5-2016

Quickest Fault Detection in Direct Drive Wind Turbine Using Stator Current

Vitaly Romanov
University of Arkansas, Fayetteville

Quickest Fault Detection in Direct Drive Wind Turbine Using Stator Current

A thesis submitter is partial fulfillment
of the requirements for the degree of
Master of Science in Electrical Engineering

by

Vitaly Romanov
Tupolev Kazan National Research Technical University
Engineering Degree in Radio Electronics, 2014

May 2016
University of Arkansas

This thesis is approved for recommendation to the Graduate Council

_____
Dr. Jingxian Wu
Thesis director

_____
Dr. Jing Yang
Committee Member

_____
Dr. Roy McCann
Committee Member

Abstract

A new online method for detecting bearing faults in direct-drive wind turbines (WT), based on stator current analysis and quick change point detection techniques, is proposed. Bearing faults are common problems for machinery, such as wind turbines. Online methods of fault detection give clear advantages by allowing the diagnosis of wind turbines without physical access. Stator current signal is always available, it is perfect for online analysis, and no additional sensors are required. Faults of bearings supporting the main shaft of direct-drive WT introduce excitations into the stator current spectrum. Change-detection procedures provide a simple and yet efficient way to find that change. In the proposed method, the stator current signal is synchronously resampled, and fault signatures are extracted using FFT. Then, distribution of a fault signal is estimated. After that, the condition of bearings is tested using a quickest change detection method such as CUSUM or Shiryaev-Roberts procedures.

Contents

1 Introduction

1.1 Motivation

Wind is one of the top sources of renewable energy. At the same time, compared to other types of energy generators, wind turbines (WT) usually operate in harsher environments, and hence, are more susceptible to failures [1]. A comprehensive survey of WT subsystem faults and methods of WT faults diagnosis was done in [1, 2]. WT fault diagnosis and prediction are of high importance for the wind power industry. No matter what type of fault has occurred - fault due to exhaustion or manufacturing defect - unexpected outage is always a cause of economic loss. One way to minimize this loss is to predict and detect a malfunction before critical damage is done. With this taken into consideration, online methods of fault detection have a clear advantage over offline methods. For diagnosing some particular types of faults, they facilitate the inspection by alleviating the need of taking a WT out of operation. In addition, methods of online diagnosis allow automatic remote examination, which is crucial for WT located in hardly accessible areas. When properly implemented, methods of online fault diagnosis will allow to predict the occurrence of a fault beforehand and identify the malfunctioning unit.

Wind turbines can be classified into two types: WT with a gearbox, and direct-drive WT. Over the recent years the design of direct-drive WT became more attractive than before. WT of this type do not have one of the complex subsystems, the gearbox, which theoretically can make this turbine design more reliable. However, higher cost and increased weight [3] of direct-drive wind turbines makes them not as attractive as WT of a conventional design with a gearbox. Moreover, according to [4], particular types of faults occurs more frequently in direct-drive wind turbines. However, something might have changed over the last ten years.

Wind turbines are complex systems with different subsystems. Each subsystem may be subject to several different types of faults. According to [5–8], bearing faults constitute a significant portion of failures in WT. The study of this type of failure is not new [5, 9–13].

Bearings are parts of different subsystems of a wind turbine, such as the pitch system, yaw system, gearbox, generator, main shaft, and rotor. The most commonly used type of bearings are ball bearings [1]. Bearing faults can be classified into four groups: inner race, outer race, cage, and rollers faults. Many defects, start with faulty bearings.

Different methods are used to identify WT compartments failures, including analyzing mechanical vibrations, acoustic emissions, temperature, oil parameters and electrical signals. The latter has started to get more attention recently due to its advantages [8, 14–17]. Electrical signals, such as current signal, are available remotely, do not require any additional sensors, and can be analyzed in real time. Since faults of a bearing are associated with mechanical defect, they introduce excitation at particular frequencies [9, 17]. For some bearings these excitations can be captured by electrical signal in the form of harmonics that modulate the amplitude of the generated frequency [18]. Over the years different processing techniques were used to extract bearing fault signatures from electrical signals such as synchronous sampling, envelope analysis, Hilbert transform, FFT, etc.

Most WT operate in variable speed conditions, which results in a current signal with non-stationary frequency. Synchronous sampling is aimed at solving this problem [18]. The fault frequencies that depend on rotation speed appear as constant values after applying this method, which simplifies further analysis. Synchronous sampling showed to be reliable and computationally efficient.

Hilbert transform is a classical time-domain analysis method, and it can be used to extract instantaneous amplitude and phase from the current signal. This method is used to demodulate the current signal and extract fault signatures [19, 20].

The Envelope of a signal is its amplitude modulating component. Since, as it was described above, bearing fault results in amplitude modulation of the stator current, the signal envelope contains information required for bearing condition diagnosis [21]. Besides Hilbert transform, envelope can be obtained using other demodulation techniques.

FFT reveals the power of different harmonics in signal spectrum and can serve as a fault

signature extraction tool [17, 18, 22, 23]. Indeed, for the case of faulty bearings, where faults appear as excitation on particular frequencies, this method has proven itself to be useful.

Besides methods that serve solely for feature extraction, there are techniques aimed for bearing fault diagnosis. Some work was done to construct simple threshold based methods for bearing fault detection in [17]. Wavelet analysis was used to detect incipient faults of bearings [24]. In some research, statistical methods were applied to detect bearing faults, as well as other types of faults [25–27]. Quantitative analysis of mean, variance, RMS, crest-factor, kurtosis, etc. is usually applied. Often, signals from additional sensors are used. However, it may be hard to identify a faulty component using this technique, depending on what signals are used, since different faults can result in similar changes of signal's statistical properties. A maximum likelihood method for detecting bearing faults was proposed in [28]. A class of methods, referred to as Bayesian methods, exists. In these methods, system state is modeled through Hidden Markov Model (HMM), and data can be used, to calculate probability of system states [29, 30]. Such methods require a training stage. Also, several papers devoted to another statistical fault detection method, based on particle filtering, were found [31–33]. Such methods are considered as computationally intensive. Besides the statistical methods listed above, a lot of research has been done regarding artificial intelligence (AI) methods for fault detection in different components of WT [34–36]. The main disadvantage of all AI techniques is high computational intensity, and then need for training.

1.2 Objectives

No recent research was found featuring simple yet powerful change detection techniques for bearing faults detection. New components, introduced by a fault, change statistical properties of the observed signal. If it is possible to estimate probability density function (PDF) of signal for normal and faulty states, a change detection procedure, such as Shiryaev-Roberts procedure or CUSUM control charts, can be applied. The problem of fault detection will become a problem of hypothesis testing, where it is assumed that a WT operates normally first, then, at

some moment of time, change from normal to faulty state occurs. The key for success for this method is the ability to obtain probability distributions for normal and defective states, which, as will be shown later in this thesis, can be achieved.

The change-point detection techniques are not new. The study of these techniques' properties was done in [37–46]. Over the years, the theory of change detection was developed, and theoretically efficient methods were designed.

In the center of attention of this thesis is the detection of bearing faults in direct-drive WT with permanent magnet synchronous generator (PMSG). In [8, 17, 47, 48] this type of WT was used to generate and record data, which contain several types of bearing fault signatures. The data recorded for [17] were used in this thesis, to design a new method for automatic detection of bearing faults. Since different faults, such as the blade imbalance, aerodynamic asymmetry, rotor eccentricity, torque bending moment, etc. [17] also produce excitations at some characteristic frequencies, the proposed method can be applied to detect these other types of faults as well, as long as their fault frequencies can be identified beforehand.

1.3 Outline

The rest of the thesis is sectioned in the following way. In the Section 2 the overview of possible WT faults is done, and the methods of fault extraction and signal processing, used for fault detection, are discussed. The distinction between direct drive and gearbox WT is done. Since WT is a complex system, faults of different subsystems are reviewed. Particular attention is paid to the methods of bearing faults signatures extraction, and bearing faults detection techniques. Limitations of popular techniques are given.

In the Section 3 the theory of change detection is given, and different change detection procedures are introduced. First, the change detection problem is defined. Then, a brief overview of the history of change detection techniques is given. The general form of a stopping rule for change detection procedures is defined. The overview of several major change detection techniques is done. Particularly Shiryaev, CUSUM, and Shiryaev-Roberts proce-

dures are given, and their performance limitations are stated. The problem of a mismatched post change distribution was faced during the fault detection algorithm design. To justify the performance of change detection techniques in case of a post-change distribution mismatch, a review of CUSUM and Shiryaev-Roberts procedures performances for the mismatched post-change distribution case is done.

In Section 4, the newly proposed method of bearing fault detection is explained. In conclusion, the summary of the new method is given. The method of fault signature extraction is based on synchronous resampling and FFT analysis. The fault signatures are combined in feature vectors. It was found that components of feature vectors are Gamma distributed, and after a fault occurs, parameters of the distribution change. After the analysis of distribution is done, the hypothesis for a change detection procedure are formulated. The methods of quickest change detection are applied to identify the presence of a fault. Different approaches to perform the change detection are introduced, and their performance is compared.

2 Defects of Wind Turbines and Methods of Their Identification

2.1 Introduction

By the type of transmission, the wind turbines can be split into two different categories: wind turbines with a gearbox, and direct-drive wind turbines. The mechanical design of these two types of WT is significantly different. Direct-drive wind turbines usually use special type of generators that can work on low RPM speed. The stator current signal of a gearbox WT has additional signals excited by wheels and bearings of the gearbox. That suggests a higher level of noise in the stator current. Direct-drive wind turbines, on the other hand, have hub with the wind blades directly connected to the generator. Now we are going to discuss basic differences between the gearbox and direct drive wind turbine designs. The main subsystems will be shown. Further, in the section 2.3, the main types of faults appearing in wind turbines are described. After possible faults of different WT subsystems are described, it is important to understand the most common signal processing techniques that are used for fault features extraction and fault detection. The section 2.4 is split into two parts. The first one describes feature extraction techniques for different types of faults. The second one tells about fault detection methods. Our goal is to review existing fault detection methods and understand, which of them can be used for bearing fault detection.

2.2 Comparison of Direct-Drive and Gearbox Wind Turbine Designs

Wind turbines are complicated machinery, that consist of many subsystems, which together make harvesting of the wind energy possible. At the same time, as the number of system components increase, the reliability usually tends to decrease.

A conventional design of a horizontal wind turbine is shown of Figure 1. Keeping the auxiliary subsystems aside, this type of wind turbines consists of several major components: the wind blades, shaft, gearbox, and generator. Wind induced low speed rotation of the wind blades. However, most of conventional design generators require high RMP value in order

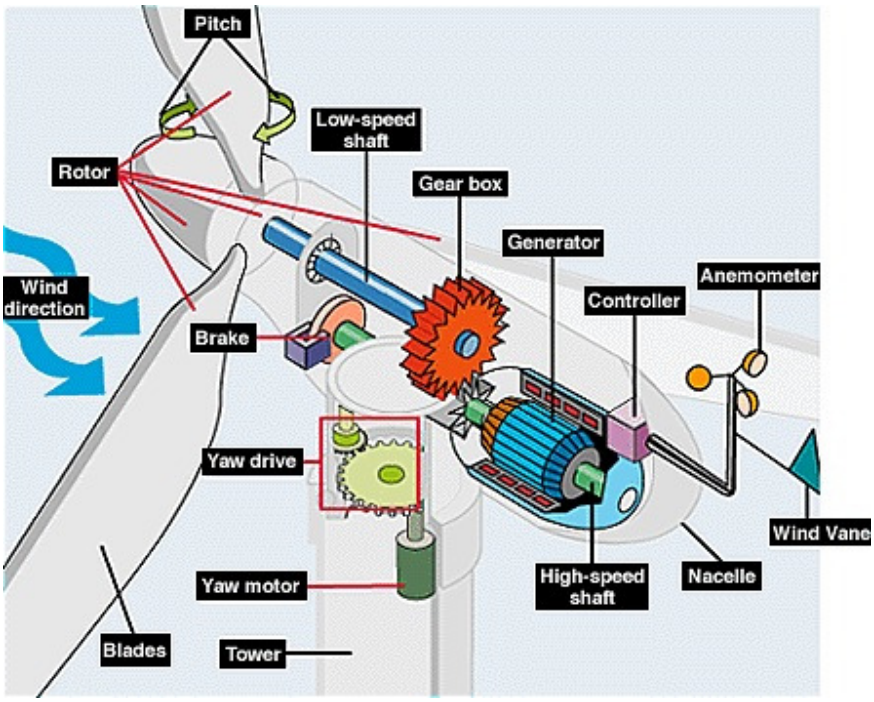

Figure 1: Conventional wind turbine [49]

to be efficient. The gearbox is designed to convert low speed rotations of the wind blades into high speed revolutions of the generator shaft. The relation of rotation speed before and after the gearbox for some turbines can be such as 15-20 RPM for the hub and about 1800 RPM for the generator [3]. Wheels and bearings of the gearbox are subject to high stress. In some designs, the gearbox can be equipped with an adjustable speed drive or continuously variable transmission. As expected, such a complicated device is prone to failures. One of the solutions to increase the reliability of a wind turbine is to eliminate the gearbox. Direct-drive wind turbines are designed to achieve that.

A schematic diagram of a direct-drive wind turbine is shown in Figure 2. Modern direct-drive wind turbines often use a permanent magnet to generate the electric power. No gearbox is used, which simplifies the design. This type of WTs used to be heavier and more expensive. However, latest advancements in direct-drive WTs allowed to achieve a reasonable reduction of the cost and weight [3]. The shaft, on which the wind blades are located, is supported by bearings and is directly connected to the generator. The generators that use permanent mag-

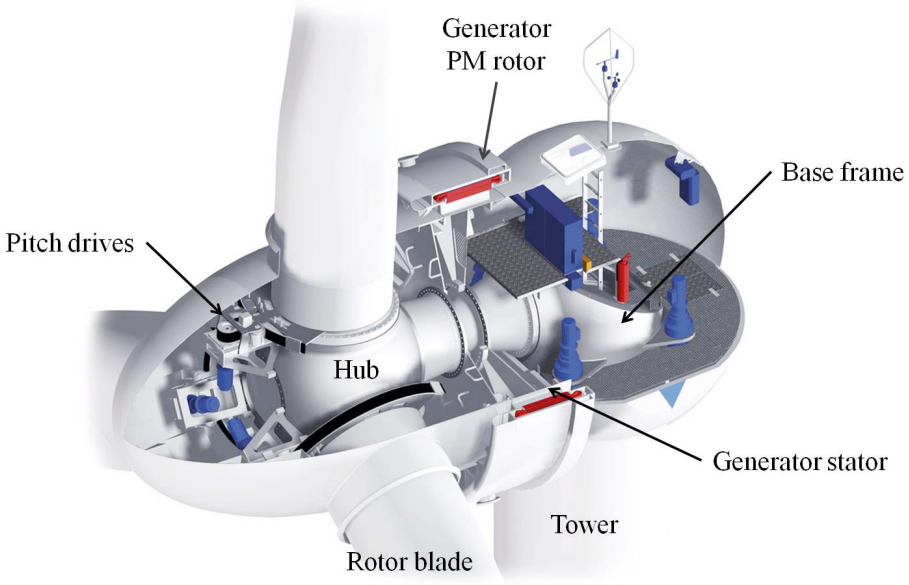

Figure 2: Direct-drive wind turbine [50]

nets (instead of a coil) are classified as permanent magnet synchronous generators (PMSG), where synchronous signifies that rotor and magnetic field rotate with the same speed. Direct-drive WTs also have many subsystems that ensure proper operation. Nevertheless, there is less mechanical stress on the components directly responsible for generating electricity.

Presence of a large number of components in a WT requires large number of sensors that are used to monitor condition of WT compartments and make adjustments for improving the efficiency of power generation. Some WTs are equipped with built-in condition monitoring system, which gives broad opportunities for online diagnosis. Some common faults of WT components are given further.

2.3 Types of Wind Turbine Faults

Failure of some WT subsystems can result in outage, loss of efficiency and money. It is important to be able to detect malfunctions before the critical damage is done, and even more loss is taken. A comprehensive review of the major subsystems faults was done in [1]. A summary of this paper is given further.

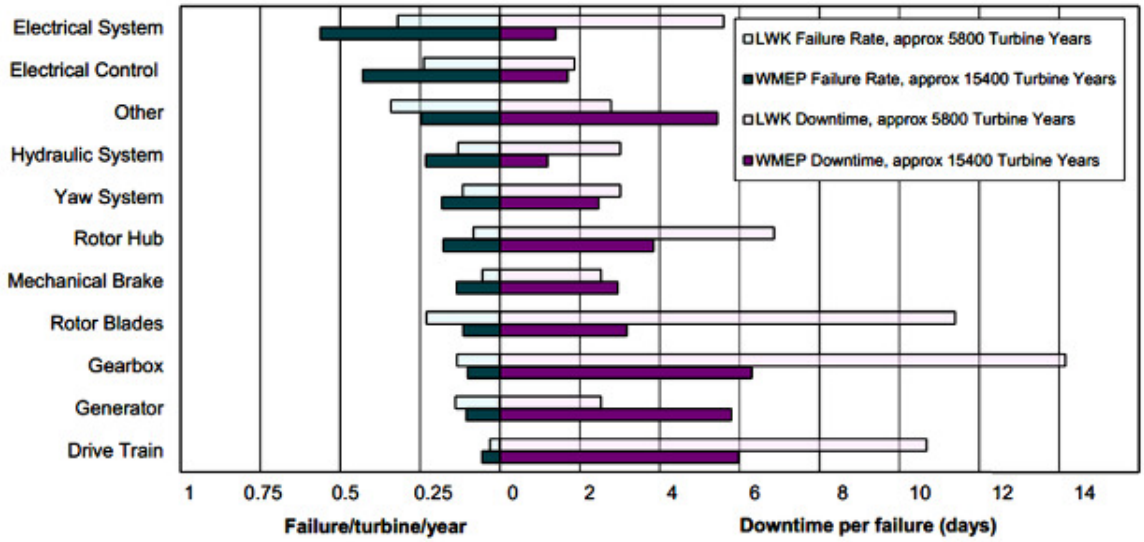

Figure 3: Failure frequencies of major WT subsystems and corresponding downtime [51]

Some subsystems of a WT are shown in Figures 1 and 2. The faults in those subsystem result in specific changes in operation parameters. Different techniques are used to detect fault in different subsystems. Now we are going to discuss some of the main subsystems of WT, the types of faults they can experience, what those faults are caused by, and how to detect them.

Rotor Hub and Blades

The rotor of a wind turbine consists of the hub with blades. Some of possible defects include rotor asymmetries, fatigue, cracks, increased surface roughness, and blades deformation. The rotor asymmetry can be caused by a pitch angle mismatch. Fatigue of the wind blades is usually caused by material aging. In long term, the fatigue can cause cracks. Increased surface roughness is usually a consequence of harsh environment. Deformation of blades can be caused by an unbalanced loading and reduced stiffness of material.

The defects, which are related to structural change, can be detected using acoustic emission sensors. If the material is worn out to a high degree, abnormal vibrations will appear in the signals acquired from vibration sensors. Also, cracks, increased surface roughness, rotor asymmetries can cause excitations on characteristic frequencies in rotor vibration signals.

These vibrations further affect the generator, and can be detected by analyzing the current signal.

Gearbox

The gearbox failure is one of the most common problems in conventional wind turbines. It constitutes up to 20% of WT's downtime. The faults of the gearbox can be caused by manufacturing and installation errors, misalignment, torque overloads, material wear, fatigue, poor lubrication and others. Most gearbox failures start with a broken bearing. The broken bearing particles cause damage of gears and tooth breaks.

The most common way to detect gearbox malfunctioning is by vibration analysis. Some other methods, such as acoustic emission, current and temperature based analysis methods has become more popular lately. Signals from vibration sensors installed on the gearbox are usually used. Excitations on characteristic frequencies will appear in those signals in case of some malfunction. Characteristic frequencies are expressed through (1).

$$f_{gb} = \left\{ f \left| f = \sum_{i=1}^{I} l_i f_{sh,i} \pm \sum_{j=1}^{J} m_j f_{m,j};\, l_i, m_j = 0,1,2,... \right. \right\}, \tag{1}$$

where $f_{sh,i}$ is the rotating frequency of the ith shaft in the gearbox; $f_{m,j}$ is the gear meshing frequency, I and J are numbers of shafts and gear pairs in the gearbox.

