A HIERARCHICAL APPROACH TO AUTOMATED IDENTIFICATION OF ANOMALOUS ELECTRICAL WAVEFORMS

By

Aaron Wilson

Donald R. Reising
Assistant Professor
(Chair)

Thomas D. Loveless
Assistant Professor
(Committee Member)

Abdelrahman A. Karrar
Associate Professor
(Committee Member)

Robert W. Hay, P.E.
Sr. Electrical Engineer
(Committee Member)

A HIERARCHICAL APPROACH TO AUTOMATED IDENTIFICATION OF ANOMALOUS ELECTRICAL WAVEFORMS

By

Aaron Wilson

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 2019

ABSTRACT

Power utilities employ "smart" field devices capable of digitally recording electrical waveforms. The relationship between events and their recorded waveforms can be exploited for characterization of the power grids state over any period of time and facilitating the impact electrical disturbances have on equipment, subsystems, and systems. Over a period of one month, these devices record approximately 2,000 electrical disturbance waveforms. Currently, analysis of these waveforms is conducted using by-hand approaches; thus, severely limiting the analysis to roughly 2%. The analysis is done hours to days after the events occurred, which negates informed, timely corrective actions. This document presents an automated hierarchical approach capable of identifying specific events using the electrical disturbance waveforms stored using COMmon format for TRAnsient Data Exchange (COMTRADE) files. The developed approach processes a single file in 1.8 seconds and has demonstrated successful identification of 140 events with a success rate of 91%.

ACKNOWLEDGEMENTS

I would first like to thank the chair of my committee, Dr. Donald Reising, for his guidance throughout the course of this work. I wish to also thank my committee members, Dr. Abdelrahman Karrar, Dr. Thomas D. Loveless, and Mr. Bob Hay, for taking the time to assist with the completion of this thesis and for serving on my committee.

Special thanks is given to Mr. Jim Glass, Mr. Bob Hay, and Mr. Raymond Johnson of Electric Power Board (EPB) of Chattanooga for allowing access to their databases, systems, providing professional guidance, and for being proactive and supportive throughout this endeavor.

This project was funded by the Electric Power Research Institute (EPRI) Distribution Modernization Demonstration (DMD) Data Mining Initiative and the University of Chattanooga Foundation Incorporated.

Lastly I would like to thank my wife, Bethany, for encouraging me and acting as my support system throughout the last two years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

COMTRADE, COMmon format for TRAnsient Data Exchange

ED, Electrical Disturbance

EPB, Electric Power Board

DFT, Discrete Fourier Transform

IEEE, Institute of Electrical and Electronics Engineers

IPCR, IntelliRupter® PulseCloser®

LG, Line-to-Ground

LL, Line-to-Line

LLG, Line-to-Line-to-Ground

LLL, Line-to-Line-to-Line

LLLG, Line-to-Line-to-Line-to-Ground

LOS, Loss of Source

LTE, Let-Through Energy

PQ, Power Quality

RMS, Root Mean Square

ROS, Return of Source

RTN-NC, Return-to-Normal on Normally-Closed

RTN-NO, Return-to-Normal on Normally-Open

TCC, Time-Current Characteristic

LIST OF SYMBOLS

E_L, Let-through energy

x_{rms}, Root-mean-square value of signal $x(t)$

$f_k(x)$, Posterior probability of training example x given class k

π_k, Prior probability of class k

$P(\cdot)$, "Probability of" operator

$\textbf{arg max}_k [f]$, The value of k that maximizes the expression f

$X(f)$, Fourier Transform of a continuous signal $x(t)$

$X[k]$, Discrete Fourier Transform of a discrete signal $x[n]$

sgn, The Signum function

j, Complex operator equal to $\sqrt{-1}$

$A(t)$, Instantaneous amplitude

I_{SC}, Short-circuit current for recording device (RMS)

I_L, Measured current during disturbance (RMS)

k_i, Frequency bin of obtained FFT nearest to a frequency of i Hz

F_s, Sampling frequency

V_X, Source-side voltage recording

V_Y, Downstream voltage recording

CHAPTER 1

INTRODUCTION AND MOTIVATION

One of the most powerful aspects of the smart grid is the deployment and integration of automated switches such as S&C Electric Companys IntelliRupter® PulseCloser® (IPCR) fault interrupter devices. The primary function of these switches is to facilitate re-energizing and re-routing of faulted power lines. Additionally, IPCRs employ sensors that digitally record current and voltage waveform profiles that can be collected, processed, stored by the local electric utility employing them. These waveform profiles are stored in text files that comply with the IEEE COMmon format for TRAnsient Data Exchange (COMTRADE) standard (IEEE C37.111-2013) [4].

The power distribution utility, Electric Power Board (EPB) located in Chattanooga, Tennessee, has 1,200 IPCRs deployed throughout its distribution network, with approximately 350 being in a normally-open state. During a typical month of operation, EPB's IPCRs will collect roughly 2,100 anomalous electrical waveform events. Following collection, the anomalous waveforms are subsequently analyzed to determine the cause of the event(s) to facilitate the elimination or minimization of the factor(s) that led to their occurrence. Typically, the analysis of these waveforms is performed hours and even days after a particular event has occurred. Additionally, the amount of anomalous waveform data is often so great that only a small percentage, roughly 2%/42 waveforms, can be processed. Therefore, a majority of the anomalous waveform data is left unprocessed and any associated information that can be gleaned from it is lost.

The ability for utilities to characterize common electrical disturbance (ED) waveforms automatically allows for saving on labor costs. EPB estimates a cost roughly equivalent to that of employing five full-time engineers, which could cost up to $500,000 annually, would be required to analyze all incoming files. Some other benefits to having an automated classification process include:

- The ability to make system improvements based on information that would have otherwise been unavailable to the utility
- Identifying and addressing problems that may lead to asset failure
- Improving customer service by making power quality (PQ) data available to industrial customers
- Prevent potentially harmful attacks, such as directed energy, EMPs, etc.

The aim of this document is to describe the developed and employed software-based approaches which facilitate automated identification of specific events, e.g., a low-side fuse melt, using the waveform signatures stored in the IPCR-generated COMTRADE files. This "hierarchical" software approach achieved a 93% correct classification rate across 140 COMTRADE files, performing analysis at a rate of approximately 1.8 seconds per file.

Technical Motivation

As the development of artifical intelligence (AI) techniques continues to grow, the opportunity for application in the field of electrical disturbance classification also increases. The work in [5] proposes the use of a digitized fuzzy logic (DFL) classifier based on sequence component analysis of faulted waveforms. A fuzzy-logic system with

"Z & S member functions" are used to assign a waveform to a class that maps to its fault type (single phase-to-ground, two-phase, two-phase-to-ground, three-phase, and three-phase-to-ground). These member functions transform their inputs into logic values "0" or "1". Different combinations of "0"'s and "1"'s for each current phase imply a different fault type. In [6], an artificial neural network-based (ANN) approach to classifying faulted waveforms based on their sequence components. In both of the above works, line-to-ground (LG), line-to-line (LL), line-to-line-to-ground (LLG), line-to-line-to-line (LLL) and triple-phase-to-ground (LLLG) faults were simulated for analysis.

Power quality (PQ) disturbance classification has been studied in a variety of ways. These methods typically perform a transformation on the disturbed voltage signals before sending the transformed information into a classification system. In [7], combinations of higher-order statistics of the corrupted waveforms are used to classify the type of power quality disturbance encountered. The S-Transform is used to extract features from PQ waveforms in [8], which are then classified using a probabilistic neural network (PNN). The S-Transform is a time-frequency analysis tool similar to the Continuous Wavelet Transform, except the mother wavelet function has a dilation parameter that changes the side of the wavelet. The Wavelet Transform is also a popular method of classifying PQ signals, as described in [9].

The methods presented above present a challenge when implementing machine learning-based classification approaches as described. The hierarchical classification structure presented here uses waveforms captured from operational field devices deployed throughout a smart grid distribution network; thus, not all ED types are represented by a large, roughly 100 waveforms or more, set of waveforms within the power utilitys database. One advantage to the developed hierarchical approach is that it facilitates the

selection, development, and implementation of machine learning approaches based upon the fault ED type and number of waveforms comprising the data set of the corresponding fault category. Therefore, the presented approach is not limited to the selection of one particular machine learning approach that may excel at the automated identification of one, e.g., low-side fuse melts, event and perform poorly at another. This also allows for the use of simpler, i.e., less computational resources and reduced run times, classification algorithms to perform the automated identification; thus, making the presented approach more tractable for adoption and implementation by power utilities nationwide.

Moreover, there are multiple categories that needed to be defined prior to classifying individual waveform profiles into sub-groups. Additional logic is required to handle shifting of COMTRADE files into the correct category. For example, if a file read from an IPCR is a recording of a switching event, it is undesirable for this file to be processed and classified as a fault. Therefore, logic for handling these types discrepancies prior to classification of the subcategories is required. This is the heart of the hierarchical framework used to route COMTRADE files to their correct locations, and is described in the next section.

Contributions

This proposal describes a process for hierarchical classification of COMTRADE files into one of three groups:

1. Valid Data: A COMTRADE file contains valid data if there is at least one sensor recording that contains at least 100 samples that exceed a certain threshold, known as the "sensor floor". The sample number is a configurable value.

2. Switching Events: Switching events are a result of controlled changes in the network. For example, a network performs switching when re-routing of power flow is required to bypass faulted sections. A switching event recording typically depicts increases or decreases in energy in either current or voltage waveforms. Closing into circuits that operate in a "normally-closed" state will show an increase in current, whereas closing into circuits that operate in a "normally-open" state will show a decrease in current. Additionally, load increases or decreases are considered switching events.

3. Electrical Disturbances (EDs): EDs, unlike switching events, are undesired changes in the state of the network. Two ED event sub-categories of were addressed in this work:

   (a) Faults: This sub-category contains: line-to-ground, line-to-line-to-ground, and three-phase line-to-ground.

   (b) Power Quality (PQ) Disturbances: This sub-category contains: voltage sags and swells, as well as various artifacts of ED events such as harmonics and capacitor-induced effects on currents and voltages.

This hierarchical classification process allows utility engineers, such as those at EPB, to obtain information contained in COMTRADE files in a matter of minutes, rather than hours, days, or never.

The remainder of this document is organized as follows. Chapter 2 provides necessary background for IPCR operation, the COMTRADE standard, characteristics of various EDs and their artifacts, and analysis of line-to-ground faults cleared by fuses. Chapter 3 details the implementation of the material presented in Chapter 2 and pro-

vides the overall structure of the algorithms in flow-chart form. Chapter 4 gives obtained results along with discussion. Chapter 5 concludes the document and discusses potential future work and opportunities.

CHAPTER 2

BACKGROUND

This chapter provides necessary background on IntelliRupter® PulseCloser® devices, characteristics on the event types studied in this document (faults, power quality disturbances, switching events, and harmonics), fuse analysis, the Naïve Bayes classifier, analytic signals, Root-Mean-Square (RMS) envelope, and switching event detection using first-order forward differences.

IPCR Operation

Modern power distribution networks use re-closing technology for fault isolation and self-healing. The primary function of re-closers is to open the circuit on either side of a fault once it has been detected. Thus, re-closers facilitate isolation of faulted portions of

Figure 1 S&C IntelliRupter® PulseCloser® Fault Interrupter [1]

the distribution system to the smallest area possible as well as assists in preventing the drawing of high-magnitude, source currents.

Following detection of a fault, traditional re-closers will close the circuit to determine if the detected fault is still present. This "re-closing" operation is repeated three times. If the fault is detected during the first and second "test", then the re-closer will re-open. If the faults is detected during the third and final "test", then the re-closer will enter a locked out state until the fault condition has been removed and a reset initiated by power utility personnel.

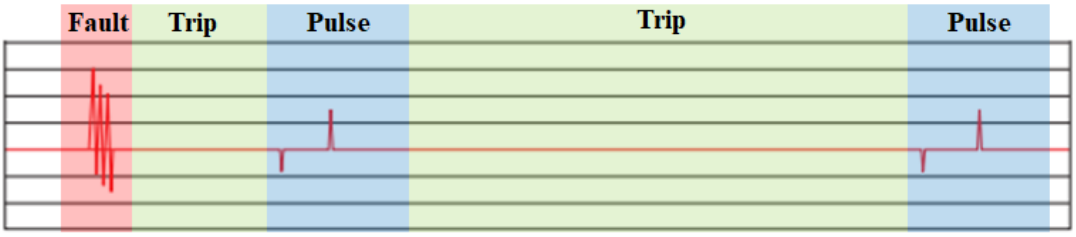
Contemporary re-closing devices, such as the IPCR (Fig. 1), provide advantages over traditional re-closing devices. These include, but are not limited to: digital current and voltage sensors for each phase, ability to integrate into a Supervisory Control and Data Acquisition (SCADA) system, and PulseClosing technology.

PulseClosing technology is particularly advantageous over traditional re-closers. PulseClosing technology, when sending a pulse into a faulted line, will allow 95% less energy than traditional reclosing technology. This helps prevent stress on equipment, e.g., transformers and generators, over time, which can otherwise lead to failures and expensive repairs. PulseClosing uses short-duration (2-8 ms) pulses of current to check for the presence of faults instead of letting large amounts of fault power back into the system [10]. Fig. 2 shows the comparison between typical re-closing and PulseClosing operation.

Fault Characteristics

Power system faults are a result of objects making contact with transmission lines in an undesired fashion. Some common causes of faults are animals, fallen or untrimmed

(a) Traditional Reclosing Current vs. Time Waveform



(b) IntelliRupter®PulseClosing Technology Current vs. Time Waveform

Figure 2 (a) Traditional vs. (b) PulseClosing Technology [2]

tree limbs, and conductor slap. Conductor slap occurs when two or more lines come in contact with each other over a span between two or more series of poles. Faults lead to problems within the affected network that include, but are not limited to: equipment damage, dangerous ground current magnitudes, and loss of power in commercial or residential areas. When a fault occurs, on one or more phases, a lower-impedance path is created leading to high amounts of current being drawn through the system. These fault currents tend to exceed maximum equipment ratings and without proper protection and control can cause irreparable or very costly damage.

