Minimally Intrusive Strategies for Fault Detection and Energy Monitoring

by

Robert Williams Cox, IV

S.B., Massachusetts Institute of Technology (2001)
M.Eng., Massachusetts Institute of Technology (2002)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2006

Author .....
Department of Electrical Engineering and Computer Science
5 September 2006

Certified by .....
Steven B. Leeb
Professor of Electrical Engineering
Thesis Supervisor

Certified by .....
Leslie K. Norford
Professor of Building Technology, Department of Architecture
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

Minimally Intrusive Strategies for Fault Detection and Energy Monitoring

by

Robert Williams Cox, IV

Submitted to the Department of Electrical Engineering and Computer Science
on 5 September 2006, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

This thesis addresses the need for automated monitoring systems that rely on minimally intrusive sensor arrays. The monitoring techniques employed in this thesis require fewer sensors because they take a different approach to the measurement problem. Specifically, these techniques use the power distribution network in the target system as a power source, a sensor array, and a communications channel. In this "multi-use" approach, the only measurement sources are a set of centrally located electrical transducers (i.e. voltage and current sensors) and a set of remotely located sensors that communicate with a central processing unit via power line modems. In general, these systems determine the status of critical loads or systems using only electrical data. Thus, remotely located sensors are only employed in order to gather information that would be difficult, if not impossible, to obtain electrically. Examples of such quantities include air exchange rates and occupancy levels in individual rooms.

This thesis describes the development and application of several critical features of the minimally intrusive monitoring systems described above. First, it presents several model-based methods that make it possible to use electrical data to detect faults in certain mechanical systems. In particular, two such models are described. The first of these is intended to be applied in systems in which an electromechanical actuator cycles its operation according to the value of some other variable, such as a pressure or a temperature. Examples include compressed air and vacuum systems. The other model is used to diagnose the impending failure of the mechanical coupling through which a motor drives an inertial load such as a pump impeller. This thesis also describes the development of a minimally intrusive airflow monitoring system that uses ozone as a tracer gas. This system fits easily into the "multi-use" framework because it relies on a network of distributed ozone generators and detectors whose operation is coordinated via power line communications. Finally, this thesis also presents and demonstrates a method for detecting the operation of various electrical loads using transient changes in the measured line voltage. This technique makes it possible to use "plug-in" sensors to determine the operating schedule of each of the various loads in a home or commercial facility. All of the techniques and methods described here are demonstrated experimentally.

Thesis Supervisor: Steven B. Leeb/Leslie K. Norford
Title: Professor of Electrical Engineering/Professor of Building Technology

for Karyn

Acknowledgments

I am indebted to my advisors, Professors Steven Leeb and Les Norford. Together, they have patiently trained and supported me over the last several years. Professor Leeb has been an excellent mentor. He has shown me by example what it means to be a great engineer and a superior teacher. I have learned many things from him over the past several years, but nothing more important than the need to see a problem from the perspective of a student.

I am also thankful for the comments provided by my thesis reader, Professor Kirtley. Several other LEES faculty members, including Professors Verghese and Perreault, provided useful suggestions. Professor Steven Shaw of Montana State University also offered several helpful insights. His pioneering efforts laid the foundation for much of this work.

Special thanks are due to the crews of the *USCGC Seneca* and the Navy's Land-Based Engineering Site. I am especially grateful to LCDR Mike Obar who went above and beyond the call of duty in helping me on the *Seneca*.

I am also thankful for the persistent efforts of several Navy and Coast Guard officers enrolled in MIT's XIII-A program. Included among those gentlemen are LCDR Steve Ramsey, LT Tom DeNucci, CDR Bill Greene, LCDR Duncan McKay, LT Jip Mosman, LT Greg Mitchell, LCDR Patrick Bennett, and LT Mark Piber.

I thank my peers in LEES, especially Chris "Mutton Chops" Laughman, Jim Paris, Ariel Rodriguez, Warit Wichakool, and Joshua Phinney, for their friendship, assistance, and sense of humor.

Finally, I wish to express my love and appreciation for my wife Karyn. She had to ride every up and down of this process with me, and she did so with the utmost patience and support. I will have to spend a lifetime returning the favor.

Essential support for this research was provided by the U.S. Navy, the Cambridge-MIT Institute, the Grainger Foundation, NEMOmetrics, and the Landsman Foundation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

FROM cell phone users to new car buyers and from commercial building owners to the United States Navy, consumers expect products to operate both *reliably* and *efficiently* ([1], [2], [3], [4]). Such demands are not new, but consumer expectations have risen as the technology needed to enable reliability and efficiency has become both less expensive and more mature. For instance, products as ubiquitous as automobiles and oil heaters are now often equipped with monitoring systems that can diagnose impending faults and promote energy efficient operation ([5], [6], [7]). Such technologically-enhanced systems offer clear benefits, but there are trade-offs. Specifically, sensing devices and wireless communications modules are becoming so inexpensive that it is common to find fault detection systems that rely on large networks of remote sensors. Unfortunately, however, as sensor count rises, there is a decrease in both the reliability and the maintainability of the sensing network. As an example, consider the case of the U.S. Navy's next generation destroyer, the DDG-1000. Initial estimates suggest that this vessel may possess as many as 200,000 sensing devices ([8]). For such a sensing network to operate as desired, power and communications must be provided to each sensor, and appropriate methods must be established to handle the inevitable occurrence of sensor faults.