In case if the current signal is used, characteristic frequencies modulate the generated power frequency, and changes appear in the current spectrum. Characteristic frequencies for current signal are shown in (2).

$$f_{cc} = \left\{ f \left| f = k f_c + \sum_{i=1}^{I} l_i f_{sh,i} \pm \sum_{j=1}^{J} m_j f_{m,j};\, l_i, m_j = 0,1,2,... \right. \right\}, \tag{2}$$

where f_c is the generator shaft rotation frequency, and k is the number of harmonic.

Sometimes defects of the gearbox can be detected by analyzing the temperature of bearings. Some information is also available in the lubrication oil. Such parameters as the tempera-

ture, viscosity, and particles can indicate the presence of a fault.

Bearing

Bearings are parts of different wind turbine subsystems. The typical faults of a bearing are increased surface roughness, fatigue, cracks, or breakage of main components: inner race, outer race, cage, rollers. The faults of a bearing are associated with individual characteristic frequencies [8]:

$$f_{\mathrm{ir}} = 0.5 \cdot N_B \cdot f_{\mathrm{r}} \left(1 + \frac{D_b \cdot \cos\phi}{D_p}\right) \tag{3}$$

$$f_{\mathrm{or}} = 0.5 \cdot N_B \cdot f_{\mathrm{r}} \left(1 - \frac{D_b \cdot \cos\phi}{D_p}\right) \tag{4}$$

$$f_{\mathrm{b}} = f_{\mathrm{r}} \cdot \left(\frac{D_b}{D_c}\right) \cdot \left[1 - \left(\frac{D_b \cdot \cos\phi}{D_p}\right)^2\right] \tag{5}$$

$$f_{\mathrm{c}} = 0.5 \cdot f_{\mathrm{r}} \left(1 - \frac{D_b \cdot \cos\phi}{D_p}\right) \tag{6}$$

where f_{r} is the rotation frequency of a bearing, D_b is the diameter of rollers, D_p is the rollers' pitch diameter, N_B is the number of rollers, and ϕ is rollers contact angle with races (see Figure 4). The defects of the inner and outer races cause eccentricity. As it is shown in Figures 5 and 6, surface defects of races induce movements of the bearing center, which essentially is eccentricity. Defects of rollers are shown in Figure 7.

The common way to diagnose a bearing fault is by analyzing vibration signals. However, methods of bearings inspection based on the analysis of electrical signals has drawn more attention recently. One more way of checking the the bearing condition is by monitoring lubrication oil parameters (temperature, viscosity).

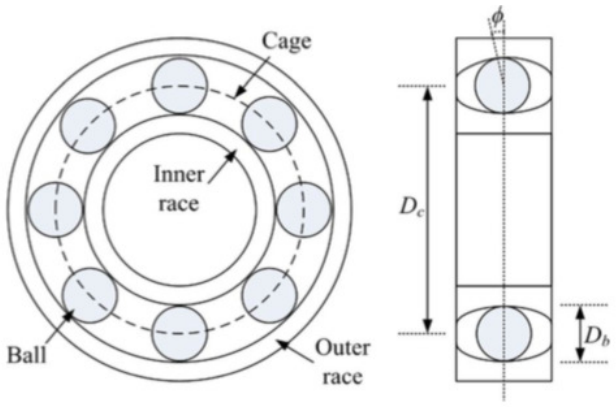

Figure 4: Configuration of a bearing [17]

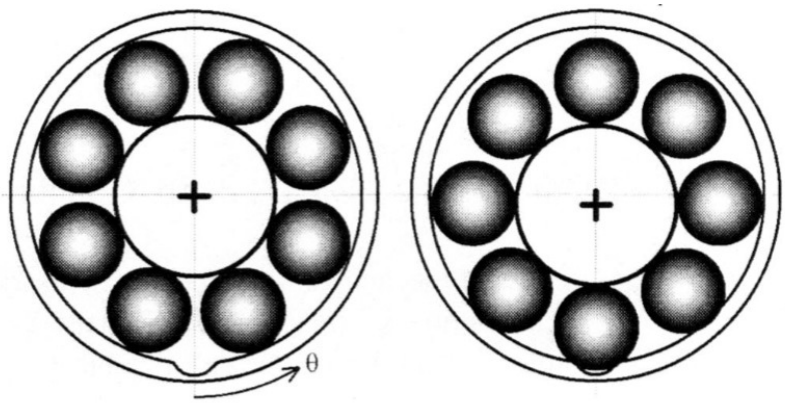

Figure 5: The effect of the outer race fault [5]

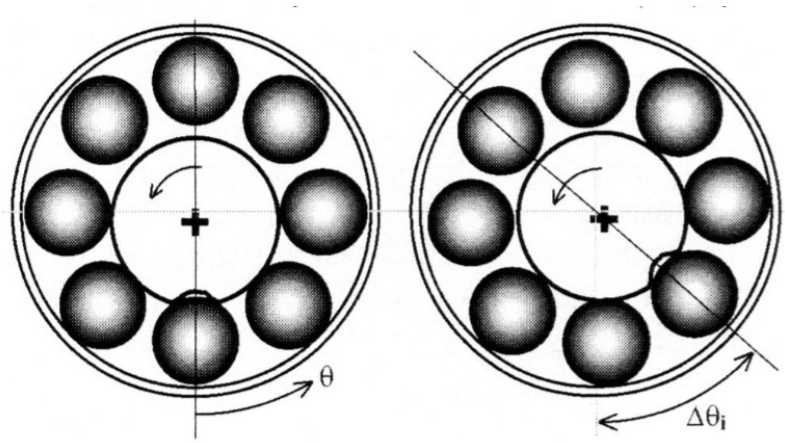

Figure 6: The effect of inner race fault [5]

Figure 7: The surface of a bearing ball (a) healthy surface (b) etched ball surface (c) grooved ball surface [10]

Main Shaft

The common defects of the main shaft are corrosion, cracks and misalignment. Due to these imperfections, vibrations can appear at particular frequencies in the rotor, gearbox and generator. Since the generator is affected, these types of faults can be detected using the current signal.

Mechanical Break

The mechanical break is a very important WT subsystem designed to prevent further degradation in case of failures or to stop WT in harsh weather conditions. The main components of mechanical brake are the disc and calipers, hydraulic system, and AC motor, powering the hydraulic mechanism. Brakes are subject to high mechanical stress, therefore, cracks and other failures due to degradation and overheating are common. However, little research was done to diagnose faults of the disc and calipers. The electric motor defects of hydraulic system are usually identified by the current signal.

Tower

The tower, as the part that supports a wind turbine, experiences high loads due to environmental conditions such as winds, rain, lightning, fire, etc. The common defects in towers are cracks and corrosions, which can be identified by analyzing tower vibrations.

Electric Machine

Electric machines includes electric generators and motors. Faults of these devices can be classified into mechanical or electrical. Electrical faults include the stator or rotor insulation damage, electrical imbalance, or short circuit. Mechanical faults include broken rotor bar, bearing failure, bent shaft, rotor mass imbalance and others.

Short circuit of coils is one of the most common faults. They usually introduce asymmetry in magnetic field, and can be determined by analyzing characteristic frequencies (7) in electrical signals from the electrical machine terminal.

$$f_w = \left\{ f \left| f = \frac{\left[k \pm \frac{n(1-s)}{p} \right]}{f_s}; \quad k = 1,3; \quad n = 1,2,...,(2p-1) \right. , \right\} \tag{7}$$

where p is the number of pole pairs, f_s is the fundamental frequency, s is the slip. This type of failure also leads to increase in the temperature of the electrical machine. Torque measurements, shaft, gearbox and electric machine vibrations can also give some information.

Electrical imbalance is another common fault. It introduces shaft vibrations, which are associated with characteristic frequencies (8) that are also present in electrical signals.

$$f_b = \{ f \,|\, f = (1 \pm 2ks) f_S; \quad k = 1,2,3,... \} \tag{8}$$

There are also frequencies associated with rotor mass imbalance that can be identified using the current spectrum. Faults of bearings in electric machines lead to amplitude and phase deviations in the current spectrum.

Power Electronic Converter

Faults of power electronics constitute the majority of faults in WTs. Moreover, the higher the capacity of a wind turbine, the more likely its electronics to malfunction. The distribution of faults for different electronic components are shown in Figure 8. Capacitors, PCBs, and

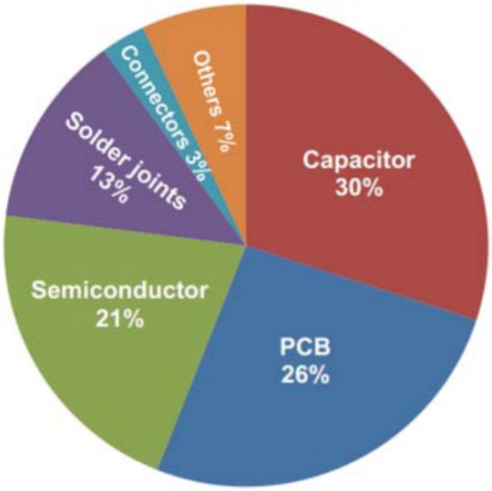

Figure 8: Failure rate distribution of different components in power electronic converters [1]

power semiconductors constitute the majority of failures. Modes of fault for capacitors include leakage, dielectric breakdown, increased dissipation, etc. Modes of fault for PCBs include corrosion, cracks, bad connections, etc. Modes of fault for semiconductor devices are often caused by thermal or mechanical fatigue.

One way to detect failures is to use thermo-sensitive electrical parameters, such as saturation voltage for transistors. However, it is often not efficient or expensive to measure them. Sometimes reference signal is used. It is difficult to locate the faulty component, because different defects can lead to the same change of output signals. Also, difference is often only seen after the damage is already done. Detection of incipient faults is problematic.

Sensors

Failure of some sensors could lead to loss of power generation efficiency, and sometimes even to the halt of a WT. Common way to identify sensor faults is anomaly search in acquired data. New samples can be compared to previous ones, or to the information from adjacent sensors. When a fault of a sensor occurs, it is difficult to identify whether the problem is in the bad sensor or there is an actual deviation in the measured system parameter.

2.4 Feature Extraction and Fault Detection Techniques

A review of WT signals and methods of their processing was done in [2]. A summary of signal processing techniques used for diagnosis of a WT condition is given further.

A variety of methods are used to detect faults in wind turbines, starting with simple FFT analysis and thresholding, and up to complicated artificial intelligence methods based on neural networks or support vector machines. A comparison of signal processing methods, including their capabilities and limitations, is shown in the Table 1. Further, particular WT fault feature extraction techniques and methods of fault detection are discussed.

2.4.1 Feature Extraction Techniques

Synchronous Sampling

Synchronous sampling is a method of handling signals that are non-stationary in frequency such as vibrations and electrical signals of wind turbines. Technically, this method does not provide fault features directly. Instead, it is used in combination with other techniques. Since most of the time characteristic frequencies depend on the rotation speed, it is hard to detect them when frequency changes. This method allows to fix the rotation speed to constant a value, which facilitates further processing by classical methods such as FFT. Other methods, that allow handling non-stationary frequencies, are Short Time Fourier Transform (STFT) and Wavelet Transform, but they do not provide the same resolution in frequency domain and are computationally more expensive.

Synchronous sampling can be performed on hardware level and by signal processing. The fist way to do that is to have a rotation sensor installed. Whenever rotation by a certain angle has occurred, the signal is sampled. This sampling method produces a signal that is non-uniformly sampled in time, but is uniformly sampled in the phase domain.

Another way to do this is to resample the signal obtained through uniform sampling in time. First, sampling points in phase domain are chosen. Then, the current phase of the

Signal processing method	Domain	Function	Resolution		Complexity	Non-stationary signals	Sampling rate
			Time domain	Frequency domain			
Synchronous sampling	Time	Signal conditioning	High	-	Low / Medium	Yes	High / medium
Hilbert transform	Time	Signal conditioning	High	-	Medium	Yes	High / medium
Statistical analysis	Time, frequency	Feature extraction	Rely on input	Rely on input	Low	Possible	Any
Envelope analysis	Time	Feature extraction	High	-	Low / Medium	Possible	High / medium
FFT	Frequency	Feature extraction	-	High	Medium	No	High / medium
Wavelet transform	Time - frequency	Signal conditioning	Medium	Medium	Low / Medium	Yes	High / medium
Model-based methods	Time, frequency	Feature extraction	Rely on input	Rely on input	Medium / high	Possible	Medium / low
ANN	Time, frequency	Feature extraction, diagnosis	Rely on input	Rely on input	Medium / high	Possible	Medium / low
SVM	Time, frequency	Feature extraction, diagnosis	Rely on input	Rely on input	Medium / high	Possible	Medium / low
Expert systems	Time, frequency	Diagnosis	Rely on input	Rely on input	Medium	Possible	Medium / low
Fuzzy logic	Time, frequency	Diagnosis	Rely on input	Rely on input	Medium	Possible	Medium / low
Bayesian methods	Time, frequency	Diagnosis	Rely on input	Rely on input	High	Possible	Medium / low

Table 1: Comparison of signal processing methods, used for WT faults feature extraction, fault detection [2]

signal is estimated. If the current phase value is close to a sampling point, estimation of the new value of the signal is performed. This method is discussed in greater details in the Section 4.3.1.

Hilbert Transform

Hilbert transform is a method that allows to extract the instantaneous amplitude $A(t)$ and phase $\phi(t)$ of a signal. It shifts the signal by 90 degrees, which is equivalent to finding the imaginary part of a real signal. It is defined as

$$\mathrm{H}(x(t)) = \int_{-\infty}^{\infty} \frac{x(\tau)}{\tau - t} d\tau. \tag{9}$$

Then the amplitude and phase are found through

$$A(t) = \sqrt{x(t)^2 + \mathrm{H}(x(t))^2},\ \phi(t) = -\mathrm{atan}\frac{x(t)}{\mathrm{H}(x(t))}. \tag{10}$$

Since many faults modulate vibration or electrical signals, the change in amplitude, found using Hilbert transform, will reveal those faults. It is equivalent to demodulating and is widely used for the fault feature extraction of important mechanical components such as gears and bearings. Hilbert transform is often combined with other signal processing methods such as FFT.

Envelope Analysis

If signal's aptitude modulated, then the envelope is a function, such as $A(t)$ in (10), that specifies the law of amplitude change in time. Most of the time, the fault features contained in the envelope can be extracted directly without further processing. Information about faults of some bearings is present in the vibration of current signal envelope. Envelope extraction techniques are done by amplitude demodulation. For example, in [52] the envelope extraction is done by signal rectification with consecutive low pass filtering. Hilbert transform can also be used as a

tool for envelope analysis.

Statistical Analysis

Statistical analysis involves such parameters as mean, variance, crest factor, kurtosis, and skewness of the signal. The analysis is often done in the time domain. First, the signal of normal condition is monitored, and the reference parameters are estimated. Then, the values of estimated parameters are continuously monitored. If a significant change occurs, and the estimated value exceeds some threshold, the fault occurrence is declared.

Statistical analysis can be applied to different signals, such as the current, or power, or vibrations. The problems with such methods is that they rarely can reveal the additional information on where exactly the fault has occurred.

FFT

Fast Fourier Transform is one of the most commonly used techniques for the extraction of fault signatures. Combined with other methods, it appears to be a powerful tool for the fault analysis. Most of mechanical faults introduce new harmonics in the measured signal that are easily revealed after FFT. Fourier transform is often used in conjunction with the synchronous resampling or envelope analysis. FFT can be applied to acoustic emission signals, vibration signals, and electrical signals.

The main drawback of FFT is that it cannot process signals non-stationary in frequency. This problem is easily alleviated by using synchronous resampling or other order tracking techniques. When used with other techniques, it can give high frequency resolution, which might be required for detection of some faults. For example, the signals of bearing faults are very weak and can only be discovered through high frequency resolution analysis.

STFT

Short Time Fourier Transform is a special time-frequency analysis technique. It allows to monitor the change of the amplitude and frequency in time simultaneously. This method was successfully used to detect generator faults such as short circuits, damage of blades, and rotor imbalance. Despite providing additional information, the frequency and time resolution are complimentary for this method. Therefore, this method is not applicable when high frequency resolution is required.

Wavelet Transform

Wavelet Transform is a tool widely used for WTs diagnosis. It allows to analyze signal both in frequency and in time. This gives an opportunity to analyze non-stationary signals. It was used to reveal different types of faults such as bearing faults, blade faults, gearbox faults, and other mechanical defects. Two types of wavelet transforms were used for the WT condition analysis: continuous wavelet transform and discrete wavelet transform. This method was also used for noise reduction and low power fault signals detection. The main shortcoming of this technique is the low resolution.

2.4.2 Fault Detection Techniques

Feature extraction methods mainly serve as tools to acquire information needed for fault detection. Several different approaches of fault detection techniques can be distinguished.

Threshold Based Methods

These methods are based on estimating some kind of parameter and comparing it to the predefined threshold. Most of statistical methods are done by comparing the mean, variance, skewness, etc. with some base level. Threshold based methods can also be formulated for other types of extracted features. Many faults introduce new harmonics into frequency spectrum.
A fault in this case can be detected by comparing the power of a certain frequency component,

obtained through FFT, with some threshold. A threshold based method for bearing fault detection was introduced in [17].

Model Based Methods

Model based methods try to describe the dynamic behavior of a WT using physical principles, trends of signals, and data-driven approaches. Once a model accurately describing the WT normal behavior for different inputs is built, it can be used to perform the diagnosis by analyzing the difference between the predicted and actual behavior. This kind of method was used to detect structural changes of the wind blades. Models, that describe behavior of WT's subsystems, are sometimes relatively easy to construct. However, it is very challenging to build a good descriptive model for a complex system, such as a wind turbine.

In [53] SCADA (supervisory control and data acquisition) signal were used design 45 normal behavior models that allow to detect abnormal behavior. Model based methods usually do not require the high sampling rate. Moreover, if model is built well, it can be used to detect incipient faults at their early stages. Despite the fact that this type of techniques can be very efficient, it has limited capability in identifying the location of the fault.

AI Methods

The latest trends in data driven analysis found their way into diagnosis of WTs. Different artificial intelligence (AI) methods, such as artificial neural networks (ANN), support vector machines (SVM), and fuzzy logic approaches, were used to detect faults in WTs. ANN were used to detect faults of bearings, gearbox, generators, power electronic components. ANN can be used for two purposes: classification and prediction. Classifications are done by training ANN on a labeled set of signals. Predictions are done by training the relationship between input and output signals of a WT, and comparing the predictions with observed values. In this sense ANN prediction is similar to model based approach, but the model itself is hidden in ANN structure. Any signals can be used as an input for ANN: raw signals, conditioned

signals, extracted features.

ANN methods are completely data driven and require no knowledge of WT dynamics. While it can be treated as a benefit of this approach, it can also be its downside, since ANN can capture the behavior only of what they have seen before. The proper application of this method definitely requires some skill. Moreover, the training phase for ANN can take an extensively long amount of time.

SVM were also used for rotor, gearbox, generator, bearings and sensors fault analysis. While SVM have better generalization capabilities than ANN, they have all the same disadvantages.

Two other fault detection approaches can be associated with AI fault detection methods: expert systems and fuzzy logic. Expert systems try to map the input signal to some modes of WT condition based on the previously observed signals and their correlation to WT health condition. Such methods were used for gearbox faults detection.

Fuzzy logic methods requires the knowledge of the failure mechanisms, and therefore are difficult to implement in practice. Nevertheless, some attempts to use them were done for the generator and pitch system diagnosis.

