AUTOMATED DETECTION AND PREDICTION OF ELECTRICAL DISTURBANCES IN A POWER TRANSMISSION SYSTEM

By

Jonathan D. Boyd

Donald R. Reising
UC Foundation Associate Professor
Electrical Engineering
(Committee Chair)

Vahid R. Disfani
Assistant Professor
Electrical Engineering
(Committee Member)

Abdelrahman A. Karrar
Professor and Department Head
Electrical Engineering
(Committee Member)

AUTOMATED DETECTION AND PREDICTION OF ELECTRICAL DISTURBANCES IN A POWER TRANSMISSION SYSTEM

By

Jonathan D. Boyd

A Thesis Submitted to the Faculty of the University of Tennessee at Chattanooga in Partial Fulfillment of the Requirements of the Degree of Master of Science: Engineering

The University of Tennessee at Chattanooga
Chattanooga, Tennessee

May 2023

ABSTRACT

As power quality becomes a higher priority in the electric utility industry, utilities simply do not have the required personnel to analyze the ever-growing amount of data by hand. This thesis presents an automated approach for the analysis of power quality phenomena within a power transmission system by leveraging rule-based analytics as well as machine learning to analyze the characteristics of the recorded data. Waveform signatures analyzed within this thesis include: various faults, motor starting, and incipient instrument transformer failure. The developed analytics were tested on 160 waveform files and yielded an average accuracy of 99%. Machine learning techniques are also used to predict voltage unbalance on the transmission system above a certain threshold, which yielded an accuracy of over 91%. This work will result in time savings for engineers as well as increased reliability of the transmission system by providing near real–time detection, identification, and prevention of disturbances.

ACKNOWLEDGMENTS

I would first like to thank Dr. Donald Reising, chair of my committee, for the direction, guidance, and expertise that he provided throughout the course of this project. Thanks also to the other members of my committee, Dr. Abdelrahman Karrar and Dr. Vahid Disfani, for their time and willingness to serve on my committee.

I would also like to extend a special thanks to Mr. Tony Murphy of the Tennessee Valley Authority (TVA) for his continual guidance during this whole project. His knowledge and expertise in the Power Quality area proved to be invaluable during the course of this research. Thanks also to the other engineers at TVA who assisted at different points along the way, including Mr. Nathan Hooker, Mr. Justin Kuhlers, Mr. Michael McAmis, and Mr. Jim Rossman.

This project was completed as part of a research grant from the Tennessee Valley Authority awarded to the University of Tennessee at Chattanooga.

Thanks to both my parents as well as my fiancé, Lesley, for their encouragement and support throughout this endeavor.

Finally, I would like to give thanks to God for blessing me with this opportunity to learn and grow in knowledge, and for the ability and endurance to complete this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A, Ampere

A/D, Analog-to-Digital

ANN, Artificial Neural Network

COMTRADE, COMmon format for TRAnsient Data Exchange

CNN, Convolutional Neural Network

CT, Current Transformer

CVT, Capacitive Voltage Transformer

DC, Direct Current

DFR, Digital Fault Recorder

EHV, Extra-High Voltage

ELM, Extreme Machine Learning

FFT, Fast Fourier Transform

FT, Fourier Transform

GT, Gabor Transform

HT, Hilbert Transform

HV, High Voltage

IEC, International Electrotechnical Commission

IEEE, Institute for Electrical and Electronics Engineers

KF, Kalman Filter

KL, Kullback–Leibler

kV, kilovolt

Mvar, Megavolt-Ampere Reactive

MW, Megawatt

PQ, Power Quality

RMS, Root Mean Square

SCADA, Supervisory Control And Data Acquisition

ST, *S* Transform

STFT, Short-Time Fourier Transform

SVM, Support Vector Machine

SVP, Space Vector Property

THD, Total Harmonic Distortion

VFD, Variable Frequency Drive

VT, Voltage Transformer

VUF, Voltage Unbalance Factor

WT, Wavelet Transform

CHAPTER 1
INTRODUCTION

1.1 Overview

The continued and increasing deployment of "smart" devices (e.g., switches, relays, etc.) within power utility generation, transmission, and distribution infrastructure has led to the recording and storage of an ever-growing amount of data. Processing and analysis of this data have been traditionally conducted by power utility personnel using "by-hand" approaches. By-hand approaches rely heavily upon the knowledge, experience, and expertise of the person or persons conducting the analysis and severely limit the amount of data that can be analyzed within a given period of time. These limitations are exacerbated when considering that: (i) power utilities are unable to dedicate personnel solely to the task of disturbance processing and analysis as well as (ii) that analysis is often conducted hours if not days after the disturbance has occurred, thus limiting its value.

The first aspect of this thesis is to classify electrical disturbances recorded by Digital Fault Recorders (DFRs) and Power Quality (PQ) monitors. These devices generate waveform disturbance records anytime the disturbance threshold of the DFR or PQ monitor is exceeded. This automated process seeks to classify each record within minutes of its occurrence so that problems with power system equipment can be corrected in a timely manner. The contributions of this portion are as follows.

- The automated process is developed and tested using real-world data whereas most other methods use simulated data. All data are recorded by smart field devices–PQ monitors and DFRs–operating in a high-voltage transmission system.
- The rule-based methods mimic the expertise of an engineer in an effort to ease the interpretation

and understanding of the classification results by power system personnel.

- The developed process uses very few functions specific to any particular programming platform. This reduces the need for expensive licenses while allowing the algorithms to be translated into other programming languages and software based on the specific needs of the power utility. This approach is adopted to facilitate the widespread use of the developed algorithms across the power industry.

- The rule-based nature of the developed process allows every threshold to be changed as needed by power utility personnel based on performance or system specifics. In this thesis, empirical thresholds are designated symbolically as τ in equations, and the corresponding numerical values are designated with **bold** lettering in sentences.

- The methods used are very detailed and will predict the actual disturbance (e.g., ferroresonance) that occurred on the power system rather than simple signal characteristics like voltage sag and swell, which are detected by existing PQ software.

The next aspect of this thesis was to predict voltage unbalance in the Extra-High Voltage (EHV) transmission system [1]. Over two years of historical line loading data were obtained through a Supervisory Control and Data Acquisition (SCADA) system and used to train an Artificial Neural Network (ANN) to predict voltage unbalance based on Megawatt (MW) and Megavolt-Ampere Reactive (Mvar) line flows. The contributions of the voltage unbalance prediction work are as follows.

- The data used for training the model are from an operational EHV transmission system. Other works such as [2] use simulated data.

- The data are obtained through a SCADA system, which has a low sampling rate and is readily available to every transmission utility.

- This approach proves that voltage unbalance can be predicted using only MW or Mvar data, which eliminates the need for three-phase current or voltage measurements that are not always

available.

- The inputs to each station's model include data from all lines and substations in the area of the studied EHV system, which accounts for all line connections.

1.2 Related Works

This section presents several works related to PQ disturbance identification as well as voltage unbalance prediction.

The authors of [3] detail a rule-based approach for categorizing PQ disturbances using the *S* Transform (ST). The data used in this approach are a mix of simulated data and real-world data from the power system. The Fourier Transform (FT) and the Short-Time Fourier Transform (STFT) have not proven to be effective in extracting the unique features of each signal. The Wavelet Transform (WT) has been used as it can extract time and frequency domain characteristics simultaneously, but it is also somewhat vulnerable to noise and computationally expensive. The ST can be thought of as a hybrid between the STFT and WT since it has time and frequency domain characteristics, but it also uses a variable window length to provide information at different resolutions. The ST has been shown to provide better noise immunity. Finally, categorization of the PQ disturbances was performed using ANNs, fuzzy logic, decision trees, and others. The ST contours highlight the distinctive features present within the original PQ disturbance signal, such as voltage sag. A set of rules is then defined to set the thresholds needed to trigger certain disturbance types. These rules rely heavily upon the knowledge of PQ experts and a data set containing distorted signals is used to determine the corresponding threshold values. The rules are designed to separate the disturbances into three categories: magnitude disturbances, transients, and signal distortion. The tests performed on the signals include positive tests and negative tests for an extra layer of classification. The authors' approach is also very portable to other applications due to the normalization of the voltage, which facilitates the use of any voltage level. The results of the work in [3] heavily favor the rule-based ST approach. This approach classified the disturbances with 98% accuracy while a traditional ANN method achieved an accuracy of 92%. The rule-based method can also

withstand a considerable level of noise in the signal. One reason for this superior accuracy is that the rule-based approach is more specialized to each type of disturbance than the ANN approach.

The authors of [4] use a machine learning approach that is augmented through the inclusion of the Kullback–Leibler (KL) divergence measure and standard deviation. The KL divergence is very efficient as it can be applied to a single cycle of the signal. The KL divergence calculates the probability that a particular cycle is a member of two or more events. Standard deviation is also used as it is very effective in the detection of PQ disturbances. These two methods are used for each cycle of the disturbed signal and compared with an ideal sinusoidal signal to capture the disturbance. After the detection phase, the classification phase is performed using a Support Vector Machine (SVM) to determine a decision boundary between disturbance types. This method proves very effective in differentiating between disturbances such as voltage sag and swell. However, voltage flicker and swell are more similar than sag and swell, so this approach likely will not function as well. Overall, this method achieved an accuracy of 94.02%.

The authors of [5] provide a novel PQ disturbance classification method. The method extracts features from the cross-correlogram of the PQ disturbances. The positive peak and two adjacent negative peaks are used as the classification features. Those three values are then fed into a fuzzy-based classification system. One drawback to the work in [5] is its use of simulated data, thus classification accuracy may change when real-world data are used. The two types of correlation are cross-correlation and auto-correlation. Cross-correlation measures the strength of similarity between two different signals, while auto-correlation is the cross-correlation of a signal with itself. The work in [5] calculates the cross-correlation response between an ideal signal with a disturbed one to detect the disturbance. A fuzzy logic classifier is used to allow for uncertainty in a logic system. The rules in the fuzzy system are designed by human experts, so the system is only as good as those who designed it. The system used in this approach is the Mamdani-type inference system with three inputs and one output. Eighteen linguistic variables are used for the output membership function to determine the PQ disturbance classification. This classifier was tested using seventy generated signals and achieved an accuracy of 100%. The accuracy remained

100% even when noise was added to the test signals.

The authors of [6] present an approach for classifying PQ disturbances using a deep Convolutional Neural Network (CNN). A data set is obtained by simulating the mathematical models of PQ disturbances. Sixteen types of disturbances are included in the training data, including voltage sag and swell, harmonics, notches, flicker, and oscilliatory transients. Each disturbance contains 48,000 samples, making 768,000 total samples in the training data set. The first step of the training process of the CNN is automatic feature extraction through deep learning. The weights in each training layer are then updated through a closed-loop feedback system so that no manual operation will be needed. Several techniques are also employed to prevent overfitting of the CNN. To test the model, 10-fold cross validation is performed, so 10% of the data is held out from training to be used for validation. The results of this approach are very accurate, yielding up to 99% accuracy.

The author of [7] provides a comprehensive review of the methods used to classify PQ disturbances. Several signal processing techniques are listed for feature extraction of the waveforms, including FT, WT, ST, Hilbert Transform (HT), Kalman Filter (KF), and Gabor Transform (GT). Each one of these techniques has different variations that the author describes. Some intelligent classification techniques involving machine learning are then introduced, including ANN, SVM, fuzzy logic, Extreme Machine Learning (ELM), and deep learning. The results of using methods are then compared to determine the relative accuacy of each one. The drawbacks of the PQ disturbance classification methods described thus far are (i) the use of simulated data rather than actual data and (ii) the use of machine learning or advanced signal processing techniques, which makes it difficult for humans to understand the decisions made by the automated process. The developed process presented in this thesis uses 100% real-world data as well as simple signal processing techniques that can be easily understood by the engineers or other utility personnel reviewing the results of the automated process. The most complex signal processing technique used in this thesis is the Fast Fourier Transform (FFT).

Several methods exist for quantifying voltage unbalance, but very few are capable of predicting voltage unbalance before it occurs. The authors of [8] developed an algorithm that detects

voltage unbalance using the Space Vector Property (SVP), which transforms three voltages into a single complex variable. The three phases of voltage are not always available in modeling tools, so that is one drawback of this method. The algorithm sums the instantaneous values of all three voltages. The authors use a sum of zero to indicate the three-phase voltages are balanced, but summing to zero does not guarantee that the voltages are balanced. Thus, the SVP is then compared to a reference space vector to determine whether or not voltage unbalance occurs. The algorithm is tested using five cases, and the presence of voltage unbalance is correctly predicted for all five cases.

The authors of [2] estimate voltage unbalance using data simulated at the power distribution level. A load flow is performed using the Newton-Raphson method to generate the three-phase voltages. These are then transformed into sequence voltages, which are needed to calculate the voltage unbalance percentage. Probabilistic estimation of voltage unbalance is then performed using Monte Carlo simulation. The random variation of the power factor at different busses highlights which busses are the sources of the voltage unbalance. These methods are then tested on an operational power distribution system, and the consideration of different loading conditions is used to determine the expected level of voltage unbalance. The contributions of individual voltage unbalance sources are summed to determine the amount of unbalance on any bus.

The authors of [9] present a voltage unbalance detection approach for three-phase induction motors using an ANN. The ANN is trained using a data set containing one hundred samples collected over nine days from an operational three-phase induction motor. The ANN is constructed using a feed-forward structure, which is the most common structure. During training, the unbalanced voltages are labeled as "-1" and the balanced voltages are labeled as "1". The performance of this model is measured using Mean Squared Error and Root Mean Squared Error. The trained ANN correctly detects voltage unbalance with an accuracy of 100%. While this approach does use real-world data, the data set is comparatively small and only encompasses one motor. A data set from a larger system would be needed to fully determine the accuracy of this approach.

1.3 Problem Statement

The amount of data collected by PQ monitors and DFRs has become too much for utility personnel to analyze manually. Every time a disturbance is recorded, a waveform is stored in a database and is typically analyzed only on an "as-needed" basis, which often happens days after the disturbance has occurred when it is too late for corrective actions to be taken.

Voltage unbalance is another issue plaguing electric utilities as this unbalance can have adverse effects on industrial loads, leading to customer complaints toward the utility. Current software tools are able to measure voltage unbalance using past data, and some may even be able to measure it in real-time. However, system operators can do very little to mitigate voltage unbalance in real-time. Voltage unbalance is best mitigated in the transmission line outage planning process, which occurs weeks or months prior. State estimation software is used to approximate line flows resulting from a line outage. However, this software does not have three-phase voltage measurements modeled, so it cannot predict when a line outage will cause voltage unbalance.

1.4 Objectives

- Create an automated, rule-based process to classify different electrical disturbances based on waveform signature.
- Leverage machine learning techniques to develop a tool to predict voltage unbalance on the EHV transmission system based on MW and/or Mvar measurements when three-phase voltage or current measurements are not available.

1.5 Thesis Outline

The remainder of this work is outlined as follows. Chapter 2 provides background on the classification of electrical disturbances as well as voltage unbalance prediction. Chapter 3 describes the methodology and provides the results of the electrical disturbance analysis and classification, and Chapter 4 provides the methodology and results for the voltage unbalance prediction work. Finally, Chapter 5 lists some conclusions and opportunities for future work.

CHAPTER 2
BACKGROUND

The waveform analysis portion of this thesis leverages data recorded by DFRs and PQ monitors. Throughout the transmission system studied in this work, hundreds of these devices monitor voltage and current waveforms and are set to trigger when certain thresholds are exceeded. These thresholds are specific to each device and the equipment that they monitor. When voltages or currents exceed the threshold for what is considered "normal", the device triggers a disturbance that is recorded at a very high sampling rate, typically 256 samples per cycle. These disturbance waveforms are then transported back to a central location where they can be automatically analyzed by the techniques described in Chapter 3 of this thesis.

2.1 Waveform Signature Classification

2.1.1 Calculating Nominal Values

The first task in voltage or current waveform processing is to calculate nominal values from the data itself. The sampling frequency is calculated by,

$$F_s = \frac{N}{t_e - t_1}, \tag{2.1}$$

where F_s is the sampling frequency in Hertz (Hz), N is the number of samples in the time vector, and t_1 and t_e are the first and last values of the time vector, respectively. After the sampling frequency is calculated, the nominal number of samples in each cycle is determined by,

$$N_c = \frac{F_s}{F_n}, \tag{2.2}$$

where N_c represents the number of samples per cycle, and F_n is the nominal frequency of the power system, which is assumed to be 60 Hz throughout the entirety of this thesis.

Generally, PQ disturbance records include several cycles of the voltage or current waveform that are captured prior to the disturbance starting. The DFRs–that recorded the data used in this thesis–are normally set to record fifteen cycles of data before a disturbance. The nominal peak values of voltage and current waveforms are determined using these "pre-disturbance" cycles for each processed waveform. In this thesis, the first cycle in the disturbance record is used to determine these nominal scalar values denoted as (i) $\hat{V}_q$ for nominal peak voltage, (ii) $\hat{I}_q$ for nominal peak current, (iii) $\bar{V}_q$ for nominal Root Mean Square (RMS) voltage, and (iv) $\bar{I}_q$ for nominal RMS current. The magnitudes of the voltage and current waveforms are compared to their corresponding nominal values to normalize the data with respect to the power system's particular voltage or current level. For instance, a sudden increase in current at one place on the system will look different than at another place on the system, and this normalization allows the same analysis techniques to be used in both places.

2.1.2 Root Mean Square

A waveform's RMS is one characteristic used in the classification of electrical disturbances and is given by,

$$\bar{x} = \sqrt{\frac{1}{N_w}\sum_{i=1}^{N_w} |x[i]|^2}, \tag{2.3}$$

where x is the analog waveform, N_w is the size of the RMS window, and $\bar{x}$ is the RMS calculation of the analog waveform [10]. Unless otherwise stated, the size of the RMS window was set at the nominal number of samples in each cycle, N_c. The size of this window can be changed based on the level of granularity needed to classify a particular disturbance.

RMS is useful for determining if the waveform value is non-zero. In the instantaneous case, the sinusoidal waveform will cross zero every half-cycle, so it is more difficult to tell whether the value remains near zero. A waveform's RMS is used in disturbances such as motor starting where

the current increases over time.

2.1.3 Differentiation

A waveform's derivative is one of the most common calculations used in electrical disturbance characterization. Equation (2.4) represents the first derivative with respect to the number of samples.

A positive first derivative indicates the waveform is increasing, and a negative first derivative indicates the waveform is decreasing. This fact is used to detect the presence of peaks or spikes within a waveform. The maximum or minimum of a peak or spike corresponds to the first derivative changing sign (i.e., going from positive to negative or vice versa). A peak refers to one of the maxima or minima present in a normal sinusoidal waveform. A spike refers to an increase in the waveform outside of the normal sinusoidal behavior. Fig. 2.1 shows a spike in current occurring at the red circle.

A change in the first derivative's sign is calculated by,

$$x'(n_1) \times x'(n_2) < 0 \tag{2.4}$$

where x' is the first derivative of the analog waveform, n_1 is the sample before the first derivative's sign changes, and n_2 is the sample after the sign changes. Multiple sign changes over a short time interval (e.g., one cycle) provide a strong indication that a disturbance is present within the waveform being processed.

The second derivative is used to determine the change in the slope of a curve. A sudden increase in the second derivative shows as a sudden increase in slope and can indicate the point at which a fault begins. Fig. 2.1 provides a representative illustration showing the use of the second derivative in determining the start of a fuse fault. The red circle indicates the point at which the second derivative is higher than an empirical threshold, thus indicating a sudden increase in the slope of the curve. The third derivative is used to detect a shift in a curve's slope.