Faults can be characterized as symmetrical or unsymmetrical faults. Symmetrical faults occur when all three phases make contact with each other, or when all three phases are shorted to ground (Fig. 3d). Due to all three phases being affected, the system remains balanced. Unsymmetrical faults occur when a single phase becomes shorted to ground (Fig. 3a), two phases make contact and create a closed circuit (Fig. 3b), or

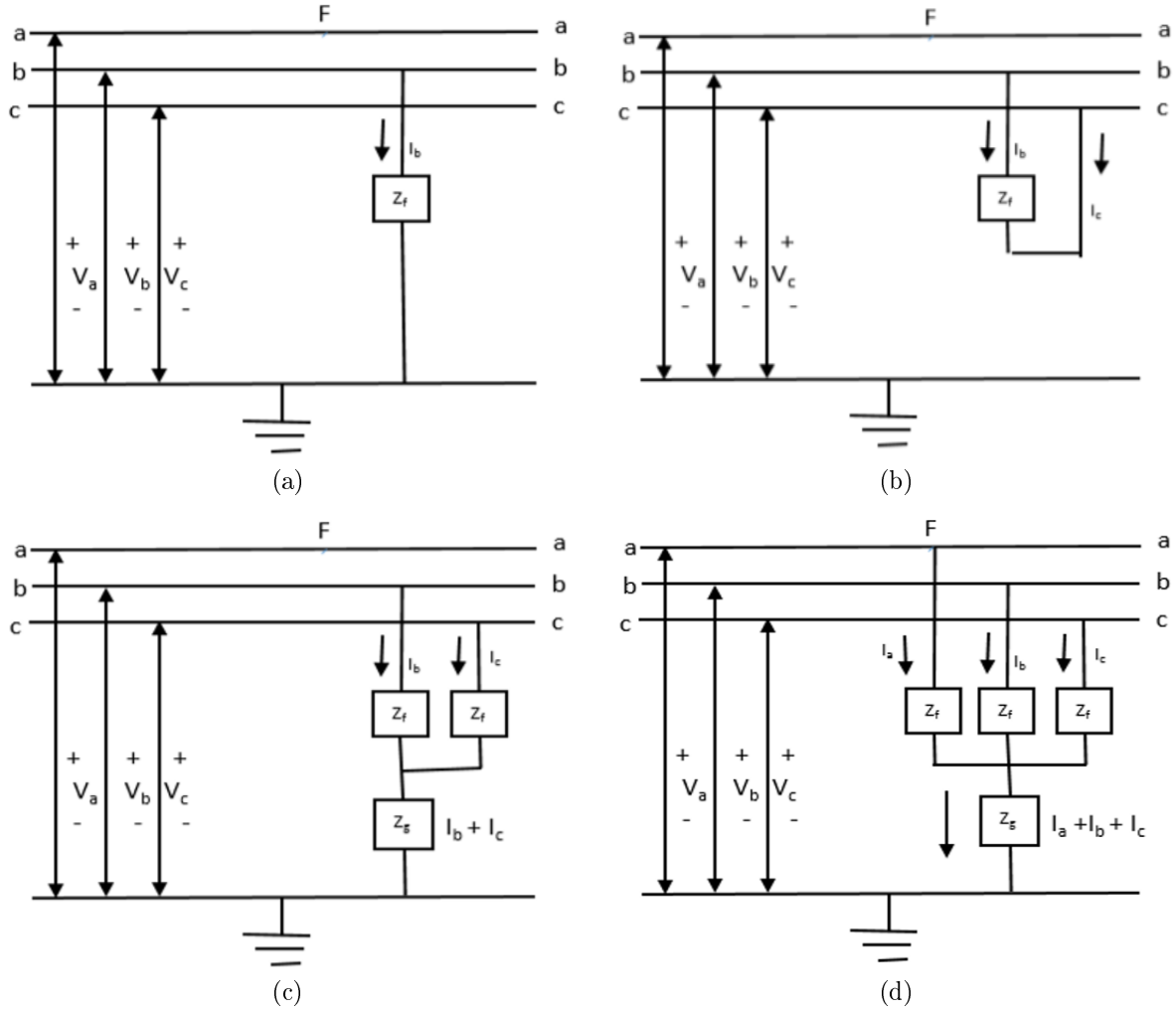


Figure 3 (a) Single-phase fault, (b) phase-to-phase fault, (c) double-phase-to-ground fault, (d) three-phase fault, [3]

two phases are both shorted to ground (Fig. 3c).

Fuse Analysis

Fuses are designed to break the flow of dangerous levels of current during faulted conditions. The S&C Positrol® fuse design employs helically-coiled silver elements designed to break at the rated current, absorb mechanical vibration, and thermal shock without causing a significant amount of damage [11].

Fuses are characterized by their respective Time-Current Characteristic (TCC) curves. TCC curves plot a fuse's minimum melting and maximum clearing times, in seconds, versus the RMS current allowed during those times. After a fuse has melted, the fault duration and RMS current value can be calculated from the IPCR recording and plotted as an (c, t) pair on the TCC curves. If the (c, t) point falls between the two curves corresponding to the same fuse size, then it is assumed that that was the size of the melted fuse. Fig. 4 shows the TCC curves for the fuse sizes that are deployed throughout power distribution network of EPB. For a given rated fuse size, the left-most (dashed) and right-most (solid) curves are designated as the "minimum-melt" and "maximum-clear" curves, respectively.

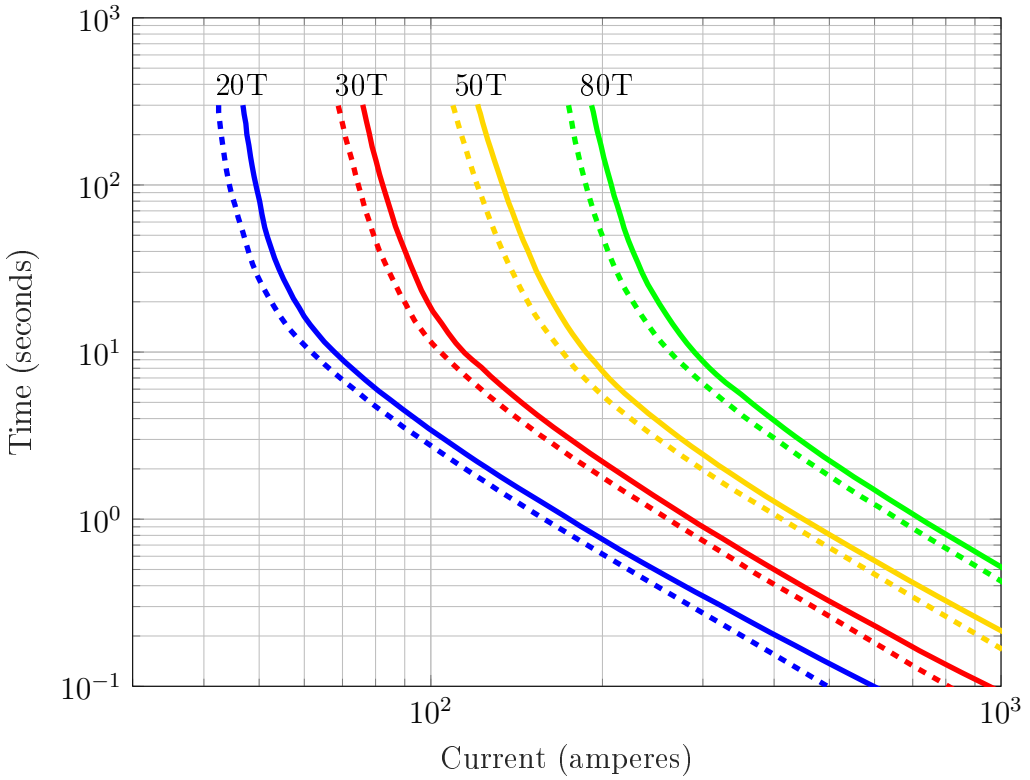


Figure 4 TCC Curves for S&C Positrol® "T" speed fuses in which the dashed curves correspond with the "minimum melt" rating and the solid curves correspond with the "maximum clear" rating

Another approach to characterizing fuses is by the amount of fault energy that is "let through", which is designated here as the Let-Through Energy (LTE). Given a high-current fault (e.g., greater than 600 amperes) that starts at time t_I and is cleared by a fuse at time t_C, then the LTE is given by [12],

$$E_L = \int_{t_I}^{t_C} I_{rms}^2 dt = I_{rms}^2 (t_C - t_I) = I_{rms}^2 t. \tag{1}$$

where I_{rms} is the RMS value of the current between times t_I and t_C, and $t = t_C - t_I$.

The Naïve Bayes Classifier

Fuse events are classified using a Naïve Bayes classifier, where the input feature is the event's LTE. The machine learning classifier known as Naïve Bayes is a probabilistic classifier based on Bayes' Theorem. Bayes' Theorem states that the probability of class label G given knowledge of training data X can be calculated using the posterior probability of X given G and the prior probabilities of X and G. The general form of Bayes' Theorem is given as [13],

$$P(G = k|X = x) = \frac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l}, \tag{2}$$

where $f_k(x) = P(X = x|G = k)$ is the posterior probability of training sample x given class k, $\pi_k = P(G = k)$ is the prior probability of class k, $x \in \mathbb{R}^p$, and the prior probability of training sample x is given by,

$$P(X = x) = \sum_{l=1}^{K} f_l(x)\pi_l. \tag{3}$$

Naïve Bayes assumes that each of the class density functions, $f_k(x)$, are products of marginal densities, i.e., a given class $G = k$,

$$f_k(X = x) = \prod_{j=1}^{p} f_{jk}(x_j). \tag{4}$$

Substituting (4) into (2) results in,

$$P(G = k|X = x) = \frac{\pi_k \prod_{j=1}^{p} f_{jk}(x_j)}{\sum_{l=1}^{K} f_l(x)\pi_l}. \tag{5}$$

Given a set of training data $\hat{X}$, the corresponding classes can be estimated by,

$$\hat{G} = \arg\max_k \left[\pi_k \prod_{j=1}^{p} f_{jk}(x_j) \right]. \tag{6}$$

The denominator in (5) is a scale factor; thus, it is neglected in (6) for computational efficiency.

Analytic Signals

Computing the LTE of a LG fault involves knowing where the "inception" and "clear" sample points lie in digital waveform. The analytic signal method was used to find these points. The "analytic" representation of a real-valued signal is a complex-valued one in which the imaginary component is simply the real-valued component shifted in phase by 90 degrees. The imaginary component is calculated via the Hilbert Transform, which introduces a 90° phase delay to all frequency components of the original signal. The Hilbert transform $\hat{x}(t)$ of a real-valued signal $x(t)$,computed by $\hat{x}(t) = x(t) \circledast h(t)$, has impulse response [14],

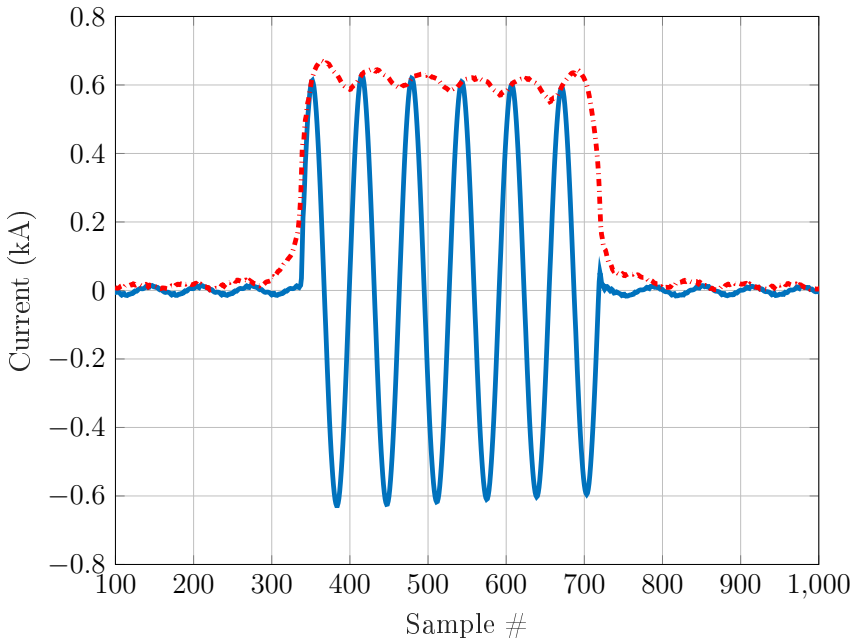


Figure 5 Fault current (blue, solid line) versus instantaneous amplitude (dashed, red line)

$$\hat{X}(f) = X(f)H(f) = -j\text{sgn}(f)X(f) = \begin{cases} -jX(f), & f > 0 \\ 0, & f = 0 \\ jX(f), & f < 0 \end{cases}, \tag{7}$$

where $H(f) = (-j)\text{sgn}(f)$ and sgn is the signum function. The time-domain impulse response of $\hat{x}(t)$ is then,

$$\hat{x}(t) = x(t) \circledast \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau. \tag{8}$$

The analytic signal is constructed as $\tilde{x}(t) = x(t) + j\hat{x}(t)$. The analytic representation of real-valued signals facilitates analysis using instantaneous information such as amplitude, phase, and frequency. In this work, only the instantaneous amplitude is used. The instantaneous amplitude of a complex-valued signal $\tilde{x}(t)$ is,

$$A(t) = \sqrt{x^2(t) + \hat{x}^2(t)}. \tag{9}$$

When $x(t)$ is a sinusoidal signal, $A(t)$ will follow the peaks of $x(t)$, but a significant portion of the oscillatory behavior of $x(t)$ will be diminished. Fig. 5 shows an example of a line-to-ground fault current waveform and its instantaneous amplitude overlaid.

Power Quality Disturbance Characteristics

Power quality (PQ) disturbances refer to changes in a voltage waveform's peak-to-peak range (i.e., amplitude) and frequency. Two very common power quality (PQ) disturbances are voltage sags and swells. Sags and swells can be harmful to industrial, commercial, and household electric loads. A voltage sag is defined as a momentary lapse in voltage with RMS values in the range of 0.1-0.9 per-unit (p.u.). An RMS voltage value of 1.1 p.u. or greater is considered a swell [15]. Fig. 6 provides a representative illustration of a voltage sag (Fig. 6b) and swell (Fig. 6c) in relation to a normal voltage waveform.