Bayesian Methods

The idea of Bayesian methods is to analyze WT signals probabilistically. This requires some training stage, and the accuracy strongly depends on the amount of training data available. Plus, modeling signals this way requires the storage of the previous samples. In general, the more samples are used, the higher the accuracy of the fault detection algorithm. In [29] a bearing fault detection algorithm, based on hidden Markov models (HMM) was proposed. The feature extraction consisted of the envelope analysis, after which HMM was trained to detect bearing faults.

In [33] a particle filtering algorithm was used to detect faults of a bearing. The idea of the method is to define a model of normal behavior. The extracted features undergo various

transformations before they can be used for particle filtering method.

2.5 Conclusions

In this section the common faults of WTs were described. It was revealed that vibration and electrical signals are the most common for fault detection, and those are the primary signals for detecting bearing faults. Popular signal conditioning techniques and fault feature extraction methods were reported. Due to the nature of the defect, bearing faults introduce new vibrations into observed vibration or electrical signals. The common way to extract those vibrations is by envelope analysis and, further, Fourier transform. If signal is non-stationary, the synchronous resampling can be applied. Methods such as wavelet transform and STFT are not very helpful for bearing fault diagnosis due to their low resolution. Additionally, statistical analysis of envelope can be used to extract fault signatures of a bearing [54].

The bearing fault detection techniques most commonly perform some testing, where the result of the test is compared with threshold. It can be a test of fault harmonics magnitude, or test of some statistical parameter, or even test of the probability that fault has occurred. AI methods and Bayesian methods were used to design bearing fault detection procedures. Constructing such methods require a lot of training data that describe all possible bearing conditions, which in reality is hard to obtain.

We propose the use of change point detection techniques for identifying bearing faults. While fault signal is tested against some baseline distribution, this method does not require any training stage. In the next section the theory of change point detection procedures is given.

3 Change Detection Theory

3.1 Introduction

A certain class of problems, called change point problems, considers a change of a process state at an unknown moment of time, usually form one state to another. Information in the form of process observations and characteristics of states, that needs to be distinguished, is available. The goal, generally, is to detect an alteration of the process state as soon as possible. The sequential approach is usually used, where observations are obtained one at a time. In sequential testing problems data arrive gradually as testing proceeds. As new data arrive, one of two decisions can be made in a sequential change detection procedure: "the change has occurred" or "continue the observation". As long as the process is in the original state, observations continue. As soon as shift between states is surmised, a corrective action is expected. In order to reduce the number of all possible change detection procedures, the upper bound on false alarm is considered. This makes the procedure to collect more information, until some level of certainty will be reached. A review of the history of change detection procedures was done in [44].

The description of a change detection problem provided above suggests that the problem of bearing fault detection can be treated as a change detection problem, where the observed random process, i.e. current or vibration signal, can be in two different states: signal is normal, and fault signature is present in the signal.

In order to use a change detection procedure for bearing fault detection, we review the theory of change detection first. In this section the definition of change detection problem is given. Then, several change detection techniques are described, and some of their characteristics are given.

3.2 Review of Change Detection Procedures

There are several change-point detection procedures as of today. The most common are Shiryaev, Shiryaev-Roberts, and CUSUM procedures. Shiryaev procedure was introduced in [37], proven to be asymptotically optimal in the i.i.d. case for geometric prior distribution of the moment of change, and shown to be asymptotically optimal in the general case by Tartakovsky [40] in the Bayesian context of minimizing the average detection delay. Although the performance of this procedure is good, it requires the knowledge of the distribution of the moment of change, which is rarely available in reality. The other two change detection procedures are optimal in minimax context (minimizing worst case detection delay), and were shown to be optimal for some cases in the Bayesian context [40]. The main advantages of these two other procedures is that they do not require the knowledge of the prior moment of change distribution. In [43] it was proven that Shiryaev-Roberts procedure is the best in terms of minimizing integral average detection delay in the case, where the change occurs in distant future. Numerical comparison of Shiryaev-Roberts and CUSUM procedures was done in [42]. The fact, that Shiryaev-Roberts procedure is not optimal in general, was proven in [45] by a counterexample.

The review of change point detection history was done in [44]. The first attempt to solve change detection problem was done by Shewart in 1931, who only investigated the shift in mean for a normal distribution. Consecutively, several techniques were introduced, but none of them could actually claim to be optimal. Girschik and Rubin (1952) and Shiryaev (1963) considered a problem in the Bayesian context and tried to minimize the average detection delay. In 1954 Page introduced his cumulative sum (CUSUM) procedure, inspired by Wald's sequential probability ratio test (SPRT). In 1966 Roberts introduced a procedure of his own that later got known as Shiryaev-Roberts procedure. In the following part a general form of a quickest change detection procedure is formulated and the three procedures introduced above are reviewed.

3.3 Problem Formulation

The quickest change detection problems can be formulated as follows. A random process $X(t)$, $t = 1,2,...$ is observed sequentially, one sample at a time. Define $\mathbf{x}^{k:n} = [x_k,...,x_n]$, where $x_i = X(i)$. Let $\mathscr{F}_x^n = \sigma(\mathbf{x}^{1:n})$ be the sigma algebra generated by $\mathbf{x}^{1:n}$. Here n denotes the moment of the latest observation. The process density f_0 changes to f_1 at an unknown moment of time $1 \leq \lambda \leq \infty$, where $\lambda = \infty$ refers to the case, where the change never occurs. The change moment λ is a random variable with prior distribution $\pi_k = \mathbf{P}\{\lambda = k\}$, $k = 0,\ 1,\ 2,...$. The change of the density is declared to be detected the first time the detection procedure raises an alarm. The goal is to minimize the average detection delay (ADD) $\nu - \lambda$ subject to the average run length (ARL) to false alarm or probability of false alarm (PFA), where ν is the moment, when the first alarm is raised. The functions $\mathrm{ADD}(\delta)$ and $\mathrm{ARL}(\delta)$ are two operation characteristics of a sequential change detection procedure. Let $\mathbf{P}_k$ and $\mathbb{E}_k$ correspond to the probability measure and corresponding expectation, when the change occurs at the moment of time $\lambda = k$. $\mathbf{P}^\pi$ is the average probability measure, defined as $\mathbf{P}^\pi(\Omega) = \sum_{k=0}^{\infty} \pi_k \mathbf{P}^k(\Omega)$, and $\mathbb{E}^\pi$ denotes the expectation with respect to $\mathbf{P}^\pi$ [40].

3.4 Detection Procedure Definition

Change point testing is a sequential procedure, that can be defined as a mapping from observed sequence $\mathscr{F}_x^n$ to a positive integer $\nu \leq n$

$$\delta : \mathscr{F}_x^n \to \{\nu : \nu \leq n\}, n = 1,2,... \tag{11}$$

The procedures can be characterized by following functions. The first one is average detection delay

$$\mathrm{ADD}(\delta) = \mathbb{E}^\pi(\nu - \lambda | \nu \geq \lambda). \tag{12}$$

where ν is detected moment of change [38]. The second one is average run length to false

alarm (ARL)

$$\mathrm{ARL}(\delta) = \mathbb{E}_{\infty}(\nu), \tag{13}$$

where $\mathbb{E}_{\infty}$ is the expectation in the case, when the change never occurs. In contract with ARL, PFA requires the knowledge of the prior change distribution [39]

$$\mathrm{PFA}(\delta) = \mathbf{P}_{\infty}(\nu < \lambda). \tag{14}$$

Further we will consider three change detection procedures: Shiryaev, CUSUM, and Shiryaev-Roberts. Despite the fact, that each of them can be considered optimal in different contexts, their performance will be shown with respect to the following metrics. Define the class of change detection procedures, for which PFA is bounded by α

$$\Delta(\alpha) = \{\delta : \mathrm{PFA}(\delta) \leq \alpha\}. \tag{15}$$

From this class, consider those, that minimize ADD

$$\rho \quad = \quad \underset{\delta \in \Delta(\alpha)}{\arg\inf} \mathrm{ADD}(\delta). \tag{16}$$

For the sake of further use, define likelihood ratio

$$\Lambda_{k:n} = \prod_{l=k}^{n} \frac{f_1(x_l|\mathbf{x}^{1:l-1})}{f_0(x_l|\mathbf{x}^{1:l-1})} = \prod_{l=k}^{n} \lambda_l. \tag{17}$$

3.5 Change Detection Procedures

3.5.1 Shiryaev procedure

Shiryaev [37] was one of the first to consider the change detection problem in Bayesian context (16), with penalty expressed in terms of ADD. The performance of the general case was

analyzed in [40]. Consider two hypotheses:

$$H_0 : n < \lambda,\ H_1 : n \geq \lambda. \tag{18}$$

Likelihood ratio of two hypotheses H_0 and H_1 given the observation vector $\mathbf{x}^n$ is evaluated.

$$\begin{aligned} S_n &= \frac{p\left(H_1|\mathbf{x}^n\right)}{p\left(H_0|\mathbf{x}^n\right)} \\ &= \frac{p(\lambda \leq n|\mathbf{x}^n)}{p(\lambda > n|\mathbf{x}^n)} \end{aligned}$$

Adaptation of Shiryaev procedure for the problem, specified above, will result in the following stopping rule:

$$\delta_s = \inf\{n \geq 1 : S_n \geq A\} \tag{19}$$

where A is some threshold. This stopping rule returns the value of n, whenever the probability in the numerator for a given sample $\mathbf{x}^n$ is A times larger than the probability in the denominator

$$S_n = \frac{\mathrm{p}(\lambda \leq n|\mathbf{x}^n)}{\mathrm{p}(\lambda > n|\mathbf{x}^n)} \geq A. \tag{20}$$

Thus, when the ratio S_n is larger than the threshold A, probability $p(\lambda > n|\mathbf{x}^n)$ will serve as a posterior probability of false alarm, since it gives us the probability of H_0 being true for a sample $\mathbf{x}^n$, while hypothesis H_1 is accepted as positive. If $\mathrm{p}(\lambda > n|\mathbf{x}^n) \leq \alpha$, then

$$\begin{aligned} \mathrm{PFA}(n) &= \mathbb{E}_k[p(\lambda > n|\mathbf{x}^n)] \\ &\leq \mathbb{E}_k[\alpha] \\ &= \alpha. \end{aligned}$$

That gives the value of the threshold

$$A = \frac{1-\alpha}{\alpha}, \tag{21}$$

that allows to detect the the change quickly while bounding PFA by α.

The test statistic for change detection can be transformed to the following form

$$\begin{aligned} S_n &= \frac{p(\lambda \le n, \mathbf{x}^n)}{p(\lambda > n, \mathbf{x}^n)} \\ &= \frac{p(\mathbf{x}^n|\lambda \le n)p(\lambda \le n)}{p(\mathbf{x}^n|\lambda > n)p(\lambda > n)} \\ &= \frac{p(\mathbf{x}^n|\lambda \le n)p(\lambda \le n)}{p(\mathbf{x}^n|\lambda > n)p(\lambda > n)} \\ &= \frac{\sum_{k=1}^{n} f_0(\mathbf{x}^{1:k-1}) f_i(\mathbf{x}^{k:n}) p(\lambda = k)}{f_0(\mathbf{x}^n) \sum_{k=n+1}^{\infty} p(k = \lambda)} \end{aligned} \tag{22}$$

The measure of the detection lag is the average detection delay (ADD)

$$\begin{aligned} \mathrm{ADD}(\nu) &= \mathbb{E}^{\pi}(\nu - \lambda | \nu \ge \lambda) \\ &= \frac{\mathbb{E}^{\pi}(\nu - \lambda)^{+}}{\mathbf{P}^{\pi}\{\nu \ge \lambda\}} \\ &= \frac{1}{\mathbf{P}^{\pi}\{\nu - \lambda\}} \sum_{k=0}^{\infty} \pi_k \mathbf{P}^{\pi}\{\nu \ge k\} \mathbb{E}_k\{\nu - k | \nu \ge k\} \end{aligned} \tag{23}$$

Here $a^{+} = \max(0, a)$. Intuitively, the smaller the upper bound on PFA is, the longer time it takes to detect the change, e.g. if $\mathrm{PFA}(\delta_s) = \alpha$, then the $\mathrm{ADD}(\delta_s)$ is minimized. In the asymptotic setting, where $\alpha \to 0$, Shiryaev procedure was proven to be optimal for i.i.d. case with geometric prior distribution in [37] and for the general case with respect to the average detection delay, and uniformly asymptotically optimal in the sense of minimizing the conditional expected delay $\mathbb{E}_k(\nu - \lambda | \nu \ge \lambda)$ for every change moment $\lambda = k,\ k = 1, 2, \ldots$ [40].

3.5.2 CUSUM

CUSUM procedure was introduced by Page in 1954 [55]. Essentially it appears to be a repeated sequence of sequential probability ratio tests f_0 versus f_1, which raises the alarm the first time SPRT exist on the side of f_1. Given previous definitions, consider the conditional

density $f_0(X_n|X_1,...,X_{n-1})$ for $n < \lambda$ and $f_1(X_n|X_1,...,X_{n-1})$ for $n \geq \lambda$, where f_0 denotes the distribution before the change, and f_1 denotes the distribution after the change. Assume that change has occurred at the moment of time k. SPRT can be written as

$$\begin{aligned} B \leq \frac{f_0(X_1,...,X_{k-1}) f_1(X_k,...,X_n|X_1,...,X_{k-1})}{f_0(X_1,...,X_n)} \leq A \\ B \leq \frac{f_1(X_k,...,X_n|X_1,...,X_{k-1})}{f_0(X_k,...,X_n|X_1,...,X_{k-1})} \leq A \\ B \leq \prod_{i=k}^{n} \frac{f_1(X_i|X_1,...,X_{i-1})}{f_0(X_i|X_1,...,X_{i-1})} \leq A \end{aligned} \tag{24}$$

Here for hypothesis H_0 :

$$X_{1:n} \quad \sim \quad f_0$$

and for the hypothesis H_1 :

$$X_{1:k-1} \sim f_0,\ X_{k:n} \sim f_1 \tag{25}$$

It the case of change detection, we are not interested in the left side of the test (24). Also, there is n possibilities, where the change λ could have occurred within $\mathbf{x}^n$. By looking for the maximum value of the likelihood ratio for different $\lambda = k,\ k = 1,2,...n$, we can get the most powerful test for the given value of n.

The stopping rule of CUSUM procedure is given by

$$\delta_c = \inf\{n \geq 1 : C_n \geq A\}, \tag{26}$$

where

$$C_n = \max_{1 \leq k \leq n} \Lambda_{k:n} \tag{27}$$

This procedure was proven to be optimal in the general case [39] for non-i.i.d. random variables in the context of minimizing the worst case detection delay $\sup_k \mathbb{E}_k(\nu - k|\nu \geq k)$. Tartakovsky considered asymptotic properties of this change detection procedure in class $\Delta(\alpha)$

defined by (15) [40]. It was shown, that the probability of false alarm is bounded by

$$\mathrm{PFA}(\delta_c(A)) \leq \bar{\lambda} A^{-1} \tag{28}$$

where $\bar{\lambda} = \sum_{k=1}^{\infty} k \cdot \pi_k$ is the mean of the moment of change. Choosing $A = \bar{\lambda}/\alpha$ guarantees (28). Moreover, setting $A = 1/\alpha$ gives a valid threshold, since

$$\mathrm{PFA}(\delta_c(A)) \leq \alpha \leq \bar{\lambda}\alpha \tag{29}$$

It was also found, that CUSUM procedure is asymptotically optimal in minimax and Bayesian context for some distributions [40]. It was shown, that if $A = 1/\alpha$, and some additional conditions are satisfied, then

$$\mathrm{PFA}(\delta_c(A)) \leq \alpha,\ \mathrm{ADD}(\delta_c) \underset{\alpha\to 0}{\asymp} \frac{|\log\alpha|}{D_{10}}, \tag{30}$$

where D_{10} is the Kullback-Leibler divergence between distributions f_1 and f_0, when samples are independent.

3.5.3 Shiryaev-Roberts procedure

The procedure, that later became known as Shiryaev-Roberts procedure (SRP), was first introduced in [56]. The stopping rule is

$$\delta_{sr} = \inf\{n \geq 1 : R_n \geq A\} \tag{31}$$

where

$$R_n = \sum_{k=1}^{n} \frac{f_{\lambda=k}(X_1,...,X_n)}{f_{\lambda=\infty}(X_1,...,X_n)} \tag{32}$$

Here k is the assumed moment of change, n is total number of samples acquired at this moment. For i.i.d. case the expression above can be simplified to

$$\begin{aligned} R_n &= \sum_{k=1}^{n} \prod_{i=1}^{k-1} \frac{f_0(X_i)}{f_0(X_i)} \prod_{i=k}^{n} \frac{f_1(X_i)}{f_0(X_i)} \\ &= \sum_{k=1}^{n} \prod_{i=k}^{n} \frac{f_1(X_i)}{f_0(X_i)} \end{aligned}$$

This procedure is not strictly optimal and was proven to be only asymptotically optimal in the sense of minimizing worst case delay $\sup_k \mathbb{E}_k(\nu - k | \nu \geq k)$. Just as the CUSUM procedure, properties of SRP in the Bayesian context were also studied in [40]. The results were similar to those for the CUSUM procedure. The value of the threshold A is based on the constraint on ARL to false alarm $\mathbb{E}_\infty[\nu] = A$. The exact value for A is hard to obtain. SRP was proven to be optimal asymptotically for minimizing stationary average detection delay[42]. It was shown, that if $A = 1/\alpha$ and some additional conditions satisfied, then

$$\mathrm{PFA}(\delta_{sr}(A)) \leq \alpha, \ \mathrm{ADD}(\delta_{sr}) \underset{\alpha \to 0}{\asymp} \frac{|\log \alpha|}{D_{10}}, \tag{33}$$

where D_{10} is the Kullback-Leibler divergence between distributions f_1 and f_0, when samples are independent [40].

The comparison of the average detection delay for Shiryaev-Roberts and CUSUM procedures is shown in Figure 9.

3.6 Multiple Hypothesis Testing

Consider the following problem. For some a random process $x_1, x_2, \ldots$ at an unknown moment of time λ the change of the distribution occurs to one of M post-change distributions with known parameters, which are further referred to as alternative hypotheses. The probability p_i that the change occurs to a particular distribution $i = 1, 2, \ldots \mathrm{M}$ is also given. The problem is to detect the change of the distribution as fast as possible and decide, which of the post-change

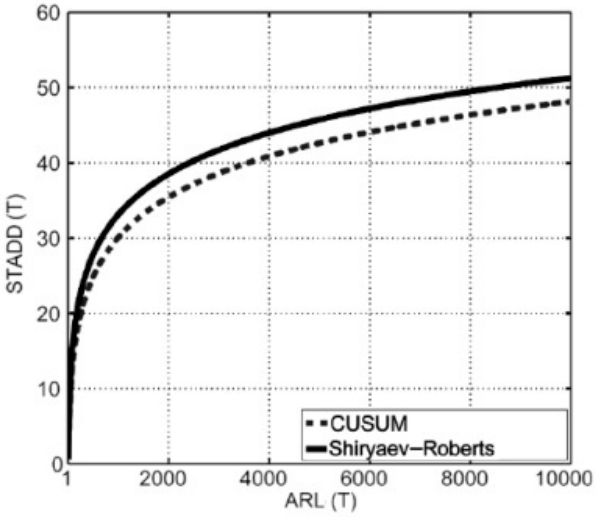

Figure 9: ADD (STADD on the plot) vs. ARL. T is the value of the threshold. [42]

distributions is present, while bounding the PFA by α and probability of misdetection (PMD) by γ.

