Data Driven Approach to Non-stationary EMA Fault Detection and Investigation Into Remaining Useful Life

William S. Craig
Data Driven Approach to Non-stationary EMA Fault Detection and Investigation Into Remaining Useful Life

by

William S. Craig

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Mechanical Engineering

Supervised by

Associate Professor Dr. Jason Kolodziej
Department of Mechanical Engineering
Kate Gleason College of Engineering
Rochester Institute of Technology
Rochester, New York
August 2014
Data Driven Approach to Non-stationary EMA Fault Detection and Investigation Into Remaining Useful Life

by

William S. Craig

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Mechanical Engineering

Approved by:

Dr. Jason Kolodziej
Associate Professor
Department of Mechanical Engineering

Dr. Agamemnon Crassidis
Associate Professor
Department of Mechanical Engineering

Dr. Marca Lam
Senior Lecturer
Department of Mechanical Engineering

Dr. Alan Nye
Associate Department Head, Professor
Department of Mechanical Engineering
Dedication

To my wife, Anna, whose sacrifices have allowed me to pursue my dreams.
Acknowledgments

I would like to thank my advisor Dr. Kolodziej for his initial encouragement to pursue a master of science degree, and for his guidance and support through a project that took me out of my comfort zone and pushed me to develop as an engineer and researcher. I would also like to thank my parents for their strong love and support through the challenges of the research process. I was very fortunate to have Dr. Lam and Dr. Crassidis serve on my committee, as they provided consistent encouragement through the project. Rob Kraynik and Jan Manetti of RIT’s machine shop were an immense help during the machine rebuild process and provided a wealth of knowledge and support. Dr. Boedo and Dr. Kerekes both provided insight integral to the success of the project, and Diane Selleck was a joy to speak with and made sure that all administrative processes flowed smoothly.
Abstract

Data Driven Approach to Non-stationary EMA Fault Detection and Investigation Into Remaining Useful Life

William S. Craig

Supervising Professor: Dr. Jason Kolodziej

Growing interest in using Electromechanical Actuators (EMAs) to replace current hydraulic actuation methods on aircraft control surfaces has driven significant research in the area of prognostics and health management. Non-stationary speeds and loads in the course of controlling an aircraft surface make fault identification in EMAs difficult. This work presents a time-frequency analysis of EMA thrust bearing vibration signals using wavelet transforms. A relatively small EMA system is designed and built to allow for simple, quick, and repeatable component replacement. A simulated signal is developed to test four potential faults in the system. Classification is performed using an artificial neural network (ANN), which yields over 99% accuracy. Indentation faults from moderate and heavy loads are seeded in thrust bearings, which are then tested to generate data. The ANN achieves 95% classification accuracy in a two class scenario using healthy and moderately indented bearings. A three class test is executed using thrust bearings at each level of damage to perform preliminary remaining useful life (RUL) testing, where an ANN is able to identify the fault severity with an accuracy of 88%.
# Contents

Dedication ................................................................. iii

Acknowledgments ......................................................... iv

Abstract ................................................................. v

## 1 Introduction ......................................................... 1
  1.1 A More Electric Aircraft ........................................ 1
  1.2 Electromechanical Actuators ................................... 2
  1.3 EMA Health Management Research ............................... 3
  1.4 Wavelets in Fault Detection .................................... 9
  1.5 Research Approach .............................................. 12

## 2 Test Stand Development ........................................... 14
  2.1 Initial Test Stand ................................................ 14
  2.2 Test Stand Redesign ............................................. 18
  2.3 Data Collection: First Iteration ................................. 25
  2.4 Servomotor Acquisition ......................................... 27

## 3 Analysis Approach .................................................. 31
  3.1 Wavelets .......................................................... 31
  3.2 Initial Simulation ................................................ 43
  3.3 Updated Simulated Signal ...................................... 51
  3.4 Maximum Energy-to-Shannon Entropy Ratio ................... 56
3.5 Features and Classification ........................................ 61

4 Data Collection ........................................................... 69
  4.1 Setup and Healthy Testing ......................................... 69
  4.2 First Bearing Fault - Mild Single Point Defect ............ 74
  4.3 Second Bearing Fault - Moderate and Severe Single Point
      Defects .............................................................. 80
  4.4 Third Bearing Fault ................................................ 82

5 Classification ............................................................. 89
  5.1 Fault Detection ..................................................... 89
  5.2 Remaining Useful Life ............................................. 93

6 Conclusions ............................................................... 96

Bibliography ............................................................... 99

A Design Drawings ....................................................... 103
List of Tables

2.1 Initial System Bill of Materials ................................. 15
2.2 Final System Bill of Materials ................................. 30
3.1 Base Wavelet MESER Comparison ............................. 58
3.2 Classification Methods on Simulated Data ..................... 67
5.1 Classification Methods on Experimental Data ............... 93
5.2 Classification Methods on Experimental RUL Data ........... 94
## List of Figures

1.1 Data-driven fault identification flow chart ............................................. 4  
2.1 Initial Benchtop EMA Design ................................................................. 14  
2.2 System Redesign Concept ................................................................. 21  
2.3 Redesigned and Assembled Bench top EMA ........................................ 24  
2.4 Velocity Decrease Under Load With Voltage Command ..................... 26  
2.5 Accelerometer Grounding Improvements, Consistent Scale .......... 27  
2.6 Finalized Bench Top EMA Fixture ......................................................... 29  
2.7 Consistent Velocity Under Load Using Servomotor .......................... 29  
3.1 Morlet Wavelet ......................................................................................... 31  
3.2 Wavelet Scaling and Shifting ............................................................... 33  
3.3 Continuous Wavelet Transform Over a Short Time Window .......... 34  
3.4 Comparison of FFT to CWT on Stationary Signal ............................. 36  
3.5 Comparison of FFT to CWT on Chirp Signal .................................... 37  
3.6 Comparison of Wavelet and STFT Time-Frequency Resolution .... 38  
3.7 Five Level Discrete Wavelet Decomposition Using Daubechies 4 Wavelet ................................................................. 40  
3.8 Five Level Discrete Wavelet Reconstruction Using Daubechies 4 Wavelet ................................................................. 41  
3.9 Wavelet Denoising .................................................................................. 42  
3.10 Frequency Spectrum From Using the BLDC Gear Motor ................ 44  
3.11 Individual Simulated Spall Pulse .......................................................... 45
4.10 Frequency Spectrum Analysis on 100 kHz Oiled HRC Fault Bearing Signal ................................... 79
4.11 CWT of Oiled HRC Fault Bearing ........................................... 80
4.12 Moderate and Severe Single Point Defects ......................... 81
4.13 Wavelet Analysis of Bearing Faulted With Dremel Tool .... 82
4.14 Press Fault Using Six Ton Load ........................................... 83
4.15 Press Fault Using 2.5 Ton Load .......................................... 84
4.16 Z Channel Vibration Data from Third Fault Using 2.5 Ton Load .................................................. 85
4.17 Wavelet Analysis of Third Fault Using 2.5 Ton Load ........ 86
4.18 Press Fault Using 1.25 Ton Load ........................................ 87
4.19 Z Channel Vibration Data from Third Fault Using 1.25 Ton Load .................................................. 87
4.20 Wavelet Analysis of Third Fault Using 1.25 Ton Load ........ 88
5.1 Healthy vs 1.25 Ton Press Fault Z Channel Vibration ....... 90
5.2 Healthy vs 1.25 Ton Press Fault Z Channel Vibration Wavelet Analysis ............................................. 91
5.3 Remaining Useful Life Z Channel Vibration Wavelet Analysis 95
A.1 Front Support Drawing ......................................................... 103
A.2 Rear Support Drawing ......................................................... 104
A.3 Nut Plate Drawing ............................................................. 105
A.4 Ball Screw Drawing ............................................................ 106
A.5 Anaheim Automation Motor Mount ................................. 107
Chapter 1

Introduction

1.1 A More Electric Aircraft

Electromechanical actuators (EMAs) are the next step in creating a “More Electric Aircraft”. Recently they have been integrated in minor applications of the Boeing 787 aircraft, such as spoiler and trim actuation, landing gear breaks, and the environmental control system [1]. Traditionally, hydraulic actuators have been trusted with the performance and reliability requirements of aircrafts’ primary control surfaces due to their robust hydraulic power system and high power density. While proven over time to be a very successful method of actuation, hydraulics come with a cost. These systems use a central hydraulic pump which must maintain over 3,000 pounds per square inch of pressure constantly, regardless of demand. They also require a heavy and inflexible infrastructure which has the potential to leak dangerous and corrosive fluid. This fluid may incapacitate everything connected to the system, the leak can be very challenging to identify, and be difficult to reach once it is finally found [2].

The problems facing full hydraulic systems initially led to the development of electrohydrostatic actuators (EHAs). The EHA has self contained hydraulic fluid which is pumped using an electrically driven pump-motor to drive the ram in the same manner as a standard hydraulic actuator. These
have undergone significant research in the past fifteen years, which has led to a strong understanding of the technology. Currently, they serve as backup systems for the Airbus 380 and Boeing 787, which eliminates the need for full central hydraulic second and third level redundant systems [2].

More recently, there has been increased interest in EMA technology. The EMA is smaller, lighter, less complex, and easier to repair than an EHA. With no fluid under load, they also tend to be stiffer, and they are more efficient because they do not suffer any windage losses or pump inefficiencies. Additionally, they are better for long term storage since they have no leak potential. Furthermore, [3] found performance gains when comparing EMAs to standard hydraulic actuators, showing that stall force for the EMA was higher and that the closed loop frequency response improved. Despite the multitude of advantages, EMAs are not used on primary control surfaces due to their potentially disastrous failure modes. This has spurned recent research in the direction of fault detection and condition based maintenance.

1.2 Electromechanical Actuators

Actuators are safety-critical components on aircraft, and their failure can have devastating consequences, as in the tragedy of Alaska Airlines MD-83 Flight 261 accident, where an EMA jackscrew failure led to a crash in the Pacific Ocean. In order to eliminate these failures, aircraft system maintenance is typically completed using a very conservative time-based methodology derived from a statistical estimation. While often effective, if the true operating conditions differ from the predicted usage, failures can happen sooner than anticipated, leading to unexpected repair, time out of service, and in extreme cases, catastrophic failure. When the life of the system is captured by the statistical estimation it is often repaired much sooner than
necessary, which induces unnecessary time and component costs. Condition-based maintenance addresses this problem by monitoring a specific system in order to determine the optimal time to perform maintenance. This involves a sensing mechanism and a prescribed tolerance for that system, which is becoming more reasonable as sensing technology and computing power grow [4].

1.3 EMA Health Management Research

Health management for EMA systems has been a relatively recent development, as most initial research was focused on design improvements [5]. Health management studies have monitored many aspects of the EMA system, including temperature, motor current draw, and vibration, in order to best identify the condition of the system. In addition to identifying sensors for the system, researchers must choose whether to use a model based, data driven, or hybrid approach to then classify the system. The model based approach has the advantage of tracking individual components in the system, thereby identifying faults from a physical perspective in a very intuitive manner. A model is designed to understand how each of the components interact; for example, what to expect from the position of the actuator output when the current draw increases. This can allow for a very detailed and accurate fault identification, giving the operator the ability to see a particular component’s deviation from healthy in real time; however, it is very difficult to capture all of the relationships within a system with such high accuracy, and any attempt requires very powerful computing. Moreover, a very accurate model must be specialized to its particular system, either the particular product from the manufacturer, or for even higher accuracy, the specific system being tested.
A data driven approach, on the other hand, knows very little about the system. Sensor data can be used directly, but is often transformed using statistics into what are referred to as features, which are then fed into a classification tool. The tool uses the features given to assign the system to a particular class, as seen in Fig. 1.1. The class can be as simple as healthy or unhealthy, can classify healthy against several different types of faults, can classify the severity of a particular fault, or even potentially perform a combination and classify the severity of several different faults. Beyond knowing very little about the fault, another potential problem with the data-driven approach is that it often requires a significant amount of data to successfully train the classifier, though once the classes have been identified, execution is much faster and simpler.

It is also common to perform a hybrid approach, utilizing the benefits of both the data driven and model based approaches. In this method, a more general model is often used to identify the subsystem which is faulted, at which point a specialized data-driven method can be used to quickly and accurately identify the fault. Many different research efforts have been undertaken to diagnose EMA faults using each of these methods.

![Figure 1.1: Data-driven fault identification flow chart](image-url)
In 2009, Balaban and Stoelting [6] from Moog and NASA monitored EMA load, position, several temperatures, motor current and velocity, 3-phase currents, and vibration at four points. With this information, they were able to investigate a nut return channel ball jam, ball screw spall, nut thermocouple drift, nut thermocouple bias, Z accelerometer scaling, and x accelerometer complete failure using a data-driven method. Here, the data outputs were statistically transformed into features, which were then given to an artificial neural network (ANN) for classification. With this method, they were able to achieve 91% successful classification using a total of six faults over eight different conditions.