Table 1 Odd Harmonics Current Limits for Systems Rated 120 V - 69 kV

	Individual Harmonic Order					
$\boldsymbol{I_{SC}/I_L}$	$\boldsymbol{h < 11}$	$\boldsymbol{11 \leq h < 17}$	$\boldsymbol{17 \leq h < 23}$	$\boldsymbol{23 \leq h < 35}$	$\boldsymbol{35 \leq h}$	TDD
< 20	4.0	2.0	1.5	0.6	0.3	5.0
20 - 50	7.0	3.5	2.5	1.0	0.5	8.0
50 - 100	10.0	4.5	4.0	1.5	0.7	12.0
100 - 1000	12.0	5.5	5.0	2.0	1.0	15.0
> 1000	15.0	7.0	6.0	2.5	1.4	20.0

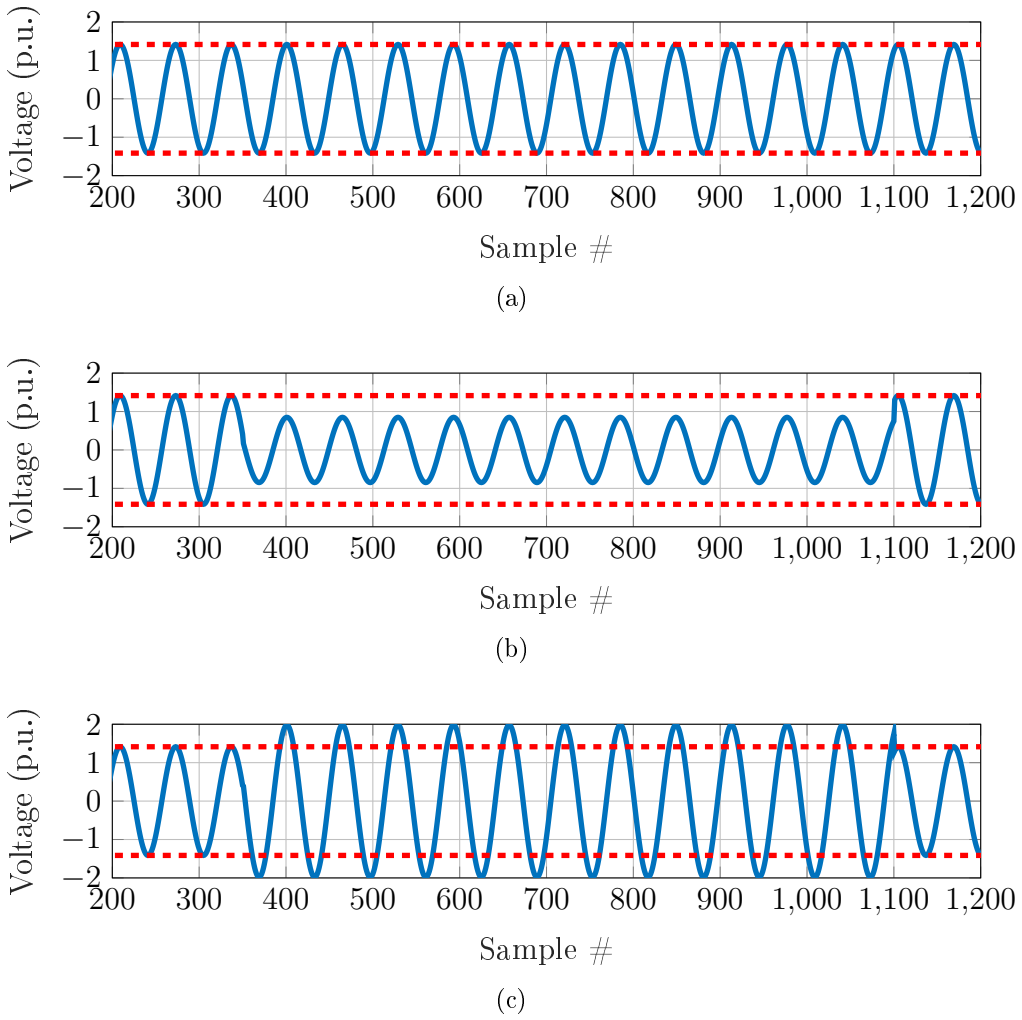


Figure 6 (a) Normal operating voltage at 1.0 p.u., (b) Voltage sag of 0.6 p.u., (c) Voltage swell of 1.4 p.u. Voltage waveforms (solid, blue line), and ±1 peak boundaries (dashed, red lines)

Harmonic Characteristics

Aside from harmful changes in voltage amplitude, changes in the frequency of a voltage waveform can also be problematic. Typically, changes in the frequency of a voltage waveform is due to harmonics. Harmonics of currents or voltages contain frequencies at multiples of the fundamental frequency, which is 60 Hertz (Hz) in the United States.

Table 2 Even Harmonics Current Limits for Systems Rated 120 V 69 kV

$\boldsymbol{h=2}$	1.0
$\boldsymbol{h=4}$	2.0
$\boldsymbol{h=6}$	3.0
$\boldsymbol{8 \leq h < 11}$	4.0
$\boldsymbol{11 \leq h < 17}$	2.0
$\boldsymbol{17 \leq h < 23}$	1.5
$\boldsymbol{23 \leq h < 35}$	0.6
$\boldsymbol{35 \leq h}$	0.3
TDD	5.0

Harmonics are typically caused by non-linear loads, the most common of which are various electronic converters that perform AC-to-AC, AC-to-DC, DC-to-AC, and DC-to-DC conversion, and variable-frequency drives. The presence of harmonics can lead to harmful effects such as higher core losses in transformers, I^2R losses in transmission lines with frequency-dependent impedance, premature circuit breaker trips and fuse melts due to increased RMS current values [16].

IEEE Standard 519-2014: "IEEE Recommended Practice and Requirements for Harmonic Control in Electric Power Systems" outlines the harmonic analysis approach used within the power industry and adopted by this work [17]. Harmonic calculations of current waveforms require a Point of Common Coupling (PCC). In the case of this work, each IntelliRupter® is considered its own PCC. Table 1 and Table 2 detail the harmonic limits for a given ratio of the rated line-to-ground short-circuit current, I_{SC}, and the RMS current value of the corresponding disturbance, I_L. Another definition for I_{SC} is the Available Fault Current (AFC), which corresponds to the short-circuit LG rated current value for the IPCR that recorded the event.

The harmonic values presented in Table 1 and Table 3 are quantified as percentages

of the fundamental frequency. Computing the harmonic components of a signal first requires the computation of that signal's Fast Fourier Transform (FFT). The FFT is a computationally-efficient method for computing the Discrete Fourier Transform (DFT), which gives the frequency spectrum (content) of a signal. The FFT returns a set of discrete points, known as "bins", each of which relates to the frequency of a the signal under analysis by:

$$k_f = \left[\frac{f * N}{F_s}\right] \tag{10}$$

where k_f is the corresponding frequency bin nearest a frequency of f Hz computed using an FFT of length N with sample rate F_s, and $[\dots]$ represents a "nearest-integer" operation.

Each harmonic amplitude value is first extracted from the FFT bins nearest each harmonic frequency (120 Hz, 180 Hz, etc.), then normalized with respect to the magnitude of the fundamental. Given an arbitrary digitized waveform $x[n]$, the harmonic components h are mathematically expressed as,

$$h[k] = \frac{|X[k]|}{|X[k_{60}]|}, k = k_{60}, k_{120}, k_{180}, \dots, \tag{11}$$

Table 3 Harmonic Voltage Distortion Limits

Bus Voltage at PCC	Individual Harmonic Distortion (%)	Total Voltage Distortion THD (%)
$\leq$ 1 kV	5.0	8.0
1.001 kV to 69 kV	3.0	5.0
69.001 kV to 161 kV	1.5	2.5
$\geq$ 161.001 kV	1.0	1.5

where $X[k]$ is the DFT of waveform $x[n]$ and is calculated by [14],

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}. \tag{12}$$

Switching Characteristics

Switching events in power systems are a result of controlled changes to the flow of power within the distribution network. This can be done manually, by operators in the field, or by the IPCR's themselves. Typically, load current may be re-routed via switching from one area to another to facilitate equipment salvage and/or repair, fault isolation, as well as meeting general load forecasting requirements.

There are seven switching categories studied in this effort. The seven switching categories are:

1. Load Shifting: Load shifting occurs when both sets of voltage sensors are reading voltage at normal operation, and the current sensors detect a deviation from its previous load value; either an increase or a decrease.

2. Energizing: Energizing occurs when all current sensors and one directional set of voltage sensors (either upstream/source or downstream/load) start in a "below sensor floor" state and energize back into a state that denotes normal operation. Sensor floor is a pre-determined value at which everything below is considered noise. For voltage sensors, this value is defined as 0.1 p.u., and for current sensors the value is set at 8.0 Amperes.

3. De-Energizing: De-energizing is the opposite of energizing in that the current and upstream or downstream voltage sensors start in the normal operating state and fall below sensor floor.

4. Return-to-Normal: A "return-to-normal" operation is when an IPCR returns to its normal operating condition after operating in another state. A return-to-normal event may happen when IPCRs belong to a Normally-Closed (NC) state or a Normally-Open (NO) state. When an IPCR returns to a NO state, the current waveform will decrease from a load state to below sensor floor. When an IPCR returns to a NC state, the current waveform will increase from sensor floor to a load state.

5. Source Return: Source return is characterized by an increase in voltage waveforms from below sensor floor. The two sub-cases for source return are:

   (a) Primary Source Return (PSR): The upstream voltages return to normal operation from sensor floor.

   (b) Alternate Source Return (ASR): The downstream voltages return to normal operation from sensor floor.

6. Loss of Source (LoS): Loss of source events occur when all of the IPCR sensors decay to below sensor floor.

7. Return of Source (RoS): Return of source events occur when all of the IPCR sensors return to normal operation from below sensor floor.

Root-Mean-Square Envelope

Throughout this research, the RMS envelope was used to facilitate threshold-based detection of voltage sags and swells, switching events, and faults to facilitate categorization of each COMTRADE file by the algorithms comprising the developed hierarchical approach. Similar to a moving average calculation, the RMS envelope is generated using

a moving rectangular window and the RMS value calculated for the discrete waveform values corresponding to the window's position. Figure 7 provides a representative illustration of an RMS envelope compared to the voltage sag waveform it was calculated. Mathematically, the RMS value at sample index k of an arbitrary digital signal $x[n]$ under a computational window containing N values can be obtained by [18],

$$x_r[k] = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1} x^2[k-n]} \tag{13}$$

The result is a much smoother waveform; thus, allowing for easier use of threshold-based techniques. For example, in Fig. 7, the RMS envelop facilitates automated determination of the discrete time values corresponding to the start and end of the voltage sag. Performing such detection on the sinusoid itself would lead to the threshold being satisfied twice over the course of just a single cycle of the waveform.

Switching Event Detection

Switching events are characterized by increases or decreases in current and/or voltage. First, the points at which the RMS current waveforms increase or decrease must be determined. These points are are designated herein as "transition points". The transition points for a given RMS envelope, $x_r[n]$, are approximated using a forward finite difference. The RMS envelope current waveforms are normalized to be in the interval $[0, 1]$. The normalized waveform is given by,

$$\bar{x}_r[n] = \frac{x_r[n] - \min[x_r]}{\max[x_r] - \min[x_r]}. \tag{14}$$

This normalized waveform is then compared with a threshold. In this work the

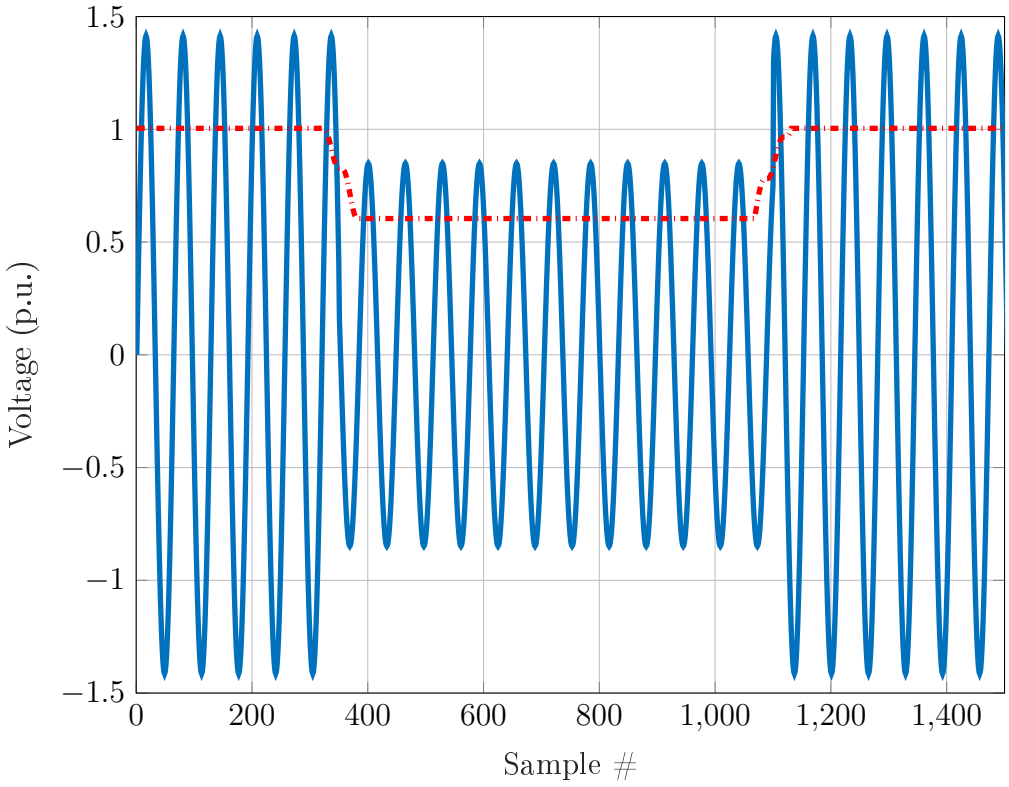


Figure 7 RMS envelope (dashed, red) superimposed on a voltage sag (solid, blue)

threshold was empirically set to a value of 0.2. Each sample of the normalized waveform, $\bar{x}_r[n]$, is compared against the threshold and a new vector generated. This new vector is of identical length to $\bar{x}_r[n]$ and its entries are either a '1' or a '0'. A '1' in the n^{th} position of this vector indicates that the n^{th} value of $\bar{x}_r[n]$ is above the threshold and a '0' indicates otherwise.

A first-order forward difference calculation is then calculated using this vector of zeros and ones. Let the vector of zeros and ones be denoted as $y[n]$, then the forward first-order difference calculation is performed simply be computing the difference between successive elements in the vector. This is effectively performing a first derivative

approximation using finite differences and a Δh value equal to 1 sample [19],

$$t[n] = y[n+1] - y[n]. \tag{15}$$

Performing a forward first-order difference on a vector of zeros and ones yields a vector of zeros and ± 1's. For example, let a current that goes from normal operation to decreasing below the threshold be denoted as $\tilde{y}[n]$ with entries around the transition point of,

$$\tilde{y}[n] = [\ldots, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, \ldots]. \tag{16}$$

The transition point is where $\tilde{y}[n]$ changes from a 1 to a 0. If the current were increasing, then $\tilde{y}[n]$ would be a string of 0's followed by a string of 1's near the transition point. Performing a forward first-order difference calculation on (16) results in,

$$\tilde{t}[n] = [0, 0, 0, 0, 0, 0, 0, 0, -1, 0, 0, 0, 0, \ldots] \tag{17}$$

The exact location of the transition point corresponds to the -1 entry in (17). Figure 8 provides a comparative illustration between $\tilde{t}[n]$ and the normalized RMS envelope of a representative current waveform that falls below the sensor floor.