Originally, multiple hypothesis testing was introduced as an extension of SPRT for the case, where no change is considered [57]. The extension for the change testing procedure can be done through the use of Shiryaev procedure. The first step is to detect the change of the distribution, bounding the PFA by α. The next step after detecting the change is to decide, which of the alternative distributions is present, while keeping probability of misdetection bounded by γ. Since the decision is made by using a sequential testing procedure, some stopping rule is required. One of the ways is to evaluate the probability $\mathrm{p}(\beta = i|\mathbf{x}^n)$ for $i = 1,...,\mathrm{M}$, and compare it to some threshold. The decision is made when at least one of probabilities $\mathrm{p}(\beta = i|\mathbf{x}^n)$ triggers the stopping rule. The one that gives the minimum probability of error is chosen as the true hypothesis.

$$\hat{\beta} = \underset{i}{\operatorname{argmax}} \{\mathrm{p}(\beta = i|\mathbf{x}^n) \geq B\} \tag{34}$$

When the probability $\mathrm{p}(\beta = i|\mathbf{x}^n)$ is evaluated, we can divide the sample space $\mathbf{x}^n$ into two events: $\beta = i$ and $\beta \neq i$. Since crossing the threshold B results in accepting the corre-

sponding hypothesis, the event $\beta \neq i$, while $\hat{\beta} = i$ is accepted as true, serves as an error event. Hence $\mathrm{p}(\beta \neq i|\mathbf{x}^n)$ is the posterior probability of detecting hypothesis i incorrectly. Since we desire to have the probability of misdetection bounded by γ, setting the threshold in such way that

$$\mathrm{p}(\beta \neq i|\mathbf{x}^n) \leq \gamma \tag{35}$$

ensures the upper bound

$$\mathrm{p}(\beta \neq i) \leq \gamma. \tag{36}$$

The corresponding value of B is

$$B = 1 - \gamma. \tag{37}$$

Test statistic for multiple hypothesis change detection procedure can be transformed to the following form

$$\begin{aligned} p(\beta = i|x^n) &= \frac{p(\mathbf{x}^n|\beta = i)p(\beta = i)}{p(\mathbf{x}^n)} \\ &= \frac{p(\mathbf{x}^n|\lambda \leq n, \beta = i)p(\beta = i|\lambda \leq n)p(\lambda \leq n)}{p(\mathbf{x}^n|\lambda \leq n)p(\lambda \leq n) + p(\mathbf{x}^n|\lambda > n)p(\lambda > n)} + \\ &\quad \frac{p(\mathbf{x}^n|\lambda > n, \beta = i)p(\beta = i|\lambda > n)p(\lambda > n)}{p(\mathbf{x}^n|\lambda \leq n)p(\lambda \leq n) + p(\mathbf{x}^n|\lambda > n)p(\lambda > n)} \\ &= \frac{p(\mathbf{x}^n|\lambda \leq n, \beta = i)p(\beta = i)p(\lambda \leq n)}{p(\mathbf{x}^n|\lambda \leq n)p(\lambda \leq n) + p(\mathbf{x}^n|\lambda > n)p(\lambda > n)} \end{aligned}$$

From the notion of the stopping rule it follows that

$$\begin{aligned} \text{PMD}(n) &= \mathbb{E}[p(\hat{\beta} \neq i|\mathbf{x}^n)] \\ &= \mathbb{E}[1 - max\{p(\beta = i|\mathbf{x}^n)\}] \\ &\leq \mathbb{E}[1 - (1-\gamma)] \\ &= \mathbb{E}[\gamma] \\ &= \gamma \end{aligned}$$

When the stopping rule (19) is used to detect the change, it will ensure PFA $\leq \alpha$. Hence, incorporating stopping rules (19) and (34) together will give the desired result.

3.7 Model Mismatch

Very often it might happen that the post change distribution is not perfectly known. In that case, the change detection performance will be affected. In [58] the study of the effect of mismatched post-change model was done. It was assumed that the original distribution f_0 is perfectly known, while the true distribution after the change f_1 in the test is replaced with $\tilde{f}_1$. The impact of such swap on the ARL, PFA and ADD for CUSUM and Shiryaev-Roberts procedures was investigated. Further, some results of that research are shown.

Consider the mismatched density $\tilde{f}_1$. The likelihood ratio transforms into the form

$$\tilde{\lambda}_i = \frac{\tilde{f}_1(x_i|\mathbf{x}^{1:i})}{f_0(x_i|\mathbf{x}^{1:i})}, \tag{38}$$

and

$$\tilde{\Lambda}_{k:n} = \prod_{i=k}^{n} \tilde{\lambda}_i. \tag{39}$$

First, the results for SRP are shown. In case of the mismatched test, the stopping rule is defined as

$$\tilde{\delta}_{sr} = \{n \geq 1 : \tilde{R}_n \geq A\} \tag{40}$$

where $\tilde{R}_n = \sum_{i=1}^{n} \tilde{\Lambda}_{k:n}$. ARL of SRP with mismatched post-change distribution satisfies

$$\mathrm{ARL}(\tilde{\delta}_{sr}) \geq A \tag{41}$$

The PFA of the SRP is upper bounded by

$$\mathrm{PFA}(\nu) \leq \min\left\{\frac{\bar{\lambda}}{A}, 1\right\}, \tag{42}$$

where $\bar{\lambda} = \sum_{k=1}^{\infty} k \cdot \pi_k$ is the average moment of change. The new ADD is expressed through

$$\mathbb{E}(\tilde{\nu}_{sr} - \lambda \,|\, \tilde{\nu}_{sr} \geq \lambda) \underset{\alpha \to 0}{\preceq} \frac{\frac{|log\ \alpha|}{1-\alpha}}{D_{10} - \tilde{D}_{11}}. \tag{43}$$

Essentially the same results are obtained for CUSUM procedure, with the only difference that PFA is strictly bounded by $\frac{\bar{\lambda}}{A}$.

3.8 Conclusions

In this section the theory of change detection procedures was given. The bounds on the PFA and ADD were defined. It was shown that the CUSUM and SRP are the most desired for practical application, since they do not require the knowledge of prior change point distribution. It was shown, that for the fixed value of ARL, SRP has a slightly longer detection delay, than CUSUM procedure. The PFA for SRP procedure does not have a strict upper bound in the case of post change distribution mismatch. This suggests that its performance for the mismatched case is worse than one of CUSUM. In the next section we are going to introduce the new bearing fault detection algorithm. In the analysis of the results it will be shown that, indeed, in the case of the absence of true post distribution knowledge, SRP is inferior to CUSUM in terms of detection delay (DD) for a given PFA.

4 Quickest Bearing Fault Detection Algorithm for Direct Drive Wind Turbine

4.1 Motivation

A review of existing methods of fault detection was done in the Section 2.4, where some bearing faults detection methods were described. It is known that faults of bearings, that support the shaft of the generator in a direct-drive WT, introduce excitations in stator current spectrum. The current signal is non-stationary in frequency, but this problem can be solved by applying synchronous resampling. After that, FFT can be used to extract bearing fault signatures. When a fault signal appears in the current spectrum, the distribution of the magnitude of particular harmonics will change. If the form of distribution is known, and parameters of the distribution are available, the quickest change detection methods can be applied. These types of methods are designed to detect the change in the distribution with the minimal delay while keeping the PFA bounded by some value α. Further, we describe the experimental setup and the feature extraction procedure. Then, the distribution of the features is studied, and the fault detection procedure is designed.

4.2 Experimental Setup

Now, we are going to describe the new bearing fault detection algorithm. Stator current signal from a 160-W Southwest Wind Power Air Breeze direct-drive PMSG WT was used in this thesis to test the proposed fault detection algorithm. The stator current signal was obtained in [17], where another bearing fault detection algorithm was described. The wind turbine, that was used to record the data, has six pole pairs ($p = 6$), it was operating in variable wind speed condition, where the wind was generated in a wind tunnel with varying speed from 0 to 10 m/s (Figure 10). The shaft of the stator was rotating in the speed range of $6 - 13$ Hz. The stator current signal was recorded using National Instruments acquisition hardware (Figure 11) with sampling frequency $f_{\text{so}} = 10$ kHz. The signal was recorded in 100-second-long files every 20 minutes. The total operation time of the WT constituted about 25 hours. The test bearing

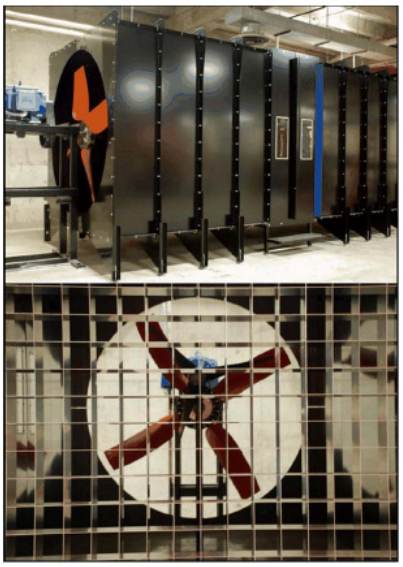

Figure 10: Wind tunnel with the test wind turbine. [17]

(7C55MP4017) was pretreated by removing the lubricant, in order to accelerate degradation. One of two bearings, that were supporting the shaft, was removed to simulate eccentricity. The test wind turbine stopped rotating at the end of the experiment due to the broken bearing cage. The influence of the broken cage on the stator current signal is shown in Figure 14. Excitations appeared on the characteristic frequencies.

4.3 Feature Extraction

The feature extraction procedure is inspired by [17]. In that paper, synchronous resampling and FFT of the stator current signal were used.

4.3.1 Synchronous Resampling

The key for fault detection is to extract fault signatures, which are located on particular frequencies relative to fundamental frequency f_1. Fundamental frequency is defined as the main harmonic generated by the WT. The spectrogram of unprocessed stator current signal is shown in Figure 15. Sampling frequency is $f_{so} = 10$ kHz, FFT size is 10^5 samples, which results in the frequency resolution of 0.1 Hz. The frequency of the main harmonic is subject to high

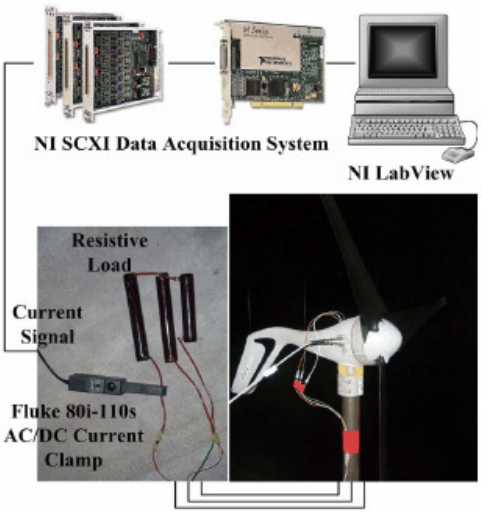

Figure 11: Sensing and data acquisition system for the test wind turbine. [17]

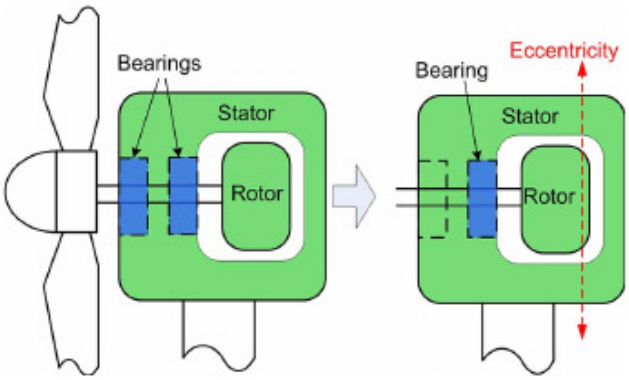

Figure 12: Emulation of a rotor eccentricity in the test wind turbine [17]

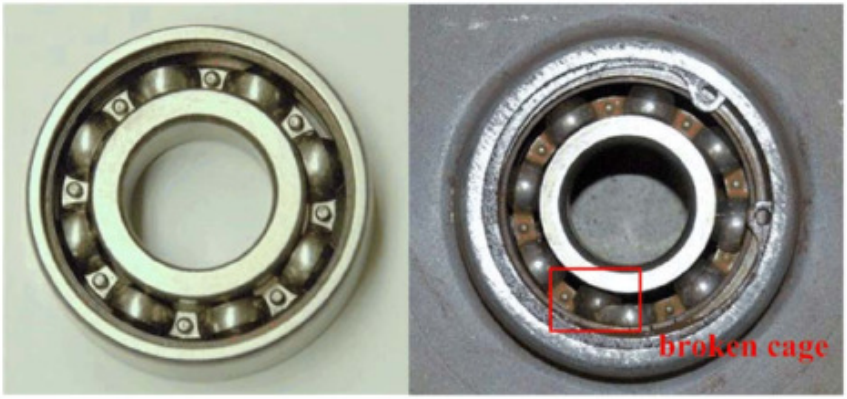

Figure 13: The test bearing before and after the experiment [17]

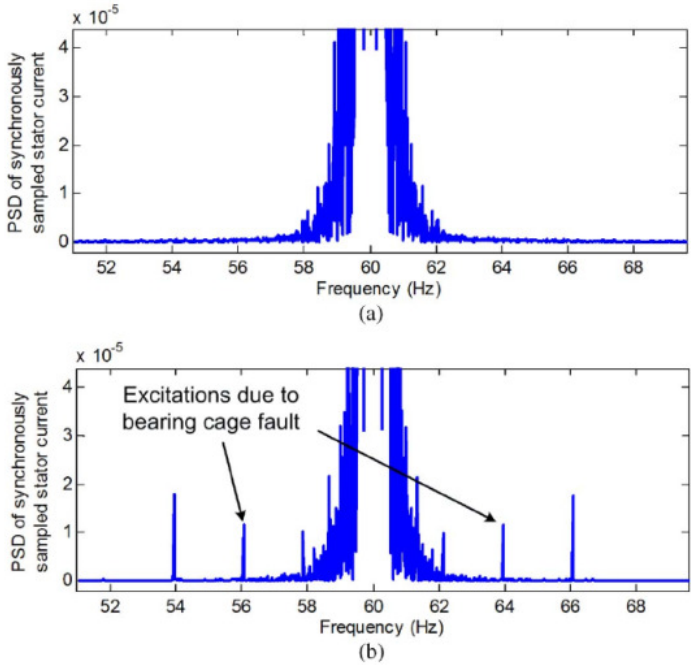

Figure 14: Comparison of the PSDs of the synchronously resampled stator current signal for (a) healthy bearing and (b) faulty bearing case [17]

variability.

The extraction of fault signatures (55) from a signal with non-stationary frequency is rather difficult, especially since this is done after performing FFT. The signal to noise ratios of excitations on fault frequencies for different bearing faults are small. In order to reduce the impact of the noise, high frequency resolution is required. Higher resolution virtually means more samples are needed to perform FFT. The frequency resolution of Figure 15 is 0.1 Hz, which results in the sample size of $f_S \cdot 10$ required for performing such transformation. Signal frequency may change by the time enough samples are acquired, and that can result in still poor frequency resolution.

In order to resolve the problem of fundamental frequency change the method of synchronous resampling can be applied. For non-uniform sampling techniques current signal can be sampled only when rotor position changes by a certain angular value $\Delta\phi$, $\frac{360}{\Delta\phi} = N$, to provide N samples per revolution (Figure 16). Synchronous resampling is a technique, designed to bring alternating signal frequency to a desired value, e.g. to change non-stationary fundamental

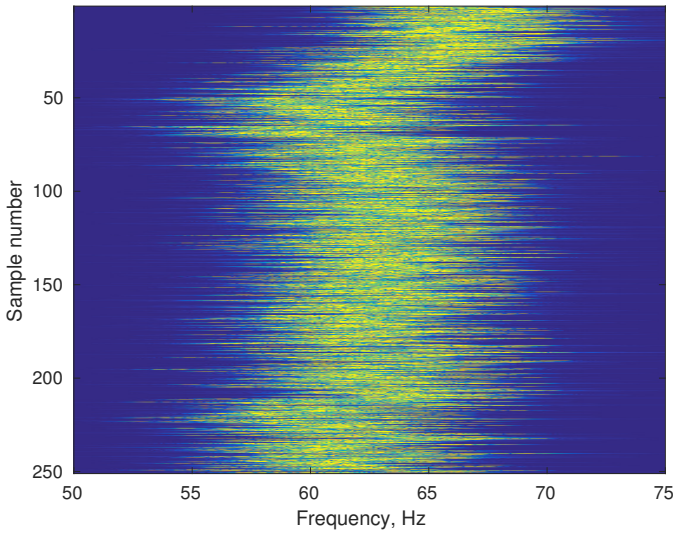

Figure 15: Spectrogram of the unprocessed stator current signal.

frequency to a constant value [18].

The method of synchronous resampling is based on estimating the instantaneous phase of the signal. Assume the WT generates a signal with a single frequency component. This single frequency is modulated by the wind and defects of mechanical components. In that case, the signal model has the following form

$$C(t) = (\mathrm{c} + A(t)) \cdot \cos(2 \cdot \pi \cdot f(t) \cdot t + \omega_i). \tag{44}$$

Here $C(t)$ is the stator current signal, c is some constant, $A(t)$ is the envelope of the signal, $f(t)$ is the non-stationary fundamental frequency, and ω_i is the initial phase.

In order to extract the phase of the signal, its quadrature component needs to be found. This can be achieved by using the Hilbert transform, which lags the signal by 90 degrees. After performing this step, we will have

$$C_h(t) = (\mathrm{c} + A(t)) \cdot \sin(2 \cdot \pi \cdot f(t) \cdot t + \omega_i). \tag{45}$$

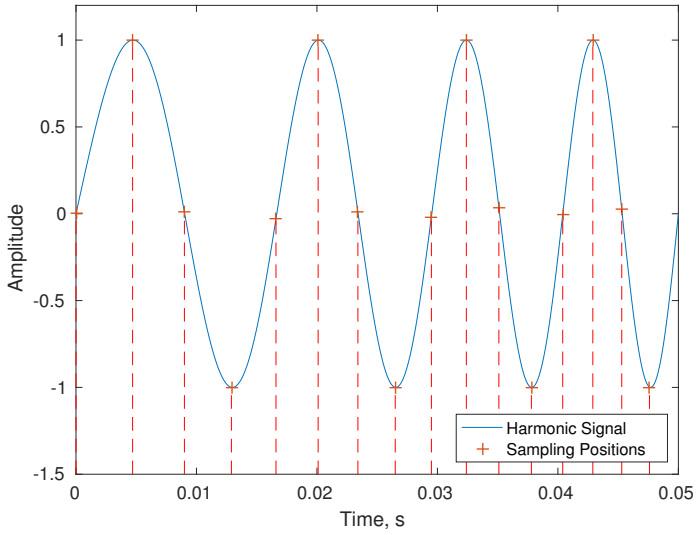

Figure 16: Non uniform sampling.

The instantaneous phase can be found by applying the arctangent function

$$\begin{aligned} \omega &= \operatorname{atan}\left(\frac{C_h(t)}{C(t)}\right) \\ &= \operatorname{atan}\left(\frac{\sin(2\cdot\pi\cdot f(t)\cdot t+\omega_i)}{\cos(2\cdot\pi\cdot f(t)\cdot t+\omega_i)}\right) \\ &= 2\cdot\pi\cdot f(t)\cdot t+\omega_i. \end{aligned} \tag{46}$$

Hilbert transform method performs well only when the harmonic of interest has the largest power in the signal. To reduce the energy of the larger order harmonics, with frequencies $k\cdot f(t), k=2,3,...$, signal needs to be filtered first. A low-pass filter will suffice in this case. The whole procedure of angle extraction is shown in Figure 17.

Further, the synchronous resampling algorithm is demonstrated in Figure 18. The instantaneous phase of the signal, obtained from (46), is analyzed. Define $C_j,\ j=1,2,...$ as a sample of stator current signal, N as the desired number of samples per signal period after resampling procedure, f_s as the desired sampling frequency after resampling, and a set of phases $\underline{\omega}=\{\omega_1,\omega_2,...,\omega_N\}$, where $\omega_{i+1}-\omega_i=\frac{2\pi}{N}$. Let $\omega_j=\frac{C_j^{\mathrm{H}}}{C_j}$ be the instantaneous phase of the jth sample of the original stator current signal C. C_j^{H} is the value of the jth sample of

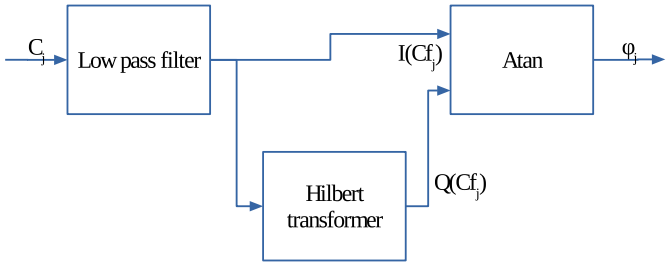

Figure 17: Signal phase extraction.