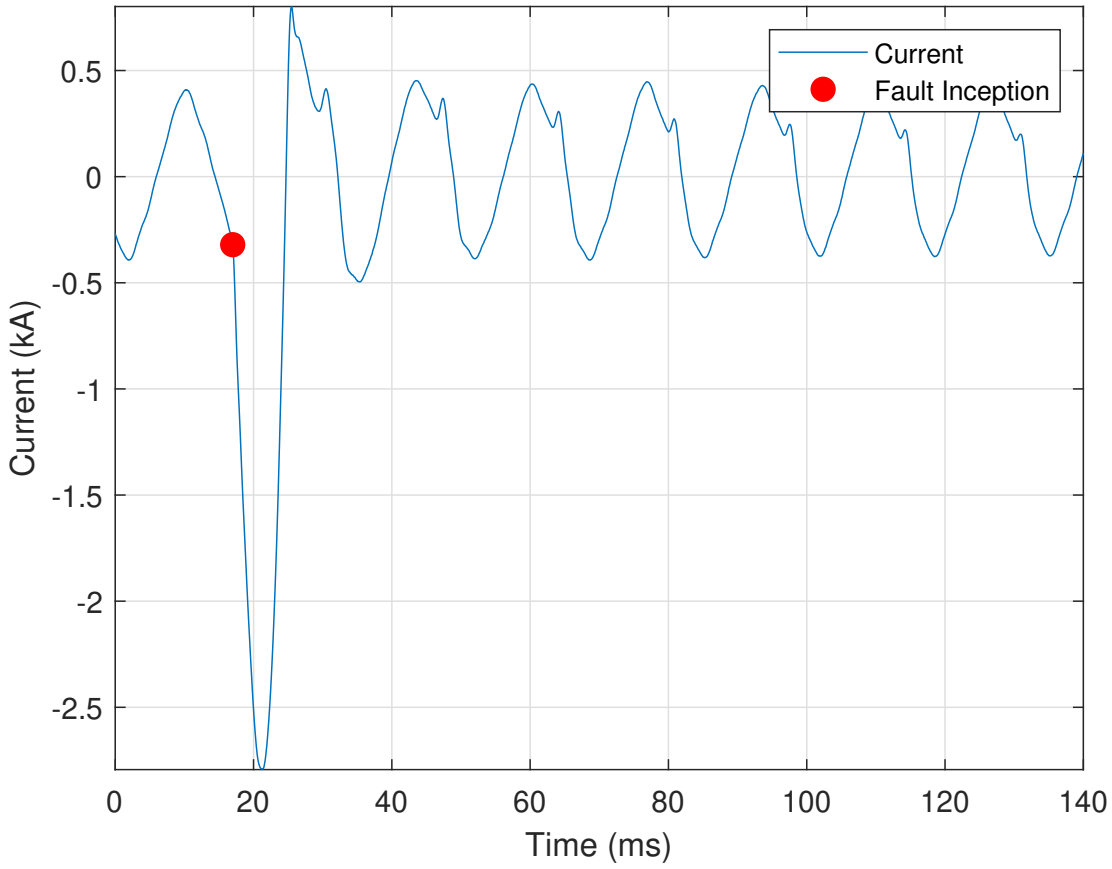


Figure 2.1 Fuse fault showing second derivative test [11]

2.1.4 Harmonic Ratios

Harmonics can be key indicators of particular disturbances within a transmission system (e.g., current transformer saturation, harmonic resonance, etc.). Harmonic analysis is facilitated through the calculation of the harmonic ratio, which is useful in determining the dominant frequency components within a waveform. The n^{th} harmonic ratio is calculated by,

$$H_n = \frac{|X_n|}{|X_1|}, \tag{2.5}$$

where X is the FFT of x, $|X_1|$ is the magnitude of the fundamental frequency (a.k.a., 60 Hz), and $|X_n|$ is the magnitude of the n^{th} multiple of the fundamental frequency [12].

2.1.5 First Cycle Comparison

A useful disturbance detection approach is to compare the waveform's first cycle with each of its remaining cycles within the disturbance record. After the first cycle is selected, it is replicated to construct an ideal waveform that is of the same length as that of the recorded waveform from which the first cycle was extracted. The generated ideal waveform is then subtracted from the recorded waveform. The time indices where this difference is very high indicate the start of a disturbance. Fig. 2.2 illustrates the application of this approach in detecting the start of a capacitor switching disturbance within a recorded voltage waveform. Fig. 2.2 shows the voltage waveform with the capacitor switching disturbance portion of the waveform highlighted and the result of the difference calculation overlaid. Where the difference calculation is highest corresponds with the

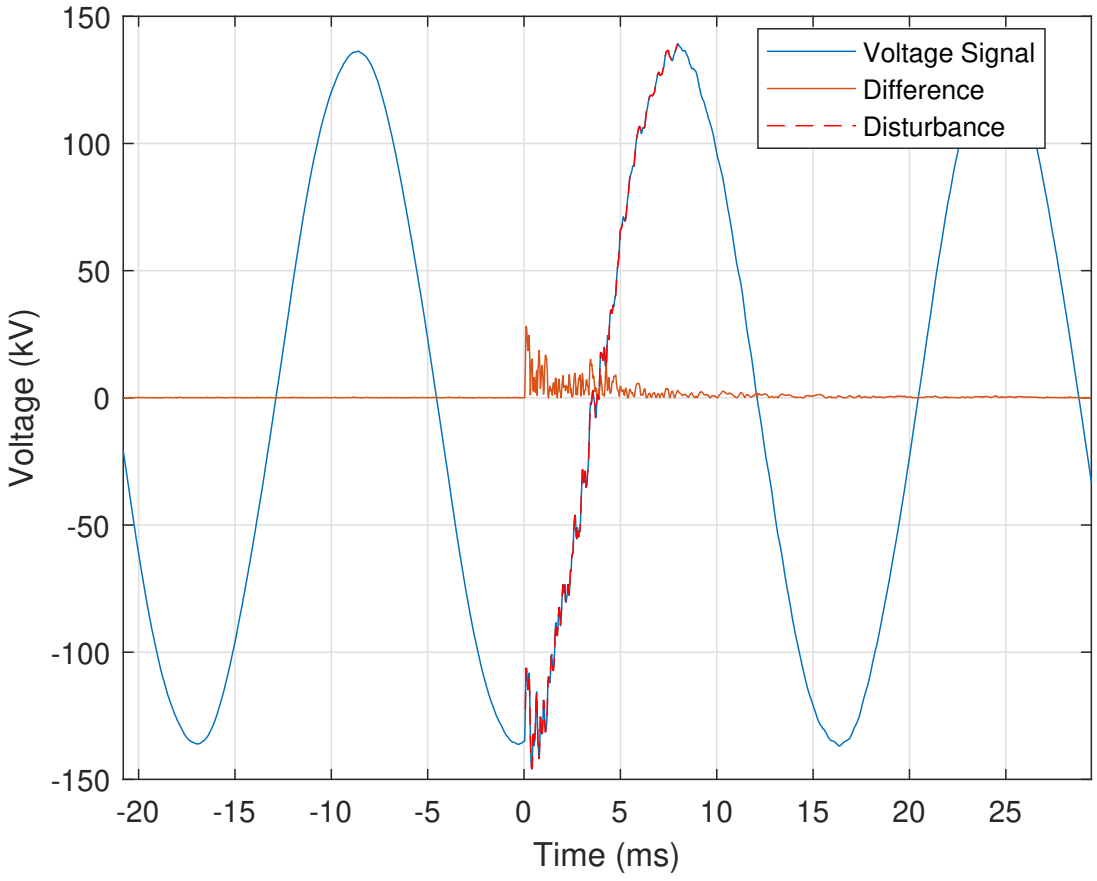


Figure 2.2 Voltage waveform showing disturbance during capacitor switching [11]

start of the capacitor switching disturbance, which is assigned a start time of zero milliseconds.

2.2 Voltage Unbalance

Voltage unbalance is a measure of the asymmetry between the voltages of a three-phase power system [13]. When power is generated, the resulting voltages–on each of the three phases–are equal in magnitude and 120° apart in phase angle [14]. However, operational system voltages can become unbalanced with respect to magnitude, phase angle, or harmonic distortion levels. One major driving factor of voltage unbalance is the presence of single-phase loads, which draw power from one phase and not the others. An unequal allocation of these loads will lead to greater unbalance. The fact that many distribution lines are single-phase is a contributor to these uneven loads. Another voltage unbalance cause is when transmission lines are not completely transposed [13]. Voltage unbalance has a negative effect on the power system, which will incur more losses and heating effects under unbalanced conditions [14]. This is due to the system not being in a state to respond to emergency load transfers. The unbalance can also have negative effects on large commercial or industrial customers operating large equipment. The effects are particularly severe on induction motors, power electronic converters, and adjustable speed drives. The next section explains how voltage unbalance is measured for the purposes of this thesis since several methods exist. The general design of the ANN used in this work is then introduced in the following section.

2.2.1 Quantifying Voltage Unbalance

In order to predict voltage unbalance, it must first be quantified. The International Electrotechnical Commission (IEC) quantifies voltage unbalance using a metric known as the Voltage Unbalance Factor (VUF) [15]. The equation for the VUF percentage is given by,

$$u_2 = \frac{|\boldsymbol{V}_2|}{|\boldsymbol{V}_1|} \times 100\%, \tag{2.6}$$

where u_2 is the percent VUF, $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are the positive and negative sequence voltages, respectively. The calculation for $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ has its basis in the theory of symmetrical components. The

Fortescue transform translates voltages from the phase domain to the sequence domain by making use of the $\boldsymbol{A}$ matrix [13], which is given by,

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \boldsymbol{a} & \boldsymbol{a}^2 \\ 1 & \boldsymbol{a}^2 & \boldsymbol{a} \end{bmatrix}, \tag{2.7}$$

where

$$\boldsymbol{a} = -\frac{1}{2} + j\frac{\sqrt{3}}{2}. \tag{2.8}$$

The sequence voltages are then calculated by,

$$\begin{bmatrix} \boldsymbol{V}_0 \\ \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \end{bmatrix} = \Big[\boldsymbol{A}\Big]^{-1} \begin{bmatrix} \boldsymbol{V}_{ab} \\ \boldsymbol{V}_{bc} \\ \boldsymbol{V}_{ca} \end{bmatrix}, \tag{2.9}$$

where $\boldsymbol{V}_0$, $\boldsymbol{V}_1$, and $\boldsymbol{V}_2$ are the zero, positive, and negative sequence voltages, respectively, and $\boldsymbol{V}_{ab}$, $\boldsymbol{V}_{bc}$, and $\boldsymbol{V}_{ca}$ are the three phase-to-phase voltage phasors, respectively. The issue with the Fortescue transform is its reliance on knowing voltage magnitude and phase to calculate the sequence voltages, and phase angle is generally not available in SCADA data. Thus, in this thesis, an equivalent calculation is used in lieu of Equation (2.6). This alternate calculation is defined in [15] and defined as,

$$u_2 = \sqrt{\frac{1-\sqrt{3-6\beta}}{1+\sqrt{3-6\beta}}} \times 100\%, \tag{2.10}$$

where

$$\beta = \frac{|V_{ab}|^4 + |V_{bc}|^4 + |V_{ca}|^4}{(|V_{ab}|^2 + |V_{bc}|^2 + |V_{ca}|^2)^2}. \tag{2.11}$$

Equation (2.10) removes the VUF calculation's phase angle component and allows the use of SCADA-measured phase-to-phase quantities.

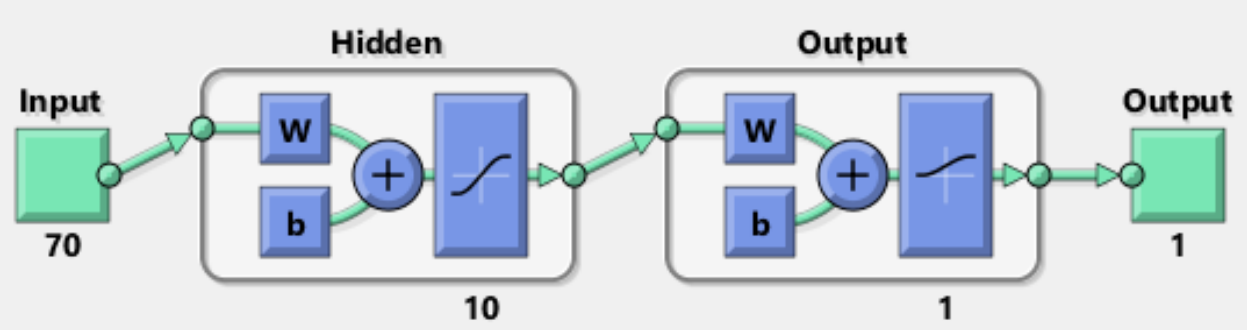


Figure 2.3 Representative diagram of an Artificial Neural Network [18]

2.2.2 Artificial Neural Network Design

An ANN performs pattern recognition by mimicking the neurons and synapses of the human brain using a collection of interconnected nodes (a.k.a., artificial neurons). Similar to the human brain, trained ANNs can recognize complex nonlinear input-output relationships [17]. The ANNs used in this thesis are feed-forward networks–the most popular neural network architecture–trained using supervised learning [18]. In supervised learning, labeled input data is used to train the ANN so it learns the non-linear patterns and relationships between the inputs and a desired output or set of outputs. In this thesis, the labeled input data is the SCADA-collected MW, Mvar, or MW and Mvar values along with their voltage unbalance status (a.k.a., labels), and the desired output is the prediction of voltage unbalance or the lack thereof. All inputs corresponding to voltage unbalance are assigned a label of '1' and all others are assigned a label of '0'.

The block diagram in Fig. 2.3 shows the adopted ANN architecture. Seventy data vectors are input into the ANN that is constructed with ten hidden feature-extracting layers, one output layer, and an output vector that contains the two possible class predictions of voltage unbalance or not [17].

CHAPTER 3
WAVEFORM SIGNATURE ANALYSIS

3.1 Methodology

This section describes the methodologies developed and employed for the categorization of specific electrical disturbances based on waveform signature.

3.1.1 Current Transformer Saturation

The first analyzed electrical disturbance is Current Transformer (CT) saturation. A CT is commonly used in relaying or metering applications in high-voltage (HV) circuits by producing an alternating current in its secondary winding that is proportional to the current it is measuring on the HV system. These low-voltage, low magnitude currents are then used as inputs to various instrumentation [19]. CT saturation occurs when the primary current is so high that its core cannot handle any more flux. This results in inaccurate replication of the current waveform on the secondary winding, which can cause protection relays to operate improperly. A key indicator of CT saturation is a change of slope as the current crosses zero each half-cycle. This change in slope is commonly referred to as "kneeing". Fig. 3.1 shows a representative illustration of "kneeing"–between 280 ms and 320 ms–within a CT's current waveform.

The following criteria are used to determine if the electrical disturbance is CT saturation. These criteria are: (i) current exceeding **fifteen** times the continuous current rating of the CT, (ii) presence of DC offset, (iii) the DC offset returning to normal (a.k.a., 0 Hz) during the fault, (iii) inconsistent spacing between zero crossings, (iv) the current waveform's third derivative is high, (v) high second harmonic current, and (vi) high third harmonic within the current. A mix of these criteria determines the likelihood of CT saturation as described at the end of this section.

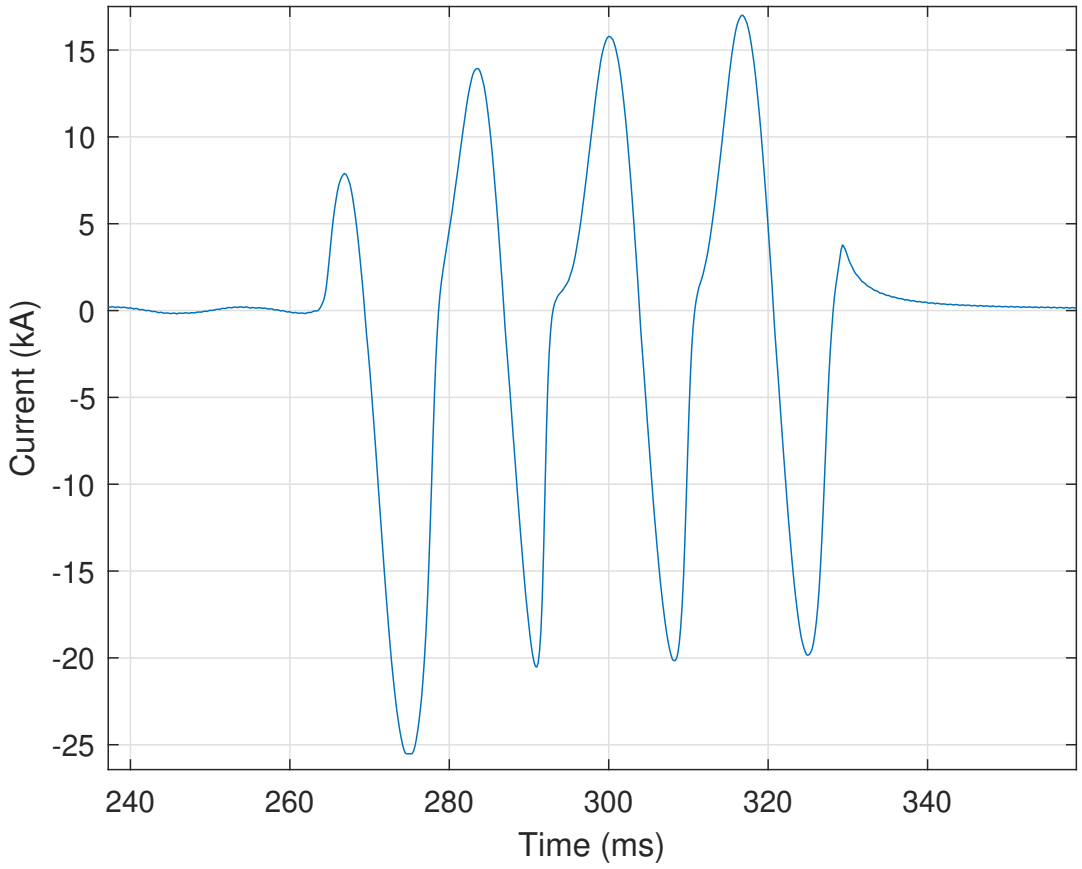


Figure 3.1 A representative illustration of "kneeing" within a current waveform during a CT saturation disturbance [11]

The first step is to determine the presence of a fault or not. Processing continues if a fault is detected and moves to the next disturbance otherwise. For this thesis, a fault means that an abnormal flow of current has occurred causing the protective relay(s) to operate and trip the breaker(s). The presence of a fault is determined using the CT ratio defined in the COMTRADE [20] configuration file. The CT ratio is,

$$R_{\text{CT}} = \frac{I_P}{I_S}, \tag{3.1}$$

where R_{CT} is the turns ratio of the CT, I_P is the rated continuous primary current, and I_S is the rated continuous secondary current. All investigated CTs used a continuous rated current of 5 Amperes (A) on the secondary side of the CT. For instance, if the CT ratio $R_{\text{CT}} = 240$, then the rated

continuous current would be 1,200 A on the primary side and 5 A on the secondary side.

If the current exceeds **fifteen** times the continuous current rating of the CT, then a detected fault is high enough to be CT saturation. This threshold was selected based upon recommendations of PQ engineers to ensure only abnormally high faults are selected since extremely high currents are generally indicative of CT saturation. Faults that do not meet this threshold will have a lower chance of being CT saturation. The threshold for CT saturation is given by,

$$\frac{I(n)}{I_P} > \tau_{\text{CT}}, \tag{3.2}$$

where I is the instantaneous current being analyzed, I_P is the rating of the CT on the primary side, $\tau_{\text{CT}} = 15$ is the CT saturation threshold, and $n = 1, 2, \ldots, N$. The CT saturation threshold was set based on inputs from power utility personnel but can be changed based on local criteria. If the current waveform is clipped at the peaks, the CT saturation threshold is lowered to **fourteen** times the continuous current rating of the CT. When the waveform is clipped, the CT does not accurately replicate the full magnitude of the current, so the saturation threshold is lowered to account for this loss of data. This threshold is also able to be customized based on the needs of the utility.