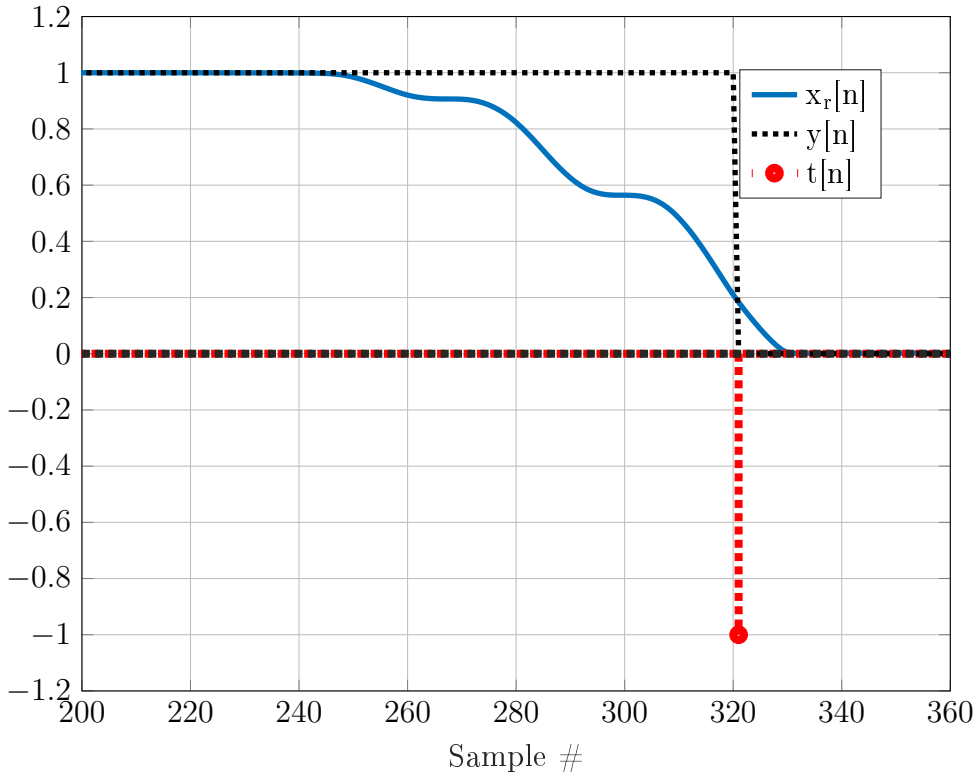


Figure 8 Normalized RMS current envelope (blue, solid line) and resulting transition point approximation (dotted, red line)

CHAPTER 3

METHODOLOGY

This chapter provides a description of the hierarchical process and algorithms that comprise it. The developed hierarchical approach categorizes COMTRADE files into one of four possible categories: invalid data, switching event, power quality, and electrical disturbance. Following this general categorization, further processing, analysis, and classification is performed that is tailored to the specific category to which the COMTRADE file was initially assigned. More detailed flowcharts may be found in Appendix A. The following sections describe the check for valid data (Pass 1), check for switching events (Pass 2), checks for faults/PQ (pass 3), and fuse forensics (pass 3), respectively.

Main Process Flow

The developed hierarchical process performs the categorization and classification of COMTRADE files using MATLAB 2017a, but is initiated by Windows Powershell 3.0. In addition to MATLAB's built-in functions, the Signal Processing Toolbox is used. PowerShell searches for new COMTRADE files within the database and creates an ordered list of these new files for subsequent categorization and analysis. Currently, Powershell performs this search every twenty-four hours. The constructed list is stored using a text file format. Each entry within this list contains the:

- Task Identification Number (Task ID): Task ID's are unique numbers assigned to each file within the list. The Task ID number increases as the list grows. This is simply a number used by the Main Process algorithm for tracking each file as they move through the process as well as their final categorization results.

- Circuit: This is the name of the circuit that the recording device operates on. This name is comprised of alphanumeric characters. The format of the circuit name is utility-specific. The device name is important for harmonics computations. As described in the "Harmonic Characteristics" section of Chapter 2, each circuit has its own LG AFC value used for I_{SC} in Table 1. These AFC values are stored in a file and are queried on based on circuit name.

- Device: This is the name of the device that made the recording. For the presented work, all of the devices are IPCR's. The device names are also combinations of alphanumeric characters. This name can indicate any field device that is capable of digitally recording anomalous electrical waveforms and storing them in COMTRADE compliant files.

- Phase Orientation: For both sets of voltages and currents, all recording devices number the individual phases numerically, e.g., 1, 2, or 3, by sensor number. However, electrical phases A, B, and C are not always connected to sensors 1,2, and 3, respectively. This field provides the mapping between the recording device's sensor number and the phase letter. Additionally, a mapping is provided for upstream and downstream voltage sensors: IPCR's record upstream and downstream voltage sets separately and each set is stored as either V_X or V_Y. This field provides a mapping in the form of "XY" if the upstream voltages and downstream voltages

are stored in V_X and V_Y, respectively. For example, an IPCR recording phases A, B, and C, with sensors 1, 2, and 3 in that order, with upstream voltages stored in V_X and downstream voltages stored in V_Y will have a field value within the list of "ABC-XY".

- Event Identification Number (Event ID): Event ID's are unique numbers assigned to COMTRADE files. The Event ID differs from the Task ID in that the Task ID is only used within the developed hierarchical process, whereas Event ID's are assigned to COMTRADE files within the database that is searched by the Powershell program. Due to this difference, Event ID's may not be listed in numerical order within the list.

- Date and Time: This field contains the date and time that the recording was made by the field device. The date and time entries are recorded in local time and have a format of YYYY-MM-DD HH:MnMn:SS, where Y represents "Year", M represents "Month", D represents "Day", H represents "Hour", Mn represents "Minutes", and S represents "Seconds".

- COMTRADE Filename: This is the name of the COMTRADE file itself and is created by EPB's databases and is comprised of the device name and event ID in the form of "DEVNAME-EVENTID.DAT"

Powershell initializes the rest of the main process following completion of the list of new COMTRADE files. A block diagram of the main process is shown in Fig. 9. Following initialization of the MATLAB based process, each entry within the list is read and the corresponding COMTRADE files loaded into the program. As described in the "Contributions" section of Chapter 1, each COMTRADE file then undergoes a series

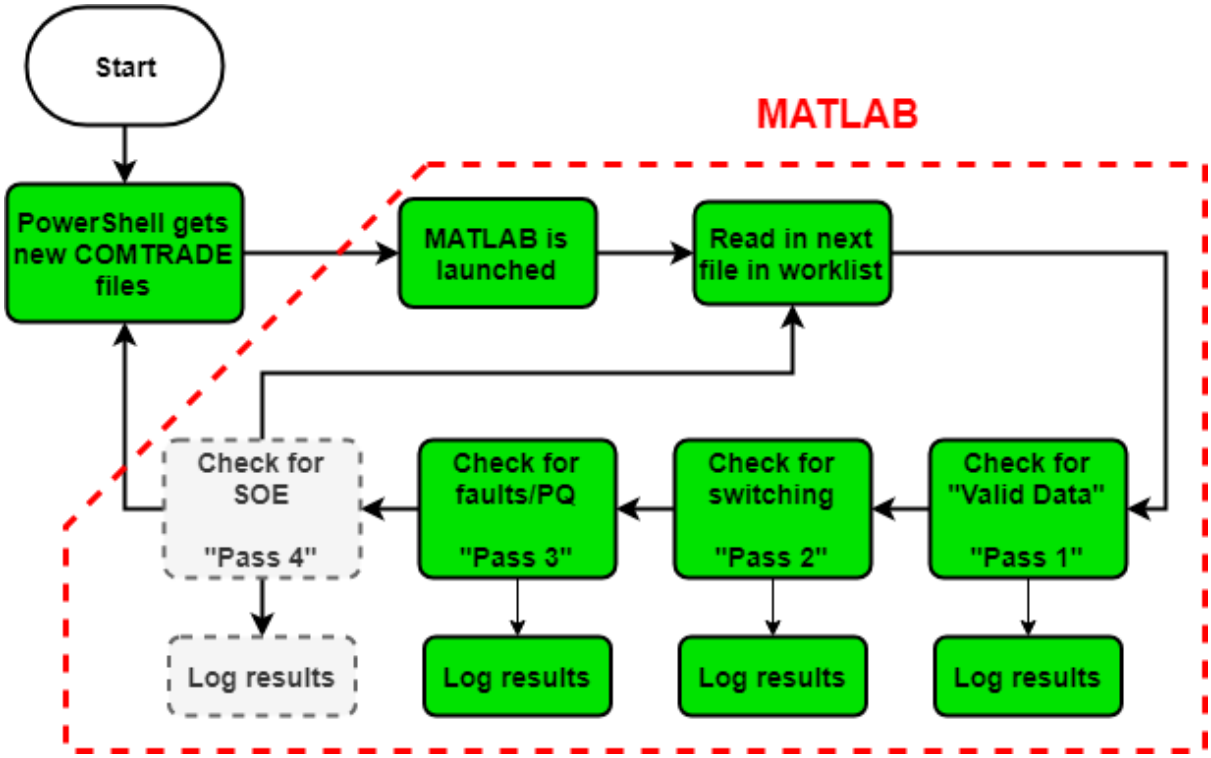


Figure 9 Main Process Flowchart

of three "passes": valid data, switching event, and faults and/or PQ, Fig. 9. As shown in Fig. 9, a fourth pass is present and denoted by the dashed block. This fourth pass is denoted as Sequence-of-Events (SoE) and is not implemented within the developed hierarchical process, but left to future work. The SoE pass is intended to handle events that span two or more COMTRADE files; thus, all of the files are required to facilitate analysis and categorization of the event.

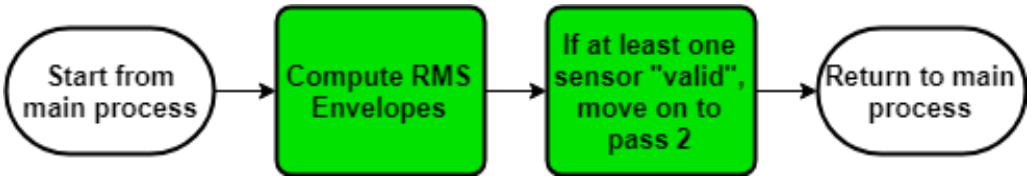


Figure 10 Main Pass 1 Flowchart - Check for "Valid Data"

Pass 1: Check for "Valid Data"

The purpose of this pass is the identify COMTRADE files that do not contain useful data. Figure 10 provides the general approach implemented within this pass. The lack of useful data occurs when the recorded waveform values fall below a threshold, defined by EPB, known as the sensor floor. For voltage recordings, the sensor floor value is 0.1 p.u. For current recordings, this sensor floor value is set at 2 A. This pass prevents unnecessary processing of COMTRADE files that only contain recordings of sensor floor waveforms.

For a given COMTRADE file, the check for "valid data" is performed by computing the RMS envelope of every recorded voltage and current waveform, as described in the "RMS Envelope" section of Chapter 2. Each RMS waveform is subsequently compared against their respective voltage or current sensor floor threshold value. RMS envelope values that exceed the threshold are assigned a true logical value and all others a false logical value. If at least one hundred true logical values are identified for at least one recorded waveform, then the COMTRADE file is designated as containing "valid" data and is passed to Pass 2: Switching Events for further processing. If this case is not met, then the a log entry is created identifying the COMTRADE files as not containing useful data. Figure 11 shows an example of a COMTRADE file recording containing invalid data.

Pass 2: Check for Switching Events

The purpose of this pass is to check for switching events to facilitate switching event specific analysis and identification. A high-level overview of this pass is shown in Fig. 12. The specific switching events that can be identified by the developed hierarchical

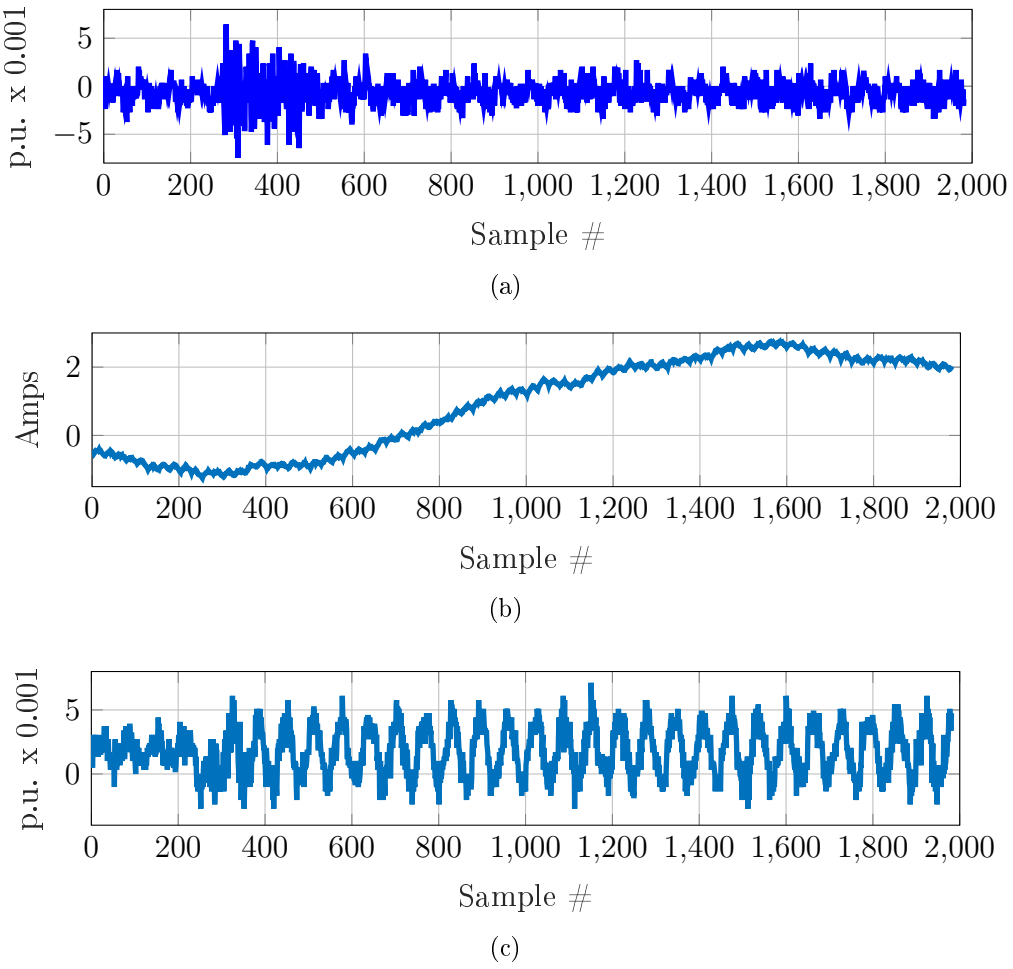


Figure 11 Single-phase recording of "invalid data": (a) Source-side voltage recording, (b) Current recording, and (c) Downstream voltage recording

approach are: load shifting, energizing, de-energizing, return-to-normal, source return (primary and alternate), loss of source, and return of source. A detailed description of each of these switching events is provided in the "Switching Characteristics" section of Chapter 2. Switching events are a result of controlled changes in the network and are typically characterized by an increase or decrease in current and/or voltage.