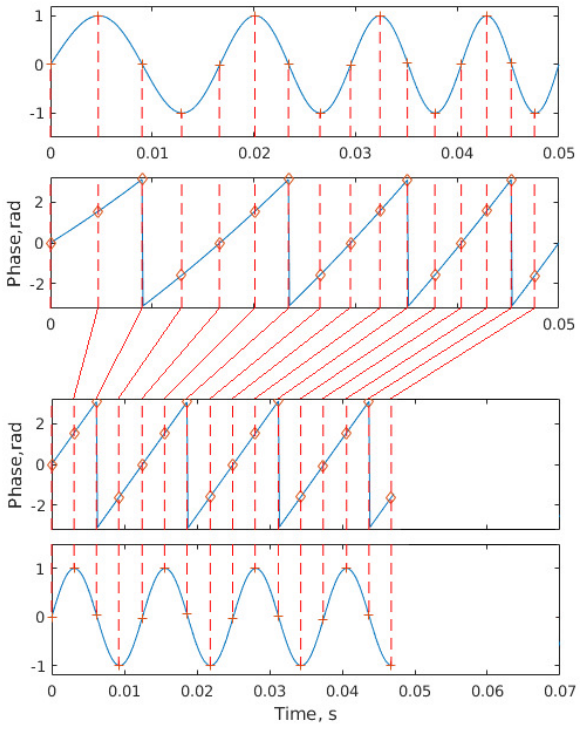

Figure 18: Synchronous resampling.

the stator current after Hilbert transform. Given f_s and N, resampled signal central frequency (fundamental frequency) is calculated from $f_1 = \frac{f_s}{N}$. To be successful in detecting faults, the inequality $f_s > 2 \cdot f_{fault}$ should hold. Define $\omega_n \in \underline{\omega},\ n = 1,...,N$ as the next phase value we are seeking. Any ω_n from the set $\underline{\omega}$ can be chosen as the initial one, since that does not have much of an influence on the final result. The full synchronous resampling algorithm is shown in Figure 19. After each new sample ω_j is obtained, the comparison

$$\omega_{j-1} \leq \omega_n < \omega_j \tag{47}$$

is done. If ω_n does not belong to this region, the next sample ω_{j+1} is drawn, otherwise resampling is performed. Given the assumption $f_{so} \gg f_s$, the linear interpolation is used to estimate the signal at the new sampling position. It is done by using the following two equations

$$\gamma = \frac{\omega_n - \omega_{j-1}}{\omega_j - \omega_{j-1}} + j - 1, \tag{48}$$

$$C_r = \frac{\gamma - (j-1)}{j - (j-1)}(C_j - C_{j-1}) + C_{j-1}. \tag{49}$$

Combining (48) and (49), we get

$$C_r = \frac{\omega_n - \omega_{j-1}}{\omega_j - \omega_{j-1}}(C_j - C_{j-1}) + C_{j-1}. \tag{50}$$

A certain part of the samples is discarded through the resampling process. As it is possible to see from Figure 20, fundamental frequency of stator current has a constant value of $f_1 = \frac{f_s}{N} = \frac{1920}{32} = 60$ Hz, and, moreover, excitations are clearly visible on frequencies 54, 56, 58, 62, 64, 66 Hz.

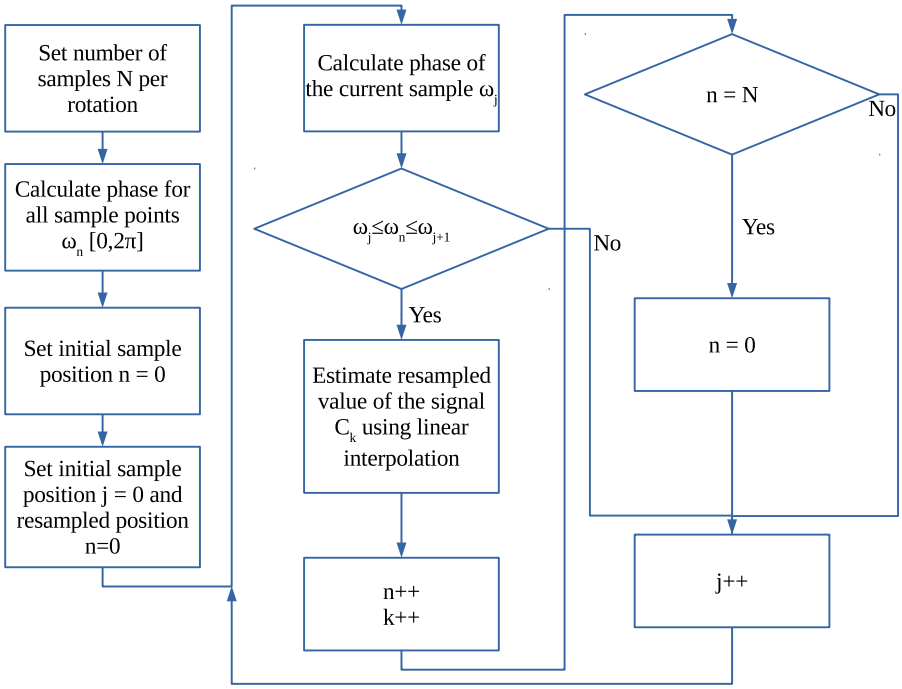

Figure 19: Synchronous resampling algorithm.

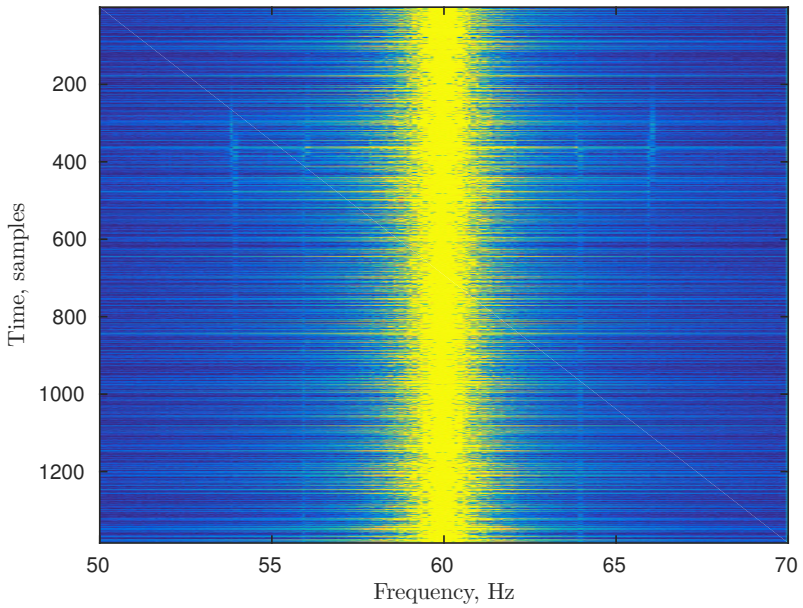

Figure 20: Spectrogram of the resampled stator current

4.3.2 Theoretical Frequencies of Bearing Fault Signal

It is known that faults of some bearings introduce harmonics on particular frequencies into the stator current of a WT. It is our desire to be able to detect them in the noise. There are four possible fault modes for bearings: cage fault, rollers defect, inner race fault, outer race fault. Each one is characterized by a particular frequency of the corresponding fault signal. The frequencies can be found from the equations [8]:

$$f_{\mathrm{ir}} = 0.5 \cdot N_B \cdot f_{\mathrm{r}} \left(1 + \frac{D_b \cdot \cos\phi}{D_p}\right), \tag{51}$$

$$f_{\mathrm{or}} = 0.5 \cdot N_B \cdot f_{\mathrm{r}} \left(1 - \frac{D_b \cdot \cos\phi}{D_p}\right), \tag{52}$$

$$f_{\mathrm{b}} = f_{\mathrm{r}} \cdot \left(\frac{D_b}{D_p}\right) \cdot \left[1 - \left(\frac{D_b \cdot \cos\phi}{D_p}\right)^2\right], \tag{53}$$

$$f_{\mathrm{c}} = 0.5 \cdot f_{\mathrm{r}} \left(1 - \frac{D_b \cdot \cos\phi}{D_p}\right). \tag{54}$$

where f_{r} is bearing rotation frequency, D_b is rollers diameter, D_p rollers pitch diameter, N_B is the number of rollers, and ϕ is rollers contact angle with races (see Figure 4).

The bearing under investigation is supporting the shaft, that connects generator with wind blades. For that reason, the vibrations excited in the bearing will affect the stator current, by modulating the amplitude of fundamental frequency. Frequencies of fault signals in stator current can be calculated using (55).

$$f_{\mathrm{fault}} = f_1 \pm k \cdot f_{\mathrm{bf}}, \tag{55}$$

where $f_{\mathrm{bf}} \in \{f_{\mathrm{ir}},\ f_{\mathrm{or}},\ f_{\mathrm{b}},\ f_{\mathrm{c}}\}$. The exact frequencies of fault do not match with those theoretically estimated using (55), most likely, due to absence of additional parameter (ϕ in (51), (52), (53), (54) is usually assumed to be 0 [17]). Nevertheless, this estimation is good, and the problem can be solved by considering signals on adjacent frequencies.

Faulty compartment	Modulating frequency, Hz	Fault frequency f_{fault} for $k = 1$, Hz
Outer race, f_{or}	30.3	29.7, 90.3
Inner race f_{ir}	49.69	10.31, 109.69
Cage f_{c}	3.78	56.22, 63.78
Rollers f_{r}	2.28	57.72, 62.28

Table 2: Fault frequencies

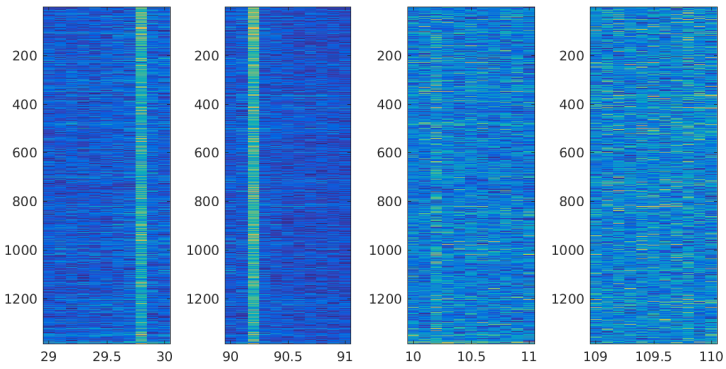

Figure 21: Spectrograms around frequencies, corresponding to inner and outer race faults. Horizontal axis represents frequency, vertical - time in samples. Relatively high power harmonics due to shaft eccentricity, caused by one bearing removed, are observed.

Now we are in position to calculate frequencies of fault. Due to resampling parameters, that were used during synchronous resampling procedure, $f_1 = 60$ Hz. For the WT used to obtain stator current, $f_{\text{r}} = \frac{f_1}{p}$, where $p = 6$ is the number of pole pairs. The parameters of the bearing supporting the shaft are $D_{\text{b}} = 8$ mm, $D_{\text{c}} = 33$ mm, $N_{\text{b}} = 8$. Calculated frequencies of fault signals are presented in the Table 2.

4.3.3 Bearing fault feature vector

The faults of a bearing in direct-drive wind turbine stator current are signified by excitations on particular harmonics in the frequency spectrum. FFT is the major tool to extract those excitations. We consider the magnitude of frequency spectrum components. This alleviates the need of some phase control for elements of the spectrum, which otherwise might be required.

The spectrograms of fault frequencies are shown in Figures 21.

In Figure 21 spectrograms around frequencies, corresponding to outer and inner race

faults are shown. Harmonics around frequencies 10, 30, 90, and 110 Hz are present. These harmonics are caused by eccentricity due to one bearing being removed during the experiment.

In Figure 20 excitations due to fault of the bearing are visible on frequencies corresponding to bearing cage and rollers defects. There are clear excitations visible near those frequencies at samples from 250 through 550. In fact, excitations remain in place, with smaller amplitude and are not captured with this plot.

The feature extraction process is described as follows. Denote $\mathbf{x}$ to be a feature vector of the fault under investigation, $\mathbf{x}^l$ is the vector of lower sideband fault frequencies, $\mathbf{x}^u$ is the vector of upper sideband fault frequencies, $w = \{2,4,...\}$ is a number of additional amplitudes of harmonics around f_{fault} included in the feature vector ($w = 10$ is used in this thesis). Each of the sidebands provides enough information, to identify the presence of fault. Therefore, $\mathbf{x}$ is a vector of size $w+1 \times 1$, and it is constructed in the following way. Denote $p = w+1$ as the length of the feature vector. If A_{ij} is the magnitude of the ith sample on jth frequency index of FFT spectrum, where $j = 0,1,...,N-1$, j_{fault} is the index, such that $|f_s \frac{j_{\text{fault}}}{N} - f_{\text{fault}}|$ is minimized, i.e. j_{fault} is the index, corresponding to the frequency closest to calculated f_{fault}, then the feature vector has the following configuration

$$\mathbf{x}_i^{\text{l}} = \left[A_i\left(j_{\text{f}}^{\text{l}} - \frac{\text{w}}{2}\right),...,A_i\left(j_{\text{f}}^{\text{l}}\right),...,A_i\left(j_{\text{f}}^{\text{l}} + \frac{\text{w}}{2}\right)\right], \tag{56}$$

$$\mathbf{x}_i^{\text{u}} = \left[A_i\left(j_{\text{f}}^{\text{u}} - \frac{\text{w}}{2}\right),...,A_i\left(j_{\text{f}}^{\text{u}}\right),...,A_i\left(j_{\text{f}}^{\text{u}} + \frac{\text{w}}{2}\right)\right], \tag{57}$$

$$\mathbf{x}_i = \left[\mathbf{x}_i^{\text{l}}\right]^{\text{T}} \text{ or } \left[\mathbf{x}_i^{\text{u}}\right]^{\text{T}}. \tag{58}$$

From this moment on, denote $j = 1,...,p$ to be the index of the feature vector components. The components of feature vector $\mathbf{x}$ are denoted by x_j.

It is well known, that generator will introduce noise in the sidebands of the fundamental frequency f_1. If f_{fault} is close to f_1, such are $f_{\text{fault,c}}$ and $f_{\text{fault,r}}$, it will result in correlated non-Gaussian noise between the feature vector components. On the other hand, if f_{fault} is far

enough from f_1, such are $f_{\text{fault,ir}}$ and $f_{\text{fault,or}}$, then the noise will be white Gaussian. According to the data acquired in [17], which are used to design fault detection procedure in this thesis, the noise in $\mathbf{x}_{\text{c}}$ and $\mathbf{x}_{\text{b}}$ is correlated, and the noise in $\mathbf{x}_{\text{or}}$ and $\mathbf{x}_{\text{ir}}$ is white Gaussian.

The key to detecting fault signal is the ability to estimate noise and find signal behavior nonconforming with noise distribution. As it is shown in the next part, the distribution of feature vector components x_j, $j = 1, ..., p$ can be identified.

4.4 Distribution of Feature Vector Components

Let's consider the distribution of feature vector components x_j when no signal is present and only noise is observed. The signals for x_j are extracted from FFT frequency spectrum. FFT generally gives in-phase and quadrature components of the signal, which are usually represented by a complex value. For that reason, we are going to discuss, how the magnitude of a complex random variable is distributed. Consider the complex Gaussian random variable

$$z \sim \mathcal{N}_c\left(\underline{\mu}_z, \sigma^2\right), \tag{59}$$

where $\underline{\mu}_z$ is a vector of mean values for real and imaginary parts. Let real and imaginary parts have equal variance $\sigma^2/2$. In case of zero mean Gaussian noise, the magnitude

$$|z| \sim \text{Rayleigh}(\sigma_r). \tag{60}$$

The probability density function (PDF) of Rayleigh distribution is

$$f_R(|z||\sigma_r) = \frac{|z|}{\sigma^2} e^{-|z|^2/(2\sigma_r^2)}, \tag{61}$$

where the parameter σ_r can be estimated using

$$\sigma_r = \mathbb{E}[|z|]\sqrt{\frac{2}{\pi}}. \tag{62}$$

Another case of magnitude $|z|$ distribution is when $\underline{\mu}_z \neq \underline{0}$. For non-zero mean complex Gaussian variable

$$|z| \sim \text{Rice}(K, \Omega). \tag{63}$$

The PDF of Rice distribution is

$$\bar{f}_{Ri}(|z||K, \Omega) = \frac{K+1}{\Omega} e^{-K-\frac{(K+1)|z|^2}{\Omega}} I_0\left(\sqrt{\frac{K(K+1)}{\Omega}}|z|\right), \tag{64}$$

where $I_0(x)$ is the zero order Bessel function of the first kind, and parameters of the PDF can be estimated as

$$\Omega = \mathbb{E}[|z|^2], \tag{65}$$

$$K = \frac{\mathbb{E}[|z|]^2}{\mathbb{E}[|z|^2] - \mathbb{E}[|z|]^2}. \tag{66}$$

Both of these distributions can be generalized by Gamma distribution. Consider

$$|z| \sim \text{Gamma}(\alpha, \beta). \tag{67}$$

The PDF of Gamma distribution is

$$f_G(|z|, \alpha, \beta) = \frac{\beta^\alpha |z|^{\alpha-1} e^{-|z|\beta}}{\Gamma(\alpha)}, \tag{68}$$

where $\Gamma(\alpha)$ is Gamma function. Parameters α and β can be estimated as

$$\alpha = \frac{\mathbb{E}[|z|]^2}{\mathbb{E}[|z|^2] - \mathbb{E}[|z|]^2}, \tag{69}$$

$$\beta = \frac{\mathbb{E}[|z|^2] - \mathbb{E}[|z|]^2}{\mathbb{E}[|z|]}. \tag{70}$$

By analyzing the available data, it was verified that the signal of feature vector components follows Gamma distribution. Consider the feature vector for outer race fault $\mathbf{x}_{\text{or}}$. In the absence

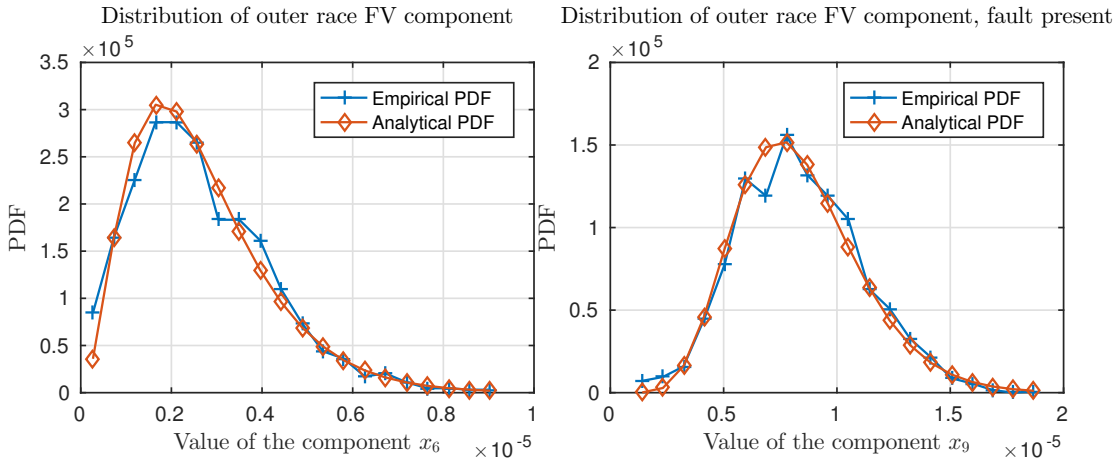

Figure 22: Empirical and analytical Gamma distribution of outer race feature vector components when: only noise is observed (on the left) and fault signal is present (on the right)

of the fault signal, components x_j have independent Gaussian zero-mean distribution, which results in Rayleigh distributed magnitude. When the fault signal is present the distribution of corresponding components appears to be Rician. Both of these distributions are generalized by Gamma distribution as well. The comparison of empirical PDF and theoretical one was done using estimated parameters (69) and (70). The results are shown in Figure 22.