Byington et al. [7] from Impact Technologies used a model based approach, where they monitored friction damping coefficient, local gear stiffness, torque constant, and motor temperature to classify gear slipping and bearing seizure faults. In addition to fault diagnosis, they were interested in determining the remaining useful life (RUL) of components to estimate how much longer a faulted component could remain in service. With their dynamic modeling approach, Impact was able to identify when one or more faults were developing as well as the rate of each. This allowed for identification of the most threatening fault and subsequent RUL calculations. With this method, they were able to achieve 70% correct overall classification of fault and approximate RUL, with a high majority of incorrect classifications occurring at low severity. Later, Smith et al. [4] worked on another model-based approach. After designing and building a test stand, they used temperature, vibration, position, and torque measurements to design and verify their model. Faults were then seeded in the model in order to test the accuracy of the diagnostic software, which showed very promising results when looking at winding shorts, loss of power to the motor, position sensor
Narasimhan et al. [1] investigated twelve EMA fault conditions including nut jam, motor failure, spall, and many sensor faults using a hybrid architecture applied to the Flyable Electromechanical Actuator (FLEA) system. The FLEA was designed to fly inside an aircraft and receive real-time input commands and load profiles, also experiencing the environmental conditions in the flying aircraft. The FLEA consists of one faulted actuator, one healthy actuator, and one dynamic load actuator, all connected to the same rigid output beam. This allows for seamless transition between healthy and faulted testing. The FLEA contains a load cell, accelerometer, rotary encoder, linear potentiometer, voltage sensor, and current sensor to characterize the condition of the system. These are fed into a model that uses a Z-test to monitor the deviation of the sensor from the nominal value. Once a significant difference is identified, the model is able to identify a group of most probable faults, at which point a data-driven diagnosis tree can be used to isolate the particular fault. Using this method, NASA was able to achieve 96% accurate classification over all of the faults. In 2011, Balaban et al. [5] furthered this research by adding an RUL test. Here, the researchers chose to investigate the jam fault in order to determine how long a pilot might have to land a plane under typical flight demands once the jam has developed. When the return channel in the EMA was jammed, the motor was faced with a significant increase in friction, which caused the motor temperature to increase, destroying the windings, leading to a short circuit and motor failure. Through their testing, NASA was able to predict end of life to within approximately 8% error.

Bodden et al. also investigated remaining useful life, where an abrasive contaminant was mixed with the EMA bearing lubricant and applied to the
ball screw, thrust bearing, cluster gear bearing, output bearing, and motor bearing. This effectively led to accelerated failure, however, the components did not wear at the same rate, even after optimal concentrations of contaminant were chosen. It was also seen that the change in frequency content as the system degraded was not characteristic of actual component failure. Additionally, bearing degradation led to unintended failure of unmonitored components, further complicating the results. This work made it clear that seeded fault RUL testing can be very difficult, particularly when attempting to limit the number of sensors required. Bodden et al. also found taking vibration measurements on the data to be ineffective since the motor was non-stationary and operated at a constantly varying angular velocity, which led to frequency smearing in a fast Fourier transform (FFT) of the raw data, making it impossible to identify frequencies of interest.

Huh et al. [8] focused on the gears within an EMA, damaging one of the forty teeth on the spur gear in an EMA. A dynamic model of the gear train was developed, and direct spatial domain re-sampling was performed on vibration data in order to synchronize the data to angular position, which served as a solution to the frequency smearing problem experienced in [9]. After that, the data was synchronous time-averaged to damp out random components, further denoising the data. This technique successfully produced a kinematic error estimator which clearly displayed the gear damage on a frequency plot.

Romeral et al. [10] approached the problem of non-stationarity through the use of the discrete wavelet transform (DWT) which is able to perform a frequency analysis across the time domain. They monitored the current of an EMA brushless motor under healthy, short circuited, and shaft eccentricity faults. Using the energy from a different DWT frequency ranges, they
were able to see clear differences between each of the conditions.

In 2008, Baybutt et al. [11] from Impact Technologies, Lockheed Martin, and Parker Aerospace aged input power converter field effect transistors (FET), H-Bridge FETs, and motor power line plastic capacitors in order to assess electrical component RUL. Voltages, currents and temperatures were monitored to assess the condition of the systems, and a probability density overlap technique was used to visually separate the healthy data from the degraded data.

Zhou et al. [12] focused on electrical sensing, where they performed a review of stator current-based fault detection techniques in general motor applications in order to show the effectiveness of bearing fault diagnostic techniques without adding sensors beyond the standard sensor array necessarily used for electrical protection. Their investigation revealed a variety of techniques, such as neural-network clustering, wavelet packet decomposition, and an extended Park’s vector approach, which effectively diagnosed single point defects in motor bearings using stator current as the primary feature from the data.

Two of Dr. Kolodziej’s previous master’s students, Rahulram Sridhar and Anthony Chiroco III, also performed research investigating fault detection in EMAs. Sridhar [13] utilized motor velocity error and torque load as inputs to a linear difference model to characterize data from a Moog Max-Force EMA. His model was designed to output motor current, and the model parameters were used as features to characterize system health. He also chose to use principal component analysis (PCA), which reduced the number of features, uncorrelated the features, and maximized the variance of the transformed features along the feature axes. Using a Bayesian classifier,
this method yielded classification accuracies up to 100%. Chiroco [14] performed work on the same system with a data-driven approach, using EMA position, motor velocity, motor phase A current, and motor bearing housing accelerometer data. Chirico performed a re-sampling of the data based on angular position to eliminate frequency smearing in the non-stationary system, similar to [8]. He then performed a power spectral density frequency analysis, which was filtered, then categorized into fewer discrete bins. The bins were passed through a PCA algorithm to choose the most effective features, which were classified using a Bayesian methodology to yield 99% correct classification when using position, velocity, and vibration data.

1.4 Wavelets in Fault Detection

The problem of non-stationarity has been approached with a variety of methods. Applied in [10] and discussed in [12], wavelet analysis has seen rather limited application to EMAs, and strong promise to identify many possible faults from the research performed on other rotating systems. Wavelets are a signal analysis technique that is able to extract information from both the frequency and time domains, allowing for visualization of a frequency response which is dynamic in time. A discussion on wavelets in much more depth can be found in Section 3.1.

The time-frequency nature of wavelets has proven very useful in many gear and bearing fault detection applications. Wang and Gao [15] investigated a localized point defect on a rolling element bearing using a combination wavelet and FFT analysis. The defect had a very weak amplitude and short duration in the raw signal, and by first processing the data in a wavelet transform, then using a portion of the wavelet output as the input to the FFT, the defect was seen clearly, despite not being obvious in either the wavelet
or FFT alone. In another similar experiment, Holm-Hansen et al. [16] first built a custom base wavelet derived from the impulse frequency of the ball bearing being tested. The customized base wavelet was then used to extract time-frequency information from the data. A portion of the wavelet decomposition was then fed into an FFT, which showed a much greater ability for the method to identify the location of the defect on the bearing than a more traditional base wavelet.

Al-Raheem et al. [17] chose to use the Laplace base wavelet to analyze roller bearing faults. The base wavelet was optimized with a genetic algorithm, then the optimal scale was found by taking the kurtosis of the wavelet transformed data signal. After processing through the optimal wavelet scale, the root mean square, standard deviation, and kurtosis were taken in the time domain, and a ratio of the wavelet power spectrum peak frequency to the shaft frequency and the ratio of the wavelet power spectrum maximum amplitude to the overall amplitude were taken in the frequency domain. These five features were used as inputs to a feed-forward multilayer perceptron ANN which consisted of three layers and showed very strong classification ability.

In 2009, Wang et al. [18] from the University of Electronic Science and Technology in China used accelerometer data from a motor bearing to diagnose inner race and outer race faults through wavelet analysis. They started with a wavelet decomposition, which was followed by a wavelet lifting scheme, that simplified the signal. Even further simplification was obtained by using empirical mode decomposition, which also kept the signal in frequency bands. A singular value decomposition was then utilized to obtain the intrinsic characteristic of the signal, whose values were classified using a Hidden Markov Model, which yielded 100% classification accuracy.
in a motor test stand case study. Diehl et al. [19] introduced a plan to investigate gear faults using dynamic gearbox models and harmonic wavelet transforms. The dynamic gearbox model was able to simulate varying degrees of chipped gear tooth wear, as well as broken tooth, spalled tooth, and cracked tooth faults. The healthy harmonic wavelet transform was directly subtracted from the unhealthy transform, which yielded clear differences, showing promise for future testing.

Hee and Leong [20] investigated turbine blade faults using the rotor dynamic wavelet map and the blade passing energy packet. A gas turbine was modified in order to assess blade creep, rotor eccentricity, and rubbing. One second of accelerometer data was split into 23 sections representing the system’s 23 cycles per second, which was then synchronous time averaged to reduce noise using many one second samples. A continuous wavelet analysis was performed on the averaged data to produce the rotor dynamic wavelet map. The blade passing energy packet was formed by taking the averaged signal, filtering out operating speed harmonic components, dividing it into twelve signals to represent the 12 blades, and using the root mean square value of each segment as the blade’s value. Experimental testing using both methods yielded clear signatures for each of the faults.

Hussain and Gabbar [21] at the University of Ontario Institute of Technology studied wind turbine gearbox degradation, specifically in an oil loss scenario. Raw vibration data maintained consistent magnitude. However, using the vibration steepness index, which analyzed the localized slope of the vibration signal, a trend was identified corresponding to the gearbox degradation, which was then de-noised using a wavelet transform. Using several different modeling techniques, predictions of future health were also characterized, which could be translated into RUL estimates.
Kankar et al. [22] used a complex Gaussian wavelet applied to vibration data in order to assess inner race, outer race, and ball bearing spall faults. The vibration signal was initially fed through an autocorrelation function, which identified the repetitive nature of the signal, yielding a 90% reduction in dimension with negligible loss of significant content. The complex Gaussian wavelet decomposed the resulting signal, and kurtosis, standard deviation, and skewness were taken from the optimal wavelet scale. The statistical features were classified using support vector machine, which resulted in 99% correct class assignment.

Ohadi and Heidari [23] at the Amirkabir University of Technology used vibration data to study several different gear conditions in an automotive gearbox. Accelerometer data was resampled according to angular position to remove non-stationary characteristics, which was then fed into a Morlet continuous wavelet analysis. Kurtosis, skewness, and standard deviation were taken from two of the 64 wavelet scales, that were then used as inputs to a three layer ANN. Principal component analysis was also used, reducing the ANN inputs from six to two, which improved classification results by about 12%, from 86% to 98%, and reduced ANN training time by 33%.

1.5 Research Approach

While there is clearly an abundance of research in the areas of electromechanical actuators and wavelets, the intersection of the two has seen limited study despite the effectiveness of the wavelet analysis in non-stationary signal diagnostics. Radial ball bearings have undergone significant wavelet-based research, and are an integral component of EMA design, but beyond radial bearings, EMAs are heavily dependent on thrust bearings to absorb actuation loads, which have seen significantly less dedicated research. This
The investigation proposes the design, manufacturing, and assembly of a relatively small, breadboard mounted EMA assembly capable of simple, quick, and repeatable component replacement with the intent of seeded fault testing on the thrust bearing which supports the ball screw. The thrust bearing is of interest due to its aforementioned importance to the function of the EMA, as well as its inexpensive and easy replacement on the test system that allows for insurance of repeatability. Many previously discussed and researched EMA faults such as a jam or motor bearing fault would require or lead to the destruction of the motor, which in this testing is considered irreplaceable. Analysis is to be done using wavelet transforms as the primary frequency domain analysis tool, followed by a pattern recognition scheme capable of accurate and robust classification.
Chapter 2

Test Stand Development

2.1 Initial Test Stand

Initially, the experimental work planned to use a test stand, seen in Fig. 2.1, which had previously been created during an undergraduate independent study. The system was designed with a brushless DC (BLDC) motor and controller, powered by a 24V power supply. The motor spins a ball screw, whose rotational motion is translated to linear motion through a ball nut that

Figure 2.1: Initial Benchtop EMA Design
supports a plate. The plate’s linear motion pushes against springs, which are supported by rails on each side and limited by the rear rail supports. The front support houses a spherical roller thrust bearing capable of enduring radial and thrust loads, which is used to locate the ball screw and absorb the load from the ball screw due to the springs. The rear aluminum support contains a radial bearing that is used for locating the ball screw and aiding in low-friction rotation. Angular position sensing capabilities are provided by an optical encoder connected to the ball screw and located on the front support. The bill of materials can be seen in Table 2.1.

System feedback and control is facilitated through Simulink and dSpace. Simulink was used to design the control system, feedback loops, and profiles for the motor controller. Once this had been established, the program was compiled and used in dSpace. The dSpace program supplies a graphical user interface (GUI), where the user can, at its simplest level, establish the display variables, control variables, and variables to be plotted. The data acquisition was also controlled through dSpace. With many BNC inputs and outputs, as well as several digital I/O ports, it has the ability to interface with a multitude of systems. In this case, a dSpace BNC output of between

<table>
<thead>
<tr>
<th>Component</th>
<th>Model</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC Motor</td>
<td>BLWS235S-36V-4000</td>
<td>Anaheim Automation</td>
</tr>
<tr>
<td>Motor Controller</td>
<td>MDC150-050301</td>
<td>Anaheim Automation</td>
</tr>
<tr>
<td>Ball Screw</td>
<td>3/4” Diameter, 13/64” Lead (190-9101)</td>
<td>Thomson</td>
</tr>
<tr>
<td>Ball Nut and Flange</td>
<td>3/4” Diameter, 13/64” Lead (8107-448-016)</td>
<td>Thomson</td>
</tr>
<tr>
<td>Ball Bearing</td>
<td>High-Load Steel Open, 1/2” ID (1616)</td>
<td>RBC Nice</td>
</tr>
<tr>
<td>Thrust Bearing</td>
<td>Steel Tapered-Roller Bearing, 1/2” ID (A4138)</td>
<td>Timken</td>
</tr>
<tr>
<td>Shaft-mount Optical Encoder</td>
<td>2500 Counts, 1/2” ID (E6-2500-500-I-S-D-D-B)</td>
<td>U.S. Digital</td>
</tr>
<tr>
<td>Springs</td>
<td>0.64” ID, 8.8” length,18 lb/in (10004)</td>
<td>CSC Springs</td>
</tr>
<tr>
<td>Motor Coupling</td>
<td>1/4x1/4 (SPX-4-4-F)</td>
<td>Ruland</td>
</tr>
<tr>
<td>Hardware</td>
<td>Steel Mounting Plate, Brackets, Rails</td>
<td>Misc</td>
</tr>
</tbody>
</table>

Table 2.1: Initial System Bill of Materials
zero and five volts is used to control motor speed, with another binary BNC voltage output used to control direction.