First, the RMS envelope is calculated for every current and voltage waveform within the COMTRADE file as described in the "RMS Envelope" section of Chapter 2. Following generation of the RMS envelopes, the transition points for every envelope is

computed using the forward first-order difference in (15), described in the "Switching Event Detection" section of Chapter 2. From these points, it can be determined if each waveform's RMS envelope is increasing or decreasing. The various switching cases are characterized as:

- Load Shifting: Load increasing/decreasing occurs when both sets of voltage sensors remain at normal operating levels but the current waveforms increase or decrease. Currents never dip below sensor floor level, as this is a simple adding or subtracting of load. After computing the RMS envelopes with a window size of $N = 64$ of the voltage and current waveforms, a power calculation is performed at the 3^{rd} and 3^{rd}-to-last cycle. At a sampling rate of 64 samples per cycle, the last sample

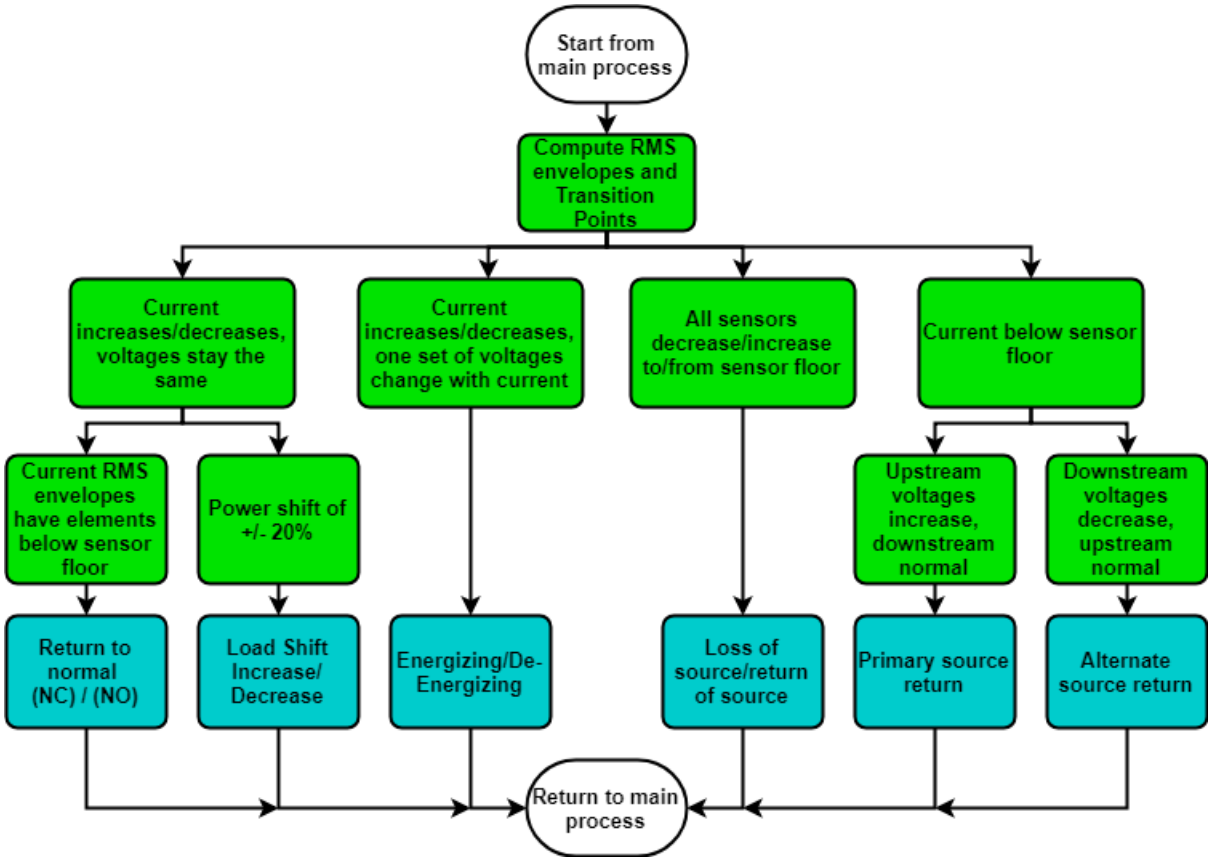


Figure 12 Main Pass 2 Flowchart - Check for Switching Events

of the third cycle will have an index of $3 * 64 = 192$, and the 3^{rd}-to-last cycle will have an index of $(M - 3) * 64$, where M is the number of cycles contained in the COMTRADE file. Most COMTRADE files sampling at 64 samples per cycle will have approximately 30 cycles per COMTRADE file. However, this number is not fixed and has to be calculated dynamically as:

$$M = \left\lfloor \frac{number\ of\ sample\ in\ of\ file}{64} \right\rfloor \tag{18}$$

where $\lfloor \ldots \rfloor$ denotes a "flooring" or "round-down" operation.

An average power calculation at each sample index is calculated by [20]:

$$P[n] = V_{RMS}[n] I_{RMS}[n]. \tag{19}$$

The 3^{rd} and $3rd$-to-last cycles of $P[n]$ are extracted at $n = 192$ and $n = (M - 3) * 64$, respectively.

If the transition points indicate an increase in current, a comparison of the power quantities at the 3^{rd} and 3^{rd}-to-last cycle are used to determine if this was an increase in load, or not a switching event. This is important as load current may sometimes drift above its normal operating point, however not due to a load shift. It was defined by EPB that if the power value at the 3^{rd}-to-last cycle, P_{M-3}, exceeds the power at the 3^{rd} cycle, P_3, by 20%, an load increase has occurred. This shift in power, ΔP_{inc}, is calculated as

$$\Delta P_{inc} = 100\% \times \frac{|P_{M-3} - P_3|}{P_3} \tag{20}$$

If ΔP_{inc} does not exceed 20%, this event is classified as not switching.

Similarly, if the current transition points indicate a decrease in current, the 3^{rd} and 3^{rd}-to-last power calculations are again performed. In the case of a load decrease, if the percentage of decrease in power between cycles 3 and $(M-3)$, ΔP_{dec}, meets or exceeds 20%, where

$$\Delta P_{dec} = 100\% \times \frac{|P_3 - P_{M-3}|}{P_{M-3}}, \tag{21}$$

this event is classified as a load decrease. Otherwise, it is classified as not switching.

- Energizing: Energizing occurs when the current waveforms and one set of voltage waveforms increase from below sensor floor to a normal operating level. The mean values of all samples in the RMS envelopes prior to the transition points is first calculated. If the mean values for both the current RMS envelopes and one set of voltage RMS envelopes before the transition points lie below the sensor floor, the event is classified as energizing.

- De-energizing: De-energizing is the exact opposite; the current waveforms and one set of voltage sensors decrease to below sensor floor from a normal operating level. The mean values of all samples in the RMS envelopes after the transition points is first calculated. If the mean values for both the current RMS envelopes and one set of voltage RMS envelopes after the transition points lie below the sensor floor, the event is classified as de-energizing.

- Loss of Source: A "loss-of-source" event occurs when all nine sensors decay to below sensor floor. If the mean values of all RMS envelopes after their respective transition points are below sensor floor, this event is classified as a loss of source

event.

- Return-of-source: A "return-of-course" event occurs when all nine sensors return to normal operating levels from below sensor floor. If the mean values of all RMS envelopes prior to their respective transition points are below sensor floor, this event is classified as a return of source event.

- Source Return: A "primary source return" or "alternate source return" event occurs when the upstream voltage waveforms or downstream voltage waveforms, respectively, return to a normal operating level from below sensor floor. This case also requires transition points that indicate an increase in voltage. An event is classified as primary source return when the mean values of the source-side RMS voltage envelopes prior to their transition points are below sensor floor. Similarly, an event is classified as alternate source return if the mean values of the downstream RMS voltage envelopes prior to their transition points are below sensor floor.

- Return-to-Normal: A "return-to-normal" event occurs when both sets of voltage waveforms remain in normal operating levels throughout the duration of the COMTRADE file, and the current waveforms increase from below sensor floor in the case of a normally-closed (NC) device, or decrease to below sensor floor in the case of a normally-open (NO) device. More specifically, when a device operates in a "normally-open" state, it is not passing load through it, whereas a "normally-closed" device is passing load through it. A "return-to-normal" event on a normally-closed device is the result of an IPCR closing back into a circuit after a fault or some other form of disturbance has been cleared. To be classified as a

return-to-normal (NC) event, the transition points of the current RMS envelopes must indicate an increase, and the mean values prior to the transition points must lie below sensor floor. Similarly, for a return-to-norm (NO) event, the transition points of the current RMS envelopes must indicate a decrease, and the mean values after the transition points must lie below sensor floor.

- Not switching: If none of the above conditions are satisfied, the event is tagged as not switching and is moved on to Pass 3.

Pass 3: Faults & Power Quality

If a COMTRADE file is identified as containing valid data, but was not identified as one of the switching events within Pass 2, then the file moves on to Pass 3. Figure 13 provides a simplified flow chart of Pass 3. Pass 3 analyzes the file for faults and Power Quality (PQ) events. As with Pass 2, this pass begins with the calculation of the RMS envelope for every waveform. Following calculation of the RMS envelopes, the RMS envelopes of the current waveforms are checked for a fault. A fault is present within the current waveform if its RMS envelope exceeds a threshold of 600 A RMS for one half-cycle or more. This value was chosen because EPB's deployed IPCR protections employ a 600 A RMS phase current "pickup" value.

For the case when two or more current waveforms contain faults, it must be determined whether this represents separate single-phase faults or the affected phases are simultaneously faulted. The use of a "stair-step" provides a simple method by which to distinguish the single-phase fault case from the other. Using the RMS envelope, a binary vector of zeros and ones is constructed in which the n^{th} entry is a one if the n^{th} value of the corresponding current RMS waveform is greater than the fault threshold of 600 A RMS. This binary vector is generated for each of the current waveforms and the resulting

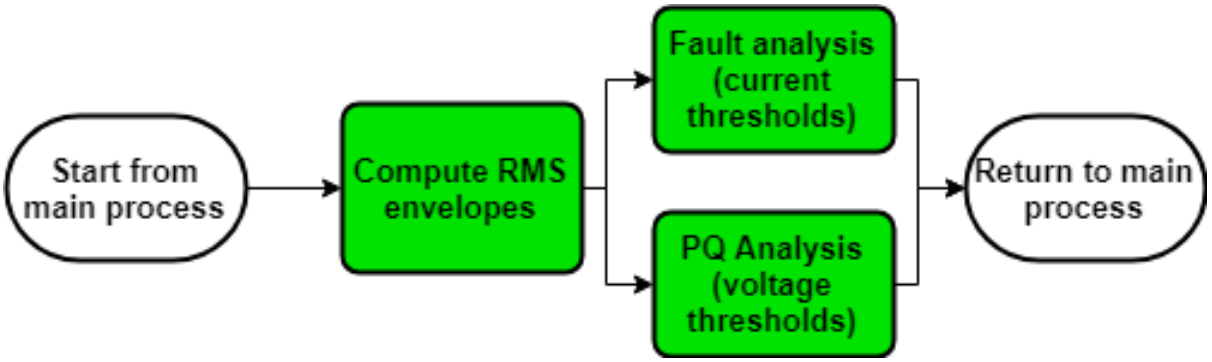


Figure 13 Overview flowchart of Pass 3: Check for Faults and Power Quality

vectors summed together element-wise. If $I_1[n]$, $I_2[n]$, and $I_3[n]$ represent these true/false fault vectors for current sensor recordings 1, 2, and 3, respectively, the resultant fault vector can be computed as,

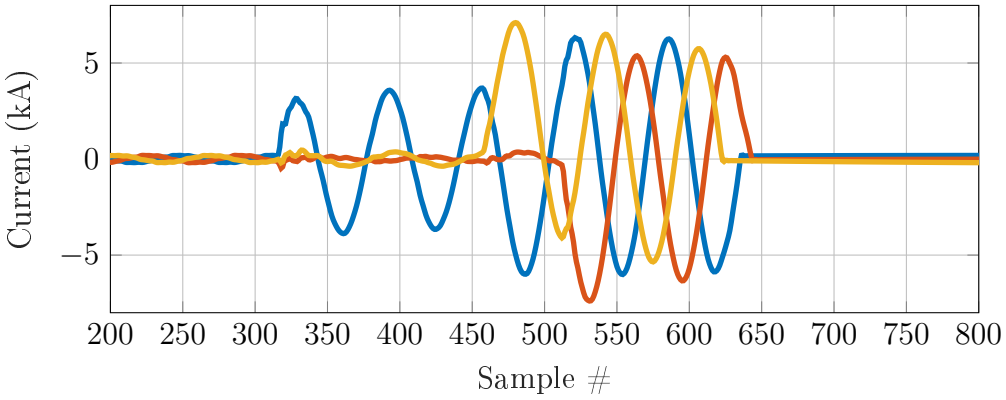
$$I_f[n] = \sum_{k=1}^{3} I_k[n], \;\; n = 1, 2, 3, \ldots N, \tag{22}$$

where N is the total number of samples. If the sum of these vectors is two or three for any point or series of points, then a line-to-line or three phase fault has occurred, respectively. Figure 14 provides a representative illustration of a three phase fault case. If the sum of these vectors results in one fourth of a cycle's (16 samples at a rate of 64 samples per cycle) worth of consecutive samples equal to 1, then the single-phase fault case is identified and fuse forensics is performed

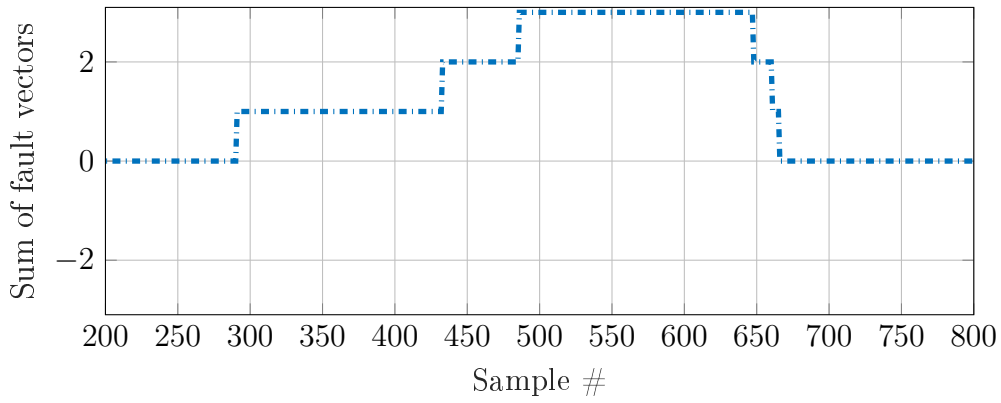
.

Pass 3: Single-Phase Faults - Fuse Forensics

A COMTRADE file undergoes Fuse forensics when a single-phase fault is detected. Instead of the RMS envelope, fuse forensics is performed using the normalized, instantaneous amplitude, (9), of the faulted current waveform. If $A(t)$ represents the instanta-



(a)



(b)

Figure 14 (a) Three phase fault (solid lines), and (b), sum of fault vectors (dashed line)

neous amplitude of the faulted current waveform, it is normalized as:

$$A_n(t) = \frac{A(t) - \min[A]}{\max[A] - \min[A]}. \tag{23}$$

Normalization ensures that all of the instantaneous amplitude values are within the interval of [0,1] as well as uniformity across all potential faulted waveforms. This allows for easier threshold-based detection independent of the load current value. The normalized, instantaneous waveform is compared with a threshold value of $A_n(t) = 0.4$ (unitless). If any of the normalized, instantaneous amplitude values exceed this threshold value it is flagged as true and false otherwise. Then the discrete time entries

corresponding to the first and last true values are taken as the inception and clear points of the fault, respectively.

After the inception, t_I, and clear, t_C, points are determined, then the LTE of the original faulted waveform between these two points is calculated using (1). When calculating LTE, the RMS value of the load current is subtracted from the RMS fault current value to ensure that the LTE calculations are independent of variable load currents across different events.

A Naïve Bayes model was trained using 397 LTE values across seven classes: 20T, 30T, 40T, 50T, 65T, 80T, and 100T. The set of LTE calculations are randomly scrambled to avoid inadvertently biasing classifier training. Approximately 25% of the overall amount of LTE values were selected for using in training the Naïve Bayes model. The remaining LTE values were used for classification and are each designated as $E_L^?$ when being compared with the developed model. A new $E_L^?$ is assigned to the class, k, which resulted in maximizing (6). An example plot of a fuse forensics report generated from the hierarchical process on a real fault is given in Appendix B.