Now consider the feature vector for cage fault $\mathbf{x}_c$. The noise signal for this feature vector is correlated between different components and cannot be described neither by Rayleigh nor Rice distribution. Nevertheless, Gamma distribution fits the signal. The comparison of empirical PDF and theoretical one is shown in Figure 23. Now we are going to discuss the methods of constructing hypotheses H_0 (no fault is present) and H_1 (fault has occurred).

4.5 Constructing Hypotheses H_0 and H_1

4.5.1 Motivation

The signals of components of feature vector x_j follow Gamma distribution. The parameters of Gamma distribution's PDF, α and β, can be estimated using (69) and (70). When the values of these parameters for hypotheses H_0 (no fault is present) and H_1 (fault has occurred) are known, it is of no difficulty to construct change detection test. However, we need to remember that

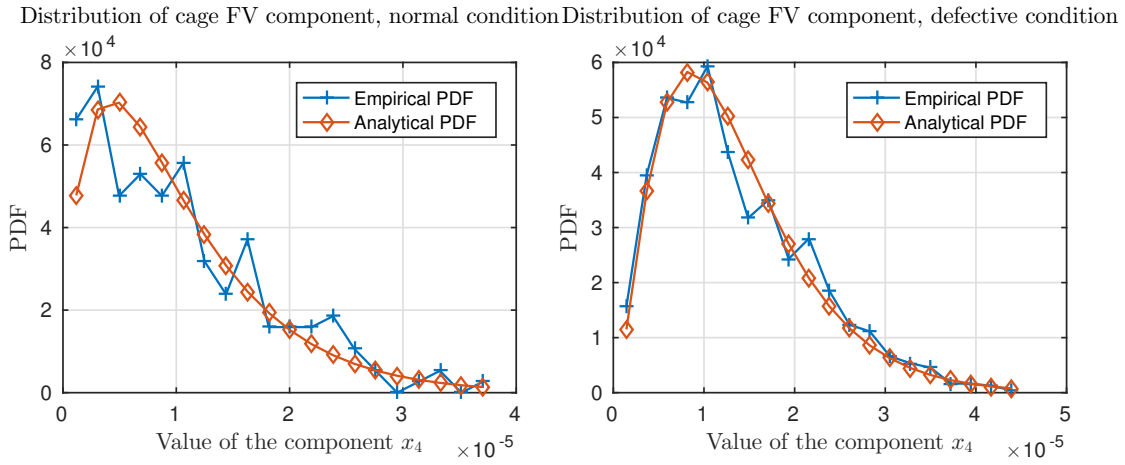

Figure 23: Empirical and analytical Gamma distribution of cage feature vector components when: only noise is observed (on the left) and fault signal is present (on the right)

real (not simulated) signals are often non-stationary. When parameters of the signal fluctuate in time, it is hard to construct proper hypotheses. Moreover, it is not clear, how exactly distribution of fault signal differs from the one of noise. One of the ways to approach the problem of non-stationary parameters is to perform some sort of normalization. Also, we can check, how the value of distribution parameters for noise change over time, and whether the trend of change is the same for the fault signal.

Further the problem of normalization is discussed, the change of distribution parameters is studied, and, eventually, the hypotheses H_0 and H_1 are defined.

4.5.2 Energy Normalization

The stator current signal is non stationary. First, it is non-stationary in frequency, due to operation in variable wind speed condition, second, it is non-stationary in power. Even in the case, when only white Gaussian noise is present, the change of noise power over time can occur. One of the approaches to detect fault signal in these conditions, is to consider some reference variable, where we assume no signal can be present. The intuition behind this approach is that the power of the fault signal should be larger than the power of the adjacent harmonics. We expect all the components of the feature vector that contain only noise, to have a similar

behavior, and hence, similar distribution parameters.

The estimators of Gamma distribution parameters are described by (69) and (70). The estimators of first and the second moment of a component x_{ij} of a feature vector $\mathbf{x}_i$, where i is a sample time, can be constructed as in (71) and (72)

$$\mathbb{E}[x_{ij}] = \frac{1}{n} \sum_{l=i-n+1}^{i} x_{lj}, \tag{71}$$

$$\text{var}[x_{ij}] = \frac{1}{n-1} \sum_{l=i-n+1}^{i} x_{lj}^2 - \mathbb{E}[x_{lj}]. \tag{72}$$

The total power can be defined as

$$e_{ij} = \mathbb{E}[x_{ij}]^2 + \text{var}[x_{ij}]. \tag{73}$$

By normalizing signals of x_j one can spot the difference of parameters α and β between cases where fault signal is present and absent. The normalization procedure is shown in (74).

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{e}} \tag{74}$$

4.5.3 Relationship Between Parameters α and β of Gamma Distribution

In this section we study the relationship between parameters of Gamma distribution. Denote

$$\mu_{ij} = \mathbb{E}[x_{ij}], \ \sigma_{ij}^2 = \text{var}[x_{ij}]. \tag{75}$$

Combining (75) with (69) and (70) one will get

$$\alpha_{ij} = \frac{\mu_{ij}^2}{\sigma_{ij}^2}, \ \beta_{ij} = \frac{\mu_{ij}}{\sigma_{ij}^2}. \tag{76}$$

Using (73) and (76) the power e_{ij} of the signal, corresponding to some component x_j, can be expressed through parameters α_{ij} and β_{ij}:

$$e_{ij} = \frac{\alpha_{ij}^2}{\beta_{ij}^2} + \frac{\alpha_{ij}^2}{\beta_{ij}}. \tag{77}$$

The normalization procedure (74) will result in change of the distribution parameters:

$$\tilde{\alpha}_{ij} = \frac{\mu_{ij}^2}{\sigma_{ij}^2},\ \tilde{\beta}_{ij} = \frac{\mu_{ij}\sqrt{e_{ij}}}{\sigma_{ij}^2}. \tag{78}$$

After normalization, the power of $\tilde{x_j}$ is 1. From the equation

$$\frac{\tilde{\alpha}_{ij}^2}{\tilde{\beta}_{ij}^2} + \frac{\tilde{\alpha}_{ij}}{\tilde{\beta}_{ij}^2} = 1 \tag{79}$$

the relationship (80) follows

$$\tilde{\beta}_{ij} = \sqrt{\tilde{\alpha}_{ij}(\tilde{\alpha}_{ij}+1)}. \tag{80}$$

The meaning of this result is that when the power of Gamma distributed random variable is fixed, the change in one of the parameters of distribution will lead to change in another one. This result will be used later in this thesis to control the level of change that we desire to detect.

4.5.4 Comparison of $\tilde{\alpha}$ and $\tilde{\beta}$ Before and After Fault Has Occurred

Wind turbine current signal is non-stationary. The parameters of distribution change over time. We want to find out, how different is the behavior of change in parameters $\tilde{\alpha}$ and $\tilde{\beta}$ for the cases when only noise is observed, and when the fault signal is present. We employed the normalization procedure (74) and estimated the values of parameters using (78). The change of parameter $\tilde{\alpha}_{ij}$ of $\tilde{x}_{ij}$ for the feature vector $\tilde{\mathbf{x}}_{i,\text{cage}}$ is shown in Figure 24. It is seen that the value of parameter $\tilde{\alpha}_{ij}$ for most of the components of feature vector $\tilde{\mathbf{x}}_{i,\text{cage}}$ change in a similar way, except $\tilde{\alpha}_{i4}$ of component x_4, where the fault signal is present. The analysis of this plot

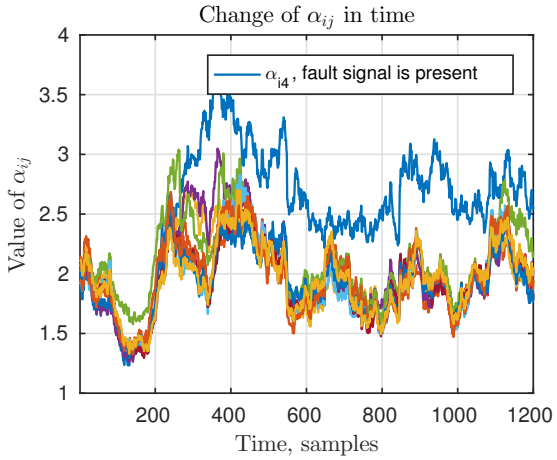

Figure 24: Change of $\tilde{\alpha}_{ij}$ in time for components of cage feature vector.

suggests the follow steps of the detection procedure:

- Select a reference variable x_k, $k = 1, ..., p$. It is assumed that only noise is present on x_k. The change of parameter $\tilde{\alpha}$ between different components x_j is observed to be similar.
- We pick x_j, $j \neq k$ and evaluate whether fault is present. If fault has not occurred,

$$H_0 : \tilde{\alpha}_{ij} \simeq \tilde{\alpha}_{ik}, \tag{81}$$

otherwise

$$H_1 : \tilde{\alpha}_{ij} > \tilde{\alpha}_i. \tag{82}$$

Select $k = 1$. Consider the ratio $\tilde{\alpha}_{i4}/\tilde{\alpha}_{i1}$, i.e. ratio of $\tilde{\alpha}$ when fault is present, and when it is not. The change of the ratio $\tilde{\alpha}_{i4}/\tilde{\alpha}_{i1}$ in time is shown in Figure 25. The average value of the ratio after fault signal is introduced is 1.4 (calculated beginning from 300th sample).

The analysis of distribution parameters for other feature vectors showed similar results. It worth noting that for outer and inner race feature vectors the magnitude of change is much larger. The results of outer race feature vector analysis are shown in Figure 26. The fault signal is observed for component x_9.

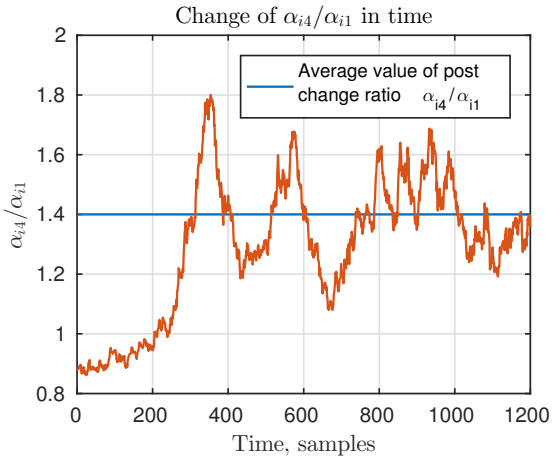

Figure 25: Change of $\tilde{\alpha}_{i4}/\tilde{\alpha}_{i1}$ shows how different the value of parameter $\tilde{\alpha}_{ij}$ is before and after fault has occurred

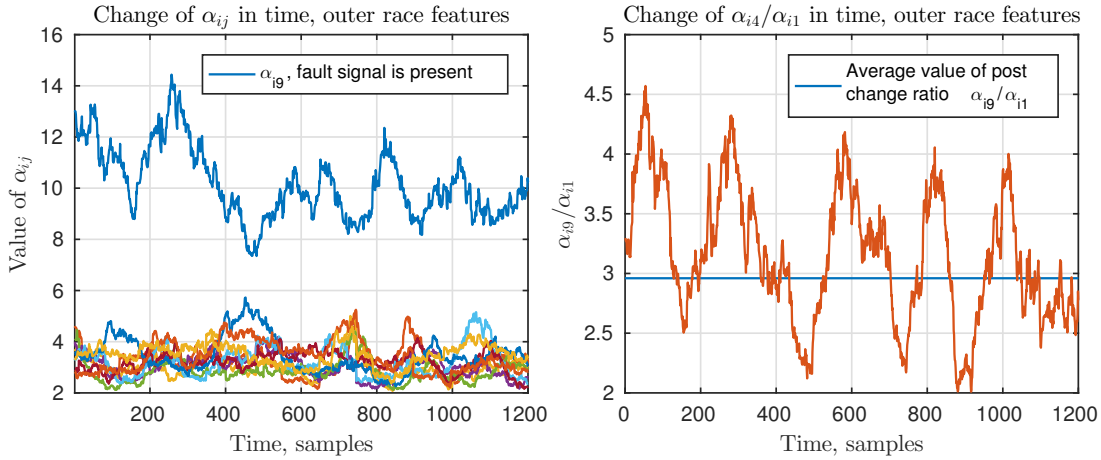

Figure 26: Change of parameter $\tilde{\alpha}_{ij}$ for components of feature vector of outer race fault (on the left) and change of the ratio $\tilde{\alpha}_{i9}/\tilde{\alpha}_{i1}$ in time (on the right).

4.5.5 Formulation of Hypotheses H_0 and H_1

We have identified that the components of feature vectors follow same family of distribution, specifically - Gamma distribution. We have shown that the parameters α and β of the distribution can be estimated using (76). Also, when the power of the signal (77) is known, and normalization (74) is performed, parameters of distribution follow the relationship (80). A new parameter (83) is introduced, which can be treated as magnitude of change in distribution parameter α we desire to detect. The analysis in section 4.5.4 suggests the following step of detection procedure. Define

$$\alpha_\Delta = \frac{\alpha_{H_1}}{\alpha_{H_0}} \tag{83}$$

where $\alpha_{H_0} = \tilde{\alpha}_{ik}$, k is the index of reference variable. We pick x_j, $j \neq k$, , and evaluate whether fault is present. If fault has not occurred,

$$H_0 : \tilde{\alpha}_{ij} \simeq \alpha_{H_0}, \tag{84}$$

otherwise

$$H_1 : \tilde{\alpha}_{ij} \simeq \alpha_{H_0} \cdot \alpha_\Delta. \tag{85}$$

The probability density for hypothesis H_0 is represented by

$$f_{ij,H_0}(\tilde{x}_{ij}|\tilde{\alpha}_{ik}, \tilde{\beta}_{ik}) = \frac{\tilde{\beta}_{ik}^{\tilde{\alpha}_{ik}} \tilde{x}_{ij}^{\tilde{\alpha}_{ik}-1} e^{-\tilde{x}_{ik}\tilde{\beta}_{ik}}}{\Gamma(\tilde{\alpha}_{ik})}. \tag{86}$$

The probability density for hypothesis H_1 is represented by

$$f_{ij,H_1}(\tilde{x}_{ij}|\tilde{\alpha}_{ik} \cdot \alpha_\Delta, \hat{\beta}_{ik}) = \frac{\hat{\beta}_{ik}^{\tilde{\alpha}_{ik}} \tilde{x}_{ij}^{\tilde{\alpha}_{ik}-1} e^{-\tilde{x}_{ik}\hat{\beta}_{ik}}}{\Gamma(\tilde{\alpha}_{ik})}, \tag{87}$$

where $\hat{\beta}_{ik}$ is calculated from (80) using the value $\tilde{\alpha}_{ik} \cdot \alpha_\Delta$.

Obviously, setting strict value of parameter α_Δ introduces mismatch of the post change

distribution model. The impact of such mismatch will be discussed in the section 4.6.5.

4.6 Application of Change Detection Theory for Bearing Faults Identification

4.6.1 Change Detection Problem Formulation

Consider a random process $x_1, x_2, ...$, where x is a scalar. Define $\mathbf{x}^{k:n} = [x_k, ..., x_n]$, let $\mathscr{F}_x^n = \sigma(\mathbf{x}^{1:n})$ be the sigma algebra generated by $\mathbf{x}^{1:n}$. Assume there is an unknown moment λ, when the distribution of the random process changes from f_0 to f_1. Let $\mathbf{P}_k$ and $\mathbb{E}_k$ denote the probability measure and the corresponding expectation, when the change point occurs at $\lambda = k$. The conditional probability density for the pre-change and the post-change distributions are $f_0(x_n|\mathbf{x}^{1:n-1})$ and $f_1(x_n|\mathbf{x}^{1:n-1})$. The change point has random value, with a prior probability distribution $\mathbf{P}(\lambda = k) = \pi_k$, for $k_\lambda = 1, 2, ...$

Quickest change detection techniques are special methods, designed to detect change occurrence from f_0 to f_1 with a small delay, while keeping probability of false alarm bounded. Change detection procedures incorporate new data sequentially as they arrive, and can be used for online change detection. Define a detection procedure as mapping to a positive integer $\nu \leq n$

$$\delta : \mathscr{F}_x^n \to \{\nu : \nu \leq n\}, n = 1, 2, ... \tag{88}$$

and ν is a change point detected by δ. Procedure $\delta(\mathbf{x}^{1:n})$ gives stopping time. This definition was used in [58].

In general case, the pre-change distribution of x_n is $f_0(x_n|\mathbf{x}^{1:n-1})$, and the post-change distribution is $f_1(x_n|\mathbf{x}^{1:n-1})$, with samples of x not necessarily independent in time, random process $x_1, x_2, ...$ has covariance matrix Σ_X, and the distributions f_0 and f_1 can have any form. In practice, many assumptions are made to simplify the model. Having independent samples much simplifies calculations for any change detection procedure.

As it was mentioned in section 3, operating characteristic functions of change detection

procedures are average detection delay [38]

$$\mathrm{ADD}(\delta) = \mathbb{E}^{\pi}(\nu - \lambda | \nu \geq \lambda), \tag{89}$$

and average run length to false alarm (ARL)

$$\mathrm{ARL}(\delta) = \mathbb{E}_{\infty}(\nu). \tag{90}$$

Probability of false alarm (FA) for change detection methods can be expressed as

$$\mathrm{PFA}(\delta) = \mathbf{P}_{\infty}(\nu < \lambda). \tag{91}$$

There is three well known change detection procedures. The first one is Shiryaev procedure [37]. While being asymptotically optimal in Bayesian context [40], it requires the knowledge of the prior distribution of change moment. The two other, CUSUM procedure, developed by Page [55], and Shiryaev-Roberts procedure, introduced by Roberts [56], on the other hand, do not require this knowledge to construct the test. The two latter procedures have similar performance.

In terms of ADD vs ARL these two procedures were compared in [42]. The comparison is shown in Figure 9 in section 3.5.3.

For the sake of further discussion define likelihood ratio

$$\Lambda_{k:n} = \prod_{l=k}^{n} \frac{f_{\bar{H}_1,l}(x_l | \mathbf{x}^{1:l-1})}{f_{\bar{H}_0,l}(x_l | \mathbf{x}^{1:l-1})} = \prod_{l=k}^{n} \lambda_l, \tag{92}$$

and denote hypotheses:

$$\bar{H}_0 : n < \lambda, \ \bar{H}_1 : n \geq \lambda. \tag{93}$$

4.6.2 CUSUM Change Detection Procedure

The stopping rule for CUSUM procedure has the following form

$$\nu_c = \inf\{n : C_n \geq A\}, \tag{94}$$

where C_n is

$$C_n = \max_{1\leq k\leq n} \frac{f_{\lambda=k}(x_1, x_2, ..., x_n)}{f_{\lambda=\infty}(x_1, x_2, ..., x_n)}. \tag{95}$$

The density $f_{\lambda=\infty}$ is the joint density of the observations, when no change ever take place, $f_{\lambda=k}$ is the joint density of the observations, when change is assumed to be $\lambda = k$. The test statistics C_n can be expressed through $\Lambda_{k:n}$

$$C_n = \max_{1\leq k\leq n} \Lambda_{k:n} \tag{96}$$

In case of i.i.d. random variable x, it is possible to calculate C_n recursively

$$C_{n+1} = \max(1, C_n)\lambda_{n+1},\ n \geq 1. \tag{97}$$

For the different cases $f_{\bar{H}_0}$is (86), and $f_{\bar{H}_1}$ is (87). The choice of the threshold A is a matter of the trade-off between ADD and ARL, and it was discussed in section 3.5.2. We simply set

$$A = \frac{1}{\alpha_{\text{FA}}}, \tag{98}$$

where α_{FA} is the desired upper bound on the PFA.