Initially the system was controlled with an open loop voltage command, requiring the user to manually define the behavior of the system. This method was used for troubleshooting basic function of the Simulink model, dSpace interface, and physical system. Once small issues were worked out, automatic control systems were developed. Position feedback was provided by the optical encoder and differentiated to give velocity feedback. These were used in a position-integral-derivative (PID) controller to control the linear position of the ball nut in an outer loop, and control the velocity of the motor in an inner loop. The motor controller itself also has a control system embedded, which could be run in open-loop or closed-loop velocity control mode. In order to have the most control over the motor, it was run in open-loop mode and the external velocity control was run through Simulink. Once basic test stand functionality had been established, data acquisition began.

While the system did function at the most basic required level, there were several issues that became clear upon data acquisition. It was apparent that the motor had difficulty maintaining a constant velocity even without load. This was double checked using the built in speed output from the motor, which gave two pulses per motor revolution. Given the much lower resolution than the encoder, it was difficult to see the more rapid sinusoidal nature of the velocity instability, so attempts to use the controller output proved to be futile. The desired top speed of the ball nut was chosen to be two inches per second, which, using the equation,

\[
\frac{nutspeed[\text{in/s}] \cdot 60}{\text{lead}} = \text{rpm}
\]

(2.1)
and knowing that the screw lead was 0.2 inches, led to the maximum motor speed requirement of 600 rpm. The motor initially in use with the stand had a maximum speed of 4000 rpm, much higher than required, which led to a lack of accuracy and torque at lower speeds. A different motor that was already in the lab, the Anaheim Automation BLWRPG112S-24V-10000-R14, is a geared motor with an initial top speed of 10,000 rpm and a gear ratio of 14:1. This proved to be a better speed match with a maximum speed of 700 rpm, also providing higher resolution for the motor velocity from the controller and higher torque output, which allowed for more consistent speed under load. The new motor did improve the output results, although the inconsistent speed output at steady state commanded velocity remained. An investigation into potential causes of motor speed inconsistency led to the discovery of many design flaws in the system.

First, the ball screw was studied, and it was found that when loosely connected to the mounting plate, the motor would wobble. This suggested that the machined portion of the ball screw was potentially out of true, though it also may have been a result of bearing misalignment or a flaw with the motor shaft. When disassembling the stand and looking for other causes, it became clear that the rails which both supported the springs and guided the plate were not straight, and that either the tapped holes in the aluminum front support or the threaded portion of the rails were misaligned as it was extremely difficult to screw the rail into place. The nut plate itself simply had through holes to align the nut and take the load, which fit loosely over the rails and occasionally rubbed on the rails. Also, each of the three rear supports connected to the mounting plate were independently attached, making it very difficult to align them together reliably. Each of the support structures except for the motor mount was connected using clearance holes
through the bottom of the mounting plate, allowing for variability in the exact placement of the structure and therefore unreliable reassembly. Furthermore, the mounting plate had no defined flatness standard, and consisted of user drilled holes which were not square. Initial attempts at disassembly and reassembly took between 45 and 60 minutes, with no method of ensuring precise reassembly. Upon discussing options with faculty in RIT’s machine shop, it was clear that a complete overhaul would be required in order to produce a precise, accurate system capable of repeatable and reliable results.

2.2 Test Stand Redesign

Due to most problems in the initial stand stemming from a lack of system squareness, flatness, and friction reduction, these became the focus of the design in the new stand. Precision is of utmost importance when designing a test stand for data collection work. If a system is not built with high precision, moving parts are very likely to introduce unnecessary noise to the system, which can make it extremely difficult to detect the signal of interest. This can also introduce meaningless signals that appear significant, leading to fruitless investigation. As mentioned above, a lack of precision also decreases the likelihood of reproducibility in results, which can lead to merely classifying a difference in system noise between healthy and faulted data, with the true nature of the signal undetected. Functionality as a moving system is one goal of building a test stand, but it takes much more effort, care, and time to create a system which is truly capable of accurate data collection. This led to two major goals in the design: first, to machine each of the components flat and square, and second, to build the frame formed by the front support, rear support, and guide rails to a very tight tolerance.
The first goal was achieved through a consistent commitment to proper machining throughout the build and taking time to insure quality. The second goal required the efforts of the first, as well as effective design decisions. Case hardened guide rails from Thompson with a straightness tolerance of one thousandth of an inch per foot replaced the old rails. A short discussion with the machinists led to the decision to replace the rear support should with one solid piece to contain both rails and the rear ball screw alignment bearing. Similarly, the front support replacement was also one solid piece for the guide rails and thrust bearing. Holes for the guide rails were cut to support a very snug fit, with screws on the outside of the front and rear supports to fix the guide rails into place, ensuring a rigid, square, flat frame.

Other preliminary designs for the new test stand were based heavily on the previously design, simply using high precision components. The ball nut plate attached to Thompson linear ball bushings designed to ride on the guide rails for straight, low friction motion. The taper roller bearing with radial and thrust support capable of thrust loads up to 550 pounds was replaced with a dedicated thrust bearing from Boston Gear capable of supporting thrust loads of 70 pounds at 500 rpm, much closer to the specifications of the system, giving more potential for forced degradation remaining useful life (RUL) testing. This bearing was a very simple design, with two washers and one ball bearing cage meant to slide onto the shaft rather than press fit, which would also allow for very easy faulting and replacement. The geared motor was used on the redesigned system, though a new motor mount was required in order to fit the new dimensions. The ball nut was unchanged, and the ball screw underwent re-machining that ensured a straight shaft, which also required a new rear support radial bearing in order to maximize the
usable length of the shaft. All of the components were mounted on a me-
chanical breadboard from Newport Company, which ensured flatness across
the system, as well as consistent, reliable mounting distances and angles.

All parts to be machined were designed using PTC Creo, a 3D parametric
feature solid modeling program with capabilities to translate 3D designs
into 2D drawings for use in machining. The front and rear supports, motor
mount, and nut plate needed to be machined in their entirety, and the ball
screw and guide rails required modification. After many discussions with
RIT’s machinists about the system and subsequent design modifications, the
concept was finalized, which can be seen in Fig. 2.2. The drawings for each
of these components can be seen in Appendix A. Steel was chosen as the
material for the front support because the thrust bearing would also be made
of steel, and this would maintain uniform material characteristics between
the two. Aluminum was chosen as the material for the rear support, motor
mount and nut plate due to its ease of manufacturing, and in the case of
the nut plate, low density for ease of acceleration. All raw materials were
purchased at Rochester Steel Surplus in Rochester, NY.

Manufacturing was performed with the help of RIT’s machine shop fac-
culty using a wide array of tools and machines. The first step was to square
each of the raw material blocks for the front support, rear support, and nut
plate, and bring them to their desired outer dimension. This was achieved
by using a Bridgeport Series I 2HP milling machine in conjunction with a
Swiss Precision Instruments dial caliper, and a file to deburr the edges of
the blocks. After the materials were brought into dimension and alignment,
component details were added. The nut plate was the first part manufac-
tured, using a ProtoTRAK Plus controlled Bridgeport Series I 2HP milling
machine, which is a computer numerical controlled (CNC) milling machine
that allowed for the precise machining of non-standard shapes and hole sizes required for the nut plate. Once the holes had been drilled using the programming technology in the mill, the linear ball bushing attachment holes were threaded, finalizing the part. Next, the guide rails were manufactured, which required drilling and tapping holes on each end with a Harrison M300 lathe, as well as chamfering on the edges using the Kalamazoo Industries belt sander.

The front and rear supports were then machined simultaneously in order to ensure an extremely accurate fit for the guide rails and ball screw. It was particularly important that the guide rails had a precise fit because they were designed to align the system and assure that everything fit and glided smoothly when running, allowing for the most accurate data acquisition possible. Simultaneous machining was possible because the front and rear support used the same height and width dimensions, which were met within five thousandths of an inch, and matched to the thousandth between the two supports. This allowed for the machine to be set at a certain position.
relative to the part, with the part location saved by a vise stop for identical machining. This way, the same exact drill bit was used at the same exact location on both the front and rear supports.

After each location was finished, the mill was moved to the next location and the process repeated. The guide rail interface holes were cut with a mill bit to slightly less than the diameter of the rail, after which they were cut very smoothly with a reamer bit to precisely the correct diameter to meet smoothness and tightness requirements. The hole for the thrust bearing on the front support was cut with a very sharp mill bit to ensure flatness of the interface between the thrust bearing and the front support. The hole was also cut with generous radial clearance as any interference would reduce data quality. The radial bearing hole in the rear support was cut to a slip fit diameter, which was tight enough to assure the bearing would roll consistently while still maintaining easy disassembly and reassembly, also ensuring that the thrust bearing would support the full load from the springs. Vertical breadboard mounting holes for the front and rear support were located by setting blind hole spotting transfer pins in the desired locations on the breadboard, then using a lead block as a hammer to drive the pins into the supports and identify the desired locations. The holes were drilled using an Electro-Mechano 601J drill press, followed by a counterbore to bring the mounting to an ideal screw height to achieve necessary thread engagement, which finished frame system manufacturing.

The ball screw was made from hardened steel, and required the machine shop’s Brown & Sharpe grinder for manufacturing. The rear of the ball screw was shortened and its diameter reduced to fit a smaller radial bearing and not protrude from the rear support. The front of the ball screw was designed to fit the motor, optical encoder, and thrust bearing, with enough
length to extend through the front support and allow for connection with each of the components. Again, each of these connections required a very precise fit, as any poor connections would reduce data quality.

Finally, the motor mount was built, which required a circular hole of precise height and diameter to locate the motor with four counter bored clearance holes to attach the motor and maintain a tight connection. The motor mount was connected to the breadboard with two screws, again located using blind hole spotting transfer pins.

With all of the parts machined, only a small number of finishing touches remained before the stand was able to be reassembled and run. Springs with an inner diameter just above the outer diameter of the rails were purchased from T.R. Goldsmith & Son. The springs were cut to fit the stand using a Dremel 4000 rotary tool. Two sets of springs were purchased, one with a spring constant of 6.9 pounds per inch, the other with a spring constant of 8.4 pounds per inch once at length. With the equation,

$$F = kx$$  \hspace{1cm} (2.2)

and knowing that there were two springs in parallel which would experience up to four inches of travel, the final load potentials for each spring were 55.2 pounds and 67.2 pounds. Spring lengths left over from the portions cut to length were used to guard against collisions with the nut and front support. The supports, nut plate, and linear ball bushings were all protected from the ends of the springs with steel washers. A mounting fixture for the optical encoder was designed to center the encoder mount around the shaft. In the course of the full disassembly, the ball nut was fully removed from the ball screw which resulted in the unpacking of all 166 ball bearings. Though there were some spaces in the channels between bearings, Thompson verified that
the number of bearings was within the tolerance of the ball nut and that it would work smoothly and without error under those conditions.

Once everything had been verified and finalized, the redesigned test stand was assembled, as seen in Fig. 2.3. The ball bearings were placed back inside the ball nut, which was loaded back onto the ball screw. The nut plate was attached to the ball nut, and the frame was assembled with the nut plate and linear ball bushings on the rails. The motor was attached to the mount, centered, and angles were checked, at which point the bushings were screwed to the nut plate. The thrust bearing was loaded by turning the ball screw until the nut plate was several inches into the loaded zone, then the motor was bolted down for the second time to make sure that the thrust bearing, rather than the motor, absorbed the full load. Very few changes were made between the finished concept and the final stand. The rubber stops proposed in the design in Fig. 2.2 were replaced with short springs, and the aluminum disks attached to the bushings were replaced with washers, also added at every other spring connection point.

Figure 2.3: Redesigned and Assembled Bench top EMA
2.3 Data Collection: First Iteration

The first testing effort undertaken after system reassembly was a simple, unloaded triangle position profile, moving at a constant velocity backwards and forwards for several iterations. While the system was very capable at moving in one direction, directional changes proved very difficult for the motor. The motor was designed to accept one input for the direction and one input for the velocity, which made it very challenging to design an automated PID controller which would smoothly transition directions. Even after an acceptable solution was achieved for unloaded motion, attempting to use the BLDC motor in a loaded application to achieve a controlled profile, particularly during the linearly reversing motion, was unsuccessful. When the motor entered into a reversing condition, the springs accelerated the motor in that direction. The Simulink controller compensated by reducing the voltage and without a brake, eventually attempt to drive the motor in the forward direction to counter the motion. This caused the motor to chatter loudly as it transitioned rapidly between direction inputs. Beyond making the data collection impossible, this would almost certainly lead to a swift destruction of the motor, so a new approach was taken.