CT saturation is also indicated by the presence of DC offset [19]. For this particular disturbance, DC offset is determined by first calculating the peak value of each cycle of the faulted section of the waveform. The peaks of the positive and negative half-cycles are then averaged together to give a value for the offset above or below 0 A. If the maximum of this value exceeds a threshold compared to the nominal peak current, then DC offset is detected in the fault as given by,

$$\frac{|I_{\text{DC}}|}{\hat{I}_q} > \tau_{\text{DC}}, \tag{3.3}$$

where I_{DC} is the maximum DC offset detected during the fault, $\hat{I}_q$ is the nominal peak current extracted from the first cycle, and $\tau_{\text{DC}} = 3$ is the empirically selected threshold for the ratio of DC offset magnitude to nominal peak current. A loss of DC offset is detected if the offset magnitude is lower at the end of the fault than at the beginning.

The number of samples between zero crossings is then compared to half the nominal number of samples in each cycle calculated using Equation (2.2) as described in Sect. 2.1.1. The zero crossing points are calculated as the indices at which the waveform changes sign (i.e., goes from negative to positive or vice versa). The number of samples between each zero crossing is calculated for every cycle by subtracting the indices accordingly. The actual number of samples between each zero crossing is compared to the nominal number (i.e., half the samples per cycle) and checked against a threshold as given by,

$$\max \left| N_{\mathrm{Z}}(k) - \frac{N_c}{2} \right| > \tau_{\mathrm{Z}},\ k = (1,2,3,\ldots,N_{\mathrm{F}}) \tag{3.4}$$

where N_{Z} is the number of samples between zero crossings, N_c is the nominal number of samples in each cycle, k is the index of each cycle, N_{F} is the total number of cycles in the faulted portion of the waveform, and $\tau_{\mathrm{Z}} = 10$ is the empirically selected threshold for the difference from nominal in the number of zero crossings.

The "kneeing" present in the waveform is detected using a third derivative test. The maximum third derivative present in the first cycle of the waveform (i.e., before the fault) is used as the nominal value. The maximum third derivative of the faulted portion of the waveform is compared to the nominal value and will be "flagged" if it exceeds a certain threshold as given by,

$$\frac{\max |I_f'''(n)|}{\max |I_c'''(n)|} > \tau_{\mathrm{D3}} \tag{3.5}$$

where $I_f'''(n)$ is the third derivative of the faulted current waveform, $I_c'''(n)$ is the third derivative of the first cycle of the current waveform, and $\tau_{\mathrm{D3}} = 5$ is the empirically selected threshold for the fault's third derivative ratio with the nominal third derivative value.

Finally, the harmonic ratios of the entire current waveform are calculated using Equation (2.5) as described in Sect. 2.1.4. A very good indicator of CT saturation is when the second and third harmonic currents exceed the thresholds of **15%** and **5%** of the fundamental frequency, respectively.

All these criteria are combined to give a confidence level for CT saturation. The confidence levels are defines as follows.

- *High confidence:* The thresholds are exceeded for the current rating of the CT and the second harmonic current. The thresholds must also be exceeded for *three* of the following: DC offset, loss of DC offset, inconsistent spacing between zero crossings, third derivative, or third harmonic current.

- *Medium confidence:* The threshold is exceeded for the current rating of the CT, but the second harmonic threshold is not exceeded. The thresholds must then be exceeded for *three* of the following: DC offset, loss of DC offset, inconsistent spacing between zero crossings, third derivative, or third harmonic current.

- *Low confidence:* The threshold is exceeded for the current rating of the CT, but the second harmonic threshold is not exceeded. The thresholds must then be exceeded for *two* of the following: DC offset, loss of DC offset, inconsistent spacing between zero crossings, third derivative, or third harmonic current.

- *Low confidence (alternative):* The threshold is not exceeded for the current rating of the CT but is for the second and third harmonics. The thresholds must then be exceeded for *two* of the following: DC offset, loss of DC offset, inconsistent spacing between zero crossings, or third derivative.

3.1.2 Analog-to-Digital Converter Clipping

An analog-to-digital (A/D) converter is a device that converts continuously varying analog waveforms into a binary or digitized sequence. Many substation electronic devices (e.g., relays and DFRs) utilize A/D converters to record voltage and current waveforms in a binary format. The range of the digitized scale is restricted by the power supply rail voltage. If the analog value results in a digitized sequence that exceeds the rail voltage, then the digitized sequence will appear "clipped" or "flat-topped" at its minimum and maximum values. For substation devices, clipping

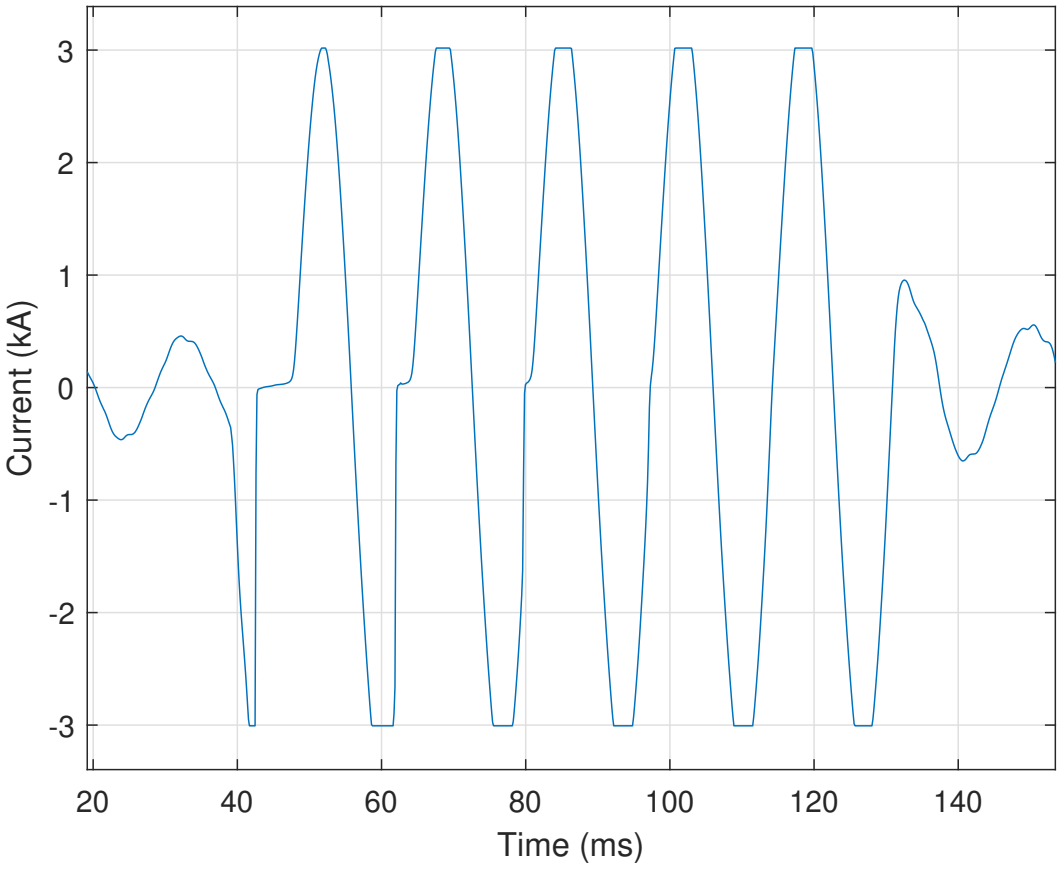


Figure 3.2 A representative current waveform showing Analog-to-Digital Converter (A/D) clipping [11]

often appears in current waveforms during fault disturbances. This results in inaccurate replication of the current waveforms, which can result in relaying misoperation. Fig. 3.2 shows the visible clipping at the minimum and maximum values of a current waveform's digitized sequence.

Clipping is indicated by the repetition of equal-magnitude samples within the digitized sequence. First, the index of the absolute maximum of the waveform is calculated. **Ten** samples before and **ten** samples after the maximum are then extracted from the waveform for analysis. If the first derivative of this extracted waveform portion is equal to zero for more than **four** consecutive samples, then A/D converter clipping is present within the waveform.

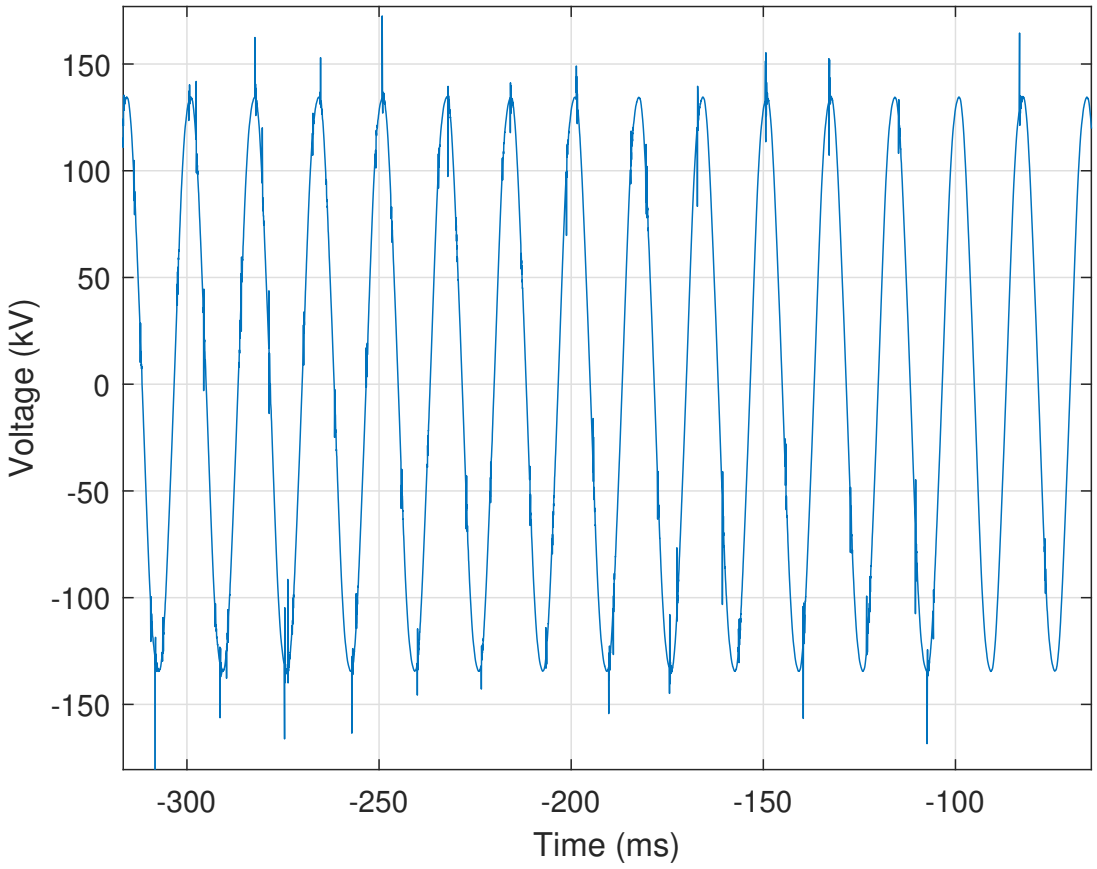


Figure 3.3 A representative voltage waveform showing transient noise due to switching [11]

3.1.3 Induced Transient Noise due to Switching

When high voltage devices–such as air-break switches–are opened to de-energize a bus section, the resulting arcing can induce high-frequency noise upon the voltage or current waveforms of the electronic monitoring equipment (e.g., a PQ monitor). Identification of this induced transient noise is used to determine where waveform chokes may need to be installed or where shielding and ground bonding integrity may need to be checked. Fig. 3.3 provides a representative illustration of this transient noise within a voltage waveform.

This disturbance is characterized by the presence of small random spikes (a.k.a., noise) throughout one or more voltage or current waveforms. Switching-induced transient noise is identified by its: (i) overall difference from an ideal waveform, (ii) harmonic content below **5%** of the

fundamental frequency, (iii) sudden spikes determined by the first derivative exceeding **10%** of the nominal peak value, (iv) persistence over **five** cycles or more, (v) occurrence averaging **once** per cycle, (vi) instances totaling **twenty** or more, and (vii) presence causing individual sample values to exceed the nominal peak waveform value occurring at least **five** times.

The first criterion is determined using the approach described in Sect. 2.1.5 in which a voltage waveform is compared to a reference waveform, which is made up of replications of the first cycle. The condition in which the difference between the actual voltage and the reference voltage exceeds a threshold is given by,

$$\frac{\bar{V}_{\Delta}}{N} > \tau_{\mathrm{N}} \tag{3.6}$$

where $\bar{V}_{\Delta}$ is the mean value of the voltage difference between the actual and ideal waveforms, N is the total number of samples in the waveform, and $\tau_{\mathrm{N}} = 30$ is the empirically chosen threshold for this ratio. If the first six criteria are met, then induced transient switching is classified with *medium* confidence. If all seven criteria are met, then induced transient switching is classified with *high* confidence.

3.1.4 High-Speed Reclosing with Tapped Motor Loads

A common practice is to employ high-speed instantaneous reclosing on faulted transmission lines. Sometimes there may be large or significant motor load served from stations tapped on the line. For this thesis, a motor load is considered significant if it is directly served from a HV transmission line (e.g., 161 kV). In such cases, the line voltage may be supported by the motors–as they spin down–so that residual voltage remains on the line by the time a high-speed breaker recloses. The residual voltage may require up to five seconds to decay in large machines [21]. Since this residual voltage is unlikely to be in phase with the system voltage, the result can be a failed reclose attempt by the line breakers as well as damage to the motors. Thus, it is important to identify lines where high-speed reclosing needs to be delayed to allow the voltage to sufficiently decay before carrying out the reclosing operation. Fig. 3.4 shows a voltage waveform in which sufficient time has passed to allow the voltage waveform to decay to a point after which the reclosing

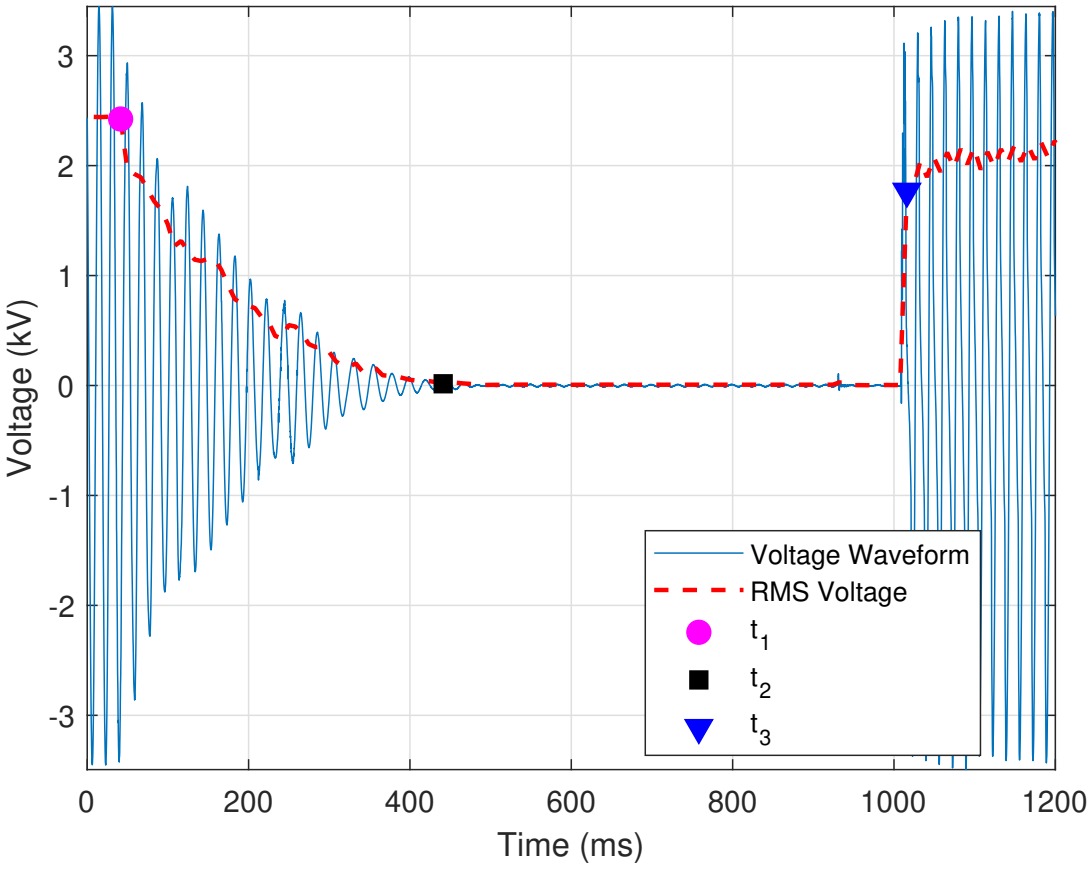


Figure 3.4 A representation of the case in which the voltage waveform *does* decay sufficiently prior to a successful reclosing operation in the presence of a tapped motor load [11]

operation was successfully completed.

For identification of this disturbance, it must be determined whether the reclosing operation is a high-speed reclosing operation. For this thesis, the reclosing operation is a high-speed one if it is "blind" (i.e., without any supervision or checks) and occurs within thirty cycles of the initial current interruption by the breaker [21]. Identification of the reclosing with tapped motor loads disturbance is achieved by determining the sample points at which the: (i) voltage waveform begins to decay, (ii) voltage waveform reaches zero, and (iii) reclosing operation occurred. The time between these three points determines whether the reclosing is a high-speed operation. In this thesis and as shown in Fig. 3.4, these three sample points are designated as t_1 (magenta circle),

t_2 (black square), and t_3 (blue triangle), respectively. The location of these three sample points is determined using the RMS waveform, which is calculated using Equation (2.3) as defined in Sect. 2.1.2 and is shown in Fig. 3.4 as a broken, red line. For this disturbance, the RMS window is set to half the number of samples in each cycle (a.k.a., $N_c/2$).

The point t_1 is the time at which the RMS voltage first decays below a threshold and is determined by,

$$\frac{\bar{V}(t)}{\bar{V}_q(t)} < \tau_{\mathrm{S}}, \tag{3.7}$$

where $\bar{V}$ is the RMS of the voltage, $\bar{V}_q$ is the nominal RMS voltage as determined from the first cycle, and $\tau_{\mathrm{S}} = 0.9$ is the empirically selected threshold for the voltage sag indicating the start of the decay. The point t_2 is determined as the time at which the voltage decays low enough to be considered approximately zero. An empirical threshold of $\tau_0 = 0.01$ was used as the threshold below which the RMS voltage must reach to be considered zero. If this condition is not met, then t_2 is the time at which the RMS voltage is at its minimum. For the process to continue, the RMS voltage must decay to below **50%** of the nominal value, which is calculated using the method described in Sect. 2.1.1.

The voltage decay portion is the RMS voltage between times t_1 and t_2 and is designated here as $\bar{V}_{\mathrm{D}}$. The median (a.k.a., middle value) of $\bar{V}_{\mathrm{D}}$ must be lower in magnitude than the voltage at time t_1 and higher than the voltage at time t_2. The mean first derivative of $\bar{V}_{\mathrm{D}}$ must also be negative to indicate decreasing slope or voltage. The maximum of the voltage decay's first derivative must also be less than a threshold to ensure that the voltage decay was not sudden. This condition is given by,

$$\frac{\max|\bar{V}'_{\mathrm{D}}|}{\bar{V}_q} < \tau_l \tag{3.8}$$

where $\bar{V}'_{\mathrm{D}}$ is the first derivative of the decaying portion of the RMS voltage, $\bar{V}_q$ is the nominal RMS voltage, and $\tau_l = 0.5$ is the empirically selected threshold for the maximum first derivative of the decaying voltage. The point t_3 is the time at which the RMS voltage increases by **30%** of nominal value in one RMS sample. This condition is determined by the first derivative of the RMS

waveform as given by,

$$\frac{\max|\bar{V}_S'|}{\bar{V}_q} > \tau_U \tag{3.9}$$

where $\bar{V}_S'$ is the first derivative of the portion of the RMS voltage after time t_2, $\bar{V}_q$ is the nominal RMS voltage, and $\tau_S = 0.3$ is the empirically selected threshold for the minimum first derivative of the reclosing voltage. Time t_3 is the point when reclosing occurs and the voltage is restored.

