40 (cit. on pp. 5, 46).

[9] T. Escobet, A. Bregon, B. Pulido, and V. Puig. *Fault diagnosis of dynamic systems*. Springer, 2019 (cit. on p. 5).

[10] W. Li, X. Cui, and M. Lu. "A robust graph optimization realization of tightly coupled GNSS/INS integrated navigation system for urban vehicles". In: *Tsinghua Science and Technology* 23.6 (2018), pp. 724–732 (cit. on p. 5).

[11] A Nawaz, S Pfeiffer, G Lichtenberg, and P Rostalski. "Anomaly Detection for the European XFEL using a Nonlinear Parity Space Method". In: *10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2018*. Vol. 51. 24. Elsevier, 2018, pp. 1379–1386 (cit. on pp. 6, 47).

[12] A Nawaz, S Pfeiffer, G Lichtenberg, and P Rostalski. "Fault Detection Method for the SRF Cavities of the European XFEL". In: *15th European Workshop on Advanced Control and Diagnosis, ACD 2019*. 2019 (cit. on p. 6).

[13] 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". In: *at-Automatisierungstechnik* 69.6 (2021), pp. 543–554 (cit. on pp. 6, 47).

[14] A Nawaz, S Pfeiffer, G Lichtenberg, and P. Rostalski. "Anomaly Detection for Cavity Signals-Results from the European XFEL". In: *9th Int. Particle Accelerator Conf.(IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018*. JACOW Publishing, Geneva, Switzerland. 2018, pp. 2502–2504 (cit. on pp. 6, 34).

[15] A. S. Nawaz, S. Pfeiffer, G. Lichtenberg, and H. Schlarb. "Self-organzied critical control for the european xfel using black box parameter identification for the quench detection system". In: *2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol)*. IEEE. 2016, pp. 196–201 (cit. on pp. 6, 19).

[16] D. M. Blei. "Build, compute, critique, repeat: Data analysis with latent variable models". In: *Annual Review of Statistics and Its Application* 1 (2014), pp. 203–232 (cit. on pp. 7, 97).

[17] W. a. Ackermann, G Asova, V Ayvazyan, A Azima, N Baboi, J Bähr, V Balandin, B Beutner, A Brandt, A Bolzmann, et al. "Operation of a free-electron laser from the extreme ultraviolet to the water window". In: *Nature photonics* 1.6 (2007), p. 336 (cit. on p. 9).

[18] M. O. Wiedorn, D. Oberthür, R. Bean, R. Schubert, N. Werner, B. Abbey, M. Aepfelbacher, L. Adriano, A. Allahgholi, N. Al-Qudami, et al. "Megahertz serial crystallography". In: *Nature communications* 9.1 (2018), pp. 1–11 (cit. on p. 10).

[19] T Tschentscher and R Feidenhans'l. "Starting User Operation at the European XFEL". In: *Synchrotron Radiation News* 30.6 (2017), pp. 21–28 (cit. on pp. 10, 11, 13).

[20] R Brinkmann, E. Schneidmiller, J Sekutowicz, and M. Yurkov. “Prospects for CW and LP operation of the European XFEL in hard X-ray regime”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 768 (2014), pp. 20–25 (cit. on pp. 10, 13).

[21] M. Ferrario, K. Flottmann, T. Limberg, P. Piot, and B. Grigorian. *Conceptual design of the XFEL photoinjector*. Tech. rep. TESLA-FEL 2001-03. 2001, p. 59 (cit. on p. 10).

[22] G Geloni, E Saldin, L Samoylova, E Schneidmiller, H Sinn, T. Tschentscher, and M Yurkov. “Coherence properties of the European XFEL”. In: *New Journal of Physics* 12.3 (2010), p. 035021 (cit. on p. 11).

[23] H. Padamsee, J. Knobloch, T. Hays, et al. *RF superconductivity for accelerators*. Vol. 2011. Wiley Online Library, 2008 (cit. on pp. 12, 17, 18).

[24] M. Pekeler. *Untersuchungen der feldbegrenzenden Mechanismen in supraleitenden Niob-Resonatoren*. Tech. rep. DESY, 1996 (cit. on p. 12).