Pass 3: Power Quality

If none of the current phases are faulted, the PQ analysis is performed. PQ analysis identifies sags and/or swells present within the voltage and current waveforms. A sag is present when the RMS voltage waveforms have at least a half-cycle number of samples (32 samples at a rate of 64 samples per cycle, or 8.3 ms) between 0.1 and 0.9 p.u. A swell is detected when an RMS voltage waveform has at least a half-cycle number of samples above 1.1 p.u. Sags and swells are only looked for in voltage waveforms, as current waveforms aren't affected in the same way due to Ohm's Law.

Pass 3: Harmonics

A particular area of interest of EPB was to be able to detect harmonics in the two cycles prior and leading up to faults and/or PQ disturbances. There is not much interest in harmonic components present during faults, as they are of lower priority in terms of potential harm than faults and PQ disturbances.

First, the LG AFC value for the device whose recording is being analyzed is obtained from the external file. This value, denoted by I_{SC} in Table 1 is an RMS current, expressed in amperes. Next, the value of I_L, also in Table 1, is calculated by,

$$I_L = \sqrt{\frac{1}{L} \sum_{n=t_s}^{t_s+L-1} I_d^2[n]} \tag{24}$$

where $L = 128$ is two cycles, t_s is the point two cycles' worth of samples prior to the start of the disturbance, and I_d is the current waveform at the disturbed phase $d = 1$, 2, or 3.

The starting point is saved from the previous fault and/or PQ analysis from Pass 3. Next, the ratio I_{SC}/I_L is computed. If this ratio falls within one of the ranges depicted in the first column of Table 1, the corresponding harmonic limits depicted in that row are used as the "thresholds". If any of the calculated harmonic values at the frequency bins, as described in the "Harmonic Characteristics" section of Chapter 2, exceed these values, a flag is raised and a log entry is created stating that the disturbed current waveform has harmonic components exceeding the limits.

Voltage harmonics are computed in the same way, using Table 3. The line-to-line bus voltage at each PCC is 12.4 kV, which equates to approximately a line-to-neutral bus voltage 7.2 kV; thus, the 2^{nd} row of Table 3 is used to determine harmonic limits.

Therefore, if any of the voltage harmonics exceed 3% of the fundamental, a flag is raised and a log entry is written stating that voltage harmonics are present.

CHAPTER 4

RESULTS AND DISCUSSION

Hierarchical Process Test Results

Testing and verification of the developed hierarchical process was conducted using 140 randomly chosen COMTRADE files for which the event contained in each file is known and verified by power personnel. This verification process took approximately 25 hours over 4 days. These COMTRADE files were placed into a worklist using PowerShell, input into the process beginning with MATLAB as described in the "Main Process" section of Chapter 3, and each individual file processed through every pass as described in Chapter 3. The logged results from the hierarchical process were then compared with the known and verified event type. The results presented here are broken down into five categories: invalid data, switching events, faults, PQ, and unclassified. Unclassified is defined as the case in which a given event was not assigned to any of the categories described in the Methodology section. Four unclassified events were purposefully included in the set of 140 files. Table 4 presents the overall classification performance results for the developed approach. Of the 140 total COMTRADE files processed, 92% of them were assigned to the correct category.

Files containing faults were either: line-to-ground, line-to-line, and line-to-line-to-line. PQ events are recorded as either a sag or a swell. Switching events contained in the dataset belonged to return-to-normal or loss/return of source. No energizing/de-energizing, alternate/primary source return, or load shifts were found for this dataset, as

Table 4 Hierarchical Classification Results

	Invalid Data	Switching	Faults	PQ	Unclassified	Total
Number of Events	2	37	68	29	4	140
Correct	1	33	67	26	3	130
Percent Correct	50	86.49	98.53	89.66	75	92.86

they are considerably rarer than return-to-normal operations. A case-by-case breakdown is given in Tables 5-7.

In Table 5, 67 out of 68 total faults were classified as either LG, LL, or LLL correctly, for an overall classification rate of 98.53%. Only one LG event was misclassified. The misclassified event contained a LG fault, but at the very end of the file, less than half a cycle of a LL fault had begun. Therefore the classifier tagged that event as a LL event. This is an example of where "SoE processing" will come into play. However, all 14 multi-phase faults were classified correctly.

In Table 6, the total correct classification rate is given as 89.66%. No "swell-only" events were found in this dataset. However, 10 out of the 29 PQ events were correctly classified as "sag and swell". This means that, in one COMTRADE file, one or more phases is sagged, and one or more phases is swelled. Three of the sagged events were misclassified. These misclassified sag events also contained current waveforms that slightly increased or decreased, and were thus classified as load shifts.

As mentioned above, energizing, de-energizing, primary source return, alternate source return, and load shifting events were not found for this dataset. Table 7 shows that the system classified the set of switching events correctly 86.49% of the time, correctly classifying 33 of 37 total switching events. In the case of one misclassified event, a ROS was classified as a load shift. The remaining misclassified events were RTN-NC

or RTN-NO events that were classified as load shifts. One possible explanation for this is shifts in sensor floor values that exceed the previously-set values. If a switching event that contains a RTN-NC has values prior to the increase in current greater than the sensor floor thresholds it will be (mis)classified as a load shift. Dynamic sensor floor "drifts" are something to be addressed in future work.

Table 5 Fault Classification Results

	LG	LL	LLL	Total
Number of Events	54	11	3	68
Correct	53	11	3	67
Percent Correct	98.15	100	100	98.53

Table 6 Power Quality Classification Results

	Sag	Sag and Swell	Total
Number of Events	19	10	29
Correct	16	10	26
Percent Correct	84.21	100	89.66

Table 7 Switching Classification Results

	LOS	ROS	RTN-NC	RTN-NO	Total
Number of Events	1	2	18	16	37
Correct	1	1	15	16	32
Percent Correct	100	50	83.33	100	86.49

Fuse Forensics Results

The LTE of 397 total fuse events representing the seven different fuse sizes of: 20T, 30T, 40T, 50T, 65T, 80T, and 100T, were used for training and validation of a Naïve

Bayes classifier model. The results are given in Table 8.

The rows of the table represent the "actual" class and the columns represent the "predicted" class. Overall, 94.67% of the validation set (approximately 298 fuse events) were classified correctly. The 20T class performs poorly relative to the other class sizes. The 20T fuse had the fewest number of events, 11, which may have contributed to the poorer percent correct classification performance. Some potential enhancements to this process to improve performance include: updating prior probabilities, π_k when new data is input to the classifier, and performing cross-validation to better train the model. Other potential sources of misclassification error are:

1. Incorrect fuse sizes being used to replace older, blown fuses

2. Partial melting of fuse links, or multiple partial meltings over time. This can lead to fuses melting and therefore interrupting a fault at a faster rate than the fuse's specifications according to its TCC curve.

A breakdown of the number of events per fuse size is presented in Table 9.

Table 8 Percent Correct - Fuse Forensics

Percent Correct (%)							
	Predicted						
Actual	20T	30T	40T	50T	65T	80T	100T
20T	87.5	12.5	0	0	0	0	0
30T	0	100	0	0	0	0	0
40T	0	8.00	88.00	4.00	0	0	0
50T	0	0	2.99	95.52	1.49	0	0
65T	0	0	0	4.44	95.56	0	0
80T	0	0	0	0	1.45	98.55	0
100T	0	0	0	0	2.44	0	97.56
Average	94.67%						

Table 9 Number of Events per Fuse Size

Fuse Size								
	20T	30T	40T	50T	65T	80T	100T	Total
Number of Events	11	55	38	90	61	90	52	397

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

Conclusions

A monthly average of 2,100 COMTRADE files are recorded by operational IPCR's within EPB's power distribution network. Manpower limitations constrain a utility's ability to analyze all events that are obtained from the field. The proposed hierarchical system correctly categorized and identified approximately 92% of the 140 files from the categories of: faults, switching, PQ, invalid data, and unclassified events.

The hierarchical system proposed in this work facilitates processing of COMTRADE files at a rate of approximately 1.78 seconds per file. This allows power utilities to reduce operational costs in terms of reduced person hours (between $250,000 and $500,000 annually). It also allows for system improvements to be made based on available (classified) data, preventing asset failure, improving customer service through availability of PQ data, and potentially preventing harmful attacks. Previously, utility engineers would need hours, or even days to process an amount of files that the developed system can process and classify in a matter of minutes. Due to the "by-hand" nature at which this analysis takes place, it typically takes a back-seat to other every-day duties performed by the engineers. This leaves a lot of unprocessed information sitting in a data-base that is not being analyzed for useful and actionable intelligence.

Future Work

All of the events studied in this research and classified using the proposed system are only a subset of the various types of Electrical Disturbances (EDs) that may take place within a power distribution network. EDs that were not studied within this work are left to future efforts, but presented here as s concise list. This list will enable future researchers to more easily develop and integrate techniques by which to process these remaining EDs within the developed hierarchical process.

- Pulse-closing events: As described in Chapter 2, IPCR's send out pulses of current to determine if a fault is still present. There are twelve unique events associated with IPCR events.

- Multi-phase grounded faults: Fault analysis in this document was only concerned with line-to-ground (LG), line-to-line (LL), and three-phase un-grounded (LLL) faults. The ability to distinguish between grounded and un-grounded faults is important for analysis as well as for public safety. The presence of fault current in the “ground" is potentially harmful for humans or animals nearby.

- Capacitor Switching and Ringing Capacitor Switching/Ringing, like harmonics, are to be treated as “artifacts”, rather than individual events; meaning, they are a reaction or consequence of some other type of disturbance that has occurred, such as a fault or switching event. Capacitors have the ability to discharge into a fault, contributing harmonics, and therefore higher losses.

- Transformer Demagnetization: Upon re-energizing a a magnetized transformer, such as during a re-closing operation, the core may become saturated, which will produce high-magnitude inrush currents. This is due to the non-linear nature of

core saturation. If the re-closing device fails to trip at the zero-crossing of the fault current, it may induce a DC bias in the post-fault current due to the inrush currents. [21]

One important feature that is to be incorporated in the future is the ability to address "drifting" sensor floor or noise levels. Sensor floor levels that drift above the set values can cause mis-classifications in both Pass 1 and Pass 2. For example, it may send events that contain invalid data on to the next portion of the process, or it may classify a RTN-NC or RTN-NO as a load shift increase or decrease, respectively.

The ability for the developed process to perform Sequence-of-Events (SoE) processing is a necessity. EDs can span multiple COMTRADE files; thus, there is a need for the development of an algorithm capable of "stitching" together multiple COMTRADE files prior subsequent processing. This stitching process must be able to track the time-stamp and IPCR ID.

The hierarchical approach presented here was implemented using MATLAB R2017a with the Signal Processing Toolbox. However, MATLAB costs $860 and $2,150 for an annual and perpetual commercial license, respectively [22]. This does not include any toolboxes and the annual license is unusable at expiration of the license. This make it difficult and even prohibitive for many power utilities to adopt the developed approach. Therefore, conversion of the MATLAB portions of the approach to an open-source language would ease adoption by other power utility companies. Some open-source languages to be considered are: Python, R, C, and Java.

Currently, the system resolves a COMTRADE file to a single category. However, for many files this is not the case. A file may contain multiple events happening simultaneously. One potential method for simplifying the current logic is to perform a large

number of "narrower" measurements, each of which would be called by its own function and would return a true or false value. Some examples of these measurement checks may potentially include:

- No source
- Voltage present in both directions
- Voltage present in source/downstream direction only
- Sag recorded by upstream device
- Sag recorded by parallel device
- Sag recorded by “this” device
- Fault recorded by “this” device
- etc.

Each check would be performed by its own function, independent of all of the others. The classifier could then make decisions based on all of the functions that returned a "true" value, rather than attempting to resolve to a single value in the existing framework. This allows for the addition and removal of individual measurement functions without affecting dependency on others. A rough flowchart of this process is given in Fig. 15.

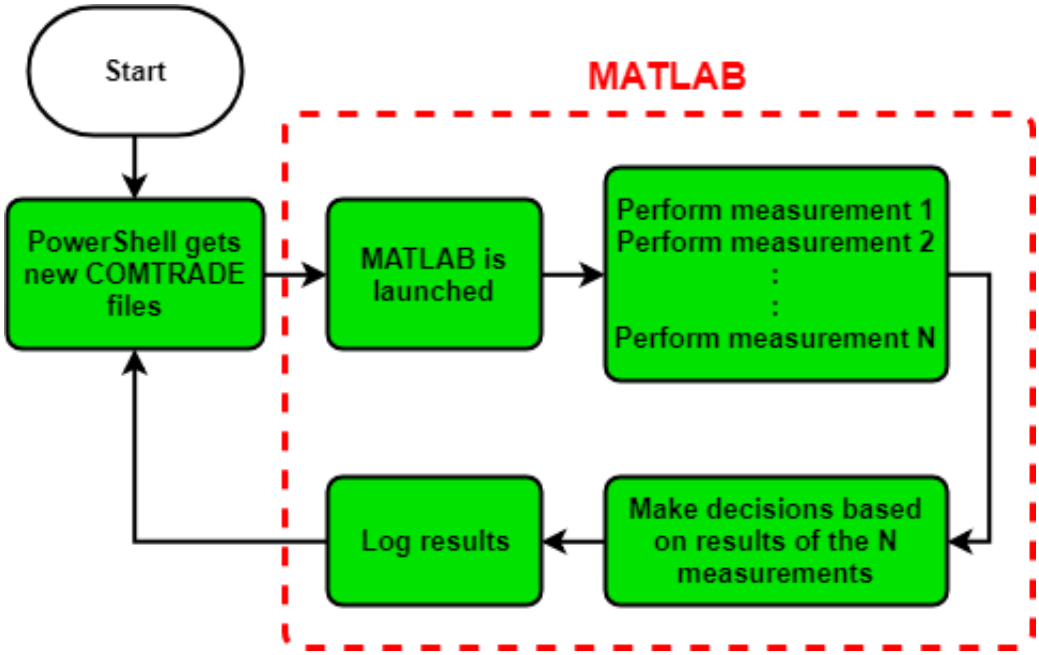


Figure 15 Potential flowchart for future code structure

REFERENCES

[1] D. Klein, "Taming the physical stress from fault current," 2016.

[2] "IntelliRupter® PulseCloser® Fault Interrupter," 2019.

[3] S. Ghimire, "Analysis of fault location methods on transmission lines," 2014.

[4] "IEEE/IEC measuring relays and protection equipment part 24: Common format for transient data exchange (comtrade) for power systems - redline," IEEE Std C37.111-2013 (IEC 60255-24 Edition 2.0 2013-04) - Redline, pp. 1–136, 2013.

[5] S. Katyara, F. Akhtar, S. Solanki, L. Zbigniew, and L. Staszewski, "Adaptive fault classification approach using digitized fuzzy logic (dfl) based on sequence components," in 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Conference Proceedings, pp. 1–7.