The criteria given thus far serve to classify the disturbance as normal reclosing with a tapped motor load. Fig. 3.4 is a normal disturbance in which there was sufficient time between t_2 and t_3. If there is not sufficient time between these two points, then the disturbance is "flagged" as

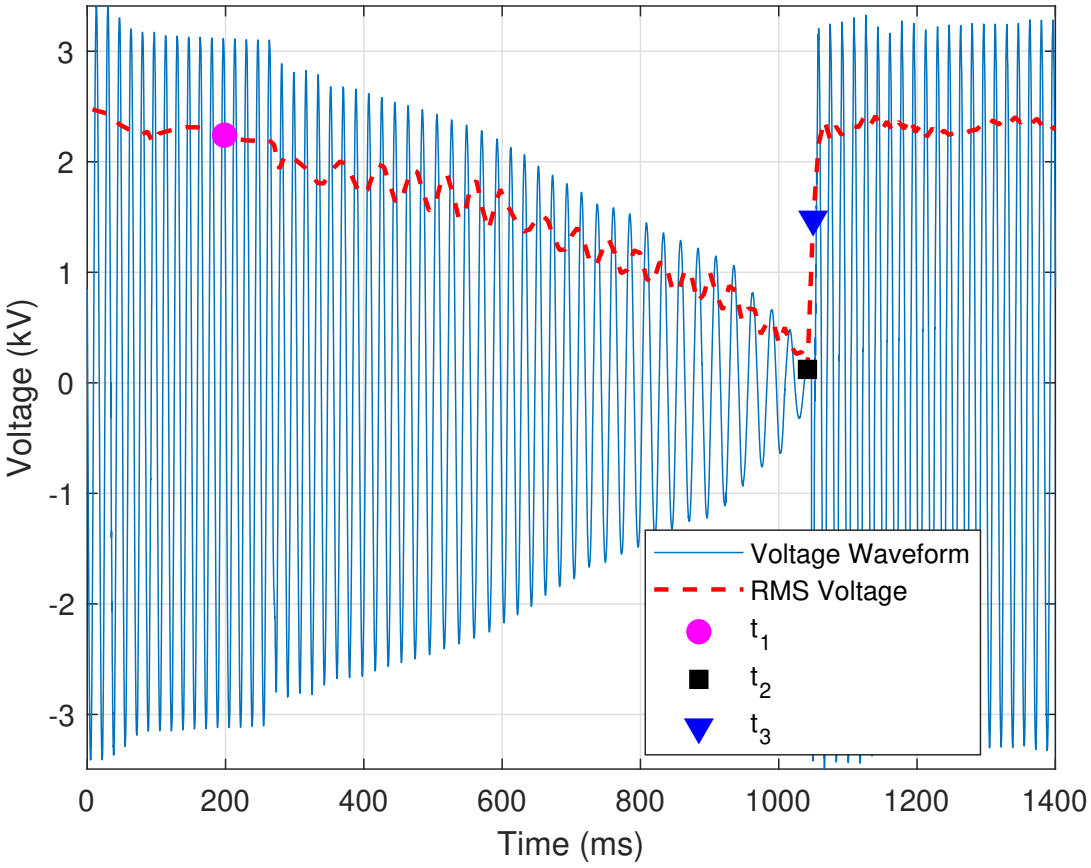


Figure 3.5 A representation of the case in which the voltage waveform *does not* decay sufficiently prior to a successful reclosing operation in the presence of a tapped motor load [11]

needing attention. The condition that defines a high-speed reclosing operation is given by,

$$t_3 - t_2 > \tau_{\text{HS}} \tag{3.10}$$

where t_2 is the time at which the voltage first decays to zero, t_3 is the time at which the voltage is restored, and $\tau_{\text{HS}} = 30$ cycles is the threshold for the minimum time the voltage must be zero before reclosing as recommended [21]. Fig. 3.5 shows a case in which the minimum time for which the voltages need to be zero is not satisfied.

3.1.5 DC Offset

DC offsets are a common occurrence in analog channels. If a DC offset is large enough, then it can negatively impact RMS calculations. A large DC offset is accounted for by re-calibration of the corresponding monitoring or recording device. Automated calculation of DC offset allows utility personnel to prioritize re-calibration of those devices associated with the largest amounts of DC offset. The DC offset disturbance is characterized by an asymmetry between the positive and negative half-cycles of a voltage or current waveform.

The presence and amount of DC offset are determined using both time and frequency domain analysis. In the frequency domain, a DC offset is present when the magnitude of the 0 Hz frequency component is greater than 50% of the magnitude at the fundamental frequency component (i.e., 60 Hz in the United States). Mathematically this condition can be expressed as,

$$\frac{X_0}{X_1} > \tau_f \tag{3.11}$$

where X_0 is the magnitude of the 0 Hz frequency component, X_1 is the magnitude at the fundamental frequency component, and $\tau_f = 0.5$ is empirically selected as the minimum ratio with respect to the fundamental frequency. Fig. 3.6 provides a representative illustration of a current waveform in which a large amount of DC offset is present from 40 ms to 90 ms. Fig. 3.7 shows the magnitude of the zeroth through the fifth harmonic of the current waveform shown in Fig. 3.6. In this case,

the 0 Hz frequency component is over two times larger than that of the fundamental frequency component (i.e., the first harmonic) and would be "flagged" as a DC offset disturbance. Interestingly, the presence of the third harmonic indicates that another disturbance is also present within the recorded waveform shown in Fig. 3.6.

If the frequency domain analysis results in the identification of a DC offset disturbance, then time domain analysis is performed as a validation step. Time domain analysis is conducted by computing the mean over each cycle within the recorded waveform. If a given cycle's mean value is zero, then there is no DC offset present within that cycle. This is because the area under the positive and negative portions of the cycle would negate each other. However, if the selected

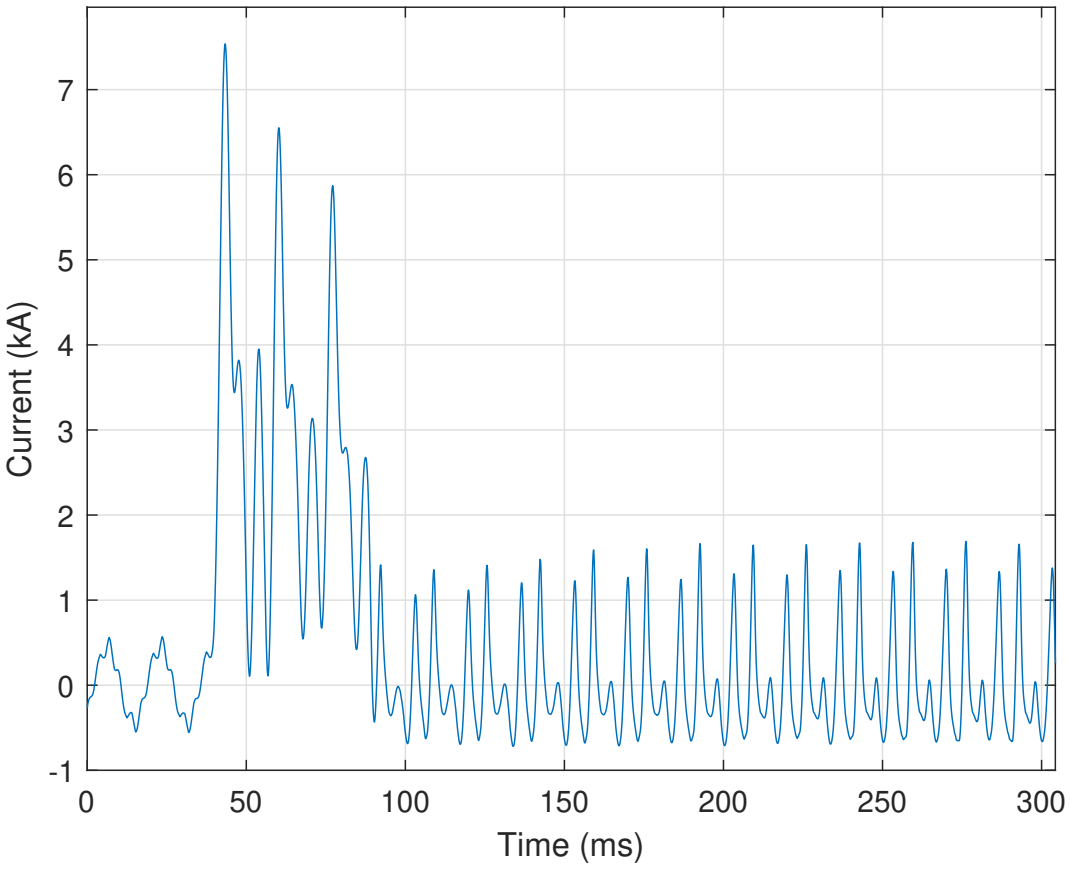


Figure 3.6 Representative illustration of a large DC offset–from 40 ms to 90 ms–within a current waveform [11]

cycle's mean exceeds 50% of the nominal waveform's peak value, then the DC offset disturbance "flag" is set once more. The amount of DC offset returned by the automated process is,

$$\arg\max_i \mu_i, \tag{3.12}$$

where μ_i is the mean value of the i^{th} cycle within the waveform being processed.

3.1.6 Motor Starting

Instantaneous increases in current may be due to faults, motor starts, transformer energizations, or other disturbances. Signatures present within the recorded waveforms can be used to

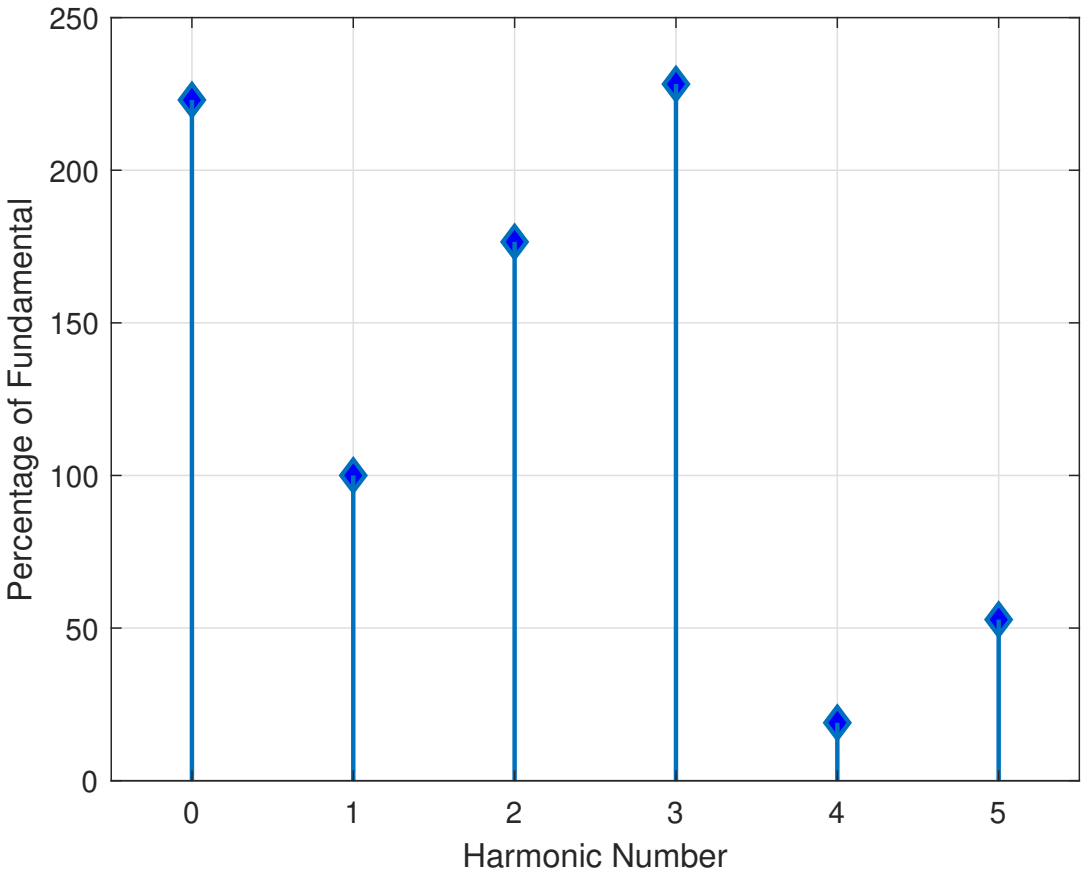
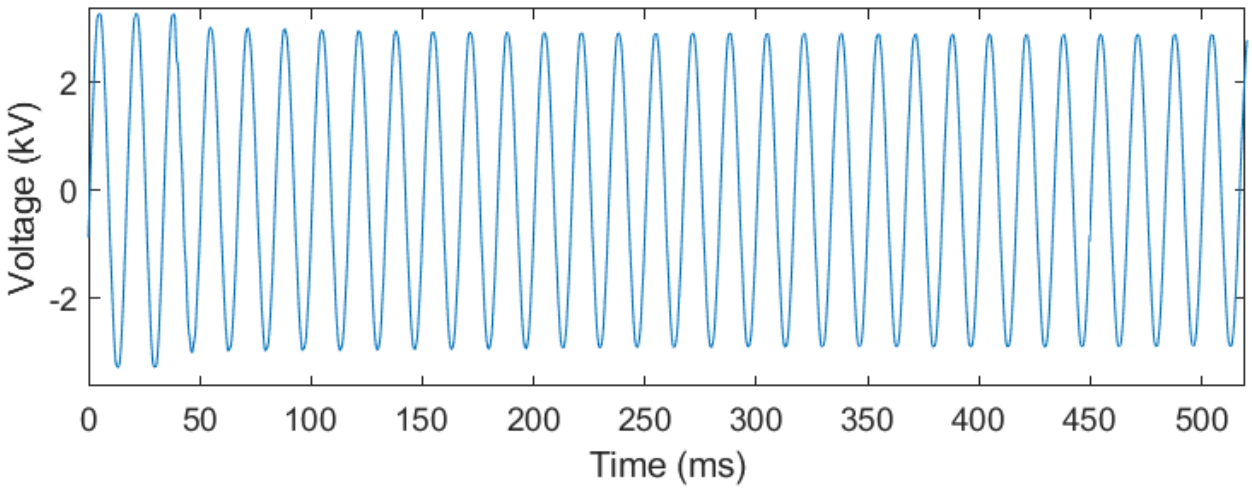
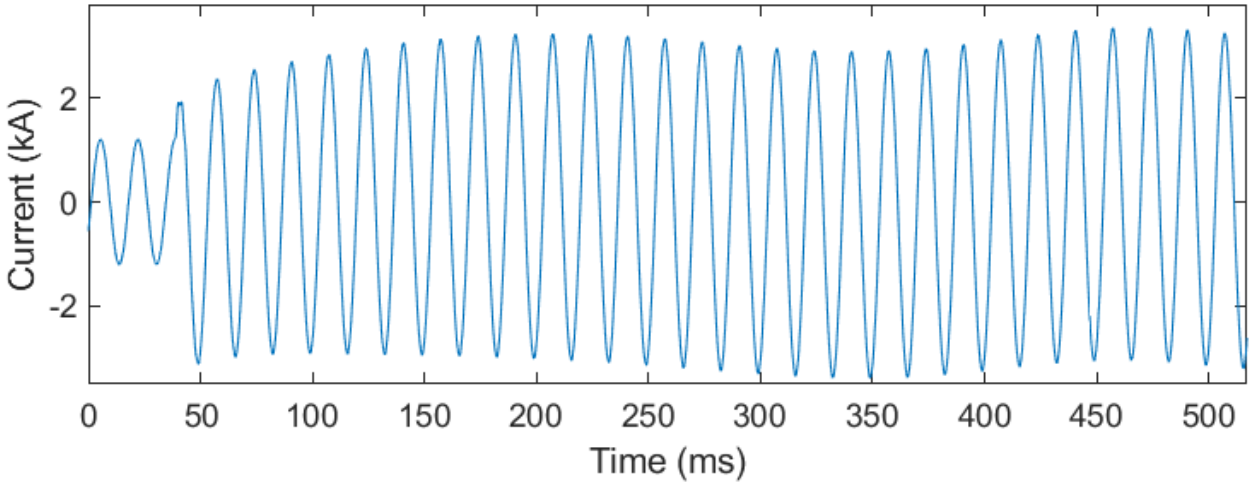


Figure 3.7 An illustration showing the zeroth through the fifth harmonic ratios of the current waveform shown in Fig. 3.6 [11]



(a) Voltage waveform



(b) Current waveform

Figure 3.8 Voltage and current waveforms showing waveform characteristics associated with a motor starting disturbance [11]

distinguish and classify each of these disturbances. PQ disturbances can then be correlated by disturbance classification. In the case of motor starting, the voltage sags and the current can increase to five to six times its rated value [22]. It is challenging to set protective relays in such a way as to enable recognition of a motor starting disturbance rather than recognizing the disturbance as a fault in the system. The automated process described in this section is developed under the assumption that the corresponding relays are properly set so they do not trip open when motor inrush current is present. Fig. 3.8a and Fig. 3.8b show representative illustrations of motor starting voltage and

current waveforms, respectively.

The automated process checks for a voltage sag below 95% of the waveform's nominal RMS value and a current spike to twice the CT's rated value determined by (3.1). If both of these conditions persist for at least ten consecutive cycles, then the first indicator of motor starting is identified. The persistence of both conditions–for ten or more consecutive cycles–distinguishes motor starting disturbances from a fault condition, which typically occurs for only several cycles before the relay trips open the breaker. Motor starting disturbances are also associated with a frequency response that is low in harmonic content. Thus, if none of the voltage or current waveforms' harmonics exceed **15%** of the fundamental frequency component's magnitude, then the second indicator of motor starting is identified. The final indicator for motor starting is that all three conditions (a.k.a., voltage sag, current spike, and harmonics below 15% of the fundamental) occur on all three phases because motors are three phase devices.

3.1.7 Variable Frequency Drive Motor Starting

Some motors utilize electronic starting (e.g., Variable Frequency Drives – VFDs) to bring the motor up to speed in a controlled manner to limit voltage supply disturbance(s). VFDs produce unique harmonic patterns, which allow these disturbances to be easily and automatically identified. When a VFD motor starts, it creates a very distinct current waveform. A representative illustration of this distinct current waveform can be seen in Fig. 3.9.