[25] B. Aune, R Bandelmann, D Bloess, B Bonin, A Bosotti, M Champion, C Crawford, G Deppe, B Dwersteg, D. Edwards, et al. “Superconducting TESLA cavities”. In: *Physical Review Special Topics-Accelerators and Beams* 3.9 (2000), p. 092001 (cit. on p. 12).

[26] D. Bafia, A. Grassellino, Z Sung, A. Romanenko, O. Melnychuk, and J. Zasadzinski. “Gradients of 50 MV/m in TESLA Shaped Cavities via Modified Low Temperature Bake”. In: *19th Int. Conf. on RF Superconductivity (SRF’19), Dresden, Germany, 30 June-05 July 2019*. JACOW Publishing, Geneva, Switzerland. 2019, pp. 586–591 (cit. on p. 13).

[27] K. Przygoda, J. Branlard, W. Cichalewski, O. Hensler, K. Kasprzak, T. Pożniak, H. Schlarb, and C. Schmidt. “Testing procedures for fast frequency tuners of XFEL cavities”. In: *6th Int. Particle Accelerator Conf.(IPAC’15), Richmond, VA, USA, May 3-8, 2015*. JACOW, Geneva, Switzerland. 2015, pp. 2991–2993 (cit. on p. 13).

[28] P Sekalsky, C Albrecht, S Simrock, and L Lilje. *Lorentz Force Detuning Compensation System for Accelerating Field Gradients Up To 35 MV/m for Superconducting XFEL and Tesla Nine-Cell Cavities*. Tech. rep. 2004 (cit. on p. 13).

[29] R Paparella. “A control and systems theory approach to the high gradient cavity detuning compensation”. In: *this conference*. 2008 (cit. on p. 13).

[30] S. Pfeiffer. “Symmetric grey box identification and distributed beam-based controller design for free-electron lasers”. PhD thesis. Deutsches Elektronen-Synchrotron, Hamburg, 2014 (cit. on p. 14).

[31] C. Schmidt. *RF system modeling and controller design for the European XFEL*. Tech. rep. 2010 (cit. on p. 14).

[32] T. Schilcher. “Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities”. Doctoral dissertation. Universität Hamburg, 1998 (cit. on pp. 14, 16, 37, 39, 92).

[33] R. Paparella. “Fast frequency tuner for high gradient SC cavities for ILC and XFEL”. PhD thesis. INFN, Milan Bicocca, 2007 (cit. on p. 16).

[34] A. Neumann. “Compensating microphonics in SRF Cavities to ensure beam stability for future Free-Electron-Lasers”. PhD thesis. 2008 (cit. on pp. 16, 17, 38).

[35] M. Hoffmann. *Development of a multichannel RF field detector for the Low-Level RF control of the Free-Electron Laser at Hamburg*. Tech. rep. 2008 (cit. on p. 16).

[36] C Pagani, R Paparella, A Bosotti, P Pierini, P Michelato, and N Panzeri. *The fast piezo-blade tuner for SCRF resonators*. Tech. rep. 2005 (cit. on p. 17).

[37] M Liepe, W. Moeller, and S. N. Simrock. “Dynamic Lorentz force compensation with a fast piezoelectric tuner”. In: *PACS2001. Proceedings of the 2001 Particle Accelerator Conference (Cat. No. 01CH37268)*. Vol. 2. IEEE. 2001, pp. 1074–1076 (cit. on pp. 17, 18, 39).

[38] R. Rybaniec, V. Ayvazyan, J. Branlard, L. Butkowski, S Pfeiffer, H Schlarb, C. Schmidt, W Cichalewski, and K Przygoda. “Real-time estimation of superconducting cavities parameters”. In: *Proc. 5th Int. Particle Accelerator Conf.(IPAC14)*. 2014, pp. 2456–2458 (cit. on pp. 19, 50).

[39] M. S. Champion, L. D. Cooley, C. M. Ginsburg, D. A. Sergatskov, R. L. Geng, H. Hayano, Y. Iwashita, and Y. Tajima. “Quench-limited SRF cavities: failure at the heat-affected zone”. In: *IEEE transactions on applied superconductivity* 19.3 (2009), pp. 1384–1386 (cit. on p. 19).