[6] M. Karimi, M. Banejad, H. Hassanpour, and A. Moeini, "Classification of power system faults using ann classifiers," in 2010 Conference Proceedings IPEC, Conference Proceedings, pp. 505–508.

[7] A. Agüera-Pérez, J. Carlos Palomares-Salas, J. J. G. de la Rosa, J. María Sierra-Fernández, D. Ayora-Sedeño, and A. Moreno-Muñoz, "Characterization of electrical sags and swells using higher-order statistical estimators," Measurement, vol. 44, no. 8, pp. 1453–1460, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0263224111001709

[8] S. Mishra, "Detection and classification of power quality disturbances using s-transform and probabilistic neural network," in 2009 IEEE/PES Power Systems Conference and Exposition, Conference Proceedings, pp. 1–1.

[9] V. Thiyagarajan and N. P. Subramaniam, "Wavelet approach and support vector networks based power quality events recognition and categorisation," in 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Conference Proceedings, pp. 1667–1671.

[10] "Detailed functional specification guide." [Online]. Available: https://www.sandc.com/globalassets/sac-electric/documents/sharepoint/documents—all-documents/information-bulletin-766-451.pdf

[11] "S&C Positrol "T" Speed Fuse Link," 2018. [Online]. Available: https://www.sandc.com/globalassets/sac-electric/documents/sharepoint/documents—all-documents/descriptive-bulletin-352-30.pdf

[12] R. H. Kaufmann, "The magic of i2t," IEEE Transactions on Industry and General Applications, vol. IGA-2, no. 5, pp. 384–392, 1966.

[13] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[14] M. F. Mesiya, Contemporary Communication Systems, 1st ed. McGraw Hill, 2013.

[15] "IEEE recommended practice for monitoring electric power quality," IEEE Std 1159-2009 (Revision of IEEE Std 1159-1995), pp. c1–81, 2009.

[16] N. Shah, "Harmonics in power systems: Causes, effects and control," 2013.

[17] "IEEE recommended practice and requirements for harmonic control in electric power systems," IEEE Std 519-2014 (Revision of IEEE Std 519-1992), pp. 1–29, 2014.

[18] S. M. Deckmann and A. A. Ferrira, "About voltage sags and swells analysis," in 10th International Conference on Harmonics and Quality of Power. Proceedings (Cat. No.02EX630), vol. 1, Conference Proceedings, pp. 144–148 vol.1.

[19] M. S. Gockenbach, "The forward difference," 2003. [Online]. Available: http://pages.mtu.edu/ msgocken/ma5630spring2003/lectures/diff/diff/node2.html

[20] T. J. O. J. Duncan Glover, Mulukutla S. Sarma, Power Systems Analysis and Design, 5th ed. Cengage Learning, 2018.

[21] D. M. Robalino, "Power transformer demagnetization," in 2016 IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI), Nov 2016, pp. 1–5.

[22] MATLAB, "Pricing and licensing."