In Fig. 3.9, each phase has two pulses per half-cycle. The number of pulses per half-cycle indicates the type of VFD (e.g., six-pulse, twelve-pulse, etc.), thus VFD motor starting disturbances are identified by counting the number of times the current waveform drops below **50%** of each cycle's maximum value. Two pulses in each half cycle of a current waveform for each phase (see Fig. 3.9) would indicate a six-pulse VFD. The number of pulses for the drive is given by,

$$N_p = \frac{3}{2} \times \text{mode}(K),\ K > 2 \tag{3.13}$$

where K is the number of times the current crosses 50% of each cycle's maximum value every

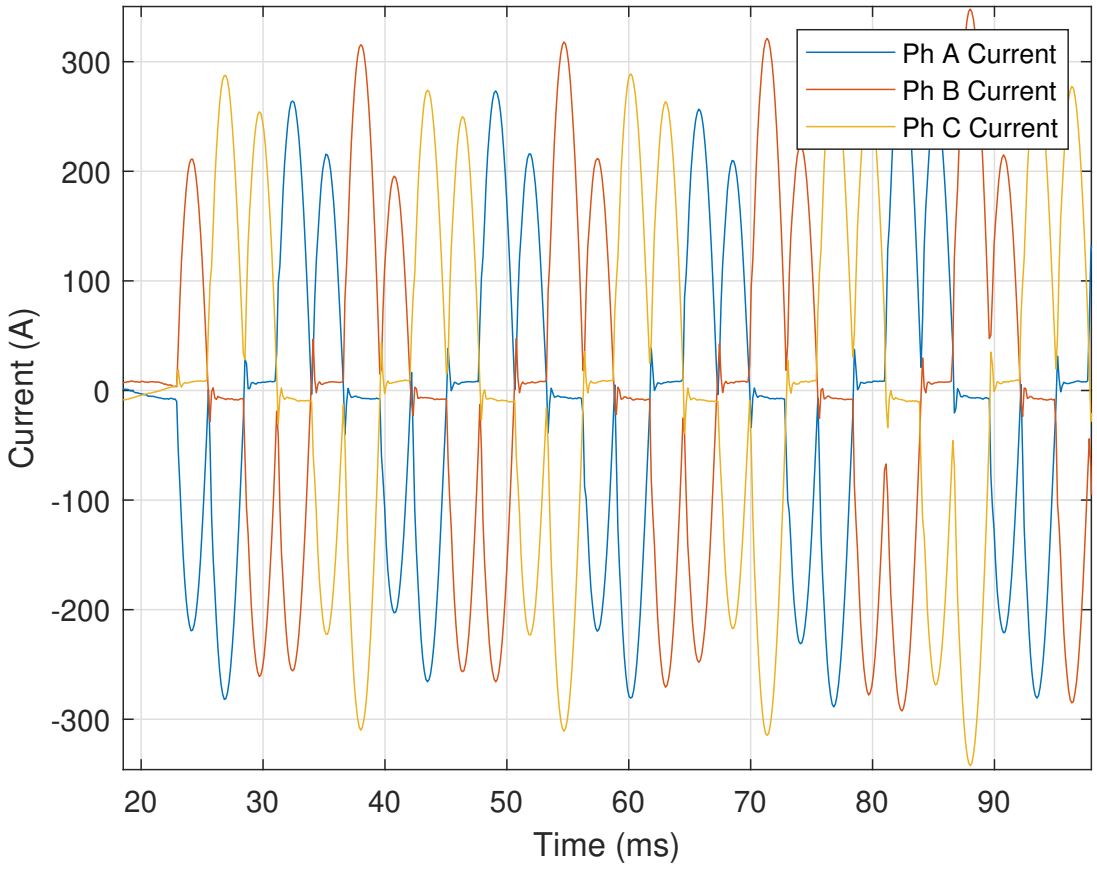


Figure 3.9 An illustration showing the distinct current waveform generated during a six-pulse VFD motor start disturbance [11]

half-cycle, and mode(K) refers to the most often occurring value of K. The current must cross the threshold more than **two** times for at least **eight** cycles during the disturbance to be considered VFD motor starting. After N_p is calculated, harmonic analysis is conducted, because VFD motor starting disturbances result in dominant harmonics on either side of an integer multiple of N_p. Fig. 3.10 shows the harmonics for the six-pulse (a.k.a., $N_p = 6$) VFD motor starting disturbance illustrated in Fig. 3.9. The fifth and seventh harmonics are the two most dominant harmonics and occur on either side of the sixth harmonic, which is equal to that of $N_p = 6$. The value of N_p is validated by ensuring that the dominant harmonics are at least **five** times larger than the value of

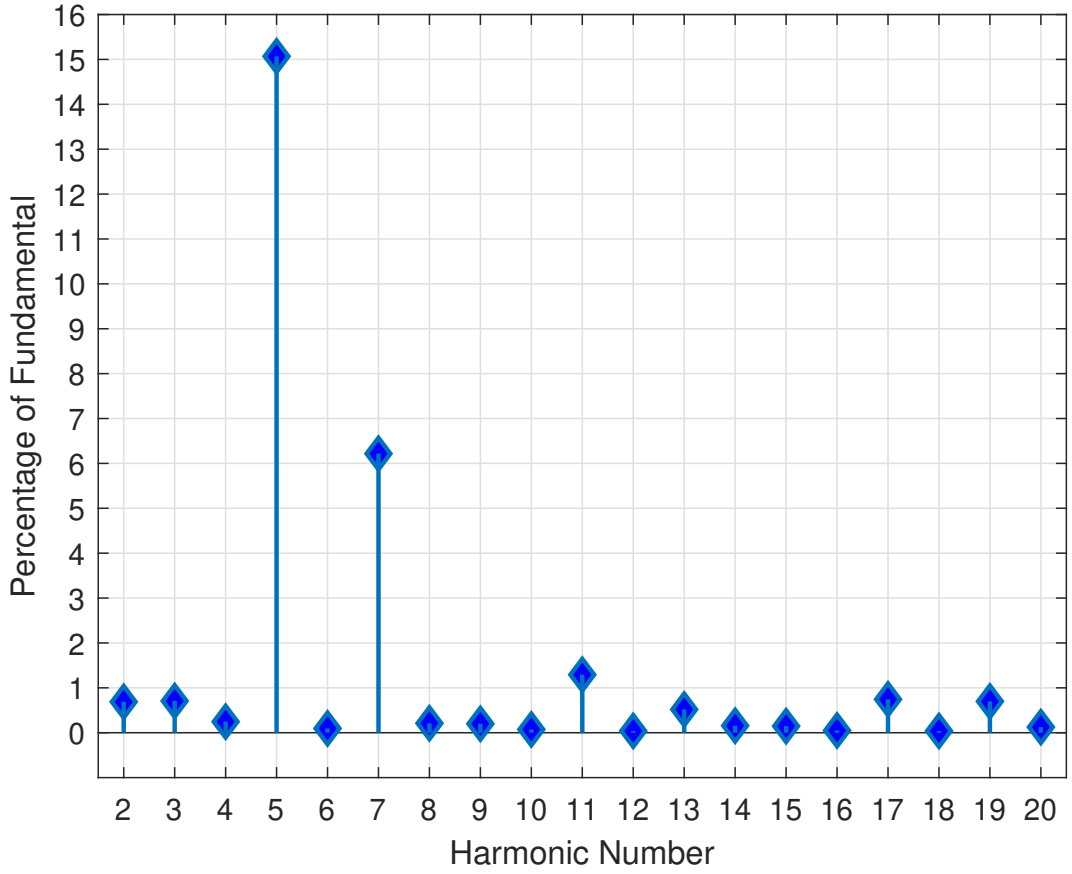


Figure 3.10 The harmonic ratios calculated from the current waveform of the six-pulse VFD motor start disturbance shown in Fig. 3.9 [11]

the harmonics at integer multiples of N_p. This validation check is performed by,

$$\frac{H_{kN_p\pm1}}{H_{kN_p}} > \tau_{\mathrm{V}},\ (k = 1,2,3,4) \tag{3.14}$$

where N_p is the number of pulses in the VFD, H_{kN_p} is the harmonic at an integer multiple of N_p, k is an integer, and $\tau_{\mathrm{V}} = 5$ is the empirically determined threshold for the ratio of the dominant harmonics with those at integer multiples of N_p. If Equation (3.14) is satisfied, then the number of predicted pulses is deemed correct. Finally, the disturbance is identified as a VFD motor starting so long as all three currents (a.k.a., phase A, B, and C) increase over the disturbance's duration.

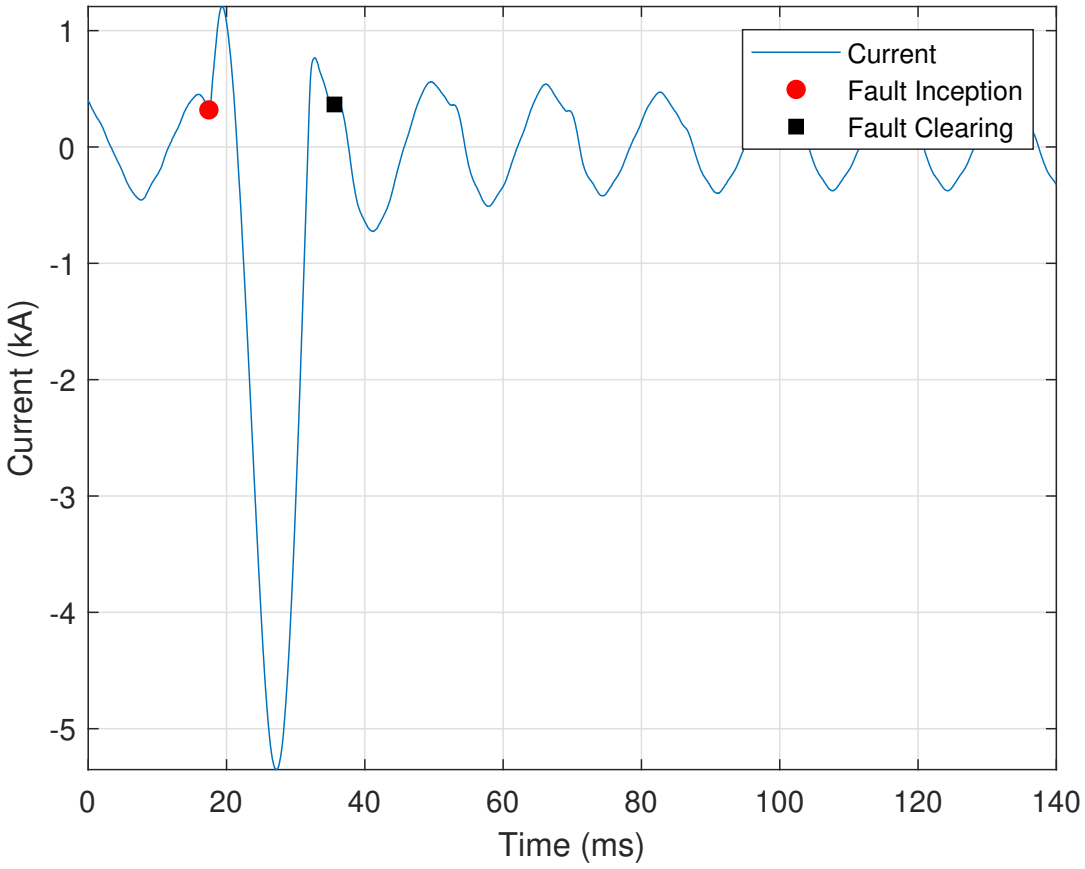


Figure 3.11 A current waveform during a fuse melting disturbance that last just over one cycle [11]

3.1.8 Melted Fuse

Unlike a breaker, a blown (a.k.a., melted) fuse requires utility personnel to physically replace it, so it is helpful to distinguish fuse faults from breaker faults. These two faults are distinguished from one another by the speed at which the fault is cleared. Breakers require between two or more cycles to clear a fault while fuses require less than two cycles. Fig. 3.11 shows an example of a fuse melting disturbance that is cleared in a little more than one cycle.

The key to automated identification of fuse melting disturbances is the accurate determination of the fault's inception and clearing points. A fuse melting disturbance occurs if the total

clearing time was less than one and a half cycles and is determined by,

$$|t_I - t_C| < \tau_c \tag{3.15}$$

where t_I is the inception point, t_C is the clearing point, and $\tau_c = 1.5$ is the threshold for the maximum fuse clearing time.

Automated identification of a fuse melting disturbance is initialized by determining if the disturbance persisted for at least a quarter of a cycle and if the current reaches at least twice its nominal value over the disturbance's duration. The cycle before and just after the portion associated with these two conditions is then analyzed one half-cycle at a time to determine the fault inception and clearing points. The three possible approaches used to determine these points are (i) a change in the sign of the first derivative, (ii) a sudden increase in the second derivative, and (iii) the current waveform's zero crossings.

The first derivative approach is implemented using Equation (2.4) as described in Sect. 2.1.3. A change in the first derivative's sign–before or after the spike in current–indicates the fault inception and clearing points. This approach is used to determine the inception and clearing points of the fuse melting disturbance shown in Fig. 3.11 where the red circle indicates the fault inception point, and the black square indicates the fault clearing point.

If the first derivative approach is unsuccessful (i.e., a change in the first derivative's sign does not exist), then the second derivative is used as described in Sect. 2.1.3. The condition for a large second derivative is given by,

$$\frac{\max |I''(n)|}{\hat{I}_q} > \tau_{\mathrm{D2}} \tag{3.16}$$

where $I''(n)$ is the second derivative of the current waveform, $\bar{x}_c$ is the nominal peak current, and $\tau_{\mathrm{D2}} = 0.02$ is the empirically selected threshold for the minimum ratio of the second derivative of the current to the nominal value. This approach was used to determine the fault inception point of Fig. 2.1 as described in Sect. 2.1.3.

If the second derivative approach is also unsuccessful (i.e., the minimum threshold is not met), then the fault inception and clearing points are assumed to be the zero crossings just before and just after the current spike, respectively. After the fault inception and clearing points are determined, Equation (3.15) is used to determine whether the fault was short enough in duration to be a melted fuse.

3.1.9 Ferroresonance

Ferroresonance is electric circuit resonance that occurs when a circuit containing a nonlinear inductance is fed from a source that has a series capacitance connected to it. In a transmission system, ferroresonance can occur when a breaker with grading capacitors is used to de-energize a bus that has magnetic Voltage Transformers (VTs) connected to it. The described scenario presents a serious safety risk to utility personnel and damage risk to equipment because severe overvoltages can occur despite the breaker being in an open state. Ferroresonance manifests in the voltage waveforms and causes the waveforms to take on a square wave-like shape or appearance. Fig. 3.12 provides a representative illustration of the square wave appearance that a voltage waveform can take on due to ferroresonance. Another characteristic of ferroresonance disturbances is that the current is normally zero during the disturbance. This is due to the line being de-energized; however, depending on the recording device's location, the current can be recorded as a nominal waveform.

Ferrroresonance disturbances are identified using three criteria: (i) a large difference between discrete samples in the voltage waveform, (ii) this behavior continuing for a certain number of cycles and often enough during that time, (iii) a significant harmonic content is present in the voltage waveform, and (iv) the current waveform is recorded as zero or a nominal waveform.

The first criterion is met if the first derivative of the voltage waveform exceeds 50% of nominal peak voltage as given by,

$$\frac{|V'(n)|}{\hat{V}_q} > \tau_{\mathrm{F}} \tag{3.17}$$

where $V'(n)$ is the first derivative of the voltage waveform, $\hat{V}_q$ is the nominal peak voltage, and $\tau_{\mathrm{F}} = 0.5$ is the empirically selected threshold for the minimum ratio of the first derivative of the

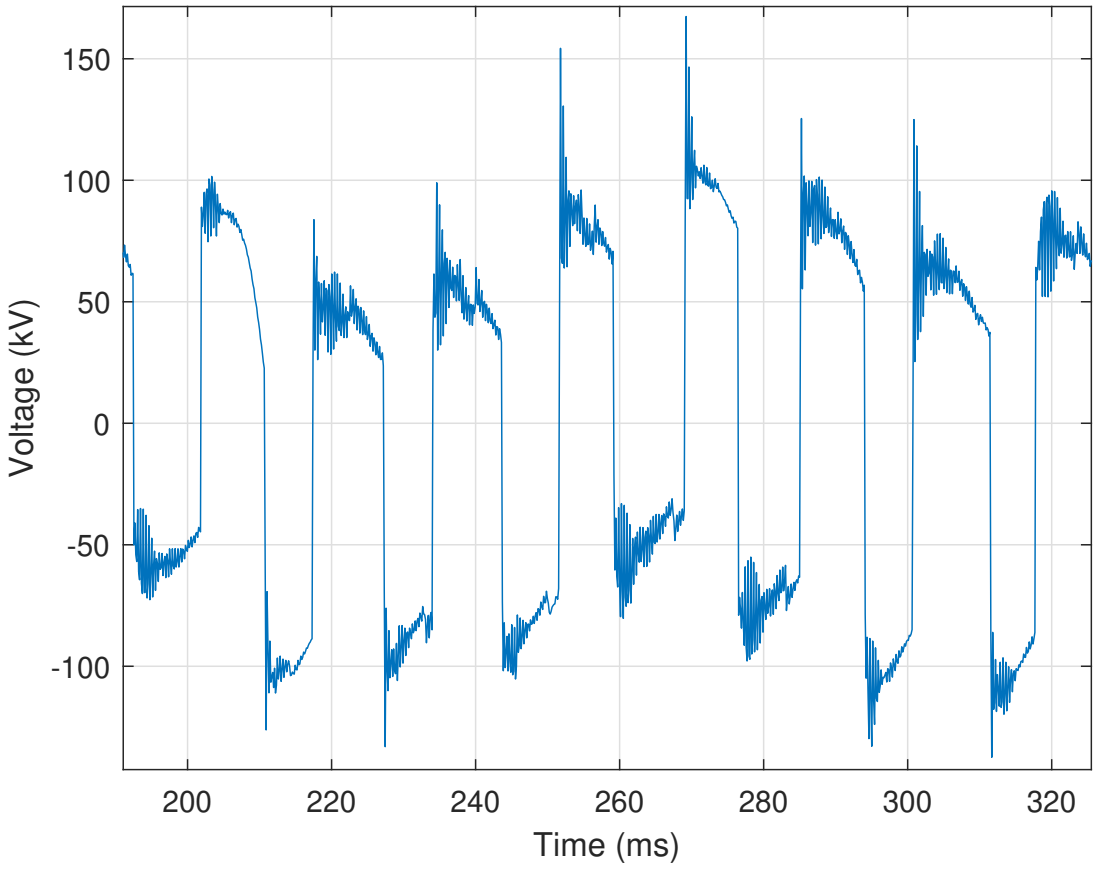


Figure 3.12 Illustration of a voltage waveform collected during a ferroresonance disturbance [11]

voltage to the nominal value. The second criterion is met when this threshold is exceeded at least **five** times, occurs at least every **three** cycles, and occurs for a length of at least **five** cycles. The third criterion is met if one of the harmonic currents is greater than **5%** of the fundamental. Finally, the fourth criterion is met when the RMS current is recorded as zero or the current waveform is nominal, which is characterized by a small number of first derivative sign changes. This nominal condition is given by,

$$\frac{N_{\mathrm{I}}}{N} < \tau_{\mathrm{I}} \tag{3.18}$$

where N_{I} is the number of first derivative sign changes in the current as calculated using Equation (2.4), N is the total number of samples in the waveform, and $\tau_{\mathrm{I}} = 0.3$ is the empirically selected

threshold for the ratio between the number of sign changes and total samples.

3.1.10 Capacitor Bank Switching

One of the most common disturbances on a power system is capacitor bank switching. Capacitor bank switching induces temporary voltage transients that can create PQ disturbances. A typical capacitor bank switching transient is characterized by a quick depression of the voltage waveform toward zero, followed by an overshoot and subsequent transient disturbance–lasting approximately one cycle–as the system returns to steady state. These voltage transients may be recorded by devices that are connected to the same bus as the capacitor bank as well as those connected to a different bus. Based upon this fact, the presented automated process is designed to identify capacitor switching for both recording device connection scenarios. Fig. 3.13 shows an example of capacitor bank switching in which a broken, red line highlights the portion of the recorded waveform associated with the disturbance.