[40] V Ayvazyan, H Schlarb, C Schmidt, O Hensler, and J. Branlard. “Superconducting Cavity Quench Detection and Prevention for the European XFEL”. In: *16th International Conference on RF Superconductivity*. DESY-2014-00617. HF-Technik Linearbeschleuniger. 2013 (cit. on p. 19).

[41] T. Wilksen, A. Aghababyan, L. Fröhlich, O. Hensler, R. Kammering, K. Rehlich, V. Rybnikov, et al. “A Bunch-Synchronized Data Acquisition System for the European XFEL Accelerator”. In: *16th Int. Conf. on Accelerator and Large Experimental Control Systems (ICALEPCS’17), Barcelona, Spain, 8-13 October 2017*. JACOW, Geneva, Switzerland. 2018, pp. 958–961 (cit. on p. 20).

[42] S. Pfeiffer, V. Ayvazyan, J. Branlard, L. Butkowski, R. Rybaniec, H. Schlarb, C. Schmidt, et al. "Virtual cavity probe generation using calibrated forward and reflected signals". In: *6th Int. Particle Accelerator Conf.(IPAC'15), Richmond, VA, USA, May 3-8, 2015*. JACOW, Geneva, Switzerland. 2015, pp. 200–202 (cit. on p. 21).

[43] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang. "The Factor Graph Approach to Model-Based Signal Processing". In: *Proceedings of the IEEE* 95.6 (2007), pp. 1295–1322 (cit. on pp. 22, 23, 25).

[44] H.-A. Loeliger. "An Introduction to Factor Graphs". In: *IEEE Signal Processing Magazine* 21.1 (2004), pp. 28–41 (cit. on p. 22).

[45] F. Pernkopf, R. Peharz, and S. Tschiatschek. "Introduction to probabilistic graphical models". In: *Academic Press Library in Signal Processing*. Vol. 1. Elsevier, 2014, pp. 989–1064 (cit. on p. 23).

[46] M. I. Jordan. *An introduction to probabilistic graphical models*. 2003 (cit. on pp. 23, 45).

[47] G. D. Forney. "Codes on graphs: Normal realizations". In: *IEEE Transactions on Information Theory* 47.2 (2001), pp. 520–548 (cit. on p. 23).

[48] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. "Factor Graphs and the Sum-Product Algorithm". In: *IEEE Transactions on Information Theory* 47.2 (2001), pp. 498–519 (cit. on p. 24).

[49] J. Winn and C. M. Bishop. "Variational message passing". In: *Journal of Machine Learning Research* 6.Apr (2005), pp. 661–694 (cit. on p. 24).

[50] T. P. Minka. "Expectation propagation for approximate Bayesian inference". In: *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 2001, pp. 362–369 (cit. on p. 24).

[51] J. Dauwels, S. Korl, and H.-A. Loeliger. "Particle methods as message passing". In: *2006 IEEE International Symposium on Information Theory*. IEEE. 2006, pp. 2052–2056 (cit. on p. 24).

[52] J. A. Bilmes et al. "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models". In: *International Computer Science Institute* 4.510 (1998), p. 126 (cit. on pp. 27, 28, 61).

[53] D. A. Reynolds. "Gaussian Mixture Models." In: *Encyclopedia of biometrics* 741 (2009) (cit. on p. 27).

[54] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". In: *Journal of the Royal Statistical Society. Series B (methodological)* 39.1 (1977), pp. 1–38 (cit. on p. 27).

[55] D. Arthur and S. Vassilvitskii. *k-means++: The advantages of careful seeding*. Tech. rep. Stanford, 2006 (cit. on p. 28).

[56] I. Kerenidis, A. Luongo, and A. Prakash. "Quantum expectation-maximization for Gaussian mixture models". In: *arXiv preprint arXiv:1908.06657* (2019) (cit. on pp. 28, 88).

[57] J. Dauwels, S. Korl, and H. Loeliger. "Expectation Maximization as Message Passing". In: *Proceedings of International Symposium on Information Theory*. 2005, pp. 583–586 (cit. on pp. 28, 60).

[58] R. J. Patton and J. Chen. "A review of parity space approaches to fault diagnosis". In: *IFAC Proceedings Volumes* 24.6 (1991), pp. 65–81 (cit. on p. 29).