A profile was designed which embraced the natural condition of the motor. This followed a constant voltage profile out to a desired loaded position at which point the voltage was reduced to zero, then given a very slightly negative voltage that allowed it to move freely under the force of the springs. Once the encoder sensed that the motor velocity had reached zero after the completion of the reverse portion of the profile, the controls reapplied a forward command to bring it to the desired loaded position. While this profile allowed for smooth motion of the EMA system, it sacrificed some data collection capability by not being able to control velocity or reverse motion.
Under voltage control, the velocity dropped approximately 25 rpm, from 115 to 90 across a 2.5 inch profile, and followed a somewhat noisy path as seen in Fig. 2.4. Once the profile was programmed, the PID controller gains were optimized, which led to the elimination of the derivative gain, as it became very noisy at any level where it was noticeable in the system. Proportional control was the largest influence at six times the integral control, which produced a responsive and stable system.

Several inconsistencies in the data signals arose as data collection progressed. The position data developed a slight positive creep, which did not exist inherent to the encoder data output device. The noise from the accelerometer signal spanned $\pm 0.03V$, which also reduced when other components were unplugged or turned off. The controller and power supply were diagnosed as the main cause of the grounding issue, which was resolved by connecting both directly to the building ground using two grounding braids. The braids eliminated the creep in the position data, reduced the
noise in the signal by a factor of five, and reduced the noise in the accelerometer by a factor of six. The improvement in the accelerometer signal can be seen in Fig. 2.5.

### 2.4 Servomotor Acquisition

While functional, the compromises in the data necessary to avoid purchasing a new motor proved too great. Velocity control algorithms both in Simulink and native to the controller were unsteady. Voltage commands were locally smoother but inconsistent against load, and most importantly the nature of the system made it impossible to run dynamic profiles similar to those encountered in real-world applications. Industrial EMAs are built with servomotors which provide smooth operation with very precise control over angular position, velocity, acceleration, and torque for much better control of system behavior. The requirements for such control involve a specialized, dedicated controller which would involve replacement of motor and controller components, as well as creating a new control scheme which
would most likely be in an unfamiliar programming language. Despite the many costs of installing such a system, in order to execute the desired profiles for this testing and allow for more advanced future work, the decision was made to invest in a servomotor for the system. First, the system requirements were calculated. With a maximum expected load of 100 pounds, the torque required from the motor was calculated using:

$$T = \frac{Fl}{2\pi\epsilon}$$

(2.3)

where $F$ is the load, $l$ is the lead of the ball screw, and $\epsilon$ is the ball screw efficiency, estimated at 90 percent. This yields a necessary torque of 3.5 in·lb, or 0.4 Nm. The speed requirement remained at 600 rpm. After several inquiries, Industrial Indexing Systems, a locally based company with strong ties to RIT, offered to very generously donate one of their systems to the project. Part of their Luminary Series, the LMC-400 controller, LD-330 driver and ESM60A motor proved to be an excellent addition to the test stand, offering all the benefits of a servomotor and exceeding the requirements with a 1.27 Nm rated torque and 3000 rpm top speed. The new shaft size of the motor required a new coupling, and a Helical coupling was chosen. The coupling has a spring type connection between the two shaft points which is rigid in the rotational direction, but has the ability to stretch or compress slightly in the axial direction. This feature worked very well for the system as it ensured that the thrust bearing supported the full load of the springs. After a motor mount was built to accommodate the new motor, the test stand was finalized, as seen in Fig. 2.6 and Table 2.2. Velocity data saw the desired improvement, and can be seen in Fig. 2.7 turning consistently at 240 rpm under a load reaching 67 pounds.
Figure 2.6: Finalized Bench Top EMA Fixture

Figure 2.7: Consistent Velocity Under Load Using Servomotor
<table>
<thead>
<tr>
<th>Component</th>
<th>Model</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Servomotor</td>
<td>ESM60A</td>
<td>Industrial Indexing Systems</td>
</tr>
<tr>
<td>Servomotor Controller</td>
<td>LMC-400</td>
<td>Industrial Indexing Systems</td>
</tr>
<tr>
<td>Motor Driver</td>
<td>LD-330</td>
<td>Industrial Indexing Systems</td>
</tr>
<tr>
<td>Ball Screw</td>
<td>3/4” Diameter, 13/64” Lead (190-9101)</td>
<td>Thomson</td>
</tr>
<tr>
<td>Ball Nut and Flange</td>
<td>3/4” Diameter, 13/64” Lead (8107-448-016)</td>
<td>Thomson</td>
</tr>
<tr>
<td>Guide Rails</td>
<td>1/2”x12” 60 Case Standard Shafting</td>
<td>Thomson</td>
</tr>
<tr>
<td>Ball Bearing</td>
<td>3/8” ID (R6)</td>
<td>SKF MRC</td>
</tr>
<tr>
<td>Thrust Bearing</td>
<td>1/2” ID (AO/SAO16)</td>
<td>Boston Gear</td>
</tr>
<tr>
<td>Springs</td>
<td>6.9 lb/in, 8.4 lb/in (RW-10C, RW-10B)</td>
<td>Lamina Components</td>
</tr>
<tr>
<td>Motor Coupling</td>
<td>1/2x1/4 (BDCAS125-16-8)</td>
<td>Helical</td>
</tr>
<tr>
<td>Mechanical Breadboard</td>
<td>22”x22” Surface</td>
<td>Newport Corporation</td>
</tr>
<tr>
<td>Raw Materials</td>
<td>Front &amp; Rear Support, Nut Plate</td>
<td>Rochester Steel Surplus</td>
</tr>
<tr>
<td>Motor Mount</td>
<td>Aluminum Bracket</td>
<td>RIT Machine Shop</td>
</tr>
<tr>
<td>Accelerometer</td>
<td>ICP Triaxial (356A16)</td>
<td>PCB Piezotronics</td>
</tr>
</tbody>
</table>

Table 2.2: Final System Bill of Materials
Chapter 3

Analysis Approach

3.1 Wavelets

The wavelet transform represents a very powerful time-frequency signal analysis technique. The wavelet itself is a signal of zero mean which must be localized in both time and frequency space. For example, the Morlet base wavelet, as seen in Fig. 3.1, is designed simply by multiplying a plane wave by a Gaussian envelope [24], and is one of several dozen types of wavelets.

Figure 3.1: Morlet Wavelet
The method by which the wavelet analyzes a range of time and frequency values is a combination of scaling and shifting. Scaling involves stretching the base wavelet in the time domain, which determines the time and frequency resolution of the signal. The frequency represented by a particular scale also varies based on the wavelet type and sampling frequency. By shifting a base wavelet over the duration of the signal, the frequency content at each time is able to be extracted from the signal. This is all done according to the equation:

\[
wt(s, \tau) = \frac{1}{\sqrt{s}} \int x(t)\psi^* \left( \frac{t - \tau}{s} \right) dt
\]

where \( s \) is the scaling factor, \( \tau \) is the shifting parameter, \( x(t) \) is the signal being processed, \( \psi(t) \) is the base wavelet, and \( \psi^*(t) \) represents its complex conjugate.

Wavelet scaling and shifting is demonstrated in Fig. 3.2. The plots on the left show the Morlet with modified amplitude for visualization purposes at three different scales, applied to a section of data corresponding to the frequency of that scale. A high scale such as in the top left plot has a large stretch, corresponding to the lower frequencies in a signal. Conversely, a low scale as in the bottom left of the figure has a very short stretch that allows it to identify high frequencies in the data. In the course of the analysis a wavelet of each scale shifts across the entire signal according to the shifting and scaling parameters. The transformation creates coefficients across the length of the signal, whose magnitude depends on how accurately the signal matches the wavelet. The 25th scale is a modest fit to data of moderate amplitude, and yields relatively low coefficient magnitude. The 14th scale is a very good fit, and fits high amplitude data, yielding very high magnitude coefficients for the first half of the signal. The sixth scale of the Morlet
wavelet compared to the signal is a very good fit, however the amplitude of the signal is small, yielding a relatively low coefficient magnitude to the plot on its right. The coefficient vectors build a matrix of coefficients, which is displayed as a scalogram.

In Fig. 3.3, the same signal from Fig. 3.2 is analyzed with a full continuous wavelet transform (CWT), shown using a scalogram. The scalogram shows how the signal responds to dynamic frequencies over the full time
Figure 3.3: Continuous Wavelet Transform Over a Short Time Window

and frequency of the analysis. In Fig. 3.3, the blue color corresponds to a low magnitude and red to the highest magnitude, using the absolute value of the CWT coefficient matrix. The scale values on the y-axis of the CWT increase from 1-48, with the highest frequency on the bottom at scale one and the lowest frequency at scale 48 on the top. The high magnitude of the wavelet analysis at scale 14 from Fig. 3.2 is clear in the scalogram, with the corresponding reduction just after 0.015 seconds. The consistent low magnitude signal at scale 25 is also clear, with a slight increase at 0.025 seconds. High frequency content at scale six is also low magnitude but
more erratic, with a peak at 0.025 seconds corresponding to expectations based on Fig. 3.2.

A very popular frequency analysis technique is the fast Fourier transform (FFT), a discrete Fourier transform (DFT) algorithm that substantially reduces computation time of the standard formula. The DFT is defined by:

\[ X(k) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi kn/N} \]

where \( N \) is the number of samples and \( k = 0, \ldots, N - 1 \) is the frequency index. While very powerful for the analysis of stationary signals, when a signal is non-stationary and varies over time, the frequency analysis is smeared across the whole frequency spectrum of the signal, yielding little to no useful information about the signal. The continuous wavelet transform’s (CWT’s) ability to shift scales and capture different frequencies over time proves to be very beneficial in this application. Fig. 3.4 shows a comparison of the FFT against the wavelet for a stationary signal. The FFT has more exact frequency and magnitude identification, showing 5 and 33 Hz at 1 and 0.75 magnitude, while the CWT shows higher magnitude at lower frequency around scale 150 and lower magnitude high frequency content around scale 25. The frequency content does not vary at all over time, therefore the FFT is the more appropriate technique in this case.

Figure 3.5 shows a non-stationary chirp signal increasing from 5 to 33 Hz. In this case, the FFT is very unclear, and reveals only the total spectrum of the signal. The CWT, on the other hand, is able to show the rate and approximate frequency of the dynamic signal over time. In a scenario where the nature of the signal is unknown, it may be most effective to first analyze the time-frequency spectrum using a wavelet analysis, then use an FFT if
the signal appears stationary.

The short-time Fourier transform (STFT) is another solution to the smearing problem. In this case, many FFTs are taken across the signal with a set fraction of the signal’s time content and the results are displayed in time and frequency, with the magnitude in the third dimension, similar to the wavelet scalogram. While the STFT is able to show a signal’s development over time, the time and frequency resolutions are dependent on each other under
the relationship,

\[ \Delta \tau \cdot \Delta f \geq \frac{1}{4\pi} \]  \hspace{1cm} (3.3)

where \( \Delta t \) represents time resolution and \( \Delta f \) represents frequency resolution [25]. This forces the user to make a compromise between frequency resolution and time resolution, parameters which are invariant through the full time and frequency plotted, as seen in Fig. 3.6. In this case, if the user wants to see high frequency fluctuations, a high time resolution must be chosen,
sacrificing frequency resolution and making it very difficult to pinpoint the frequency location of the signal, and a similar sacrifice is involved when a high frequency resolution is required, particularly since the STFT cannot resolve a feature whose time interval is longer than the window of the STFT. The wavelet analysis, however, varies throughout the frequency range, giving a high time resolution and low frequency resolution at high frequencies, and low time resolution and high frequency resolution at low frequencies. While still not perfect, this allows for the high time resolution necessary to identify high frequency signals, and gives the lower frequencies a higher frequency resolution due to their naturally lower time resolution.

The discrete wavelet transform (DWT) is another wavelet tool used to perform time-frequency analysis of a signal. The CWT uses many scales with a relatively high frequency resolution, particularly at low frequency. As the scale of a base wavelet increases at high scale, from 100 to 101 for example, the frequency of the wavelet sees a very small change, which can lead to redundant information at low frequency, as seen in Fig 3.4. The DWT looks at fewer scales by increasing by an exponential, from scales 2 to 4 to 8, for example. The exponential nature of the DWT allows the signal to

![Figure 3.6: (a) Continuous wavelet analysis. (b) STFT where $\Delta t_1 = \Delta f_1$, and $\Delta t_1 \cdot \Delta f_1 = \frac{1}{4\pi}$ (c) STFT where $\Delta t_2 = 2\Delta t_1$ and $\Delta f_2 = \frac{1}{2}\Delta f_1$; $\Delta t_2 \cdot \Delta f_2 = \frac{1}{4\pi}$](image)
be efficient while still retaining most characteristic frequency information. The DWT discretizes the scaling and shifting parameters as,

\[
\begin{align*}
  s &= (s_0)^j \\
  \tau &= k\tau_0(s_0)^j
\end{align*}
\]

which then gives the DWT base wavelet equation,

\[
dwt(j, k) = \frac{1}{\sqrt{(s_0)^j}} \int x(t) \psi^* \left( \frac{t - k\tau_0(s_0)^j}{(s_0)^j} \right) dt
\]

The parameters \( s_0 = 2 \) and \( \tau_0 = 1 \), which constitute the dyadic discretization, are commonly chosen. The wavelet input to a DWT must be orthogonal, which falls within the framework of the dyadic discretization. If a wavelet is orthogonal, that means that the product of the wavelet and itself is unity, while an inner product of the base wavelet with a shifted or scaled version of the wavelet will be zero [25]. Since the DWT requires orthogonality, some wavelets cannot be used in this manner. The DWT is used as a combination of low-pass and high-pass filters built around the architecture of the base wavelet, which decompose the signal into a lower frequency signal known as the approximation and a higher frequency signal known as the detail. Using \( s_0 = 2 \), the filters are half band filters, splitting the frequency content of the signal into two parts of equal resolution. For example, if the signal is sampled at 10 kHz, the Nyquist frequency is 5 kHz, and the first level wavelet decomposition includes an approximation signal with the content between 0 and 2,500 Hz, and a detail signal with the content between 2,500 and 5,000 Hz. In each case half of the frequencies of the original have been removed, therefore each of the decomposed signals can be down sampled by a factor of two according to Nyquist’s rule, yielding two signals each of half the original length.
Figure 3.7: Five Level Discrete Wavelet Decomposition Using Daubechies 4 Wavelet

Moving from level one to level two applies the same steps to the level one approximation signal ($a^1$) that were previously applied to the original. High pass and low pass filters are applied to the approximation signal, yielding the approximation and detail signals for the next level, each additional level containing approximately 50% of the data from the previous level. In order
Figure 3.8: Five Level Discrete Wavelet Reconstruction Using Daubechies 4 Wavelet

to preserve the energy of the down sampled signal, the magnitude of the points in the approximation and detail may be greater than the magnitude of the original signal [26]. The down sampling and change in magnitude can be seen in Fig. 3.7.