In a power transmission system, capacitor banks are simultaneously switched in on all three phases. Although Fig. 3.13 shows only a single phase, the other two-phase voltage waveforms are similar in appearance, but will not be identical due to the 120° phase difference between each of them. In other words the switching disturbance occurs at different points of the corresponding phase's sinusoidal waveform. The disturbance is located within the waveform using the first cycle as a reference as described in Sect. 2.1.5 and shown in Fig. 2.2. The condition for the difference between the actual and ideal voltage waveforms is given by,

$$\frac{|V_{\Delta}|}{\hat{V}_q} > \tau_{\Delta} \tag{3.19}$$

where V_{Δ} is the difference between the actual and ideal voltage waveforms, $\hat{V}_q$ is the nominal peak voltage value, and $\tau_{\Delta} = 0.02$ is the threshold empirically selected for this ratio. Once the presence and location of the disturbance have been determined, the disturbance's duration is calculated to ensure that it does not exceed **two** cycles. The voltage waveform's peak values must satisfy one of these two criteria: (i) one peak **2%** above the nominal value and no more than one peak **10%** above

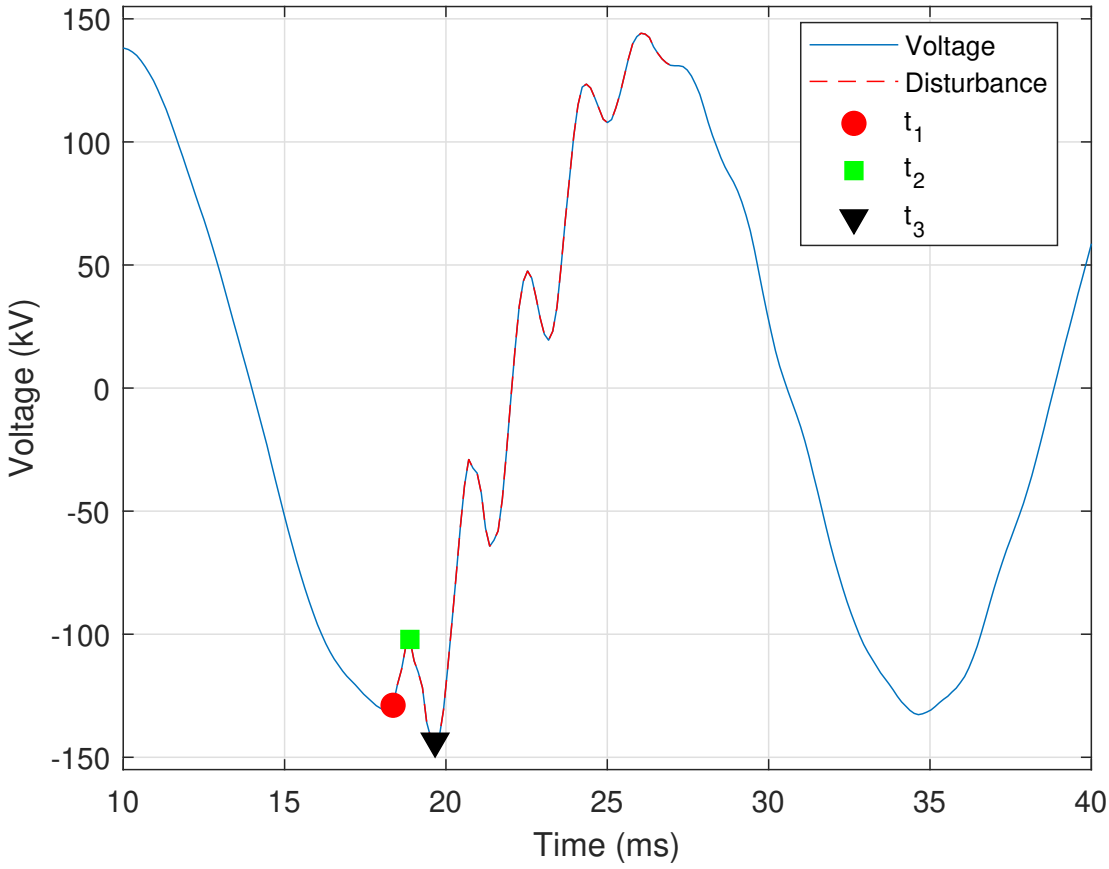


Figure 3.13 Illustration of a voltage waveform collected during capacitor switching disturbance [11]

the nominal value; (ii) exactly two peaks **10%** above the nominal value occurring in neighboring cycles.

The next step is to determine the three characteristic points highlighted on the waveform of Fig. 3.13, which are designated as t_1 (red circle), t_2 (green square), and t_3 (black triangle). These points are indicative of a capacitor-switching disturbance. First, the portion of the voltage waveform one half-cycle before and one half-cycle after the highest peak value is extracted and designated as V_{O}. The point t_1 is determined as the first point in which the voltage waveform's first derivative exceeds the threshold τ_{O}. This condition is expressed as,

$$\frac{|V'_{\mathrm{O}}(n)|}{\hat{V}_q} > \tau_{\mathrm{O}} \tag{3.20}$$

where $V'_{\mathrm{O}}(n)$ is the first derivative of the voltage waveform's overvoltage cycle, $\hat{V}_q$ is the nominal peak voltage value, and $\tau_{\mathrm{O}} = 0.02$ is the threshold empirically selected for this ratio. The first occurrence of this condition is determined to be t_1. After the magnitude of the waveform drops below **90%** of its nominal peak value, the point t_2 is the minimum magnitude of the voltage before another sudden increase occurs. The point t_3 is then determined as the time index of the highest peak of the voltage waveform V_{O}.

The location of these three characteristic points is then validated using the following three checks: (i) the voltage magnitudes at these points are the expected values, (ii) the nominal number of samples between the overvoltage and the peak prior to it, (iii) the waveform slope is reversed at t_1. For the first check, the expected voltage magnitudes at t_1, t_2, and t_3 must follow the inequality given by,

$$|V_{t_2}| < |V_{t_1}| < |V_{t_3}|, \tag{3.21}$$

where $|V_{t_1}|$, $|V_{t_1}|$, and $|V_{t_3}|$ are the voltage magnitudes at times t_1, t_2, and t_3, respectively. The second check is that the peak before must be approximately equal to $N_c/2$ samples before the overvoltage peak as determined by,

$$\frac{N_{\mathrm{PB}} - N_c/2}{N_c} < \tau_{\mathrm{P}} \tag{3.22}$$

where N_{PB} is the number of samples between the overvoltage peak and the peak before it, N_c is the number of samples in each cycle, and $\tau_{\mathrm{P}} = 0.1$ is the threshold empirically selected for this ratio. Finally, the third check was validated using Equation (2.4) described in Sect. 2.1.3. If the first derivative of the voltage waveform leading up to t_1 is of the opposite sign than the first derivative of the voltage between t_1 and t_2, then the third check is satisfied. After all these criteria are satisfied for one of the three voltage phases, the other two phases are analyzed to ensure that some form of disturbance is present.

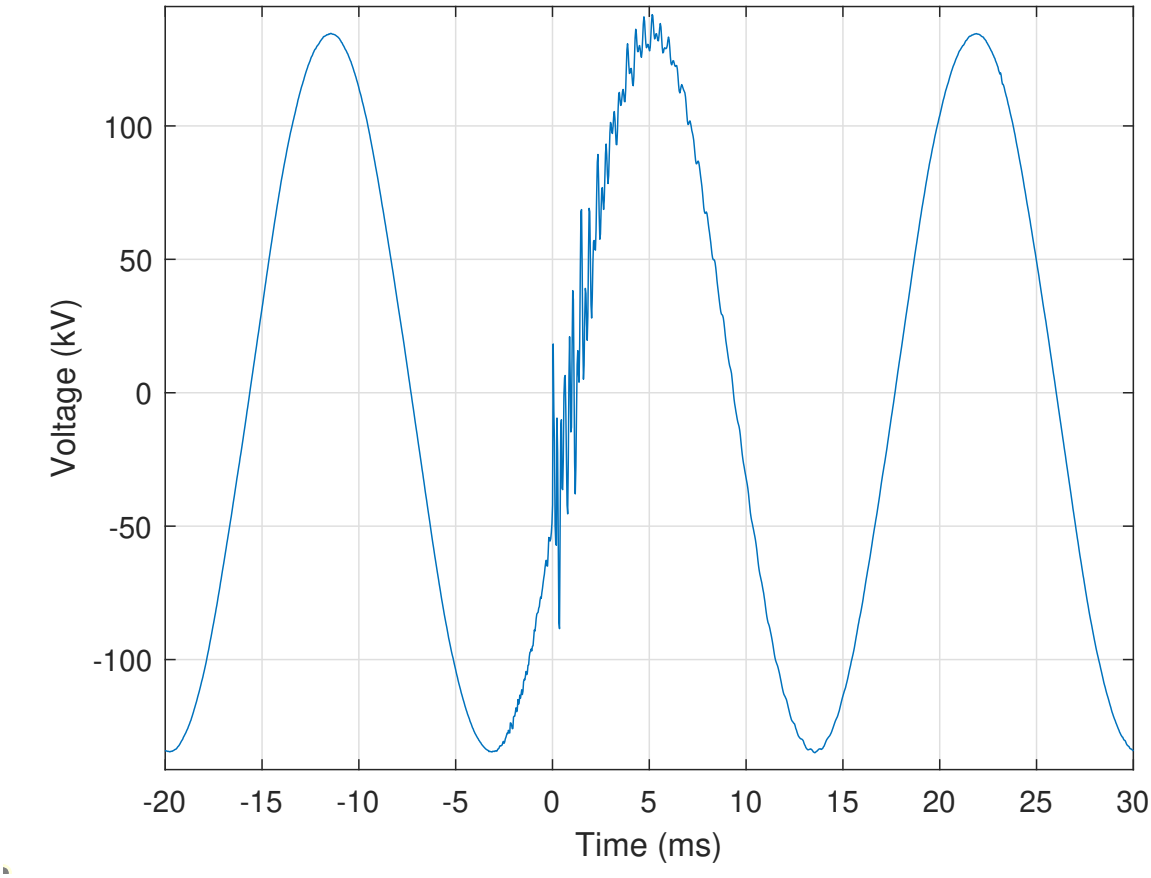


Figure 3.14 Illustration of a voltage waveform collected during a lightning strike disturbance [11]

3.1.11 Lightning Strikes

Transient overvoltages due to lightning strikes on a transmission line are typically impulses with a rise and decay time in the range of microseconds. Due to the instrument transformer's inability to pass these high frequencies and limited instrumentation sampling rates, lightning strike disturbances are not readily identified. A representative voltage waveform that includes a lightning strike disturbance is shown in Fig. 3.14.

First, the automated identification process attempts to identify the disturbance as a capacitor bank switching disturbance, see Sect. 3.1.10, and then a melting fuse disturbance, see Sect. 3.1.8. These steps are taken to ensure that a lightning strike disturbance is not incorrectly identified as either of these two disturbances. Although similar to a lightning strike both are easily distinguished

from it as well as one another. If the disturbance is not identified as a capacitor bank switching or melting fuse disturbance, then the disturbance is isolated from the overall waveform using the method given in Equation (3.19) for the isolation of the capacitor bank switching disturbance's disturbance. The disturbance isolation process is repeated for each lightning strike, and the longest strike duration is checked to ensure that it does not exceed **one** cycle. If more than **five** disturbances are isolated, then the disturbance is not identified as a lightning strike. In all of the processed data, lightning did not strike more than three times during a single recording. So long as no more than three lightning strike disturbances are isolated, then the automated process identifies the disturbance as a lightning strike and returns the number of strikes along with the disturbance's duration in seconds.

3.1.12 Harmonic Resonance

Power systems have natural frequencies that are a function of the system's inductive and capacitive impedance. When a nonlinear load on the power system–such as a VFD–generates a frequency that is a natural frequency (a.k.a., a multiple of the fundamental frequency) of the power system, then a resonance condition can result. This resonance can subject equipment to overvoltages or overcurrents, which can result in equipment failure or misoperation. Thus, it is important to detect harmonic resonance conditions quickly, so that appropriate and necessary actions can be taken to correct the problem(s). Fig. 3.15 shows an example case of harmonic resonance on an operationally recorded voltage waveform.

Harmonic resonance is characterized by the presence of high-frequency content in the voltage waveforms. Based upon this information, the automated identification process first calculates the Total Harmonic Distortion (THD) of the voltage waveform by,

$$V_{\text{THD}} = \frac{\sqrt{\sum_{i=2}^{M} |H_i|^2}}{H_1}, \tag{3.23}$$

where H_i is the i^{th} harmonic, H_1 is the fundamental frequency, $M = 100$ is the total number of

harmonics used for the calculation, and $|\bullet|$ denotes the magnitude [23]. If the THD is greater than **8%** of the fundamental frequency, then the process continues else it moves onto the next disturbance category. A value of 8% was empirically selected but can be adjusted as more data become available or based on power system specifics.

If the THD threshold is satisfied, then the automated identification process determines whether or not at least the sixth or one of the higher harmonics is more than **5%** of the fundamental frequency's magnitude. If this is the case, then the sign changes in the first derivative are calculated for each cycle using Equation (2.4) as described in Sect. 2.1.3. The number of first derivative sign changes in each cycle must be at least **10%** of the samples in each cycle N_c and also occur across **three** cycles. If all these criteria are satisfied, then the automated process identifies

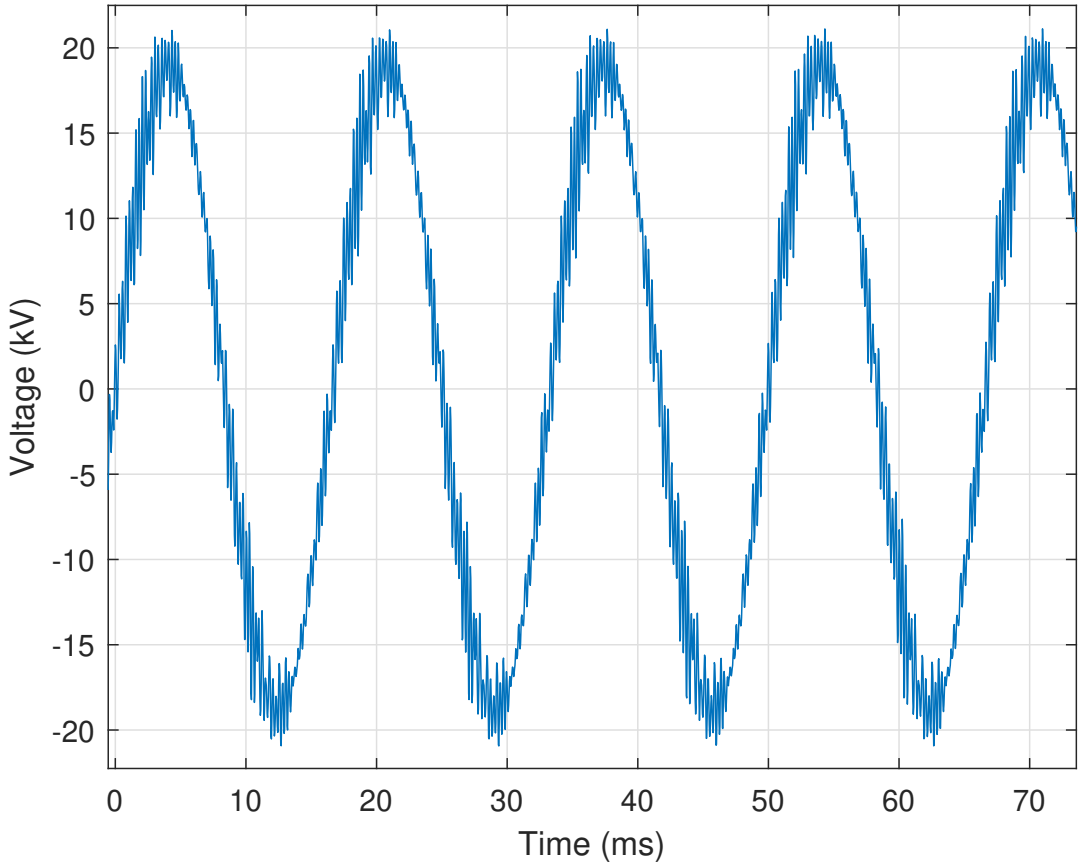
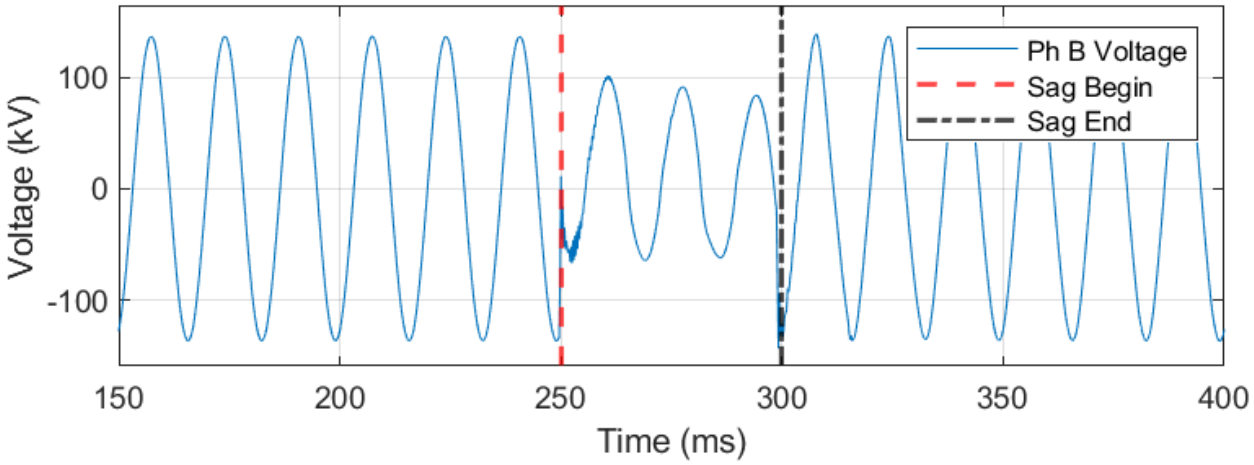
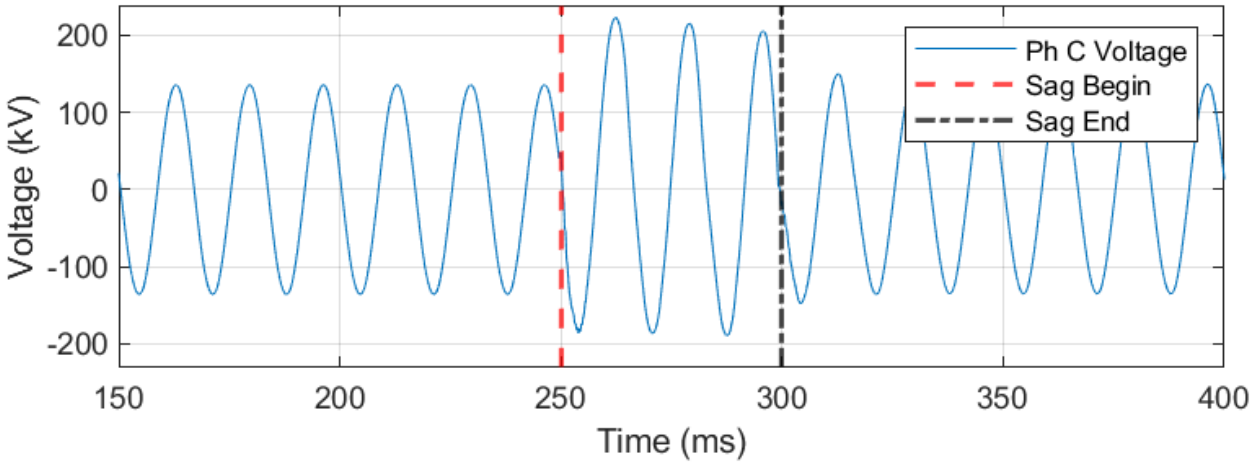


Figure 3.15 Illustration of a voltage waveform collected during a harmonic resonance disturbance [11]



(a) Phase B Voltage



(b) Phase C Voltage

Figure 3.16 Voltage waveforms indicating improper VT secondary grounding due to the simultaneous presence of voltage sag and swell on two phases [11]

the disturbance as harmonic resonance.

3.1.13 Improper Voltage Transformer Secondary Grounding

It is good design practice to use a single and solid grounding point on an instrument VT's secondary [24]. Otherwise, the result may be incorrect secondary voltage waveforms in both magnitude and angle, which can lead to the misoperation of protective relays. This can be exacerbated when faults occur on the lines protected by these relays.