[59] Z. Gao, C. Cecati, and S. X. Ding. "A survey of fault diagnosis and fault-tolerant techniquesPart I: Fault diagnosis with model-based and signal-based approaches". In: *IEEE Transactions on Industrial Electronics* 62.6 (2015), pp. 3757–3767 (cit. on p. 29).

[60] R. Isermann. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Berlin and New York: Springer, 2006. ISBN: 978-3-540-24112-6 (cit. on pp. 29, 30, 33, 46).

[61] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri. "A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies". In: *Computers & chemical engineering* 27.3 (2003), pp. 313–326 (cit. on pp. 29, 45).

[62] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin. "A review of process fault detection and diagnosis: Part III: Process history based methods". In: *Computers & chemical engineering* 27.3 (2003), pp. 327–346 (cit. on pp. 29, 45).

[63] M. Kinnaert. "Fault diagnosis based on analytical models for linear and nonlinear systems-a tutorial". In: *IFAC Proceedings Volumes* 36.5 (2003), pp. 37–50 (cit. on p. 29).

[64] E. Chow and A. Willsky. "Analytical redundancy and the design of robust failure detection systems". In: *IEEE Transactions on automatic control* 29.7 (1984), pp. 603–614 (cit. on p. 29).

[65] J. Bokor and Z. Szabó. "Fault detection and isolation in nonlinear systems". In: *Annual Reviews in Control* 33.2 (2009), pp. 113–123 (cit. on p. 30).

[66] T. Mulumba, A. Afshari, K. Yan, W. Shen, and L. K. Norford. “Robust model-based fault diagnosis for air handling units”. In: *Energy and Buildings* 86 (2015), pp. 698–707 (cit. on p. 30).

[67] S. Kumar, S Sinha, T Kojima, and H Yoshida. “Development of parameter based fault detection and diagnosis technique for energy efficient building management system”. In: *Energy Conversion and Management* 42.7 (2001), pp. 833–854 (cit. on p. 30).

[68] Y. Hu, B. Gao, X. Song, G. Y. Tian, K. Li, and X. He. “Photovoltaic fault detection using a parameter based model”. In: *Solar Energy* 96 (2013), pp. 96–102 (cit. on p. 30).

[69] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft. “Deep one-class classification”. In: *International conference on machine learning*. 2018, pp. 4393–4402 (cit. on p. 33).

[70] J Branlard, V Ayvazyan, T Jezynski, H Schlarb, W Cichalewski, and W Jałmuzna. “LLRF testing of superconducting cryomodules for the European XFEL”. In: *Proc. 3rd Int. Part. Accel. Conf.(IPAC)*. 2012, pp. 3263–3265 (cit. on p. 38).

[71] J. Branlard, V Ayvazyan, M. Grecki, H Schlarb, C. Schmidt, W Cichalewski, K Gnidzinska, A Piotrowski, K Przygoda, and W Jalmuzna. “LLRF tests of XFEL cryomodules at AMTF: first experimental results”. In: *Proc. SRF*. 2013, pp. 1126–1128 (cit. on p. 38).

[72] J. Neyman. “Outline of a theory of statistical estimation based on the classical theory of probability”. In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 236.767 (1937), pp. 333–380 (cit. on p. 38).

[73] T. Czarski, K. T. Pozniak, R. S. Romaniuk, and S. Simrock. “Cavity parameters identification for TESLA control system development”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 548.3 (2005), pp. 283–297 (cit. on pp. 38, 39).

[74] J. Jugo, A. Elejaga, and P. Echevarria. “Control algorithm tests using a virtual CW SRF cavity”. In: *arXiv preprint arXiv:1910.06227* (2019) (cit. on pp. 38, 39).

[75] *MATLAB version 9.3.0.713579 (R2017b)*. The Mathworks, Inc. Natick, Massachusetts, 2017 (cit. on p. 42).

[76] J. Bergstra and Y. Bengio. “Random search for hyper-parameter optimization”. In: *The Journal of Machine Learning Research* 13.1 (2012), pp. 281–305 (cit. on p. 42).

[77] X. He, K. Zhao, and X. Chu. “AutoML: A Survey of the State-of-the-Art”. In: *arXiv preprint arXiv:1908.00709* (2019) (cit. on p. 42).