If a five level DWT is performed, the data is stored in an array with the
signals adjacent, as $[a^5|d^5|d^4|d^3|d^2|d^1]$, seen graphically in Fig. 3.13. The approximation signals previous to $a^5$ are left out because the information contained in them was used and conserved in the development of the higher level details and approximations. The DWT can be reconstructed using the inverse of the initial transformation. If this is performed using the full DWT with an orthogonal wavelet, the full signal will be reproduced exactly. It is also possible to extract the approximation and detail signals individually or in a combination, seen individually in Fig. 3.8 corresponding to Fig. 3.7. There are many benefits and possible applications to this approach. The data in Fig. 3.8 shows a very noisy original signal, though in $a^4$ and $d^5$ a periodic pulse occurring every 0.05 seconds can be seen clearly which corresponds to a defect in the system. It is also very common in signal denoising, where the noise is often much higher frequency than the signal and can easily be extracted in the appropriate detail levels, the approximation signal for which is then reconstructed to be much smoother, as seen in Fig. 3.9. This can also be used for signal compression in the case where the majority of the signal’s

![Figure 3.9: Wavelet Denoising](image)
energy is contained at a frequency much lower than the Nyquist frequency, allowing the use of an approximation signal at a fraction of the original signal’s length. In the case of this work, the DWT is applied in order to eliminate system noise and highlight the signal corresponding to the defect frequency, choosing the most effective levels and applying a threshold on the coefficients. This is described in detail in the next section.

### 3.2 Initial Simulation

Simulated data was based on the spall fault planned for the thrust bearing. Spalling is a common bearing defect and can lead to high vibration, increased load on the motor, or in extreme cases can lead to bearing seizure. Normal fatigue can be the cause of spalling, though in most cases it is one of many other potential problems, such as overloading, misalignment, indentations, deep seated rust, electrical current damage, and smearing [27].

Spalling reveals itself in vibration data through a periodic signal based on the characteristics of the bearing and the system. In the case of a single spall on the bearing race, the balls are expected to each pass over the defect at an interval directly related to the shaft speed at a frequency known as the ball pass frequency (BPF). In a radial type bearing with different radii to the inner and outer race, there are two equations, one for an inner race defect and one for an outer race defect, as seen in Equations (3.6) and (3.7).

\[
BPFI = \frac{Nb}{2} \cdot f_r \left[ 1 + \left( \frac{Bd}{Pd} \cdot \cos\theta \right) \right] \tag{3.6}
\]

\[
BPFO = \frac{Nb}{2} \cdot f_r \left[ 1 - \left( \frac{Bd}{Pd} \cdot \cos\theta \right) \right] \tag{3.7}
\]
Where the ball pass frequencies are in Hz, \( N_b \) is the number of bearings, \( f_r \) is the shaft speed in rotations per second, \( B_d \) is the ball diameter, \( P_d \) is the pitch diameter, and \( \theta \) is the contact angle. The contact angle is formed by drawing a line from the center of the bearing out to the balls, and from the balls to the race. In a radial type bearing, this is a straight line, and the cosine function has a value of one. However, in a thrust bearing, the contact angle is \( 90^\circ \) which gives the cosine function a value of zero, leaving the value in the brackets at one for either equation. The result can be seen in Equation (3.8).

\[
BPF = \frac{N_b}{2} \cdot f_r \tag{3.8}
\]

The Boston Gear bearing chosen for testing has ten bearings and initial simulations were done with a shaft speed of 100 rpm. This gives a BPF of 8.33 Hz, or one pulse every 0.12 seconds. Using data taken from the

![Y Vibration Frequency Spectrum, Gearmotor Loaded at 100 RPM][1]

Figure 3.10: Frequency Spectrum From Using the BLDC Gear Motor
BLDC gear motor and a FFT program which performs Hanning windowing and averaging, the primary natural frequency of the system is established to be approximately 600 Hz, which can be seen in Fig. 3.10. The fault pulse developed vibrates at the natural frequency of the system and decays to one percent of its peak value by 0.006 seconds. The equation and corresponding plot are Eqn. 3.9 and Fig. 3.11.

\[ 0.3(e^{-800t} \sin(2\pi 600t)) \] (3.9)

Using a Hanning window multiplies the data of interest by a single cosine cycle shifted by \( \pi \) radians and scaled to fit the domain of the data. This gives the data at the beginning and end of the sample a value of zero, which reduces the leakage that can be encountered in FFT analyses. An FFT expects a signal to be periodic and infinitely repeating in nature, which is almost never the case. Using the Hanning window allows the data to begin and end
in the same manner, giving it the repeating characteristic desired for the FFT [28]. As this can potentially hide data on the edges of a sample, the window extracts a set fraction of the data set in each sample and is shifted across the entire set with 40% overlap, which ensures that all of the data was seen in the analysis. Each window is then averaged to use all of the data and reduce noise.

A simulated periodic signal of decaying pulses is designed using Eqn. 3.9. A bounded random scaling factor is built into the fault signal to add variability to the peaks and create more realistic results. The fault is then added to vibration data taken from the gear motor, the result of which can be seen in Fig. 3.12.

The resulting data is processed using a DWT, with the initial goal of investigating the potential of the method against a spall fault. After testing several dozen different base wavelet types, the wavelet decomposition is performed using a Coiflet 4 wavelet. A thresholding function is developed,

![Figure 3.12: Healthy Signal vs Faulted Signal with Pulses Added](image)
with the final threshold based on [29] the universal threshold rule:

\[ t = \sigma \sqrt{2 \log N} \]  \hspace{1cm} (3.10)

where \( \sigma \) is the standard deviation of the signal, \( N \) is the number of data points, and \( t \) is the threshold. The equation is modified with a constant multiplier outside the square root according to the nature of the data. The data is broken down into seven levels which create \( a_7 \) and \( d_1 \) through \( d_7 \), and the threshold of each level is taken individually. Given the potential non-stationary character of the signal, each level is further broken down into sections as an extra step to account for the variable standard deviation throughout each level. After a variety of numbers were tested through many iterations, six sections were chosen. Once the threshold is determined based on the standard deviation of the section, any DC offset is removed from the section and points below the threshold are set to zero. Furthermore, the first several layers were set to zero in their entirety, given their high noise and largely irrelevant content. The result can be seen in Fig. 3.13 which shows the full decomposition array in series at the top figure, and the resulting array after the threshold in the lower two.

Levels five through seven are chosen as the most effective for simulated pulse identification after extensive visual figure interpretation. While layers three and four better match the frequency of each individual pulse, as seen in Fig. 3.14, there are at the natural frequency of the system which causes many false identifications, which can be seen in Fig. 3.15.

Levels five through seven, on the other hand, are not a precise match for the frequency of the individual pulse, but are much more resistant to false identifications. The result can be seen in Figs. 3.16 and 3.17. Wavelets’ ability to function in time and frequency allows them to perform a local
Figure 3.13: Discrete Wavelet Transformation and Thresholding
Figure 3.14: Discrete Wavelet Reconstruction Using Levels 3 and 4

Figure 3.15: Discrete Wavelet Reconstruction Using Levels 3 and 4, Zoomed
Figure 3.16: Discrete Wavelet Reconstruction Using Levels 5 and 7

Figure 3.17: Discrete Wavelet Reconstruction Using Levels 5 and 7, Zoomed
analysis on the signal, identifying even minor discontinuities like those encountered in the simulated fault signal due to the sudden impulse in a manner that resonates through a wide frequency range, making it identifiable even outside of the pulse’s frequency where system noise is significantly reduced.

Another possible result of unexpected content is an extension of period, in the case where a regular frequency in the system is disrupted and lengthened due to an unexpected event. This has the potential to double the period of that particular portion of the signal if it peaks as the event occurred, then the vibration in the positive direction from the event overcomes the vibration in the negative from the natural course of the system. This case would also give the wavelet a strong signal in a lower frequency range. Using the level 5-7 decomposition the algorithm visually identified a high percentage of peaks, with very few false identifications, 94% correctly identified and one falsely identified in the case of Fig. 3.16.

3.3 Updated Simulated Signal

A more thorough frequency analysis was performed after data from the servomotor bench top EMA was collected which allowed for more accurate signal development. First, natural frequency tests were performed across the whole system. With the accelerometer on the front support, where it would be placed for all testing, the ball screw, ball nut, and front support were tested by tapping a piece of wood repeatedly on each piece. Wood was chosen because it excited the frequencies inherent to the component with a much lower chance of saturating the accelerometer than if a material with a much higher modulus of elasticity such as steel were chosen. The ball screw was tested by tapping in a direction perpendicular to the ball screw at
the far end of the system. The ball nut was tested by tapping on the ball nut in a direction parallel to the ball screw, and the front support was tested by being tapped in a direction parallel to the ball screw. Results of the testing can be seen in Fig. 3.18.

![Graph showing natural frequency tested at ball nut, ball screw, and front support in bench top EMA](image)

**Figure 3.18:** Natural Frequency Tested at Ball Nut, Ball Screw, and Front Support in Bench Top EMA
The ball nut contains only one significant peak, which occurs at 230 Hz. The ball screw shows several peaks below 100 Hz, as well as frequency content surrounding 420 Hz and 850 Hz in addition to a peak at 230 Hz, with relatively low content elsewhere. The FFT from the front support is clearly the most dynamic, with content growing from 400 Hz and peaking close to 700 Hz, and two peaks at 1350 Hz and 1900 Hz. Sustained content between 850 Hz and 1000 Hz can also be seen, as well as a small increase at 2500 Hz.

Data from the system in motion was also analyzed to further validate the results. A 240 rpm triangular profile was performed under load with a healthy bearing, and the resulting frequency spectrum can be seen in Fig. 3.19. Overall data from the profile showed peaks similar to what was found from the natural frequency analysis, confirming what was found and allowing for the development of a simulated signal.

Some modifications were necessary when creating the simulated signal.

Figure 3.19: Healthy Frequency Spectrum From Servomotor EMA
The low frequency content from the ball screw is negligible in the profile data, so it is ignored. Both the ball screw and ball nut contain the same peak at 230 Hz, but because it is the only peak contained in the ball nut analysis, it is attributed to the nut and scaled down in the system analysis. A 500 Hz signal is very strong in the system FFT, which is most likely a result of the front support natural frequency, and is increased from the natural frequency plot to more closely match the system FFT. The frequency content near 850 Hz seen in both the ball screw and front support is attributed to the ball screw. Other significant content in the system FFT is captured by the natural frequency plots and attributed accordingly. This led to the creation of the 240 rpm simulated signal through a series of scaled frequency components with added Gaussian noise, as seen in Eqn. 3.11,

\[
\begin{align*}
    f(x) &= \sin(x(2\pi(t + \phi_r))) \\
    \text{Support} &= 0.75 \cdot f(500) + f(720) + 0.66 \cdot f(1350) \\
                   &\quad + 0.15 \cdot f(1700) + 0.05 \cdot f(2500) + 0.2(N) \\
    \text{Nut} &= f(230) + 0.2(N) \\
    \text{Screw} &= f(420) + 0.66 \cdot f(800) + 0.66 \cdot f(900) + 0.2(N) \\
    y &= 0.02 \cdot (\text{Support} + 0.33 \cdot \text{Nut} + 0.5 \cdot \text{Screw})
\end{align*}
\]

(3.11)

where \(f(x)\) determines the frequency content, \(\phi_r\) is a random \(\phi\) value for each iteration of \(f(x)\), \(t\) is time, and \(N\) is Gaussian noise. Each individual frequency is scaled according to its contribution to its component, and each component is scaled according to its contribution to the system. The final output is then scaled to match system vibration levels. The result can be seen in Fig. 3.20 compared to actual profile data.