A key indicator of improper VT secondary grounding is when one voltage phase has sagged

while another one has swelled. Fig. 3.16 provides a representative example of this indicator in which the Phase B voltage waveform is experiencing a sag from 250 ms to 300 ms, Fig. 3.16a, while the Phase C voltage waveform experiences a swell over the same time period, Fig. 3.16b. Automated identification of improper VT secondary grounding is facilitated by determining if a voltage sag and swell simultaneously exists on two of the three voltage phases. A sag occurs when one of the voltage waveform's peaks *falls* below the nominal peak voltage by more than **5%**, and a swell occurs when one of the voltage peaks *rises* above the nominal peak by more than **5%**. The phase angle between the sagged and swelled voltage phases is calculated by,

$$\theta = \cos^{-1}\left(\frac{\mathbf{V}_\alpha \cdot \mathbf{V}_\beta}{|V_\alpha||V_\beta|}\right), \tag{3.24}$$

where $\mathbf{V}_\alpha$ and $\mathbf{V}_\beta$ are two of the three faulted voltage phasors, $\cdot$ denotes dot product, and θ is the phase angle between $\mathbf{V}_\alpha$ and $\mathbf{V}_\beta$. The phase angle is calculated between phases: A to B, B to C, and A to C. In a balanced system, the nominal angle between two voltage phases is 120° [25]. If the phase angle deviates from this 120° nominal angle by more than **5**°, then the disturbance is identified as an improper VT secondary grounding disturbance.

3.1.14 Incipient Capacitive Voltage Transformer Failure

Capacitive Voltage Transformers (CVTs) supply voltage to protective relays, so it is very important that the CVT is measuring voltage accurately. If a catastrophic CVT failure results in a complete loss of this voltage, then the affected relays detect the loss using Loss of Potential (LOP) logic and act accordingly [26]. However, relays are not equipped to detect a CVT that is showing early signs of failure by providing incorrect data but has not yet failed to provide the supply voltage. The automated identification process is designed to detect early indicators of impending CVT failures to facilitate proper actions by utility personnel or equipment. Additionally, a CVT failure poses a significant safety risk to any utility personnel who happen to be nearby when it fails. The voltage waveform shown in Fig. 3.17 provides a representative illustration of the early indicators of an impending (a.k.a., incipient) CVT failure.

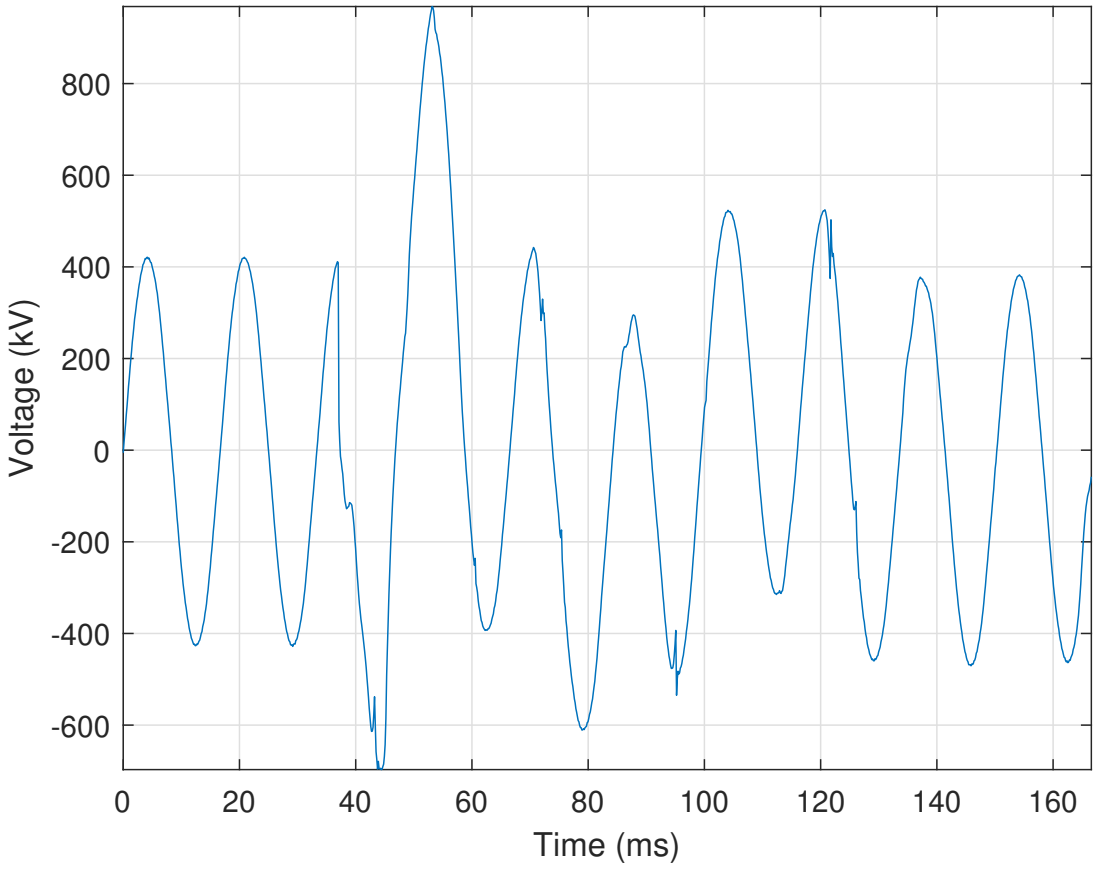


Figure 3.17 Illustration of a voltage waveform showing an incipient CVT failure disturbance [11]

The first indicator of an incipient CVT failure disturbance is that one of the voltage waveform's peaks will rise or fall by more than **10%** of the nominal peak value, and this behavior must persist for at least **three** cycles. The second incipient CVT failure indicator is that the disturbance portion of the voltage waveform will differ from its corresponding nominal waveform, by more than $\tau_{\Delta} = 0.02$ as introduced in Sect. 2.1.5 and implemented in Equation (3.19). Since CVTs are single-phase devices, incipient CVT failure would also only occur in one phase, which is a differentiating factor between other disturbances. Finally, the current waveform is analyzed to ensure that no disturbance is present since this disturbance type is specific to voltage waveforms.

Table 3.1 Automated electrical disturbance identification performance results

Disturbance Type	# Disturbances	# Correct	% Correct
CT Saturation	480	464	96.67%
A/D Clipping	960	953	99.27%
Induced Transient Noise	480	477	99.38%
High-Speed Reclosing	160	160	100%
DC Offset	960	956	99.58%
Motor Starting	160	160	100%
VFD Starting	160	160	100%
Blown Fuse	160	159	99.38%
Ferroresonance	480	476	99.17%
Capacitor Switching	160	159	99.38%
Lightning	480	477	99.38%
Harmonic Resonance	480	480	100%
VT Secondary Grounding	160	159	99.38%
CVT Failure	160	154	96.25%

3.2 Results

The performance of the developed rule-based, automated electrical disturbance identification process is assessed using a data set comprised of 160 total disturbance records that were collected by field devices operating in a power utility's transmission system. This data set contains approximately ten records for each of the discussed disturbances. The data set also contains disturbances with undisturbed voltage and current waveforms as well as single-phase and multi-phase disturbances. Each phase of every single-phase disturbance is processed, thus tripling the size of the associated disturbance's data set. False positive and false negative disturbance identifications are counted as incorrect or misidentifications. If a waveform did not contain one of the listed electrical disturbances and the automated process did not identify it as a disturbance, then it is counted as a correct result. Overall automated identification results are presented in Table 3.1 for each of the fourteen disturbance types described in Section 3.1. Table 3.1 shows the number of disturbances analyzed, the number correctly identified, and the percent correct accuracy for each disturbance type.

3.2.1 Results: Current Transformer Saturation

The accuracy of the automated process in determining CT saturation is 96.67%, which corresponds to 464 of the 480 processed waveforms being correctly identified. The CT saturation test proved to be challenging due to the complexity of this disturbance. The range of criteria used may not always be met for each CT saturation disturbance. For example, the A/D clipping waveform–shown in Fig. 3.2–appears to contain CT saturation based on the characteristic "kneeing" in the first two cycles of the fault. However, DC offset is not present and the rating of the CT was likely not exceeded, so this disturbance could be incorrectly classified. Also, for most of this testing, a CT ratio of 1,200:5 is used for each disturbance type regardless of the actual CT ratio. This was done for simplicity, but actual CT ratios from COMTRADE configuration files will be used when these tools are implemented in a production environment. When the actual CT ratio is known, then the rated current of the CT will be known and the automated process is able to accurately determine whether this rating was exceeded.

3.2.2 Results: Analog-to-Digital Converter Clipping

The accuracy in detecting the A/D converter clipping disturbance is very high achieving an accuracy of 99.27% because the process correctly identifies 953 out of 960 total, processed waveforms. The number of consecutive repeated samples threshold is set to four samples. There are some disturbances where clipping looks obvious to the human eye, but the samples that looked repeated are slightly different. Those results are counted as incorrect, even though the automated process functioned properly. Utility personnel could decide whether disturbances like these actually are a problem with the A/D converter or not. The A/D clipping detection methods should return proper results 100% of the time if the repeated samples have the exact same value. If they do not, then a very small tolerance (e.g., 10 V or 1 A) could be allowed between the magnitudes of samples that appear to be the same value.

3.2.3 Results: Induced Transient Noise due to Switching

Initial identification performance for this disturbance was poor at roughly 70%. In an effort to improve automated identification of induced transient noise from switching disturbances, the automated process was modified by incorporating a rule in which the presence of ferroresonance is checked first, then harmonic resonance, and finally induced transient noise from switching so that the three disturbances do not take place at the same time. The reason for this is purely due to the similarity with other disturbances and the lack of distinguishing characteristics in this disturbance. Also, a change was made to use the first cycle as a reference to isolate the disturbance as described in Sect. 2.1.5. These changes improved accuracy from 70% to 99.38% due to 477 out of 480 total, processed waveforms being correctly identified as induced transient noise from switching disturbances.

3.2.4 Results: High-Speed Reclosing with Tapped Motor Loads

This disturbance is correctly identified 100% in the conducted; however, there were only two disturbances in which the voltage did not sufficiently decay before reclosing since these disturbances do not often occur if utilities are aware of special settings that are needed for reclosers on such lines with tapped motor loads. Thus, a larger data set will be needed to determine the accuracy of the algorithm.

3.2.5 Results: DC Offset

The DC offset algorithm is one that is well-suited for rule-based analytics as shown by its accuracy of 99.58% (i.e., 956 out of 960 total, processed waveforms are correctly identified). The frequency analysis and cycle mean methods are very accurate at identifying DC offset. A few waveforms are falsely classified as DC offset. Waveforms such as the CT saturation example in Fig. 3.1 contain a very steep spike at the fault inception, so DC offset will be seen in that first half-cycle. Further logic could be added in future work to account for these faults so that DC offset is not detected in the first half-cycle.

3.2.6 Results: Motor Starting

Motor starting disturbances are very straightforward to identify. All 160 waveforms are correctly identified. One reason for the 100% accuracy is that the other disturbances analyzed did not have many similarities with motor starting. Transformer inrush would produce a similar waveform signature, but the differentiating factor is that motor starting is not as rich in harmonics. Motor inrush is also different from single-phase–the most often occurring–faults in that the elevated current always occurs across all three phases. For these reasons, the motor inrush classification process should be one of the most robust.

3.2.7 Results: Variable Frequency Drive Motor Starting

This disturbance type also results in 100% accuracy when tested by correctly identifying 160 out of 160 total, processed waveforms. However, the 10 VFD starting disturbances used are all from the same motor on the transmission system since these devices are not extremely common. More data will be needed to test the accuracy of the process for this disturbance type.

3.2.8 Results: Melted Fuse

The melted fuse disturbance classification accuracy is 99.38% as it correctly identified 159 out of 159 total waveforms. Melted fuse disturbances are relatively straightforward to identify due to their short duration. One incorrect classification stemmed from a disturbance containing a minor fault that was incorrectly labeled as a fuse fault. Although the fault lasted several cycles, the part of the current that exceeded the threshold was short enough to be classified as a blown fuse. The process of finding the fault inception and clearing points is very nuanced, and it may not always be 100% accurate in determining the clearing time, especially for faults that do not greatly (e.g., two times the rated current) exceed the predefined threshold.

3.2.9 Results: Ferroresonance

Ferroresonance is a unique disturbance that was classified with 99.17% accuracy by these analytics (i.e., correctly identifying 476 out of 480 total waveforms processed). In most of the data studied, the waveforms contain large gaps between samples (i.e., at least 50% of nominal peak value). A few waveforms did not have such large gaps, possibly due to the ferroresonance being less severe. These disturbances were not identified as ferroresonance, so new methods will need to be developed in the identification of these disturbances. One such method could be incorporating breaker statuses (i.e., open or closed) into the process since ferroresonance usually occurs with the breaker(s) in the open state.

3.2.10 Results: Capacitor Switching

The capacitor switching classification process correctly identified 159 out of 160 total waveforms resulting in an accuracy of 99.38%. The methods employed for this disturbance type are very detailed and are much more likely to generate false negatives than false positives. As long as the characteristic three points on the waveform (see Fig. 3.13) are present, the results should be accurate. The only capacitor switching disturbance that was missed was one in which the voltage transient occurred on the first cycle. Since the nominal peak value is taken using the first cycle as a reference, the rest of the processing becomes incorrect. This issue could be solved by using a predefined nominal peak value for each voltage level from an external data source.

3.2.11 Results: Lightning Strikes

The automated process correctly identified whether lightning was present for 477 out of 480 waveforms to achieve an accuracy of 99.38%. Originally, many capacitor switching disturbances were characterized as lightning. To remedy this, the process was updated such that the presence of lightning would only be checked if capacitor switching returned negative. The lightning detection process relies on an accurate determination of the duration of the disturbance. A short disturbance distinguishes lightning from other disturbances. A few disturbances were discovered in which the

process determined the disturbance to be longer than it was, which could be due to an outside disturbance unrelated to lightning. This phenomenon results in a few misclassifications.

3.2.12 Results: Harmonic Resonance

Harmonic resonance is difficult to distinguish from ferroresonance, so a modification was made to only run the harmonic resonance algorithm if ferroresonance has not occurred. This resulted in an accuracy of 100% with 480 out of 480 waveforms being correctly identified. There are five different harmonic resonance disturbance records in the data set for a total of fifteen voltage waveforms, so more data will be needed to test the robustness of this algorithm. One future improvement that could be made is to detect resonance under the fifth harmonic since resonance conditions can sometimes develop at those frequencies.

3.2.13 Results: Improper Voltage Transformer Secondary Grounding

The classification for this disturbance was very successful with an accuracy of 99.38% on the 160 waveforms studied. There are a large number of disturbances in which there is improper VT secondary grounding. Many of the CT saturation faults are not exactly 120° apart in their voltage phase angles, which would indicate improper grounding. This disturbance is straightforward to classify by rule-based techniques. The only issue that may occur is if inaccurate data are fed into the automated process.

3.2.14 Results: Incipient Capacitive Voltage Transformer Failure

The results for this disturbance are not as accurate as the others studied. A total of 154 out of 160 waveforms–for an accuracy of 96.25%–were correctly classified as demonstrating incipient CVT failure or not. The lower accuracy is due to the inconsistency in CVT failure disturbances. CVTs could be in different stages of their incipient failure, so the waveform signatures will not look the same. The differentiating factor though is that these disturbances are assumed to only occur one phase at a time, which improves the results.

CHAPTER 4
VOLTAGE UNBALANCE PREDICTION

4.1 Methodology

This section describes the processes used to construct a data set, preprocess the data set, and train the ANN used for voltage unbalance prediction. The section also describes the process to simulate a transmission line outage along with how voltage unbalance is predicted in that case.

4.1.1 ANN Training and Validation

First, a data set is constructed from the utility's SCADA system, which uses a central database to store data from all transmission system substations. Every four seconds, the SCADA system records each substation's MW and Mvar data for a sampling rate of fifteen samples per minute. However, voltage unbalance is not typically observed over short periods of time, thus one hundred fifty, four-second MW or Mvar values are averaged together. This represents ten minutes of SCADA-measured MW or Mvar values. This process is repeated for every transmission line and transformer associated with the selected substations. From these averaged MW and Mvar values, a total of two data sets are constructed. The first is constructed using SCADA MW and Mvar measurements accumulated over a twenty-eight-month period–from January 1, 2020 to April 30, 2022–for eight substations (see Fig. 4.1) within the utility's 500 kV transmission system. These eight substations consist of seventy transmission lines or transformers, thus there are a total of seventy, individual MW or Mvar average values per ten-minute interval for a total of 122,000 samples within this first data set. The second data set spans thirty months–from January 1, 2020 to June 30, 2022–and all forty-two transmission system substations. This second data set consists of 131,000 ten-minute averages that are each comprised of three hundred seventy-four MW or

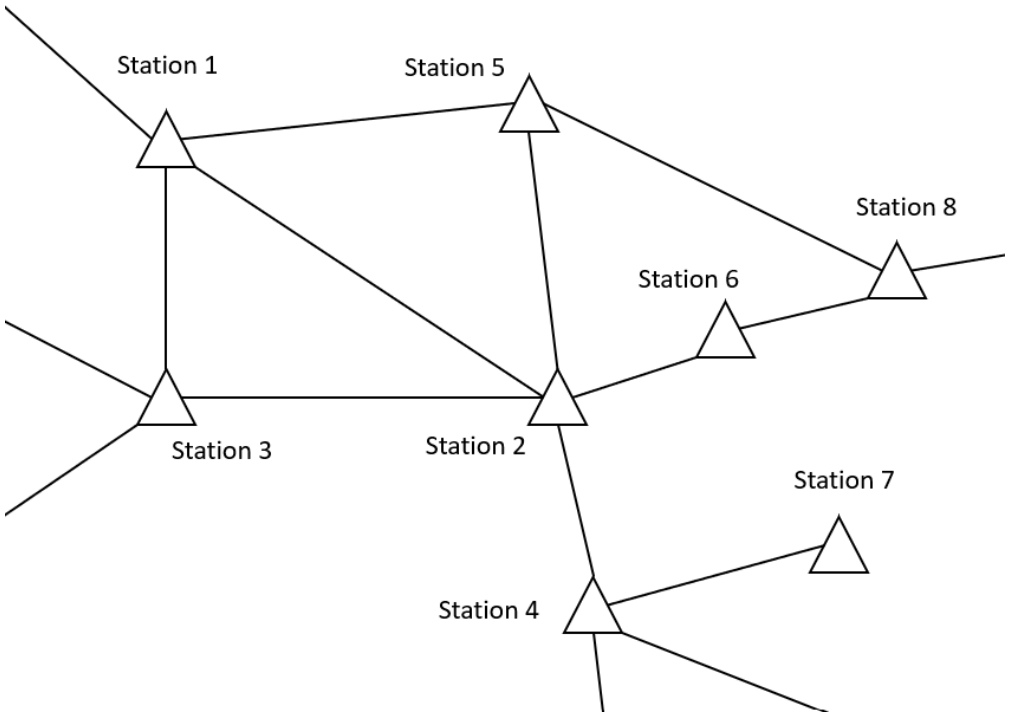


Figure 4.1 The initial, eight station section of the 500 kV transmission system studied in this work

Mvar values (i.e., one per transmission line or transformer). The results presented in Sect. 4.2.1 are generated using the normalized, ten-minute averages such that all MW and Mvar values range from zero to one.

During ANN training, each input is assigned a label corresponding to voltage balance '0' or unbalance '1'. These labels are determined using the VUF, which is calculated using Equation (2.10) for each set of three voltage measurements taken on the three phases (a.k.a, phases A, B, and C). In cases where the voltage is measured from the line rather than the bus, the percent VUF–for the lines at a given substation–is averaged together. VUF values over a threshold of 1.4% are assigned to a class label of '1' (a.k.a., unbalanced), while values under that threshold are given the class label of '0' (a.k.a., balanced). This threshold is the same as that defined in IEC 61000-3-13 as the planning level for HV systems [15]. Though the system studied here is at the EHV level, the 1.4% threshold was adopted by the utility.