[78] S. X. Ding. *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Springer Science & Business Media, 2008 (cit. on pp. 45, 56, 57).

[79] J. Gertler. “Structured residuals for fault isolation, disturbance decoupling and modelling error robustness”. In: *IFAC Proceedings Volumes* 25.4 (1992), pp. 15–23 (cit. on p. 45).

[80] S. J. Julier and J. K. Uhlmann. “New extension of the Kalman filter to nonlinear systems”. In: *Signal processing, sensor fusion, and target recognition VI*. Vol. 3068. International Society for Optics and Photonics. 1997, pp. 182–193 (cit. on p. 46).

[81] R. K. Mehra and J Peschon. “An innovations approach to fault detection and diagnosis in dynamic systems”. In: *Automatica* 7.5 (1971), pp. 637–640 (cit. on p. 46).

[82] P. S. Nag, G. kumar Silla, V. H. V. Gummadi, C. Harishankar, V. K. Ray, and C. S. Kumar. “Model based fault diagnosis of low earth orbiting (leo) satellite using spherical unscented kalman filter”. In: *IFAC-PapersOnLine* 49.1 (2016), pp. 635–638 (cit. on p. 46).

[83] X. Wei, M. Verhaegen, and T. van Engelen. “Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques”. In: *International Journal of Adaptive Control and Signal Processing* 24.8 (2010), pp. 687–707 (cit. on p. 46).

[84] B. Cai, L. Huang, and M. Xie. “Bayesian networks in fault diagnosis”. In: *IEEE Transactions on Industrial Informatics* 13.5 (2017), pp. 2227–2240 (cit. on p. 46).

[85] K. Verbert. “Fault diagnosis and maintenance optimization for interconnected systems: With applications to railway and climate control systems”. In: (2016) (cit. on p. 46).

[86] M. Basseville, I. V. Nikiforov, et al. *Detection of abrupt changes: theory and application*. Vol. 104. prentice Hall Englewood Cliffs, 1993 (cit. on pp. 46, 56).

[87] S. Zarrin and T. J. Lim. “Belief propagation on factor graphs for cooperative spectrum sensing in cognitive radio”. In: *2008 3rd IEEE symposium on new frontiers in dynamic spectrum access networks*. IEEE. 2008, pp. 1–9 (cit. on p. 46).

[88] T. van de Laar and B. de Vries. “A probabilistic modeling approach to hearing loss compensation”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24.11 (2016), pp. 2200–2213 (cit. on p. 46).

[89] D. Lu, P. Mausel, E. Brondizio, and E. Moran. “Change detection techniques”. In: *International journal of remote sensing* 25.12 (2004), pp. 2365–2401 (cit. on p. 46).

[90] V. Chandola, A. Banerjee, and V. Kumar. “Anomaly detection: A survey”. In: *ACM computing surveys (CSUR)* 41.3 (2009), pp. 1–58 (cit. on pp. 46, 69, 92).

[91] M. Heydarzadeh and M. Nourani. “A two-stage fault detection and isolation platform for industrial systems using residual evaluation”. In: *IEEE Transactions on Instrumentation and Measurement* 65.10 (2016), pp. 2424–2432 (cit. on p. 46).

[92] J. Yu. “Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models”. In: *Mechanical Systems and Signal Processing* 25.7 (2011), pp. 2573–2588 (cit. on p. 46).

[93] M. L. Schwall, J. C. Gerdes, B. Bäker, and T. M. Forchert. “A Probabilistic Vehicle Diagnostic System Using Multiple Models.” In: *IAAI*. 2003, pp. 123–128 (cit. on p. 46).

[94] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. E. Orchard, and G. Vachtsevanos. “A probabilistic fault detection approach: Application to bearing fault detection”. In: *IEEE Transactions on Industrial Electronics* 58.5 (2010), pp. 2011–2018 (cit. on p. 46).

[95] A. Ozerov, M. Lagrange, and E. Vincent. “Uncertainty-Based Learning of Gaussian Mixture Models from Noisy Data”. In: *Computer Speech and Language* 27.3 (2013), pp. 874–894 (cit. on pp. 47, 60, 62).

[96] H.-A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmai. “On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing”. In: *2016 Information Theory and Applications Workshop (ITA)*. IEEE. 2016, pp. 1–10 (cit. on p. 47).