Once an effective healthy simulation was established, four faults were injected: bearing race spall, bearing ball spall, high bearing wear, and ball
Figure 3.20: Comparison of Vibration Data from System and Simulation

screw wear. The bearing race spall is injected by increasing the ball screw and ball nut components of the signal at intervals corresponding to the ball pass frequency (BPF) of the bearing by factors of 75% and 38%, respectively. The ball spin spall is injected using the same modifications to the signal components, but at an interval corresponding to the ball spin frequency (BSF), which is approximately half the BPF. The ball screw and ball nut are expected to express content with a higher signal to noise ratio during race and ball spalls due to the relatively low baseline signal at those frequencies and the impulse experienced by those components under a spall fault. The bearing wear fault is introduced by increasing the front support component by 50% as it most likely serves to increase the general noise directly on that component. Finally, the ball screw wear fault is injected by increasing the influence of the ball screw and ball nut by 50% each for the same reasoning. Once the faults had been established, the wavelet methodology was
investigated. Continuous wavelet transforms (CWTs) were chosen as the primary tool rather than the previously examined discrete wavelets because the CWT gave a higher resolution signal in the scale domain that was clearer when attempting to visually diagnose early faults in the system data.

### 3.4 Maximum Energy-to-Shannon Entropy Ratio

First, the wavelet type had to be chosen. In [25], Yan presents an intuitive method for wavelet selection, the Maximum Energy-to-Shannon Entropy Ratio Criterion (MESER), which was chosen to rate the performance of each wavelet against the data. The energy contained in a signal is the collective magnitude of each point in that signal, defined in the discrete domain as:

$$E = \sum_{i=1}^{N} |x(i)|^2$$  \hspace{1cm} (3.12)

where $E$ is the energy, $x(i)$ is the signal amplitude, and $N$ is the length of the signal. Examining a wavelet in particular, the energy associated with each scale $s$ is:

$$E_{\text{energy}}(s) = \sum_{i=1}^{N} |W(s, i)|^2$$  \hspace{1cm} (3.13)

A wavelet transform applied to a signal yields a high value if the scale and shape of the wavelet are similar to the signal being tested. As expected, a better match yields a higher value. When considering fault analysis, wavelets that are able to extract more energy from the fault signal will be more effective at processing the signal. The criterion for wavelet selection that follows is the Maximum Energy Criterion: the wavelet that is able to extract the most energy from the defect-induced transient vibration signal is the most appropriate wavelet for performing fault extraction.
While extracting energy from a signal is valuable, energy extraction on its own may not characterize the fault impulses of a signal, as the distribution of that energy may vary significantly. A wavelet signal with consistent content could have a higher energy value than another wavelet that characterizes only the impulses. In order to avoid this error, the Shannon entropy is considered. The Shannon entropy describes the energy distribution of the coefficients as:

\[ E_{\text{entropy}}(s) = - \sum_{i=1}^{N} p_i \cdot \log_2 p_i \]  

(3.14)

Where \( p_i \) is the energy probability distribution of the wavelet coefficients, defined as:

\[ p_i = \frac{|W(s,i)|^2}{E_{\text{energy}}(s)} \]  

(3.15)

with \( E_{\text{energy}}(s) = \sum_{i=1}^{N} p_i = 1 \), and \( p_i \cdot \log_2 p_i = 0 \) if \( p_i = 0 \). Equations (3.14) and (3.15) indicate that the Shannon entropy must be bounded as:

\[ 0 \leq E_{\text{entropy}}(s) \leq \log_2 N \]  

(3.16)

where \( E_{\text{entropy}}(s) \) will be zero if all wavelet coefficients are zero except one, and \( \log_2 N \) if the energy probability distribution is the same for all wavelet coefficients. Therefore, an ideal analysis with a majority of low magnitude signals and few high magnitude peaks corresponding to fault impulses should yield a low Shannon entropy value. The corresponding indicator is the Minimum Shannon Entropy Criterion: the base wavelet which minimizes the Shannon entropy for the defect-induced transient vibration signal represents the most effective wavelet for performing fault extraction.

The Energy-to-Shannon Entropy Ratio which follows attempts to provide a metric which maximizes the energy while minimizing the Shannon
entropy of the wavelet coefficients by simply dividing the energy by the entropy:

\[ R(s) = \frac{E_{\text{energy}}(s)}{E_{\text{entropy}}(s)} \]  

(3.17)

which finally leads to the Maximum Energy-to-Shannon Entropy Ratio Criterion: the base wavelet which maximizes the energy-to-Shannon entropy ratio for the defect-induced transient vibration signal represents the most effective wavelet for performing fault extraction.

The MESER metric was applied to a one second simulated fault signal with a sampling rate of 10 kHz based on a bearing race spall and tested with 64 wavelets, the results of which can be seen in Table 3.1. Each wavelet was tested at each scale between 1 and 64, scaled so that the highest value in each

<table>
<thead>
<tr>
<th>Base Wavelet</th>
<th>MESER</th>
<th>Base Wavelet</th>
<th>MESER</th>
<th>Base Wavelet</th>
<th>MESER</th>
<th>Base Wavelet</th>
<th>MESER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haar</td>
<td>304.7</td>
<td>Sym3</td>
<td>311.5</td>
<td>Bior3.7</td>
<td>296.9</td>
<td>Rbio3.9</td>
<td>307.1</td>
</tr>
<tr>
<td>Db2</td>
<td>297.8</td>
<td>Sym4</td>
<td>319.2</td>
<td>Bior3.9</td>
<td>298.1</td>
<td>Rbio4.4</td>
<td>312.9</td>
</tr>
<tr>
<td>Db3</td>
<td>302.5</td>
<td>Sym5</td>
<td>294.4</td>
<td>Bior4.4</td>
<td>310.3</td>
<td>Rbio5.5</td>
<td>297.2</td>
</tr>
<tr>
<td>Db5</td>
<td>289.7</td>
<td>Sym7</td>
<td>287.5</td>
<td>Bior6.8</td>
<td>325.5</td>
<td>Meyr</td>
<td>267.9</td>
</tr>
<tr>
<td>Db6</td>
<td>278.4</td>
<td>Sym8</td>
<td>292.4</td>
<td>Rbio1.1</td>
<td>304.7</td>
<td>Dmey</td>
<td>270.9</td>
</tr>
<tr>
<td>Db7</td>
<td>269.7</td>
<td>Bior1.1</td>
<td>304.7</td>
<td>Rbio1.3</td>
<td>302.8</td>
<td>Gaus1</td>
<td>309.3</td>
</tr>
<tr>
<td>Db8</td>
<td>272.2</td>
<td>Bior1.3</td>
<td>299.4</td>
<td>Rbio1.5</td>
<td>341.5</td>
<td>Gaus2</td>
<td>317.5</td>
</tr>
<tr>
<td>Db9</td>
<td>279.9</td>
<td>Bior1.5</td>
<td>299.8</td>
<td>Rbio2.2</td>
<td>318.7</td>
<td>Gaus3</td>
<td>319.8</td>
</tr>
<tr>
<td>Db10</td>
<td>269.8</td>
<td>Bior2.2</td>
<td>323.4</td>
<td>Rbio2.4</td>
<td>314.2</td>
<td>Gaus4</td>
<td>320.4</td>
</tr>
<tr>
<td>Coif1</td>
<td>314.4</td>
<td>Bior2.4</td>
<td>312.1</td>
<td>Rbio2.6</td>
<td>311.3</td>
<td>Gaus5</td>
<td>329.9</td>
</tr>
<tr>
<td>Coif2</td>
<td>323.0</td>
<td>Bior2.6</td>
<td>315.1</td>
<td>Rbio2.8</td>
<td>300.9</td>
<td>Gaus6</td>
<td>313.9</td>
</tr>
<tr>
<td>Coif3</td>
<td>325.5</td>
<td>Bior2.8</td>
<td>313.4</td>
<td>Rbio3.1</td>
<td>305.9</td>
<td>Gaus7</td>
<td>343.1</td>
</tr>
<tr>
<td>Coif4</td>
<td>322.7</td>
<td>Bior3.1</td>
<td>319.0</td>
<td>Rbio3.3</td>
<td>305.5</td>
<td>Gaus8</td>
<td>315.2</td>
</tr>
<tr>
<td>Coif5</td>
<td>307.6</td>
<td>Bior3.3</td>
<td>316.0</td>
<td>Rbio3.5</td>
<td>298.9</td>
<td>Mexh</td>
<td>317.4</td>
</tr>
<tr>
<td>Sym2</td>
<td>314.9</td>
<td>Bior3.5</td>
<td>301.7</td>
<td>Rbio3.7</td>
<td>300.9</td>
<td>Morl</td>
<td>311.7</td>
</tr>
</tbody>
</table>

Table 3.1: Base Wavelet MESER Comparison
Figure 3.21: Simulated Vibration Data for Healthy and Faulted EMA Conditions
Figure 3.22: CWT Analysis on Simulated Vibration Data
coefficient matrix was one, and the highest MESER result was recorded. All of the base wavelets performed adequately, and many performed very well with the Gaussian 7 base wavelet yielding the highest MESER at 343.1.

Figure 3.21 shows the raw vibration signal output from the simulation, and Fig. 3.22 shows the subsequent CWTs for each condition using the Gaussian 7 wavelet. The simulation uses a CWT with individually chosen scales 4, 8, 12,..., 48, as they contained the most important frequency content and allowed for a much smaller matrix with much faster computation time in both sample and feature generation.

3.5 Features and Classification

Once the wavelet was chosen, the same thresholding function used previously on the DWT is applied to the CWT to highlight fault characteristics. Next, features were chosen from the wavelet analysis to be used as inputs. Several important considerations went into this decision. First, the number of features would have to be limited to simplify the classification algorithm. This allows for reliable training with practically sized data sets with faster and often more accurate results. Second, it was important to find a way to incorporate the time-frequency nature of the CWT, as this could support the classification of non-stationary and time-dependent signals. Third, data from both low and high frequency was necessary because different faults manifested themselves in different frequency ranges. In a variety of previous work ([6, 17, 22, 23]), statistical features were used to reduce the number of features in a data set while retaining the most valuable information.

The most common features used were standard deviation, kurtosis, skewness, and root mean square (RMS). Standard deviation measures the signal’s
proximity to its average, kurtosis measures the peakedness of a signal, skewness is a measurement of the asymmetry of a signal’s probability distribution, and RMS is a measurement of the magnitude of a signal. In order to use a signal from the time domain, a chosen frequency band in the wavelet transform was sampled so that it used only enough data to categorize frequency data at the scales chosen. Multiple frequency bands were incorporated by taking statistical features from both a low frequency band and a high frequency band, as well as using a sampled waveform from each. The high frequency band used a summation over the frequency range of the data from approximately 400 Hz to 500 Hz in the CWT, and the low frequency band used a summation over the frequency range of the data from approximately 50 Hz to 300 Hz in the CWT. Several feature combinations were tested both during and after the classifier was being developed, the results of which are described at the end of this section.

Once the features were established, a classification scheme was pursued. With the power and adaptability of neural networks, in addition to the success of previous work using neural networks ([6, 17, 23]), a feed-forward multilayer perceptron artificial neural network (ANN) was first pursued to classify the data. A multilayer perceptron ANN consists of an input layer, at least one hidden layer, and an output layer, a model of which can be seen in Fig. 3.23. Each layer contains a certain number of nodes which can vary widely between networks depending on what is required. Here, the nodes are represented by the shapes in each column under the name of the layer in the model. The input layer consists of the inputs to the classifier, which in this case are all of the features in each sample, as the ANN classifies one sample at a time. The model shows several possible features, the first two are statistical features and the second two a portion of the sampled time...
series, though every ANN used in this research contains more than four input nodes. The transformation from the input layer to the hidden layer takes place with the equation:

$$h_k = \varphi_1 \left( \sum_{j=1}^{m} w_{kj} x_j + b_k \right)$$  \hspace{1cm} (3.18)$$

where $h_k$ is the value of the hidden layer node, $k$ corresponds to the individual node of the hidden layer, $j$ corresponds to the index of the input, $m$ is the total number of inputs, $b_k$ is a bias element, and $\varphi_1$ is the activation function. The activation function is used to limit the amplitude of the output of a neuron, usually on the interval $[0,1]$ or $[-1, 1]$, and the bias is used to

![Artificial Neural Network Architecture](image-url)

Figure 3.23: Artificial Neural Network Architecture
modify the magnitude of the net input to the activation function [30]. Equation 3.18 is modeled graphically in Fig. 3.24. Each input is multiplied by a unique weight, after which all are summed. The bias term is added to the sum, which then passes through the activation function, yielding the value at a single node.

The size of the output layer is defined by the desired number of outputs, which in the case of the simulation is the five potential classes: healthy, bearing race spall, bearing ball spall, high bearing wear, and ball screw wear. The transformation from the hidden layer to the output layer happens in a very similar fashion to the transformation from the input to the hidden layer:

\[
y_v = \varphi_2 \left( \sum_{u=1}^{n} w_{vu}x_u + b_v \right)
\]

(3.19)

where \(y_v\) is the value of the output layer node, \(v\) corresponds to the output index, \(u\) corresponds to the hidden layer node, \(n\) is the total number of hidden layers, and \(\varphi_2\) is an activation function which may or may not be

![Figure 3.24: Model of One Node in an Artificial Neural Network](image)
the same as $\varphi_1$. In this case, a hyperbolic tangent is used as the activation function for both the hidden and output layers. This classifier outputs a value of one for the assigned class and zero for all other classes, as seen in Fig. 3.23.

For this work, the number of hidden layers and hidden nodes was chosen by training networks using 1-4 layers and 1-64 nodes, applying each combination several times, as it is common for ANNs to optimize into local minima. While many combinations displayed very good results on the simulated data, one layer with 32 nodes was the simplest network that consistently provided excellent classification.