Both data sets are split into training and "blind" testing subsets in which 80% of the data

is randomly assigned to the training set and the remaining 20% assigned to the testing set. This partitioning can be changed based on individual needs. An ANN is then created for each substation or bus using the inputs and labels in the training data set. The inputs for each substation's unique model are the ten-minute average MW and Mvar measurements for all lines and transformers in the studied region, which accounts for the interconnection of the system. Each station's model is trained using $k = 5$-fold cross-validation [16], tested using inputs that are unknown to it (i.e., not used during ANN training), and the outputs are compared to the known labels from the "blind" testing set. In this way, it can be determined whether or not the model correctly predicts balanced voltage (Class 0) or unbalanced voltage (Class 1). Each individual sample that is not classified correctly by the model is counted as an error. The percentage of correct classifications is calculated for each model by,

$$\% \text{ Accuracy} = \frac{N_s - N_e}{N_s} \times 100\%, \tag{4.1}$$

where N_s is the number of samples in the data set and N_e is the number of classification errors produced.

4.1.2 Testing in a Line Outage Study

The trained model is tested in an outage study–conducted by utility personnel–as an additional test of the developed voltage unbalance prediction approach. These outage studies involve using state estimation software to simulate what would happen in the transmission system should a line be removed from service. These studies are conducted weeks or months ahead of a scheduled line outage. Thus, a utility knows in advance how a line outage will affect the loading of the lines in the system. However, these studies are unable to predict voltage unbalance. Voltage unbalance on the EHV system is often a function of the line loading, so removing a critical line from service would cause very high power flows on the remaining lines. The ANN voltage unbalance prediction approach augments these outage studies to also include the impact of a line outage on voltage unbalance. The state estimation tool used in this case study was not sufficient in itself to predict voltage unbalance because it only outputs a single-phase voltage reading. Thus, the VUF

cannot be calculated using Equation (2.10). However, MW and Mvar readings are produced by the simulation, so these measurements are used as inputs to the developed prediction model.

The goal is to determine whether voltage unbalance resulting from a past line outage could have been accurately predicted using the trained ANN. Voltage unbalance occurred when the transmission line–connected to Substation 4 in Fig. 4.1 and connected to a substation not shown–was removed from service. This voltage unbalance event is present in the second data set, so it is assumed that the model learned the patterns present in that event. The simulation case was created on April 27, 2022 to remove the same line from service on May 4, 2022, which corresponds to the exact calendar date of the unbalance event May 4 two years prior. The system is configured as it is on April 27, 2022, except that a particular line is opened. The MW and Mvar measurements from this case study are then used as the inputs to the previously trained prediction model. The outputs of the model (either balanced or unbalanced) are then compared with the "true" values of unbalance from the event two years prior to gauge the accuracy of the model. The unbalance measurements from the previous event are just used as estimates since the impacts of opening the same line today are not fully known.

4.2 Results

A model is trained for each substation or bus, and the accuracy is calculated for each one. The results of the tests are displayed in Table 4.1 for the first data set representing the eight substations in Fig. 4.1. A substation number of "2-1" corresponds to Substation 2 and Bus 1. Table 4.1 also shows the percentage of the total data set that is above the 1.4% threshold and considered to be unbalanced. Finally, the percentage accuracy is shown in the third column for each station or bus.

The trained ANN is further tested using an outage study as described in Section 4.1.2 and the corresponding voltage unbalance prediction results are presented in Table 4.5. The first column contains the same station and bus numbers as Table 4.1, the second column lists whether or not voltage unbalance occurred during the actual event two years ago, and the third column

lists whether or not the ANN predicts voltage unbalance to be present in the simulated line outage conducted in the present day.

4.2.1 Results: ANN Training and Validation

The results of the model validation using historical data shown in Table 4.1 are shown to be very accurate with the lowest accuracy being 96.75%. The issue with some stations is that there are too few or no data points above the 1.4% VUF threshold, thus impeding the ANN's ability to accurately predict voltage unbalance at the corresponding substation or bus. A possible solution to this problem would be to either (i) find more data for model training further in the past or (ii) lower the threshold for those particular stations or busses to something lower than 1.4% since the thresholds are adaptable for each individual station or bus. For the stations with more data above the threshold, the results are still very accurate. This demonstrates the robustness of the training algorithm used, and it shows the connection between line MW and Mvar loading and voltage unbalance.

Table 4.2, Table 4.3, and Table 4.4 display the results of the tests using the second data set that covers the entire 500 kV system. The results using MW and Mvar as well as only MW are

Table 4.1 Results of voltage unbalance prediction model tested using historical MW and Mvar data for eight substations

Station-Bus	% Above Threshold	% Accuracy
1-1	1.09%	99.95%
1-2	0%	99.99%
2-1	0.19%	99.92%
2-2	3.61%	98.66%
3-1	0%	100%
3-2	0.01%	100%
4-1	0.55%	99.94%
5-1	0.12%	99.93%
6-1	20.90%	96.75%
7-1	2.01%	99.28%
8-1	11.94%	99.41%
8-2	2.71%	98.82%

Table 4.2 Results of voltage unbalance prediction model tested using historical MW and Mvar data for forty-two substations

Station Number	% Above Threshold	Average % Correct	Station Number	% Above Threshold	Average % Correct
1	2.96%	91.80%	22	0.30%	92.49%
2	0.16%	92.28%	23	11.70%	90.54%
3	0.01%	92.42%	24	0.35%	92.28%
4	0.00%	92.42%	25	0.24%	92.41%
5	0.00%	100.00%	26	19.20%	89.88%
6	11.40%	92.16%	27	0.06%	92.41%
7	1.06%	92.44%	28	0.59%	92.34%
8	1.54%	92.02%	29	2.46%	91.66%
9	0.00%	92.47%	30	0.00%	100.00%
10	0.00%	92.32%	31	0.48%	92.28%
11	0.03%	92.31%	32	0.48%	92.33%
12	0.00%	92.58%	33	0.11%	92.46%
13	0.02%	92.42%	34	0.03%	92.48%
14	0.01%	92.38%	35	0.90%	92.40%
15	0.00%	92.43%	36	0.00%	94.00%
16	0.01%	92.48%	37	0.10%	92.29%
17	0.01%	92.44%	38	0.01%	92.34%
18	0.00%	92.39%	39	0.00%	94.02%
19	0.02%	92.40%	40	0.01%	92.41%
20	0.00%	92.53%	41	0.01%	92.46%
21	0.00%	92.35%	42	6.56%	90.72%

Table 4.3 Results of voltage unbalance prediction model tested using historical MW data for forty-two substations

Station Number	% Above Threshold	Average % Correct	Station Number	% Above Threshold	Average % Correct
1	2.96%	92.04%	22	0.30%	92.45%
2	0.16%	92.37%	23	11.70%	90.59%
3	0.01%	92.51%	24	0.35%	92.48%
4	0.00%	92.56%	25	0.24%	92.34%
5	0.00%	100.00%	26	19.20%	89.44%
6	11.40%	92.01%	27	0.06%	92.59%
7	1.06%	92.49%	28	0.59%	92.40%
8	1.54%	92.21%	29	2.46%	91.67%
9	0.00%	92.66%	30	0.00%	100.00%
10	0.00%	92.53%	31	0.48%	92.37%
11	0.03%	92.52%	32	0.48%	92.49%
12	0.00%	92.41%	33	0.11%	92.53%
13	0.02%	92.51%	34	0.03%	92.59%
14	0.01%	92.55%	35	0.90%	92.52%
15	0.00%	92.74%	36	0.00%	94.06%
16	0.01%	92.49%	37	0.10%	92.57%
17	0.01%	92.59%	38	0.01%	92.60%
18	0.00%	92.60%	39	0.00%	94.10%
19	0.02%	92.55%	40	0.01%	92.53%
20	0.00%	92.46%	41	0.01%	92.53%
21	0.00%	92.50%	42	6.56%	90.21%

Table 4.4 Results of voltage unbalance prediction model tested using historical Mvar data for forty-two substations

Station Number	% Above Threshold	Average % Correct	Station Number	% Above Threshold	Average % Correct
1	2.96%	94.79%	22	0.30%	95.28%
2	0.16%	95.13%	23	11.70%	93.30%
3	0.01%	95.32%	24	0.35%	95.21%
4	0.00%	95.33%	25	0.24%	95.20%
5	0.00%	100.00%	26	19.20%	91.75%
6	11.40%	94.77%	27	0.06%	95.36%
7	1.06%	95.24%	28	0.59%	95.13%
8	1.54%	94.94%	29	2.46%	94.36%
9	0.00%	95.36%	30	0.00%	100.00%
10	0.00%	95.28%	31	0.48%	95.12%
11	0.03%	95.32%	32	0.48%	95.23%
12	0.00%	95.24%	33	0.11%	95.28%
13	0.02%	95.27%	34	0.03%	95.38%
14	0.01%	95.38%	35	0.90%	95.33%
15	0.00%	95.47%	36	0.00%	96.31%
16	0.01%	95.29%	37	0.10%	95.34%
17	0.01%	95.33%	38	0.01%	95.31%
18	0.00%	95.35%	39	0.00%	96.29%
19	0.02%	95.36%	40	0.01%	95.31%
20	0.00%	95.31%	41	0.01%	95.32%
21	0.00%	95.31%	42	6.56%	93.23%

very similar, and the model is less accurate than the previous test. This is due to the large amount of data and the number of substations monitored, so the model must discriminate between more cases. Using only Mvar data results in the highest voltage prediction unbalance accuracy.

One explanation for this could be the untransposed nature of the transmission lines. When lines are not transposed, the inductance and capacitance of the lines are different across the three phases while the resistance is essentially the same on all three. Since inductors consume and capacitors produce Mvar, this would impact Mvar flow. There is also a known relationship between Mvar and voltage. The work in [27] discusses ways to improve the VUF of a power system through different methods of injecting Mvar. Also, the fast decoupled power flow method is able to separate MW and phase angle from Mvar and voltage magnitude [28]. Thus, it makes sense that Mvar is found to be the main factor in predicting voltage unbalance.

Table 4.5 Results of voltage unbalance prediction model tested in a line outage study using state estimation software

Station-Bus	Unbalanced Before?	Unbalanced Now?
1-1	No	Yes
1-2	No	Yes
2-1	Yes	Yes
2-2	Yes	Yes
3-1	No	Yes
3-2	No	No
4-1	Yes	Yes
5-1	Yes	Yes
6-1	Yes	Yes
7-1	Yes	Yes
8-1	Yes	No
8-2	Yes	Yes

4.2.2 Results: Testing in a Line Outage Study

The developed model shows itself to be very accurate when used in a line outage study. The outputs from the simulation (balanced or unbalanced) are compared to the historical unbalance values as an approximation of the model's performance. The issue with making this comparison is that the system is not configured in the exact same way as it was in the prior event, so it is not known what the exact voltage unbalance values would be if the line were opened today. A future test would be to conduct an outage study on a line that will actually be opened in the future, predict whether unbalance occurs, and compare the results with the actual values after the line is opened. That was not feasible in this case as the line known to cause the most voltage unbalance was not scheduled for an upcoming outage.

However, the results compared to the prior event are still accurate. Station 1 Bus 1 and Station 3 Bus 1 had VUFs of 1.38% and 1.26% in the prior event, respectively. These values are very close to the 1.4% threshold, so the fact that the model predicted unbalance is not a serious issue and just makes it slightly more secure. In the prior line outage, Bus 2 at Station 1 was de-energized, so the model did not have an example within the training data as to whether or not voltage unbalance would occur based on this outage. The reason for the misclassification of Station

8 Bus 1 is not immediately apparent. It could be that there is conflicting data in the training set for what the VUF is when the line in question is removed from service. Another explanation could be that the difference in system configuration between the previous event and today could be enough to cause that bus to not have voltage unbalance present. It is still preferable that the model predicts unbalance when it will not occur rather than the opposite, thus the model is biased toward security.

CHAPTER 5
CONCLUSIONS AND FUTURE WORK

This thesis presented an approach for automated identification of electrical disturbances in a power system. The first objective in Sect. 1.4 was met as the automated process was able to analyze short, severe disturbances using rule-based analytics based on the waveform signatures in the associated disturbance records. Fourteen different disturbance types were successfully classified with an average accuracy of 99.13%. The developed processes will result in time savings for utility personnel as well as increase awareness of disturbances occurring in the power system. This process can categorize events in a matter of minutes rather than hours or days, thus providing utility engineers, operators, and managers with actionable intelligence that will enable immediate and decisive corrective action. Impending–or incipient–device failures will also be detected to enable corrective action before complete failure so that safety hazards can be removed.

Voltage unbalance is another type of disturbance analyzed in this work. The developed ANN-based process was able to predict such a phenomenon using historical SCADA data, meeting the second objective in Sect. 1.4. Eight substations in an EHV system were studied with voltage unbalance being predicted at each bus at an accuracy greater than 95%. The tests were further expanded to include all, forty-two substations, and it was found that using only Mvar data produced the most accurate results at over 91% accuracy. A line outage study was also performed in the same manner as a real-world study, and the impact of removing a line on the voltage unbalance of the same stations was determined and compared to historical data. The ability to predict voltage unbalance through Mvar flows will give utility personnel more awareness of this phenomenon when planning transmission line outages and also enable them to shift the timing of these outages based on voltage unbalance.

Future investigation can increase the number of disturbance types that can be classified and to further test the process using more data. Future work with voltage unbalance prediction could include: (i) gathering more training data to represent all cases of voltage balance or unbalance, (ii) performing an outage study before an actual line is removed from service and studying the real-world impact, and (iii) incorporating this voltage unbalance prediction model into existing state estimation software used by utilities for line outage studies.

This work serves to increase the overall reliability of the transmission system by providing engineers and operators with actionable intelligence that will enable immediate and decisive action. The voltage unbalance prediction will save utilities time and money by reducing damage to power system equipment that results from voltage unbalance and also increasing customer satisfaction by lowering damage to their equipment as well.

REFERENCES

[1] *Electric Power Systems and Equipment — Voltage Ratings (60 Hertz)*, ANSI C84.1-2020.

[2] J. Liu and J. Milanović, "Probabilistic Estimation of Voltage Unbalance in MV Distribution Networks With Unbalanced Load," *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 693–703, Apr. 2015.

[3] A. Rodríguez, J. Aguado, F. Martín, J. López, F. Muñoz, and J. Ruiz, "Rule-based classification of power quality disturbances using S-transform," *Electric Power Systems Research*, vol. 86, pp. 113–121, 2012.

[4] R. Kapoor, R. Gupta, L. Son, S. Jha, and R. Kumar, "Boosting performance of power quality event identification with KL divergence measure and Standard Deviation," *Measurement*, vol. 126, pp. 134–142, May 2018.

[5] S. De and S. Debnath, "Real-time cross-correlation-based technique for detection and classification of Power Quality Disturbances," *IET Generation, Transmission & Distribution*, vol. 12, no. 3, pp. 688–695, 2017.

[6] S. Wang and H. Chen, "A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network," *Applied Energy*, vol. 235, pp. 1126–1140, 2019.

[7] M. Mishra, "Power quality disturbance detection and classification using signal processing and soft computing techniques: A comprehensive review," *International Transactions on Electrical Energy Systems*, vol. 29, no. 8, 2019.

[8] J. Ghaeb and J. Chebil, "Prediction of Voltage Unbalance Employing Space Vector Property," *International Journal of Engineering Research and Development*, vol. 12, no. 12, pp. 65–70, Dec. 2016.

[9] M. Okelola and O. Olabode, "Detection of voltage unbalance on three phase induction motor using Artificial Neural Network," *International Journal of Emerging Trends in Engineering and Development*, vol. 4, no. 8, pp. 18–25, Jul. 2018.

[10] M. Albu and G. Heydt, "On the use of RMS values in power quality assessment," *IEEE Transactions on Power Delivery*, vol. 18, no. 4, pp. 1586–1587, Oct. 2003.

[11] J. Boyd, J. Tyler, A. Murphy, and D. Reising, "Learning from Power Signals: An Automated Approach to Electrical Disturbance Identification Within a Power Transmission System," *25th Annual Georgia Tech Fault and Disturbance Analysis Conference*, May 2022.

[12] A. Wilson, D. Reising, R. Hay, R. Johnson, A. Karrar, and T. Loveless, "Automated identification of electrical disturbance waveforms within an operational smart power grid," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4380–4389, Sep. 2020.

[13] T. Seiphetlho and A. Rens, "On the assessment of Voltage Unbalance," *Proceedings of 14th International Conference on Harmonics and Quality of Power - ICHQP*, Sep. 2010.

[14] A. von Jouanne and B. Banerjee, "Assessment of voltage unbalance," *IEEE Transactions on Power Delivery*, vol. 16, no. 4, pp. 782–790, Oct. 2001.

[15] *Electromagnetic compatibility (EMC) – Part 3-13: Limits – Assessment of emission limits for the connection of unbalanced installations to MV, HV and EHV power systems*, IEC/TR 61000-3-13, 2008.

[16] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning; Data Mining, Inference, and Prediction," Springer-Verlag, 2001.

[17] J. Basu, D. Bhattacharyya, and T. Kim, "Use of Artificial Neural Network in Pattern Recognition," *International Journal of Software Engineering and Its Applications*, vol. 4, no. 2, pp. 23–33, Apr. 2010.

[18] "Generate pattern recognition network," *MathWorks*. [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/patternnet.html. [Accessed: 30-Apr-2022].

[19] A. Hargrave, M. Thompson, and B. Heilman, "Beyond the knee point: A practical guide to CT saturation," *71st Annual Conference for Protective Relay Engineers (CPRE)*, Mar. 2018.

[20] *Common format for transient data exchange (COMTRADE) for power systems*, IEEE Std C37.111, 2013.

[21] R. Patterson and G. Pitts, "Reclosing and Tapped Motor Load," *61st Annual Georgia Tech Protective Relaying Conference*, May 2007.

[22] *Power Quality Waveform Identification*. EPRI, Palo Alto, CA: 2020. 3002019329.

[23] D. Shmilovitz, "On the definition of total harmonic distortion and its effect on measurement interpretation," in *IEEE Transactions on Power Delivery*, vol. 20, no. 1, pp. 526-528, Jan. 2005.

[24] *IEEE Guide for Grounding of Instrument Transformer Secondary Circuits and Cases*, IEEE Std C57.13.3, 2014.

[25] J. Glover, T. Overbye, and M. Sarma, "Chapter 2: Fundamentals," in *Power System Analysis & Design*, 6th ed., Boston, MA: Cengage Learning, 2017, p. 56.

[26] S. Gray, D. Haas, and R. McDaniel, "CCVT failures and their effects on distance relays," *71st Annual Conference for Protective Relay Engineers (CPRE)*, Mar. 2018.

[27] J. Ghaeb, M. Alkayyali, and T. Tutunji, "Wide range reactive power compensation for voltage unbalance mitigation in electrical power systems," *Electric Power Components and Systems*, vol. 49, no. 6-7, pp. 715–728, 2021.

[28] R. Portelinha, C. Durce, O. Tortelli, and E. Lourenço, "Fast-decoupled power flow method for integrated analysis of transmission and distribution Systems," *Electric Power Systems Research*, vol. 196, Jul. 2021.

VITA

Jonathan D. Boyd was born in 2000 in Chattanooga, Tennessee. He graduated with a Bachelor of Science in Electrical Engineering from the University of Tennessee at Chattanooga (UTC) in 2021. He expects to graduate with a Master of Science in Engineering in May of 2023. Jonathan is an avid distance runner and enjoyed competing for the UTC Mocs cross country team for five years. He is currently employed by the Tennessee Valley Authority in the SCADA & Reliability Systems group where he helps maintain the SCADA system that provides grid visibility and control to system operators in the control center.