[97] C. Herzog, E. Petersen, P. Rostalski, et al. “Iterative Approximate Nonlinear Inference via Gaussian Message Passing on Factor Graphs”. In: *IEEE Control Systems Letters* (2019) (cit. on p. 48).

[98] B. M. Åkesson, J. B. Jørgensen, N. K. Poulsen, and S. B. Jørgensen. “A generalized autocovariance least-squares method for Kalman filter tuning”. In: *Journal of Process control* 18.7-8 (2008), pp. 769–779 (cit. on p. 49).

[99] B. Efron, A. Gous, R. Kass, G. Datta, and P Lahiri. “Scales of evidence for model selection: Fisher versus Jeffreys”. In: *Lecture Notes-Monograph Series* (2001), pp. 208–256 (cit. on p. 58).

[100] D. Kolossa, R. F. Astudillo, E. Hoffmann, and R. Orglmeister. “Independent Component Analysis and Time-Frequency Masking for Speech Recognition in Multitalker Conditions”. In: *EURASIP Journal on Audio, Speech, and Music Processing* 2010.1 (2010), p. 651420 (cit. on p. 59).

[101] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science & Business Media, 2006 (cit. on p. 59).

[102] A. R. Runnalls. “Kullback-Leibler Approach to Gaussian Mixture Reduction”. In: *IEEE Transactions on Aerospace and Electronic Systems* 43.3 (2007), pp. 989 –999 (cit. on p. 65).

[103] S. Kokoska and C. Nevison. “Critical Values For The Chi-Square Distribution”. In: *Statistical Tables and Formulae*. New York, NY: Springer New York, 1989, pp. 58–59 (cit. on p. 70).

[104] D. Simon. “Optimal state estimation: Kalman, H, and nonlinear approaches. Hoboken”. In: *NJ: John Wiley and Sons* 10 (2006), p. 0470045345 (cit. on p. 91).

[105] J. Plouin, B. Baudouy, A. Four, J. P. Charrier, L. Maurice, J. Novo, B. J. Peters, and K. Liao. “Experimental study of second sound quench detection for superconducting cavities”. In: *Phys. Rev. Accel. Beams* 22 (8 2019), p. 083202 (cit. on p. 94).

[106] I. Guttman. “The use of the concept of a future observation in goodness-of-fit problems”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 29.1 (1967), pp. 83–100 (cit. on p. 97).

[107] G. I. Webb, R. Hyde, H. Cao, H. L. Nguyen, and F. Petitjean. “Characterizing concept drift”. In: *Data Mining and Knowledge Discovery* 30.4 (2016), pp. 964–994 (cit. on p. 105).

[108] R. Elwell and R. Polikar. “Incremental learning of concept drift in nonstationary environments”. In: *IEEE Transactions on Neural Networks* 22.10 (2011), pp. 1517–1531 (cit. on pp. 105, 115).

[109] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar. “Learning in nonstationary environments: A survey”. In: *IEEE Computational Intelligence Magazine* 10.4 (2015), pp. 12–25 (cit. on p. 105).

[110] A. Bellandi, Ł. Butkowski, B. Dursun, A. Eichler, Ç. Gümüş, M. Kuntzsch, A. Nawaz, S. Pfeiffer, H. Schlarb, C. Schmidt, et al. “Online Detuning Computation and Quench Detection for Superconducting Resonators”. In: *IEEE Transactions on Nuclear Science* (2021) (cit. on p. 115).

[111] A Ushakov, P Echevarria, A Neumann, et al. “Developing Kalman filter based detuning control with a digital SRF CW cavity simulator”. In: *9th Int. Particle Accelerator Conf.(IPAC18), Vancouver, Canada, Apr.-May*. 2018 (cit. on p. 115).

[112] H.-C. Yan, J.-H. Zhou, and C. K. Pang. “Gaussian mixture model using semisupervised learning for probabilistic fault diagnosis under new data categories”. In: *IEEE Transactions on Instrumentation and Measurement* 66.4 (2017), pp. 723–733 (cit. on p. 115).

[113] K. Petersen and M. Pedersen. “The matrix cookbook. technical university of denmark”. In: *Technical Manual* (2008) (cit. on p. 120).