4.6.3 Shiryaev-Roberts Change Detection Procedure

The procedure, that later became known as Shiryaev-Roberts procedure, was first introduced in [56]. The stopping rule is described by (99)

$$\nu_{sr} = \inf\{n \geq 1 : R_n \geq A\} \tag{99}$$

where

$$\begin{aligned} R_n &= \sum_{k=1}^{n} \frac{f_{\lambda=k}(x_1, x_2, ..., x_n)}{f_{\lambda=\infty}(x_1, x_2, ..., x_n)} \\ &= \sum_{k=1}^{n} \prod_{l=k}^{n} \lambda_l. \end{aligned}$$

Here k is assumed moment of change, n is total number of samples acquired. For i.i.d. case the expression above can be expressed in recursive form

$$R_{n+1} = (1 + R_n)\lambda_{n+1}, \; n \geq 1. \tag{100}$$

The densities for $\bar{H}_0$ and $\bar{H}_1$ are the same as for CUSUM procedure. The choice of the threshold A was discussed in 3.5.3. We simply set

$$A = \frac{1}{\alpha_{\text{FA}}}. \tag{101}$$

4.6.4 Multiple Hypothesis Testing for Feature Vector Components

The hypotheses for the change detection procedure (93) can be expressed through (84) and (85). The probability densities (86) and (87) are described for a particular variable of the feature vector x_j, $j = 1, 2, ..., p$. Therefore, keeping that we have a reference variable with index k, there are $p - 1$ possibilities where the fault signal can be located in the feature vector. This fact may suggest the use of multiple hypothesis testing techniques. Some of research, devoted to study of multiple hypothesis testing was done in [46, 57, 59]. Technically, the proper way to

test hypotheses is to construct a joint PDF, which in our case will have a form

$$f_{H_1}(\mathbf{x}) = \sum_{j=1, j\neq k}^{p} f_{j,H_1}(x_j) \cdot p(j). \tag{102}$$

However, to construct (102) the knowledge of $p(j)$ is required, in order to give proper weights to f_{j,H_1}. In reality those weights are unknown, and moreover, the event of fault signal being present in component x_j of the feature vector is not truly random, but rather the value of j_{fault} is simply unknown. For that reason, whichever multiple hypothesis test is used, the exact values of the performance parameters, like PFA, will be different. Nevertheless, a simple multiple hypothesis testing procedure can be constructed.

The parallel multiple hypothesis testing can be defined as follows. Consider one variable x_j, $j \in \{1,...,p\}, j \neq k$ at a time. Perform a change detection procedure δ_j for x_j, using stopping rule (94) or (99):

$$\nu_{c,j} = \inf\left\{n : C_{n,j} \geq A\right\} \quad \nu_{r,j} = \inf\{n \geq 1 : R_{n,j} \geq A\}. \tag{103}$$

Whenever one of the procedures δ_j raises alarm, the fault detection is declared at a time

$$\bar{\nu} = \min_j(\nu_j), \tag{104}$$

where ν_j is obtained using one of the rules in (103).

Another stopping rule can be designed based on the assumption that the distribution of j_{fault} is uniform, and $\mathrm{p}(j) = \frac{1}{p-1}$. This stopping rule is described as

$$\hat{\nu} = \inf\left\{n : \bar{C}_n \geq A\right\}, \quad \bar{C}_n = \frac{1}{p-1} \sum_{j=1, j\neq k}^{p} C_{n,j} \tag{105}$$

for CUSUM procedure, or

$$\hat{\nu} = \inf\{n : \bar{R}_n \geq A\}, \quad \bar{R}_n = \frac{1}{p-1} \sum_{j=1, j\neq k}^{p} R_{n,j} \tag{106}$$

for Shiryaev-Roberts procedure.

One of the advantages of this simple approach is that we do not consider the the whole feature vector **x**. When the components x_j are correlated it is challenging to construct a multivariate Gamma distribution. While considering only one variable at a time does not incorporate all available information in the testing procedure, it is shown later in the section 5 to be sufficient to detect a fault.

4.6.5 Post Change Model Mismatch

In the section 4.5.5 the PDFs for hypotheses H_0 and H_1, denoting whether fault signal is present or not, were introduced. The parameter α_Δ (83) was used to describe the PDF of H_1 (87). This parameter was defined as the magnitude of Gamma distribution parameter α change we desire to detect. It was mentioned that this approach, where α_Δ is defined beforehand inevitably introduces mismatch of the post change distribution.

In [58] the study of the mismatched post-change models was done. It was found that the assumption of such kind introduces additional detection delay, while keeping the PFA upper bound at the same level. For two procedures (CUSUM, Shiryaev-Roberts) the upper bound of FA is described in (107).

$$\mathrm{PFA}(\bar{\delta}_c(A)) \leq \frac{1}{A} \leq \frac{\mathbb{E}[\lambda]}{A}, \; \mathrm{PFA}(\bar{\delta}_r(A)) \leq \min\left\{\frac{\mathbb{E}[\lambda]}{A}\right\}. \tag{107}$$

where $\bar{\delta}_c$ defines CUSUM detection procedure with mismatched post-change model, and $\bar{\delta}_r$ defined Shiryaev-Roberts procedure with mismatched post-change model.

The analysis for increase in delay for detecting a change with mismatched post-change model was shown in section 3.7.

4.6.6 Change Detection by Using Threshold

The analysis in section 4.5.4 suggests a method of change detection, based on estimating the value of parameter $\alpha_{\Delta,ij} = \frac{\tilde{\alpha}_{ij}}{\tilde{\alpha}_{ik}}$ and comparing it with a threshold α_Δ

$$\alpha_{\Delta,ij} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha_\Delta. \tag{108}$$

The method presumably should work well, given the proper value of the threshold α_Δ is set. Probability of FA for methods based on thresholding, in general, is very small, when α_Δ is large enough. The price to pay is the detection delay being increased due to additional time needed to estimate the value of parameters. Moreover, thresholding methods are bad at detecting incipient faults that only start to appear, and the value of $\alpha_{\Delta,ij}$ has not yet reached sufficient for detection level. When the value of α_Δ is not set properly, the detection of fault through thresholding may never occur.

The comparison of detection techniques using change detection methods and thresholding is done in section 5. It is shown that change detection techniques are superior to thresholding methods, especially when the value of α_Δ is not set properly, as long as the distribution model for H_0 is set right.

5 Performance of Proposed Bearing Fault Detection Algorithm

Now we are in position to test the proposed bearing fault detection algorithms. The signal that are expected to be discovered in the WT stator current are:

- cage fault, due to broken cage
- rollers fault, due to broken cage
- outer race fault, due to eccentricity
- inner race fault, due to eccentricity

The test wind turbine was pretreated by removing one of the bearings, supporting main shaft. That resulted in eccentricity that should be present in stator current signal from the very beginning. The cage fault of bearing occurred during testing. The breakdown did not happen simultaneously, and parameters of signal distribution have changed gradually. It is assumed that the change of distribution occurred after 230th sample of feature vector for cage and rollers feature vectors.

5.1 Algorithm Parameters

Essentially, there are several parameters that can be changed to tune fault detection algorithm performance:

1. Parameter estimators (71) and (72) window length n.
2. The degree of change α_Δ
3. The value of threshold A
4. Size of FFT window N

Further the analysis of algorithm performance for different values of most important parameter α_Δ and A is done.

5.2 False Alarm Rate of Fault Detection Algorithm

The of false alarm rate for different values of parameter α_Δ is shown in Figure 27. It was plotted based on the signals from the parts of the frequency spectrum, where no fault signal is present.

The bound for probability of false alarm for CUSUM procedure is given in (107). However, this bound is valid when the detection procedure is built properly. The change detection procedure (104) cannot be classified so. Therefore, the bound of false alarm rate should be different. The plot in Figure 27 suggest that the bound is larger, but still quite close to $\frac{1}{A}$. Even when the value of parameter α_Δ changes, FA rate tends to stay on the same level. When $\alpha_\Delta = 1.1$ no FA is present. The procedure decides that change never occurred.

For Shiryaev-Roberts procedure the bound on false alarm is (107). As we can see, the bound is not strict, and can be up to one. The simulation results show that increasing α_Δ leads to reduction of FA rate. The increase of threshold A leads to decrease in FA, as expected. It follows from the Figures that CUSUM procedure is more reliable for bearing fault detection using stopping rule (104) in terms of FA rate.

The FA rate for procedures (105) and (106) are shown in Figure 28. These two procedures assume that the location of fault signal in feature vector is uniformly distributed. The FA rate is below specified upper bounds. We were unable to capture the values of FA rate lower than 10^{-3} due to limited amount of data in disposal. Though the procedures that incorporate averaging have lower FA rate, we still cannot say that they describe the real signal properly, due to uniformly distributed j_{fault} assumption.

From these results it follows that setting the threshold A to sufficient level will ensure low probability of false alarm for CUSUM procedure.

5.3 Detection Delay

Probability of false alarm is not the only important evaluation metric. Intuitively, the procedures that have lower FA rate have longer detection delay. In Figure 29 and 30 the detection

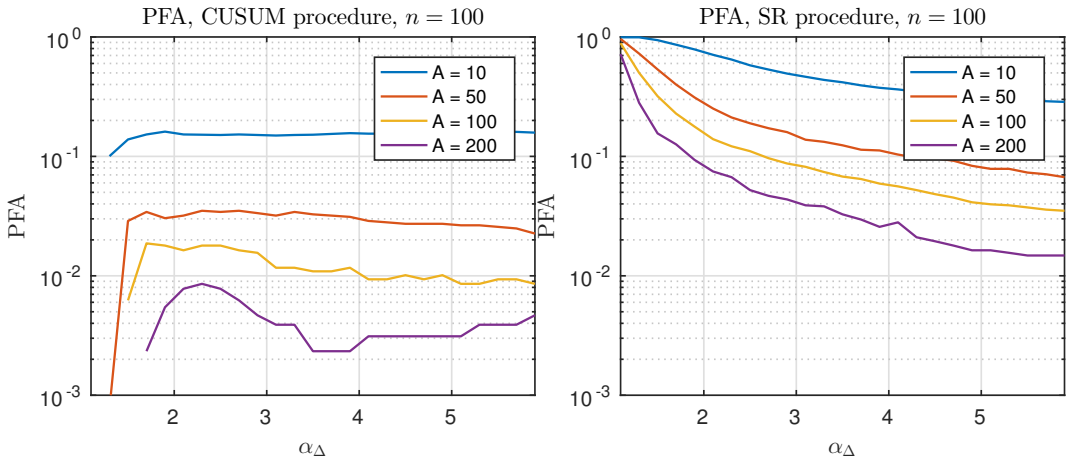

Figure 27: FA rate for: CUSUM procedure (on the left) and Shiryaev-Roberts procedure (on the right) using stopping rule (104).

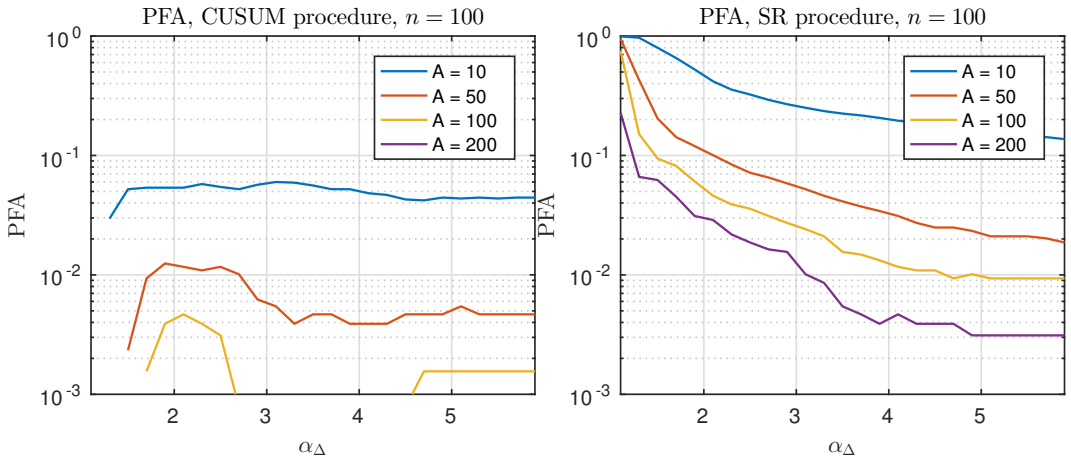

Figure 28: FA rate for: CUSUM procedure (105) (on the left) and Shiryaev-Roberts procedure (106) (on the right)

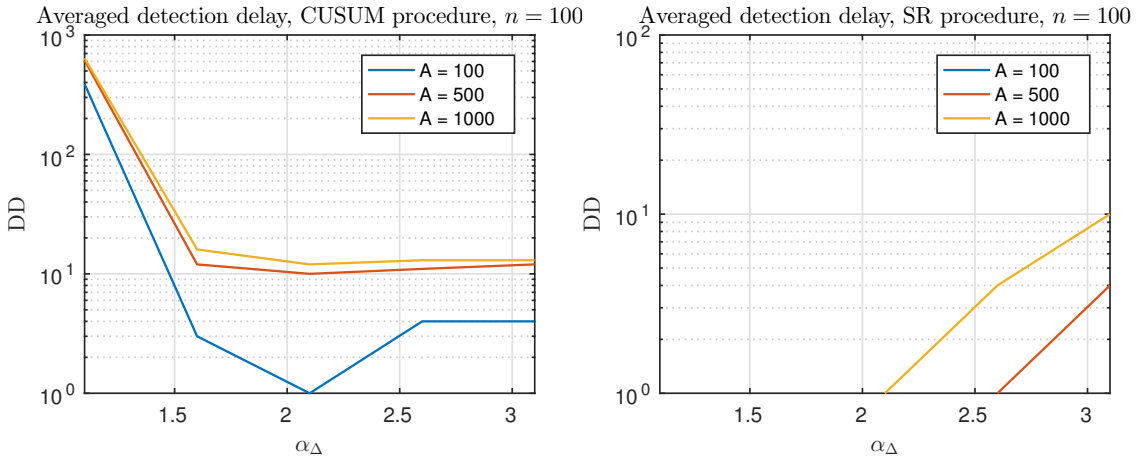

Figure 29: Averaged detection delay: CUSUM procedure (on the left) and Shiryaev-Roberts procedure (on the right) using stopping rule (104).

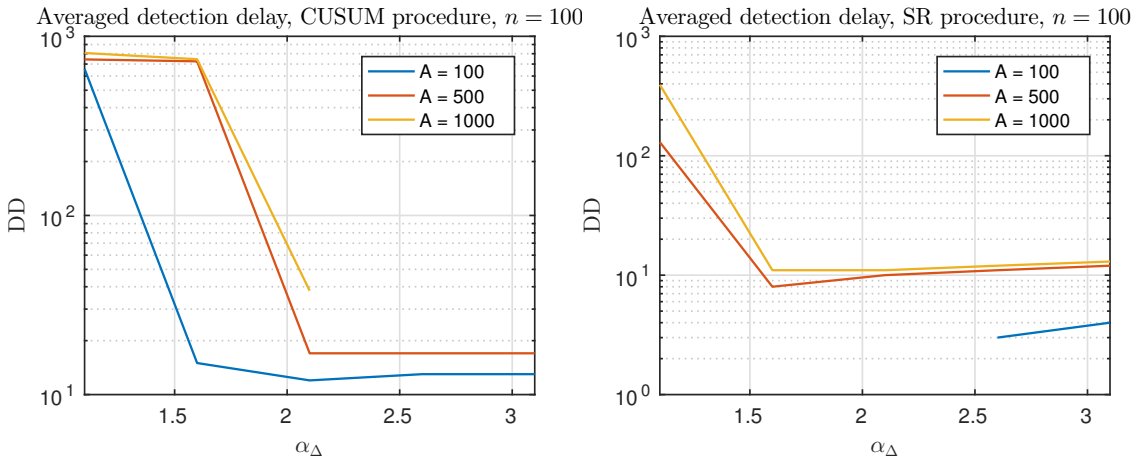

Figure 30: Averaged detection delay: CUSUM procedure (105) (on the left) and Shiryaev-Roberts procedure (106) (on the right)

delay, averaged between different types of fault is shown. As it was expected, the delay for Shiryaev-Roberts procedure is lower, which explains high FA rate. For CUSUM procedure the delay is minimized around value of $\alpha_\Delta = 2$. The true values of post-change distribution parameter is close to this value, as it was shown in section 4.5.4.

In Figure 31 the detection delays for different faults are shown. The method of thresholding, described in the section 4.6.6 was used. For the value of $\alpha_\Delta = 2$ and above no fault was detected, except for outer and inner race fault, there detection delay was 0. This demonstrated the disadvantage of using thresholding methods for fault detection. If the threshold is

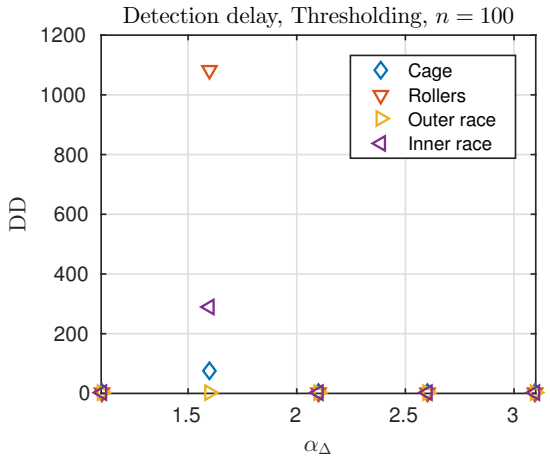

Figure 31: Detection delay for the thresholding method (108)

not set properly, the fault will never be detected.

5.4 Comparison of Change Detection Procedures

Four detection procedures, given in (104), (105), and (106), were suggested. Here we perform the direct comparison, of these procedures. The main parameter that can be used to tune change detection algorithm performance is the threshold A. Different values of the threshold were used to build curves, describing relationship between false alarm rate and detection delay. The results are shown in Figure 32.

The results show that the procedures (105) and (106) do not really give advantage over (104) by assigning some weights to different hypotheses. The intuition that weights will improve performance, is not justified. For the same value of FA rate procedure (104) gives smaller value of delay. Moreover, CUSUM procedure give a big advantage over Shiryaev-Roberts procedure for large values of FA. However, the difference becomes very small, as FA rate goes down, for all considered detection rules. The comparison of CUSUM procedure performance for different values of α_Δ is shown in Figure 33. Here the delay is reduced, as the value of α_Δ becomes larger.

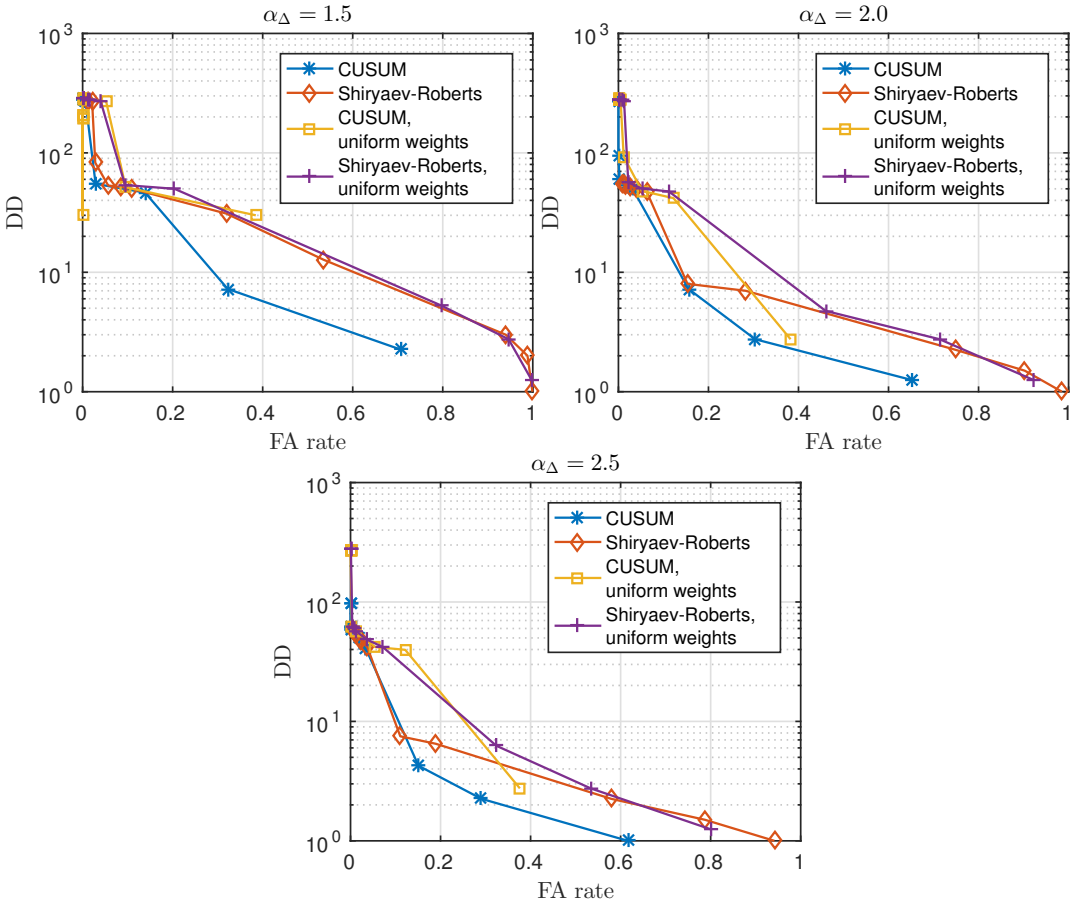

Figure 32: Comparison of change detection procedures (104), (105), and (106) for different values of α_Δ.