Data is split into training, validation, and test sections. A training data set with known class is used to initially develop the network. The ANN initially chooses arbitrary weights, which produce the first outputs and are updated using backpropagation. The Levenberg-Marquardt backpropagation method is used to minimize the mean squared error between the classifier output and known class of the data set. The weight and bias values are updated based on the error to increase the accuracy of the network. As weight and bias values approach their optimal values for the training set, the training error continues to decrease. However, if left unchecked, this leads to an overspecialization to the exact samples in the training data set and poor performance outside of that specific data set. In order to combat this, the network uses a validation data set with known class to check for overspecialization. When the error for the validation data set begins to increase, the network knows that it has reached an error minimum and ceases training. Once the network is finalized and its error is within design specifications, it can be used on data with unknown class. In the case of this work, it is tested on data with known class that has not been used in training the
Based on the feature selection requirements, the first feature set for this testing was designed using RMS, kurtosis, and standard deviation applied individually to low frequency and high frequency data from the CWT, as well as a 20 element time vector sampled from the low frequency data in the CWT for a total of 26 points. The second feature set included everything from the first as well as a 100 element time vector taken from the high frequency data for a total of 126 points. The third set used just the two statistical sets from the first and second set for a total of 6 features, and the final feature set used applied the three statistical features to each of the 12 scales between 4 and 48 produced by the CWT for a total of 36 features. Using 5,000 total samples (1,000 of each condition) with one hidden layer and 20 nodes, chosen after each feature set had been tested with a variety of ANNs, average global-minimum error in the testing data for each feature set was 0.5% with 26 features, 1.25% with 126 features, 1.25% with 6 features, and 2% with 36 features. The results indicated that the high frequency time waveform in the 126 feature set contained too many data points, the statistical features alone in the six feature set were not enough information, and the 36 feature statistical set contained too much similar information that did not effectively differentiate between classes. The 26 feature set was able to provide differentiating characteristics while maintaining a practical number of features.

With the features and dimensionality identified, the number of samples was increased to 15,000 (3,000 of each class) for more accurate classification and the ANN was compared to discriminant classifiers. The network was trained using 70% of the data set for training, 15% for validation, and
15% for testing. The discriminant classifiers chosen were a linear discriminant, which fit a multivariate normal density to each class and used a pooled covariance matrix, and a quadratic discriminant, which was similar to linear but used different covariance matrices for each class. The discriminant classifier was first chosen for comparison because it is a much simpler classifier than the ANN, which would assess whether or not the complexity of the ANN is necessary for the classification. As expected, both discriminant classifiers trained much more quickly than the ANN, requiring just over 0.1 seconds each compared to the ANN’s 30-50 seconds, though all classifiers executed very quickly, each taking less than 0.1 seconds. The discriminant classifiers were also more consistent than the ANN, but ultimately, the ANN was capable of better results. The linear discriminant classified 2% of data incorrectly, the quadratic classified 0.4% incorrectly, and the ANN could classify the data with just 0.3% error. The optimal network yielded 0.10% error for the training data, 0.09% error in the training data, and 0.3% error in the test data, and the confusion matrix for the test data can be seen in Fig. 3.25. Classes 1, 2, 3, 4, and 5 are healthy, bearing race spall, bearing ball spall, high bearing wear, and ball screw wear, respectively.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Characteristics</th>
<th>Features</th>
<th>Error [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>1 Layer 20 Nodes</td>
<td>6</td>
<td>1.25</td>
</tr>
<tr>
<td>ANN</td>
<td>1 Layer 20 Nodes</td>
<td>26</td>
<td>0.5</td>
</tr>
<tr>
<td>ANN</td>
<td>1 Layer 20 Nodes</td>
<td>36</td>
<td>2.0</td>
</tr>
<tr>
<td>ANN</td>
<td>1 Layer 20 Nodes</td>
<td>126</td>
<td>1.25</td>
</tr>
<tr>
<td>Discriminant</td>
<td>Linear</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>Discriminant</td>
<td>Quadratic</td>
<td>26</td>
<td>0.4</td>
</tr>
<tr>
<td>ANN</td>
<td>1 Layer 32 Nodes</td>
<td>26</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 3.2: Classification Methods on Simulated Data
Figure 3.25: Confusion Matrix From Simulation Test Data
Chapter 4

Data Collection

4.1 Setup and Healthy Testing

The Servomotor interfaced with an IIS internal Emerald Development Environment (EDE) software which used an unfamiliar though intuitive language. The program allowed for the very straightforward development of sinusoidal and triangular velocity profiles, which were controlled precisely with the motor’s built-in 10,000 count encoder. Initial tests were performed driving three inches of travel in each direction at 360 and 720 rpm, corresponding to 1.2 and 2.4 inches per second at the ball nut. This gave four total tests per condition that yielded quality data and operated well within motor and system specifications. Each speed had advantages, and the range allowed for fault identification optimization. The rapidly increasing load from the 720 rpm profile had potential to excite less obvious faults, though this made testing of more mechanically obvious faults appear far less controlled, as peaks from passes over the fault did not have time to damp out before the balls passed over the fault a second time. The lower speed profiles were better at showing faults slowly and in a controlled way, which was beneficial for this testing effort. As testing progressed, more load and lower speed were desired to investigate bearing faults. Travel distance was increased to four inches which increased the load on the bearing by 33%,
while still remaining in the limits of the EMA system. A set of 240 rpm profiles was added which moved the ball nut at 0.8 inches per second.

Work on profile development was accompanied by data output investigation. Accelerometer data continued to pass easily into the dSpace data collection system through three Bayonet Neill-Concelman (BNC) cables for X, Y, and Z directions. The X direction points horizontally, perpendicular to the ball screw. The Y direction points parallel to the ball screw, and the Z direction points vertically, perpendicular to the ball screw. While not a simple analog output, the Emerald system is able to interface with the dSpace data collection unit through an RS-232 digital ASCII output, which is converted to decimal using Simulink. This enabled the collection of one variable from the Emerald system. Position was chosen as it provided the most relevant information regarding system output and had the most promise to aid in diagnostic algorithm development.

After the system profiles had been established, a log book was created to record all testing parameters, particularly any nuances between tests. As testing progressed and characteristic system changes were made, new parameters were added. This facilitates consistent and accurate records of all testing performed for both current and future analysis efforts. Columns are provided for date, time, test name, bearing number, healthy or faulted condition, oil or no oil on the bearing, spring constant, distance travelled, profile, sample rate, which channels were filtered, and a notes section for other details.

Data collection on the finalized system began with a healthy bearing in order to establish baseline system behavior. The bearing consists of two 1/8 inch thick hardened and ground steel washers and ten 1/8 inch diameter hardened steel balls retained in a nylon cage. Figure 4.1 shows the healthy
bearing used on the system.

Figure 4.2 shows the 60 second test as time data and Fig. 4.3 shows a quarter second continuous wavelet transform of each profile. These indicate that all signals have strong characteristics between scales 6 and 12, and that the lower speed profiles have higher magnitude vibration between scales 12 and 24 relative to their higher frequency content. The initial testing was completed without oil, using the lighter springs, to a distance of three inches, with vibration sampled at 10 kHz and the accelerometer’s X and Y channels passing through a filter previously built in the lab. The filter is an in-line analog filter used to remove high frequency components of a signal before they reached the dSpace data collection unit. It is a three pole passive filter, with poles at 5,710 Hz, 44,833 Hz, and 93,618 Hz. This reduces any components above 5,700 Hz and eliminates frequencies above 100 kHz. Oil was not used initially due to inexperience with bearing systems, though this did not cause problems with initial testing. All profiles ran without error and showed that the system behaved as expected, which allowed testing to move forward into faulted bearings.
Figure 4.2: Healthy Vibration Signal For All Profiles
Figure 4.3: Continuous Wavelet Transforms For All Healthy Profiles
4.2 First Bearing Fault - Mild Single Point Defect

The second bearing was initially tested just as the first bearing. Two profile sets were run with the second bearing in a healthy condition in order to establish repeatability between different bearings and yielded very similar results to those of the first bearing. The second bearing was then faulted using a Clark CR-3e to perform a Rockwell C Hardness (HRC) test, the indentation from which can be seen in Fig. 4.4. This method of faulting the bearing gave precise, repeatable results. The HRC test uses a 10 kg load to seat the indenter, followed by a 150 kg major load that drives the indenter to final depth. The formula to calculate the final value of the HRC test is:

\[
R_C = 100 - \frac{d}{2\mu m}
\]  
(4.1)

where \(d\) is the depth of the indentation in \(\mu m\). With a final value of 53 HRC provided by the machine, the final depth was 23.5\(\mu m\). The diameter was also calculated using the 120° angle of the cone shaped indenter and the properties of right triangles, yielding 81\(\mu m\).

Figure 4.4: Bearing Faulted with HRC Test
Results from the HRC fault did not follow expectation. Raw data appeared identical to healthy from the X and Y accelerometer channels, while the data from the Z channel displayed periodic peaks at very high magnitude. Figure 4.5 includes data from each channel using half of one cycle of the 360 rpm triangle profile. The pulses in the Z channel appeared every
0.033 seconds, corresponding precisely to the expected BPF of 30 Hz, as seen in Fig. 4.6. This confirmed the expected nature of the fault, as well as the ability of the system to detect it, however, it was surprising that channels X and Y did not see the peaks when they were occurring at such a magnitude.

As mentioned before, the Z direction was the only channel that did not pass through the passive filter and had no protection against aliasing. The dSpace data collection unit sampled at 10 kHz, and the triaxial accelerometer was rated to 4500 Hz. Both limited the accurate frequency range to approximately 5,000 Hz, insuring that any content over that limit would be less reliable. A closer look at the signal indicated that frequencies within each pulse exceeded the 5 kHz capabilities of the system, as seen in Fig. 4.7. The plot shows very rapid vibration with several points very inconsistent with the overall behavior of the pulse, suggesting that signal is not accurately captured.
Figure 4.7: One Pulse From Z Channel Accelerometer of HRC Faulted Bearing at 10 kHz

Figure 4.8: One Pulse From Z Channel Accelerometer of HRC Faulted Bearing at 100 kHz
Frequency analysis on the signal was performed by using an Agilent Technologies oscilloscope rated up to 200 MHz, followed by testing with a National Instruments USB-4431 vibration data acquisition (DAQ) module which can record at 100 kHz, both of which indicated content far above the limits of the acquisition tools. A pulse signal comparison from the USB-4431 can be seen in Fig. 4.8. Results from the frequency analysis of 100 kHz collection using the USB-4431 can be seen in Fig. 4.9. This shows that much of the information contained in the signal exists above 20 kHz, with the highest sustained content an order of magnitude higher than the accurate range of the accelerometer at slightly below 50 kHz, rendering the data unusable. The resonant frequency of the accelerometer is above 25 kHz, which may be the reason that the data was amplified. If the bearing experienced very high frequency vibration upon contacting the HRC indent, even a small magnitude may have excited the resonant frequency of the accelerometer.

A discussion with faculty confirmed that friction characteristics of un lubricated bearings frequently cause high frequency vibration, which led to the oiling of all bearings in an effort to reduce friction and damp high frequency vibrations. Adding oil to the bearings significantly reduced high frequency content, as seen in Fig. 4.10. It was clear that considerable content remained at a 32 kHz frequency. This content was due to motor grounding issues, though it was weak enough in the accelerometer data that the passive filter eliminated its effects in the dSpace 10 kHz sampled data.

Without the oil or anti-aliasing techniques the fault was extremely apparent even in just the raw data, however, once the data was processed both of these methods the fault failed to manifest itself in either the raw data or post processing efforts using wavelets. A CWT from the oiled second bearing using data sampled at 10 kHz can be seen in Fig. 4.11. The fault was
Figure 4.9: Frequency Spectrum Analysis on 100 kHz Dry HRC Fault Bearing Signal

Figure 4.10: Frequency Spectrum Analysis on 100 kHz Oiled HRC Fault Bearing Signal
Figure 4.11: CWT of Oiled HRC Fault Bearing

expected to appear at a rate of 30 Hz corresponding to the BPF through periodic higher magnitude pulses on the scalogram similar to what was seen in the simulated data, but no consistently repeating peaks are visible in the CWT.

4.3 Second Bearing Fault - Moderate and Severe Single Point Defects

Once it had been established that the Rockwell hardness test was not capable of producing a fault accurately observable with the instruments available, a more aggressive approach was pursued. A Dremel 4000 rotary tool with a carbide wheel was used to damage another bearing. The fault stretched across the radius of the bearing washer, approximately 0.05 inches wide and
0.002 inches deep with a very rough surface, and can be seen in Fig. 4.12a.

The bearing was oiled, then tested at a 100 kHz sample rate to investigate high frequency content. The data contained very low magnitude pulses corresponding to the BPF, though the frequency of the pulses remained far above the accurate range of the accelerometer. When tested with a 10 kHz sampling rate and a low pass filter, the pulses were again eliminated. Fig. 4.13 shows the wavelet analysis of the Dremel fault signal. Content exists across the frequency spectrum of the analysis, but even a detailed visual analysis yields no consistent, repeating pulses.

Under the assumption that the damage was not significant enough, the Dremel was used again to make the fault deeper and wider, as seen in Fig. 4.12b. The fault was then approximately 0.07 inches wide and 0.005 inches deep. The 6.9 pound per inch springs were replaced with 8.4 pound per inch springs expecting that more load would increase the fault signature and expose it in lower frequencies. Despite several changes to facilitate fault identification, the results differed very little from from first tests which used the bearing faulted with a Dremel. Therefore, it is probable that the bearings are largely uninfluenced by single-point loss-of-material type defects.
of any depth. The 1/8 inch thick hardened steel washers do not flex when only one of the ten bearings is not supported, which allows the balls to roll unhindered over the defect with very minimal system disruption. With this discovery, a third method of faulting the bearings was developed.