APPENDIX A

ALGORITHM FLOWCHARTS

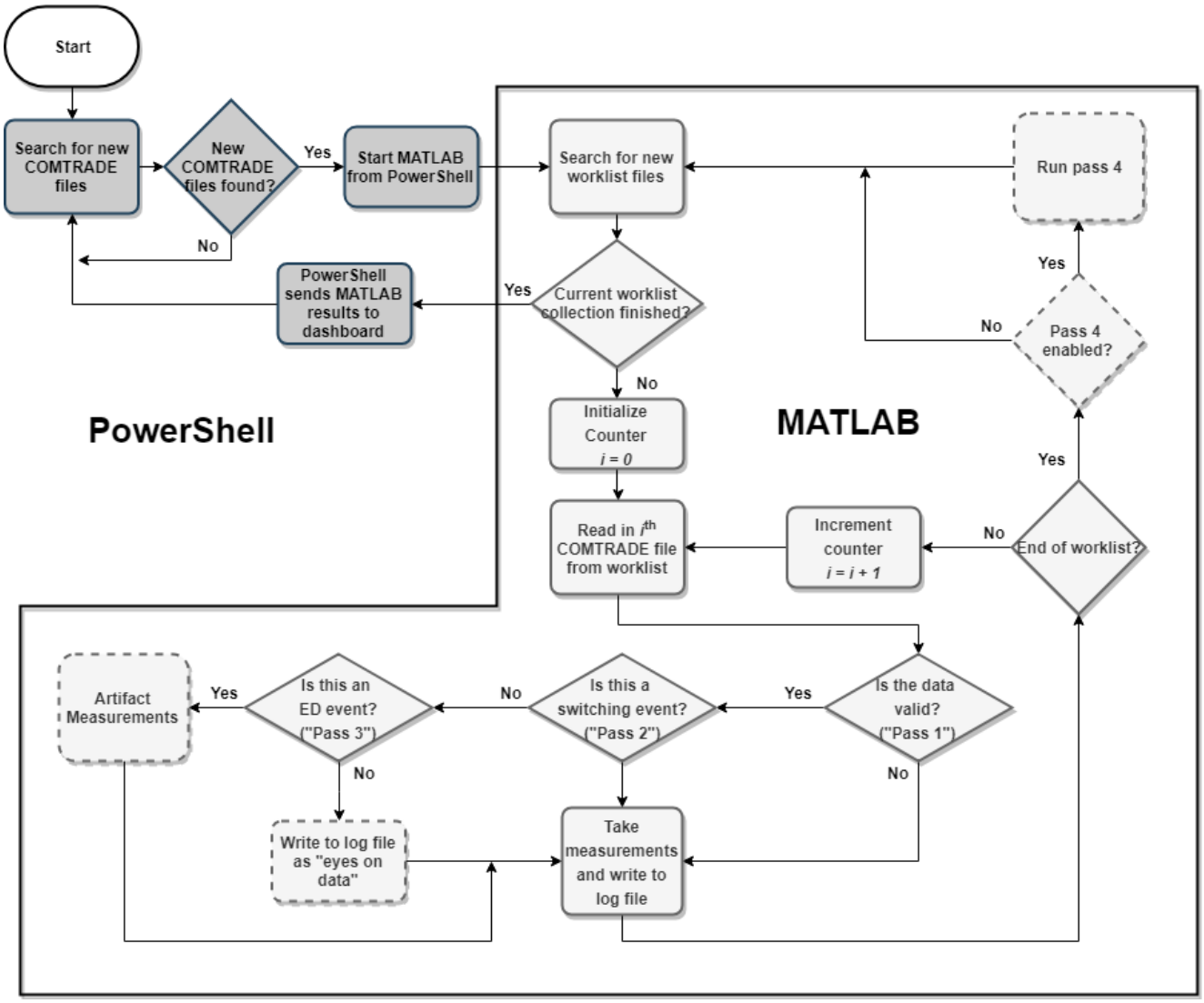


Figure A1 Detailed Flowchart of Main Process

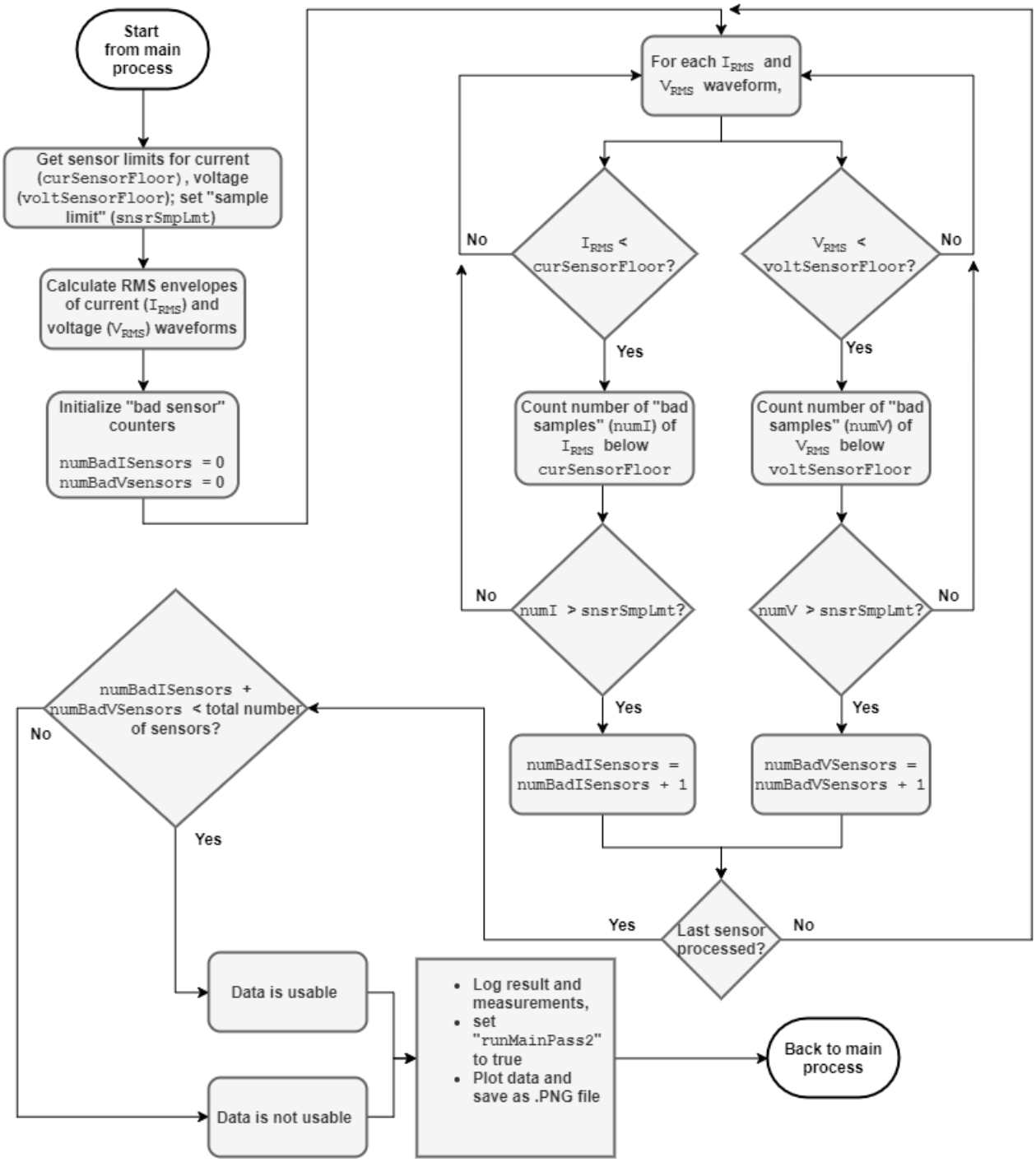


Figure A2 Detailed Flowchart of Main Pass 1

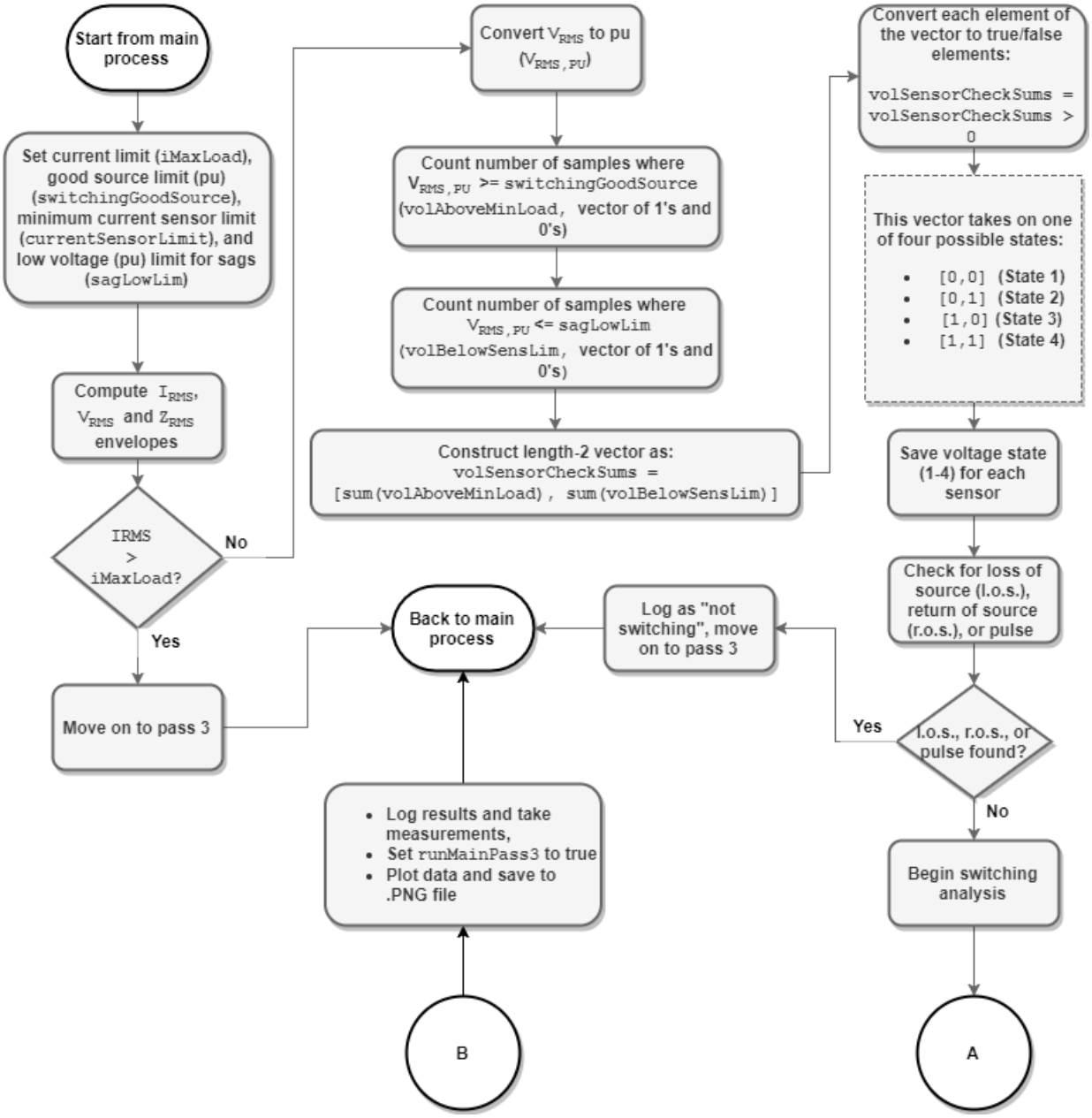


Figure A3 Detailed Flowchart of Main Pass 2 - Part 1

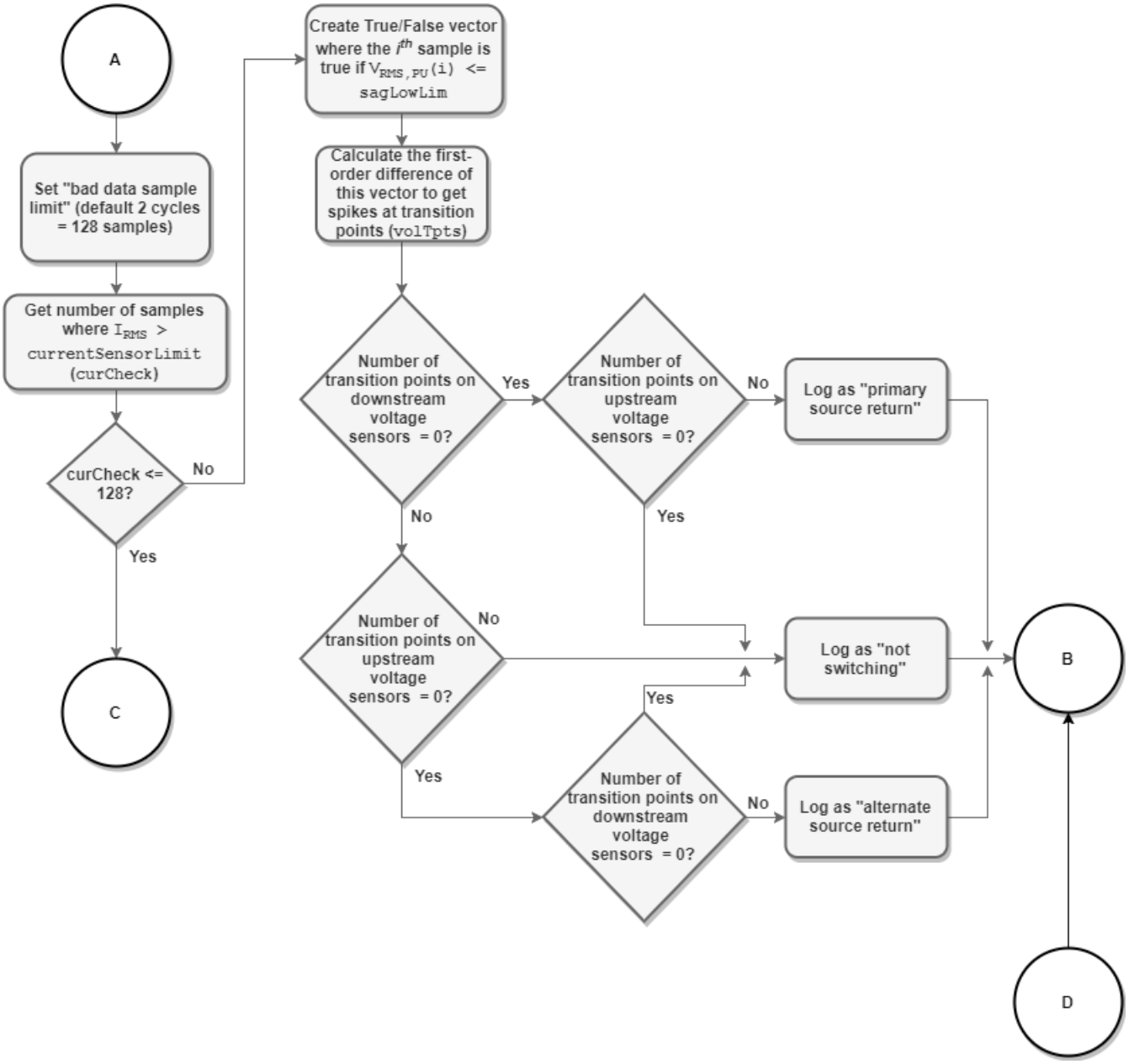


Figure A4 Detailed Flowchart of Main Pass 2 - Part 2

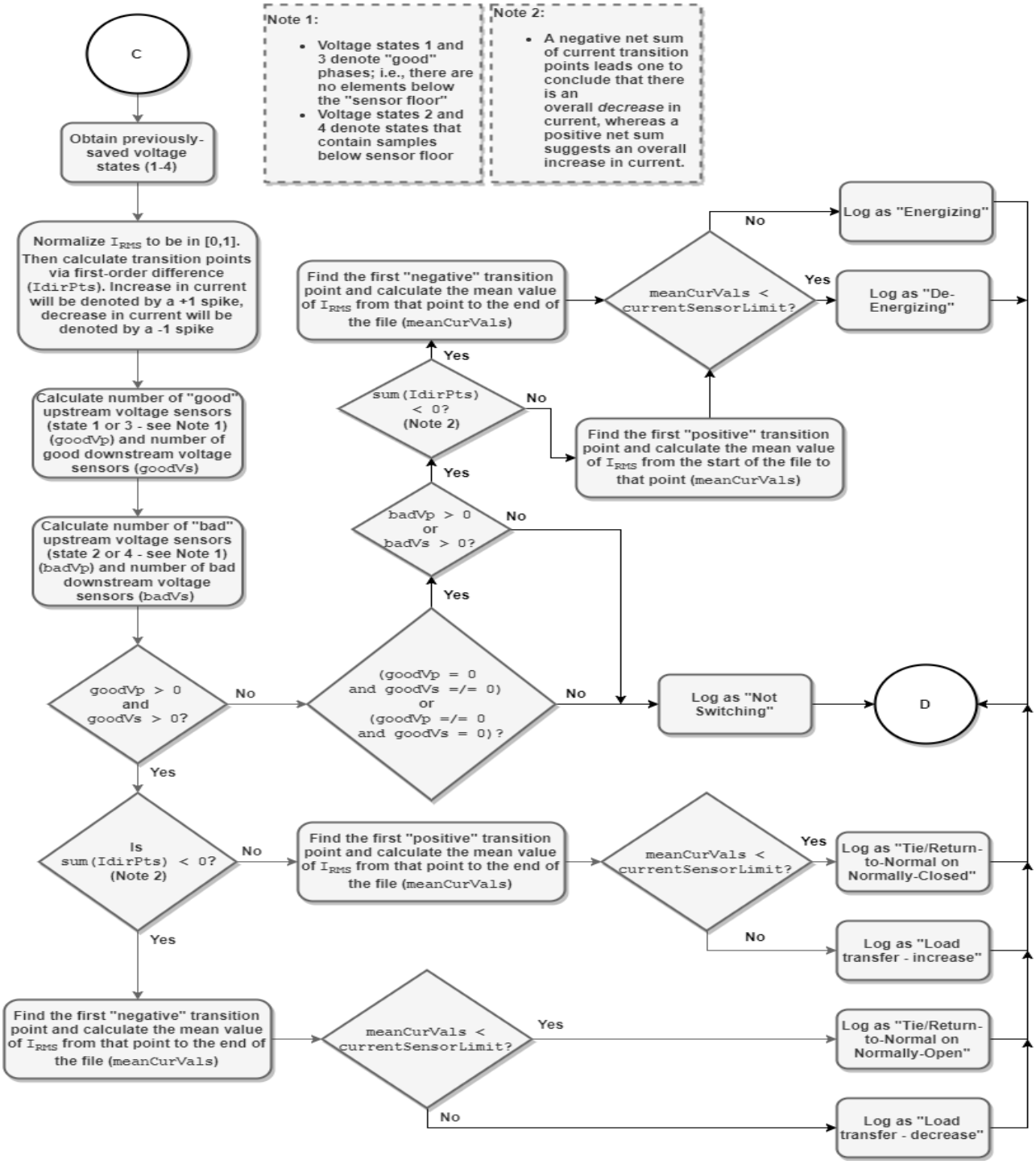


Figure A5 Detailed Flowchart of Main Pass 2 - Part 3

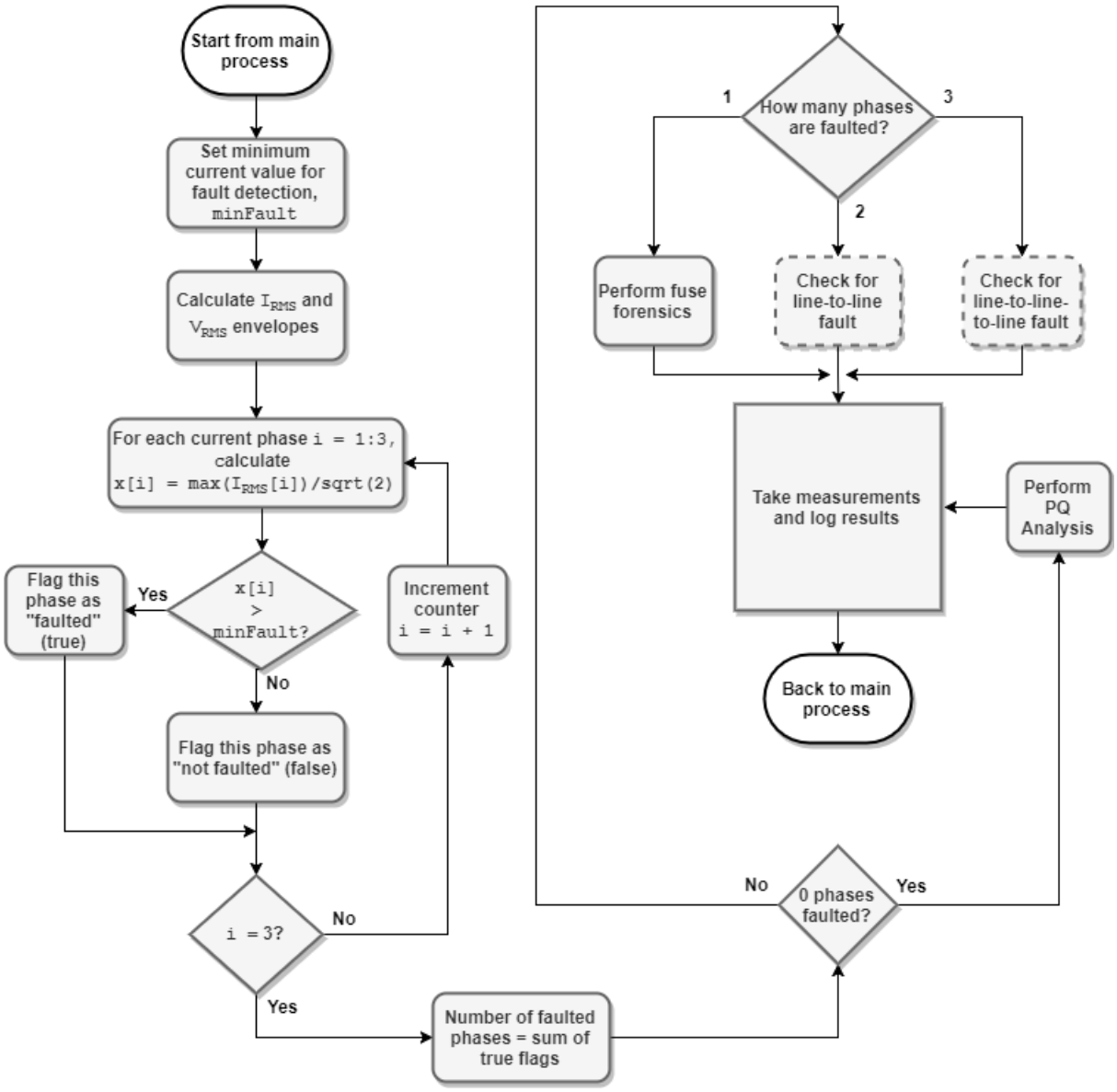


Figure A6 Detailed Flowchart of Main Pass 3

APPENDIX B

EXAMPLE FUSE REPORT PLOT

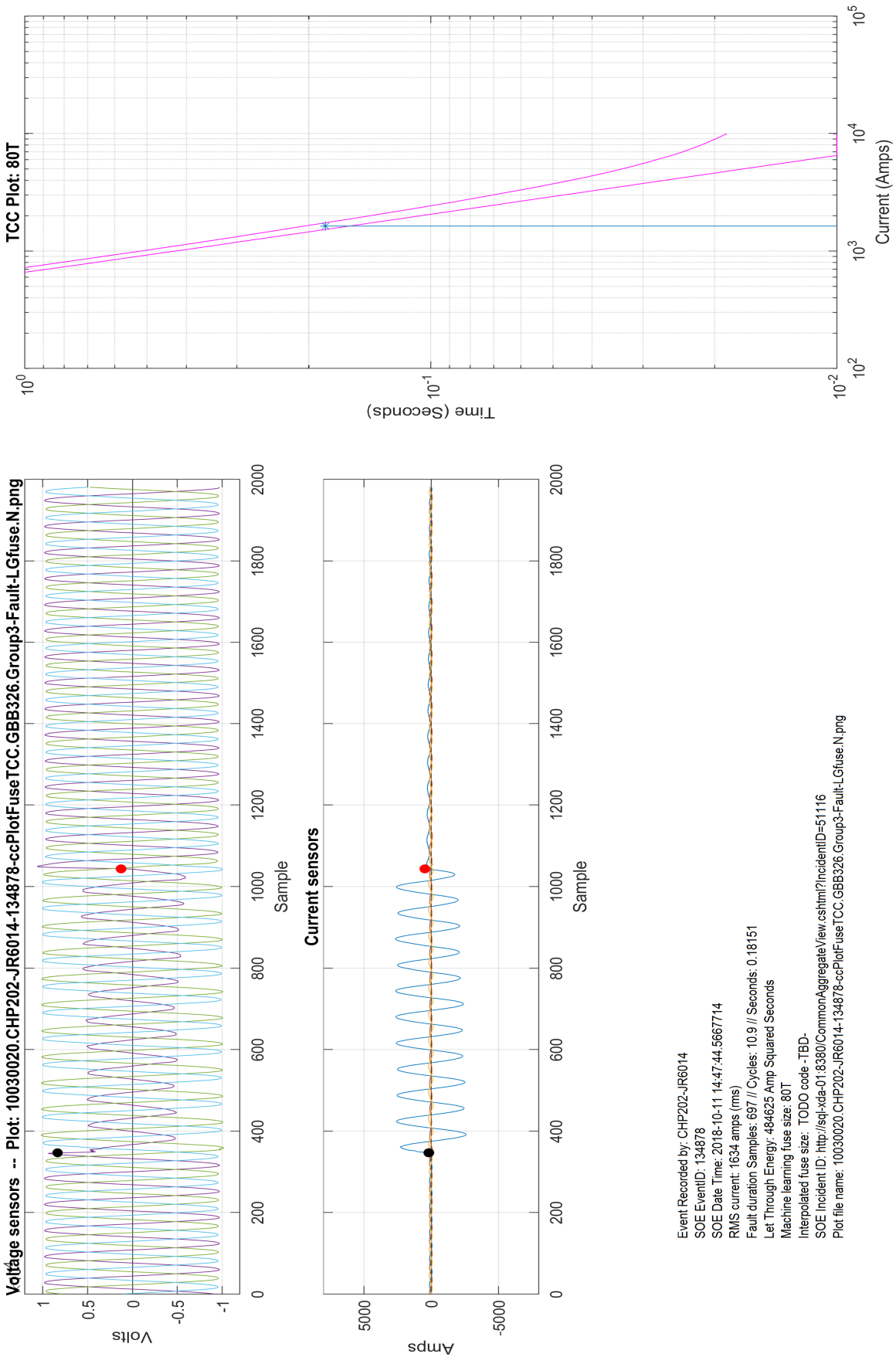


Figure B1 Example Plot of Fuse Report

VITA

Aaron J. Wilson was born in 1995 in Chattanooga, Tennessee. He attended Soddy Daisy High School, from which he graduated in 2013. He received a Bachelor's of Science in Electrical Engineering from the University of Chattanooga in 2017. He is expected to graduate with a Master's of Science in Engineering in May of 2019.