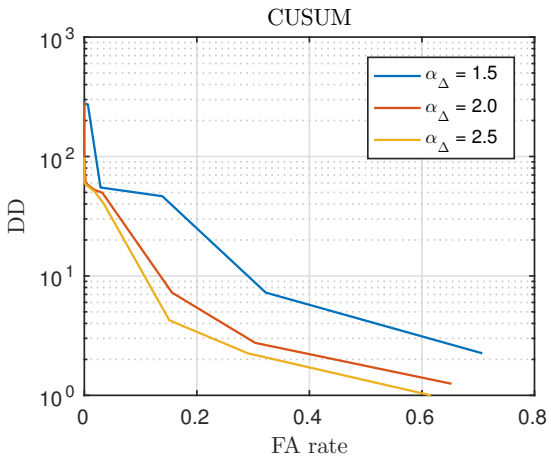

Figure 33: Performance of CUSUM procedure (104) for different values of α_Δ.

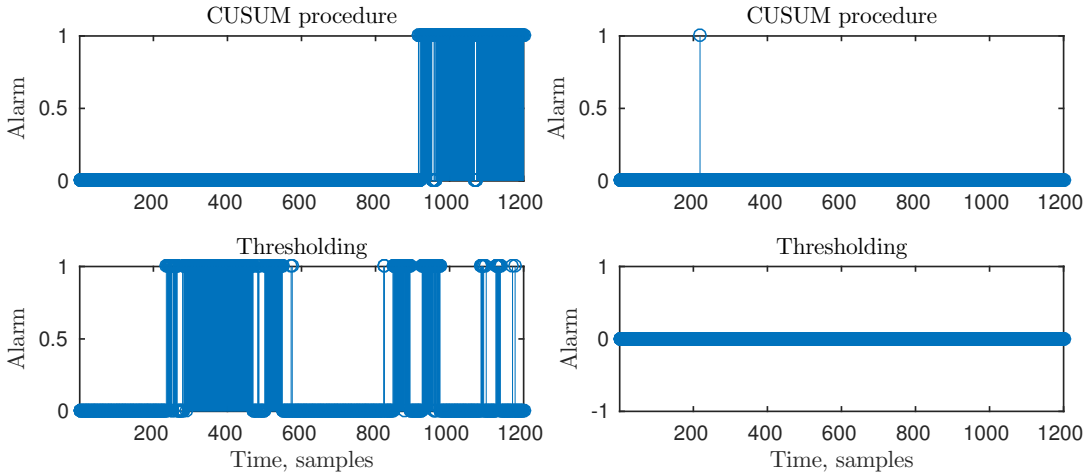

Figure 34: Alarms of detection procedure in time. Cage feature vector is used. $n = 100$, $A = 1000$: $\alpha_\Delta = 1.3$ (on the left) and $\alpha_\Delta = 2$ (on the right).

5.5 Detection Frequency

One more important parameter is how often the alarm of detection procedure is fired. The comparison of detection frequency is shown in Figure 34 for cage feature vector, and in Figure 35 for outer race feature vector. The detection procedure (105) was used. The method of thresholding fires alarm earlier, when the value of α_Δ is close to the true value. CUSUM procedure waits until the likelihood exceed the specified level of confidence before declaring the alarm. When α_Δ is set to be large, thresholding method cannot detect change, while change detection procedure still works.

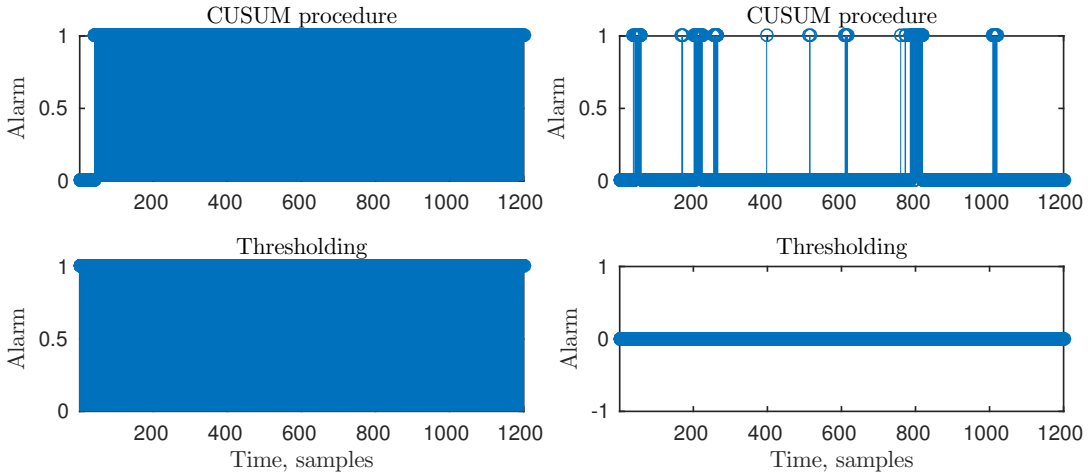

Figure 35: Alarms of detection procedure in time. Outer race feature vector is used. $n = 100$, $A = 1000$: (on the left) and $\alpha_\Delta = 10$ (on the right).

6 Conclusions

In this thesis a new bearing fault detection algorithm was proposed. The whole fault detection procedure can be split into several steps.

1. Synchronous resampling of the stator current signal is performed.
2. Fault signatures are extracted using magnitudes of FFT spectrum components.
3. Reference variable is selected and distribution parameters are estimated.
4. One of the change detection procedures is used to detect change in signal distribution, given the value of parameter α_Δ.

The application of change detection procedure was possible, because the distribution of feature vector components was obtained. It was revealed that the components of feature vector follow Gamma distribution. The change detection procedure based on CUSUM was shown to be superior over Shiryaev-Roberts procedure in terms of PFA. The methods based on assigning some weights to multiple hypotheses did not give advantage over the simple parallel multi-hypothesis change detection procedure in terms of detection delay for a given value of FA rate. The obtained results are based on a small number of observations, and larger datasets should be used, to verify observed behavior of change detection procedures, given different parameters.

Estimation of parameters introduces model mismatch, which results in longer detection delay, but the method is still able to detect fault. Even when the value of α_Δ is set to be large, change point detection procedure is able to spot the change in distribution, while the thresholding method ceases to work.

Because samples are assumed to be independent in time, and change detection procedures are designed recursively, only small amount of memory is required. The estimation of the second moment $\mathbb{E}[x_j^2]$ is done for all $j = 1, 2, ..., p$ in a sliding window, but that does not require a lot of resources.

The amount of available data did not allow us to check PFA for greater values of A. From the simulations results, it is seen, that the value of the threshold A is closely related to the upper bound on the FA rate. That suggests, that setting A to higher values will significantly reduce the chance of FA. More data is needed to check performance for larger A.

This algorithm was designed to detect excitation in the frequency spectrum. This suggests, that this method can also be used for detecting other types of fault that introduce additional harmonics into stator current spectrum, such as blade imbalance, aerodynamic asymmetry, rotor eccentricity, torque bending moment, etc.

References

[1] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis - part i: Components and subsystems," *Industrial Electronics, IEEE Transactions on*, vol. 62, no. 10, pp. 6536–6545, 2015.

[2] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis - part ii: Signals and signal processing methods," 2015.

[3] L. Morris, "Direct drive vs. gearbox: Progress on both fronts." `http://www.power-eng.com/articles/print/volume-115/issue-3/features/direct-drive-vs-gearbox-progress-on-both-fronts.html`, 2011. Online; Accessed: 2016-03-27.

[4] P. Tavnet, G. Van Bussel, and F. Spinato, "Machine and converter reliabilities in wind turbines," in *Power Electronics, Machines and Drives, 2006. The 3rd IET International Conference on*, pp. 127–130, IET, 2006.

[5] M. Blödt, P. Granjon, B. Raison, and G. Rostaing, "Models for bearing damage detection in induction motors using stator current monitoring," *Industrial Electronics, IEEE Transactions on*, vol. 55, no. 4, pp. 1813–1822, 2008.

[6] B. Lu, Y. Li, X. Wu, and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis," in *Power Electronics and Machines in Wind Applications, 2009. PEMWA 2009. IEEE*, pp. 1–7, IEEE, 2009.

[7] Z. Daneshi-Far, G. Capolino, and H. Henao, "Review of failures and condition monitoring in wind turbine generators," in *Electrical Machines (ICEM), 2010 XIX International Conference on*, pp. 1–6, IEEE, 2010.

[8] X. Gong and W. Qiao, "Bearing fault diagnosis for direct-drive wind turbines via current-demodulated signals," *Industrial Electronics, IEEE Transactions on*, vol. 60, no. 8, pp. 3419–3428, 2013.

[9] R. R. Schoen, T. G. Habetler, F. Kamran, and R. Bartfield, "Motor bearing damage detection using stator current monitoring," *Industry Applications, IEEE Transactions on*, vol. 31, no. 6, pp. 1274–1279, 1995.

[10] F. Immovilli, M. Cocconcelli, A. Bellini, and R. Rubini, "Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals," *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 11, pp. 4710–4717, 2009.

[11] B. M. Ebrahimi, J. Faiz, and M. J. Roshtkhari, "Static-, dynamic-, and mixed-eccentricity fault diagnoses in permanent-magnet synchronous motors," *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 11, pp. 4727–4739, 2009.

[12] W. Yang, P. J. Tavner, and M. Wilkinson, "Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train," *Renewable Power Generation, IET*, vol. 3, no. 1, pp. 1–11, 2009.

[13] F. Immovilli, A. Bellini, R. Rubini, and C. Tassoni, "Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison," *Industry Applications, IEEE Transactions on*, vol. 46, no. 4, pp. 1350–1359, 2010.

[14] S. Djurovic, C. J. Crabtree, P. J. Tavner, and A. Smith, "Condition monitoring of wind turbine induction generators with rotor electrical asymmetry," *Renewable Power Generation, IET*, vol. 6, no. 4, pp. 207–216, 2012.

[15] X. Gong and W. Qiao, "Imbalance fault detection of direct-drive wind turbines using generator current signals," *Energy Conversion, IEEE Transactions on*, vol. 27, no. 2, pp. 468–476, 2012.

[16] B. M. Ebrahimi, M. Javan Roshtkhari, J. Faiz, and S. V. Khatami, "Advanced eccentricity fault recognition in permanent magnet synchronous motors using stator current signature analysis," *Industrial Electronics, IEEE Transactions on*, vol. 61, no. 4, pp. 2041–2052, 2014.

[17] X. Gong and W. Qiao, "Current-based mechanical fault detection for direct-drive wind turbines via synchronous sampling and impulse detection," *Industrial Electronics, IEEE Transactions on*, vol. 62, no. 3, pp. 1693–1702, 2015.

[18] X. Gong and W. Qiao, "Current-based eccentricity detection for direct-drive wind turbines via synchronous sampling," in *Energy Conversion Congress and Exposition (ECCE), 2013 IEEE*, pp. 2972–2976, IEEE, 2013.

[19] D. Lu, X. Gong, and W. Qiao, "Current-based diagnosis for gear tooth breaks in wind turbine gearboxes," in *Energy Conversion Congress and Exposition (ECCE), 2012 IEEE*, pp. 3780–3786, IEEE, 2012.

[20] D. Lu, W. Qiao, X. Gong, and L. Qu, "Current-based fault detection for wind turbine systems via hilbert-huang transform," in *Power and Energy Society General Meeting (PES), 2013 IEEE*, pp. 1–5, IEEE, 2013.

[21] R. M. Jones, "Enveloping for bearing analysis," *Sound and vibration*, vol. 30, no. 2, pp. 10–15, 1996.

[22] L. M. Popa, B.-B. Jensen, E. Ritchie, and I. Boldea, "Condition monitoring of wind generators," in *Industry Applications Conference, 2003. 38th IAS Annual Meeting. Conference Record of the*, vol. 3, pp. 1839–1846, IEEE, 2003.

[23] S. Djurovic, S. Williamson, P. Tavner, and W. Yang, "Condition monitoring artefacts for detecting winding faults in wind turbine dfigs," in *Proc. EWEC*, pp. 16–19, 2009.

[24] X. Gong, W. Qiao, and W. Zhou, "Incipient bearing fault detection via wind generator stator current and wavelet filter," in *IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society*, pp. 2615–2620, IEEE, 2010.

[25] P. Guo and N. Bai, "Wind turbine gearbox condition monitoring with aakr and moving window statistic methods," *Energies*, vol. 4, no. 11, pp. 2077–2093, 2011.

[26] P. Santos, L. F. Villa, A. Reñones, A. Bustillo, and J. Maudes, "Wind turbines fault diagnosis using ensemble classifiers," in *Advances in Data Mining. Applications and Theoretical Aspects*, pp. 67–76, Springer, 2012.

[27] I. Antoniadou, G. Manson, N. Dervilis, T. Barszcz, W. Staszewski, and K. Worden, "Condition monitoring of a wind turbine gearbox using the empirical mode decomposition method and outlier analysis," in *6th European Workshop on Structural Health Monitoring. Proceedings available online at http://www. ewshm2012. com/Proceedings. aspx*, 2012.

[28] I. S. Bozchalooi and M. Liang, "Parameter-free bearing fault detection based on maximum likelihood estimation and differentiation," *Measurement Science and Technology*, vol. 20, no. 6, p. 065102, 2009.

[29] H. Ocak and K. A. Loparo, "A new bearing fault detection and diagnosis scheme based on hidden markov modeling of vibration signals," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 5, pp. 3141–3144, IEEE, 2001.

[30] X. Zhang, R. Xu, C. Kwan, S. Y. Liang, Q. Xie, and L. Haynes, "An integrated approach to bearing fault diagnostics and prognostics," in *American Control Conference, 2005. Proceedings of the 2005*, pp. 2750–2755, IEEE, 2005.

[31] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. Orchard, and G. Vachtsevanos, "Anomaly detection: A robust approach to detection of unanticipated faults," in *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, pp. 1–8, IEEE, 2008.

[32] M. E. Orchard and G. J. Vachtsevanos, "A particle-filtering approach for on-line fault diagnosis and failure prognosis," *Transactions of the Institute of Measurement and Control*, 2009.

[33] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. Orchard, and G. Vachtsevanos, "A probabilistic fault detection approach: application to bearing fault detection," *Industrial Electronics, IEEE Transactions on*, vol. 58, no. 5, 2011.

[34] X. An, D. Jiang, and S. Li, "Application of back propagation neural network to fault diagnosis of direct-drive wind turbine," in *World Non-Grid-Connected Wind Power and Energy Conference (WNWEC), 2010*, pp. 1–5, IEEE, 2010.

[35] Y. Wang and D. Infield, "Neural network modelling with autoregressive inputs for wind turbine condition monitoring," in *Sustainable Power Generation and Supply (SUPERGEN 2012), International Conference on*, pp. 1–6, IET, 2012.

[36] M.-S. An, S.-J. Park, J.-S. Shin, H.-Y. Lim, and D.-S. Kang, "Implementation of automatic failure diagnosis for wind turbine monitoring system based on neural network," in *Multimedia and Ubiquitous Engineering*, pp. 1181–1188, Springer, 2013.

[37] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.

[38] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, pp. 206–227, 1985.

[39] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *Information Theory, IEEE Transactions on*, vol. 44, no. 7, pp. 2917–2929, 1998.

[40] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic bayesian theory of quickest change detection," *Theory of Probability & Its Applications*, vol. 49, no. 3, pp. 458–497, 2005.

[41] Y. Mei, "Sequential change-point detection when unknown parameters are present in the pre-change distribution," *The Annals of Statistics*, pp. 92–122, 2006.

[42] G. V. Moustakides, A. S. Polunchenko, and A. G. Tartakovsky, "Numerical comparison of cusum and shiryaev–roberts procedures for detecting changes in distributions," *Communications in Statistics–Theory and Methods*, vol. 38, no. 16-17, pp. 3225–3239, 2009.

[43] M. Pollak and A. G. Tartakovsky, "Optimality properties of the shiryaev-roberts procedure," *Statistica Sinica*, pp. 1729–1739, 2009.

[44] M. Pollak, "The shiryaev–roberts changepoint detection procedure in retrospect – theory and practice," in *Proceedings of the 2nd International Workshop on Sequential Methodologies, University of Technology of Troyes, Troyes, France*, 2009.

[45] A. S. Polunchenko and A. G. Tartakovsky, "On optimality of the shiryaev-roberts procedure for detecting a change in distribution," *The Annals of Statistics*, pp. 3445–3457, 2010.

[46] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

[47] X. Gong and W. Qiao, "Bearing fault detection for direct-drive wind turbines via stator current spectrum analysis," in *Energy Conversion Congress and Exposition (ECCE), 2011 IEEE*, pp. 313–318, IEEE, 2011.

[48] X. Gong and W. Qiao, "Current-based online bearing fault diagnosis for direct-drive wind turbines via spectrum analysis and impulse detection," in *Power Electronics and Machines in Wind Applications (PEMWA), 2012 IEEE*, pp. 1–6, IEEE, 2012.

[49] M. M. Gowda, N. Mallikarjun, P. Gowda, and R. Chandrashekhar, "Improvement of the performance of wind turbine generator using condition monitoring techniques," in *2013 7th International Conference on Intelligent Systems and Control (ISCO)*, 2012.

[50] B. Courtice, "Rare earth magnets: not all new turbines are using them." `http://yes2renewables.org/2012/03/06/rare-earth-magnets-not-all-new-turbines-are-using-them/`, 2012. Online; Accessed: 2016-03-27.

[51] P. Tavner, "How are we going to make offshore wind farms more reliable?," *SUPERGEN Wind*, 2011.

[52] N. Weller, "Acceleration enveloping–higher sensitivity, earlier detection," *Orbit*, pp. 10–19, 2004.

[53] M. Schlechtingen, I. F. Santos, and S. Achiche, "Wind turbine condition monitoring based on scada data using normal behavior models. part 1: System description," *Applied Soft Computing*, vol. 13, no. 1, pp. 259–270, 2013.

[54] Y. Amirat, V. Choqueuse, and M. Benbouzid, "Wind turbines condition monitoring and fault diagnosis using generator current amplitude demodulation," in *Energy Conference and Exhibition (EnergyCon), 2010 IEEE International*, pp. 310–315, IEEE, 2010.

[55] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[56] S. Roberts, "A comparison of some control chart procedures," *Technometrics*, vol. 8, no. 3, pp. 411–430, 1966.

[57] V. Draglia, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests. i. asymptotic optimality," *Information Theory, IEEE Transactions on*, vol. 45, no. 7, pp. 2448–2461, 1999.

[58] J. Wu and J. Yang, "Quickest change detection with mismatched post-change models," *arXiv preprint arXiv:1601.06868*, 2016.

[59] T. L. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 595–608, 2000.