### 4.4 Third Bearing Fault

When a bearing is left under load in a stationary system, the balls can collectively indent the races. This can happen in a system with a high preload that is unused for a period of time, or in a system which is left in a highly loaded state when not in operation. The collective indentation creates small pits spaced exactly at each ball distance on each of the races. In the case of a thrust bearing with both races of equal diameter, the bearing is likely
to maintain this alignment. Then, the bearing compresses according to the depth of the pits each time the system rotates into a position where each ball experiences a pit in both races. As can happen with a spall on a thrust bearing, this causes vibration peaks corresponding to the BPF.

A 45-ton shop press was used to compress a bearing and create indentations around each washer. The first bearing faulted with this technique experienced a six ton load, significantly greater than the 250 pound thrust capacity at 50 rpm. The indentations have a depth of approximately 0.004 inches on each race, and can be seen in Fig. 4.14.

The effects of the new fault were immediately apparent. All three attempted runs were forced to end prematurely due severe malfunction. The balls began and stayed aligned with the indents in the washers during the first run, locking and forcing sliding contact between the washers and the system. In the second attempt the balls were purposely misaligned with the indents, however, during the profile both washers realigned and locked causing the same condition. Finally, one washer was flipped so that the balls interfaced with one indented side and one flat side to reduce the severity of fault. This also locked and caused the washers to slide against the system. This fault very clearly caused malfunctions in the system, to a degree that it was unusable. It did, however, display the potential of this type of fault in the bench top EMA system.

Figure 4.14: Press Fault Using Six Ton Load
A 2.5 ton load was used to fault another bearing in order to determine if lighter indents would display the fault in the data without locking the bearing. The indents were approximately 0.001 inches deep on each race, and can be seen in Fig. 4.15. Upon testing, this bearing yielded much better results. Indents were aligned with the balls to ensure the balls contacted indents on both sides at the same time. The results were apparent in the raw data, and the wavelet analysis showed the ability to highlight peaks that were not as clear to the naked eye. At 720 rpm the system experienced very high vibration, eliminating the need for post-processing to identify the fault. At 360 and 240 rpm, the fault was less obvious and warranted wavelet analysis to extract characteristic fault information.

Raw vibration data from triangular profile tests at 240 rpm can be seen in Fig. 4.16, and results of the wavelet analysis can be seen in Fig. 4.17 compared to a portion of the raw signal. At 240 rpm the BPF is 20 Hz, which shows very clearly in the wavelet analysis every 0.05 seconds over the course of the 0.25 second signal between scales 30 and 60, even when it is not as obvious in the raw data. These results proved the system capable of displaying a fault in a usable range, and provided an opportunity to showcase the power of wavelet analysis.
A third bearing was faulted using the shop press with a 1.25 ton load in an effort to show a more subtle fault signal and test the limits of the wavelet analysis. The indents in this case had a depth of approximately 0.0005 inches on each race, and can be seen in Fig. 4.18. This fault was more hidden in the raw data, which highlighted to an even greater extent the value of the wavelet analysis. The vibration signal from a four inch 240 rpm triangular profile is shown in Fig. 4.19. The overall magnitude of the peaks is significantly reduced compared to the 2.5 ton fault, though the wavelet analysis clearly extracts the periodic fault information every 0.05 seconds between scales 30 and 60 in Fig. 4.20.
Figure 4.17: Wavelet Analysis of Third Fault Using 2.5 Ton Load
Figure 4.18: Press Fault Using 1.25 Ton Load

Figure 4.19: Z Channel Vibration Data from Third Fault Using 1.25 Ton Load
Figure 4.20: Wavelet Analysis of Third Fault Using 1.25 Ton Load
5.1 Fault Detection

The data from the 1.25 ton press faulted bearing matches the desired characteristics for a fault signal, deviating only slightly in magnitude from the healthy signal while remaining within the frequency range detectable to the accelerometer. A direct comparison of healthy and faulted data is shown in Fig. 5.1 with the corresponding wavelet analysis in Fig. 5.2. In order to achieve a high number of samples compared to the number of features used in classification for more accurate analysis, 220 cycles are run for both healthy and faulted data. Using 0.25 seconds of raw data per sample and vibration from approximately 0.8 to 3.2 inches in each profile corresponding to three seconds of data, 2,640 samples are collected for each condition.

The ANN designed using simulation data is then applied to the experimental data using the 26 feature vector and the Gaussian 7 wavelet. While the frequency information of the simulated data was captured in its entirety by scale 48, experimental data shows lower frequency content, therefore, the classification algorithm incorporates up to scale 96. Given that the experimental and simulation are not identical, the ANN is optimized using a brute force algorithm which tests the data over a variety of layer and node combinations. Using an optimal network composed of one layer with four
nodes, the experimental data gave 95% correct classification.

A variety of other classification techniques were also investigated. First, the very simple k-nearest neighbor technique was used. In order to classify, this technique compares a test data sample to the training samples of known class. The user specifies $N$ nearest neighbor points to use for comparison, and the most common class among the $N$ samples reported is assigned as the class of the test sample.

Principal Component Analysis (PCA) was used as a feature reduction technique applied to the inputs of the ANN. Principle component analysis transforms a set of data into a new coordinate system such that the variance of the feature on the first axis is the highest, followed by the second, third, and so on. In an ideal case, this spreads the data effectively and allows the user to correctly classify the data more accurately with fewer features.

Figure 5.1: Healthy vs 1.25 Ton Press Fault Z Channel Vibration
Figure 5.2: Healthy vs 1.25 Ton Press Fault Z Channel Vibration Wavelet Analysis
This transformation can be performed by first performing a singular value decomposition on the data matrix \( X \) which produces matrices \( U \), \( \Sigma \), and \( V \),

\[
X = U\Sigma V^T
\]  

(5.1)

where \( U \) and \( V \) are unitary matrices, and \( \Sigma \) represents the singular values of the matrix. The principal component matrix \( T \) is then found by,

\[
T = U\Sigma
\]  

(5.2)

which has the same number of features and samples as the data matrix with the features ordered from greatest variance to least variance. The \( T \) matrix can then be truncated to reduce the number of features of the data set and input to the desired classifier.

A support vector machine (SVM) was also investigated, which is a discriminant function that places a linear boundary between two classes. The shortest distance from the boundary to each class is the margin, and the points on the margin are known as the support vectors. The boundary is oriented such that it maximizes the distance between itself and the support vectors. It is also possible for the machine to transform the data into a higher dimensional space, which allows the classifier to linearly separate nonlinear data, giving it an advantage against more traditional discriminant functions.

Each of these techniques are compared to the initial ANN classifier with 26 features in Table 5.1. The only technique which improves upon the original method is using PCA in conjunction with the ANN. Unfortunately, it requires an extra step to transform the data, which is not justified for such a small reduction in error.

The 5.0% error produced by the ANN with 26 features is much higher
than the error in the simulated data, particularly when considering that the simulation encountered 5 separate classes. There are several factors that contribute to this. First, the experimental data varies in magnitude over the length of the profile due to the change in load. Each profile produced 12 samples, causing the samples on each end to have very different magnitude characteristics. Although the faulted signal has more magnitude in general, a sample from the lightly loaded side of the profile has a much smaller influence from the fault due to the non-stationary load, and a relatively low magnitude overall, making it very difficult to distinguish from low and moderately loaded healthy samples. The experimental data also included noise and other system frequencies that were not captured in the simulation due to the relative simplicity of the simulation. This noise further complicated the classification effort and increased the error.

### 5.2 Remaining Useful Life

Once the classifier was proven in a two-class scenario RUL testing commenced, which investigated the ability of the classifier to determine the progression of a particular fault. Remaining useful life testing hopes to identify

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Characteristics</th>
<th>Features</th>
<th>Error [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>1 Layer 4 Nodes</td>
<td>26</td>
<td>5.0</td>
</tr>
<tr>
<td>ANN with PCA</td>
<td>1 Layer 3 Nodes</td>
<td>10</td>
<td>4.7</td>
</tr>
<tr>
<td>K-Nearest Neighbor</td>
<td>3 Neighbors</td>
<td>26</td>
<td>11.7</td>
</tr>
<tr>
<td>Discriminant</td>
<td>Linear</td>
<td>26</td>
<td>7.7</td>
</tr>
<tr>
<td>Discriminant</td>
<td>Quadratic</td>
<td>26</td>
<td>6.1</td>
</tr>
<tr>
<td>Discriminant</td>
<td>Support Vector Machine</td>
<td>26</td>
<td>5.6</td>
</tr>
</tbody>
</table>

Table 5.1: Classification Methods on Experimental Data
component condition along a trajectory of fault severity, which can be used to estimate the remaining functional life of the component. This can give an airplane an approximate time before which it needs to land, or in less safety critical instances, can help provide an optimal service window. In this testing, the fault caused by excessive stationary load was studied by using the bearing faulted with 2.5 tons in addition to the bearing which was faulted using 1.25 tons. Both bearings were also compared to a healthy specimen. Another 2,640 samples were gathered from the heavily faulted bearing to give a final total of 7,920 samples. The wavelet analysis plots of each can be seen in Fig. 5.3.

Classification is performed again by checking different combinations of layers and nodes. Increasing the nodes drastically improved the classification while increasing the layers had little impact. An ANN with one hidden layer composed of twelve nodes offered a test error of 11.8%. While this is greater than the error in the two class example, given the similarity of the two faulted classes, it is acceptable. A PCA technique was also attempted in this case, and with ten dimensions, one layer and four nodes, the least error was 17.9%, which can be seen in Table 5.2. The PCA optimizes with fewer nodes required in this case, but cannot match the low error of the system with full features.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Characteristics</th>
<th>Features</th>
<th>Error [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>1 Layer 12 Nodes</td>
<td>26</td>
<td>11.8</td>
</tr>
<tr>
<td>ANN with PCA</td>
<td>1 Layer 4 Nodes</td>
<td>10</td>
<td>17.9</td>
</tr>
</tbody>
</table>

Table 5.2: Classification Methods on Experimental RUL Data
Figure 5.3: Remaining Useful Life Z Channel Vibration Wavelet Analysis
Chapter 6

Conclusions

This research set out to investigate the use of wavelet analysis on a bench top EMA system. Due to data quality issues, the original system was rebuilt, and in the end, everything other than the ball screw and ball nut were replaced. Front and rear supports connected precisely to guide rails, ensuring system squareness. The ball nut connected to a plate, which rode on ball bushings for smooth operation. The system was mounted on a mechanical breadboard to establish flatness and serve as another measure to aid squareness. The original brushless DC gear motor was replaced with a servomotor capable of the profiles desired to replicate EMA function. The original motor coupling was also replaced, using a flexible coupling capable of accepting small axial deflection while maintaining rotational rigidity.

Simulated data was developed using natural frequencies of the system, and provided excellent classification results with the continuous wavelet analysis as the primary post-processing tool. A combination of statistical features and a reduced time signal were used as the inputs to an ANN, which achieved a classification accuracy of over 99%.

Experimental data collection proved very challenging. After attempting several spall-type faults to the system that were expected to be visible in the accelerometer data, a fault was created using a shop press which emulated the results of a static bearing under excessive preload. This final fault was
somewhat visible in the raw vibration data, but made truly obvious by the wavelet analysis. Using over 5,000 samples, the ANN classified 95% of the lightly faulted bearings correctly versus the healthy bearing. An RUL investigation was also pursued, which involved adding a more heavily faulted bearing to the classification analysis. In this case, also using the ANN classifier, 88.2% of samples were classified correctly. This allows a sample with damage of unknown severity to be classified into a level of degradation and direct an efficient maintenance plan.

There is a variety of work which can be performed on the bench top EMA system to improve upon this research. More advanced profiles can be developed and implemented using the new system. Even when using triangular velocity profiles, the non-stationary load created interesting dynamics, however, the servomotor software and hardware are capable of much more dynamic position and velocity movements. These can be used to better emulate a true EMA in flight. A more accurate profile will yield better data that is easier to translate to real-world systems.

New faults can also be injected into the system. Using a review of several failure modes, effects, and criticality analysis (FMECA) resources, Balaban et al. prepared several tables of EMA faults and failures which included probability of failure and relative criticality of the failure [6]. Ball return jam, spalling, and excessive wear of the ball nut, as well as contaminated or dry lubricant faults all rank relatively high overall in the analysis, and can be implemented in the system. The ball nut ball return jam is a potentially destructive fault, as it involves the balls becoming caught within the ball nut return channel, unable to move. When this happens, friction between the ball nut and ball screw increases drastically which places a much higher load on the motor and can lead to failure in relatively short time. The other
faults may be less likely to destroy the system, and therefore preferable for analysis in the near future.

Finally, it may also be beneficial to investigate motor current as a feature for fault identification, as it involves using sensors which are already present on the system. This reduces cost and setup required to implement a health monitoring system, helping to streamline the transition from a lab environment to an aircraft, and in [12] was shown to be effective at diagnosing single point defects.
Bibliography


Appendix A

Design Drawings

Figure A.1: Front Support Drawing
Figure A.2: Rear Support Drawing
Figure A.3: Nut Plate Drawing
Figure A.4: Ball Screw Drawing
Figure A.5: Anaheim Automation Motor Mount