This thesis describes monitoring devices and techniques that minimize the number of installed sensors. These methods take a different approach to matters of sensing, power, and communications. Specifically, these monitoring systems operate by forcing the power distribution network in the target system to function simultaneously as a power source, a sensor array, and a communications channel. In this "multi-use" approach, the primary sensing devices are voltage and current transducers installed at a central location in the power system. From this extremely small set of measurements, an off-the-shelf personal

computer can track both energy consumption and load status. Furthermore, any useful information that cannot be determined electrically can be gathered by sensors installed at electrical outlets and communicated to a central processing unit via power-line modems. This configuration is useful, for example, when estimating quantities that could be used in high-performance HVAC systems, such as air exchange rates and room occupancy levels.

The remaining sections of this chapter serve to introduce useful background information. Section 1.1 describes the basic functionality provided by most modern automated monitoring devices. It also discusses some potential trade-offs. Section 1.2 introduces the monitoring systems that were developed and applied during the course of this research. Section 1.3 summarizes the primary contributions of this thesis, and, finally, Section 1.4 outlines the contents of the ensuing chapters.

1.1 Background

The modern approach to achieving reliable and efficient operation in complex systems is to use automated supervision. Supervisory systems operate at a high level, monitoring both a process or plant and its associated controller. Figure 1-1 shows a block diagram of a typical implementation ([9]). As shown, supervisory monitors perform several important tasks, the first of which is known as limit checking. In that role, the monitor continuously tracks certain key quantities, such as temperatures, pressures, and speeds. If the value of any such quantity deviates from an acceptable range, then the supervisory control system will command either the controller or a human operator to take appropriate action. Furthermore, these systems can typically perform more sophisticated analyses, such as model-based fault detection and diagnosis. In this capacity, the supervisory monitor uses information extracted from the measured variables in order to compare the current state of the system to a model-based prediction. If unacceptable discrepancies are detected, the system will alert either the operator or the controller.

High level supervisory monitors such as those described above have been developed for many systems, including electric motors ([10], [11], [12]), automobiles ([13], [14], [5], [6]), pumps ([15], [9], [16]), gas turbines ([17]), HVAC plants ([18], [19], [20], [21]), and aircraft ([22], [23], [24]). The list of applications continues to grow, and there are many reasons for this growth. Most importantly, continuous monitoring often leads directly to reduced maintenance costs, longer equipment life, increased productivity, and reduced energy con-

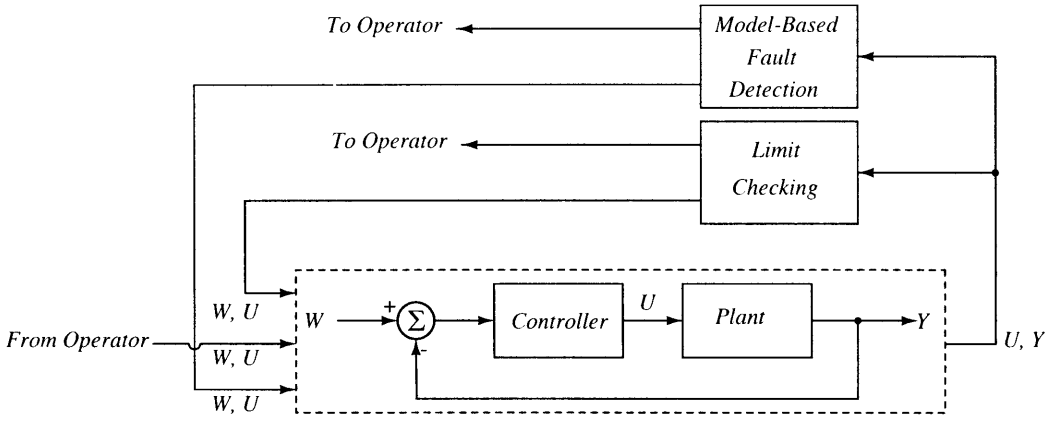


Figure 1-1: Block diagram of a typical supervisory control system. In this model, the signals labeled W are commands, the signals labeled U are inputs, and the signals labeled Y are outputs. Note that the plant does not have to be under closed-loop control.

sumption. Additionally, there are many systems in which supervisory control is an absolute necessity. Without it, for instance, aircraft with fly-by-wire systems and gas turbines would not be able to recover from potentially deadly fault conditions. Other potential applications loom on the horizon. As a result of federal fiscal constraints, for instance, future U.S. combat vessels must be able to operate with smaller crews and increased mission capability. By the Navy's own description, supervisory control will be required to meet those competing requirements ([25]).

Despite the clear benefits of automated supervisory systems, there remains the challenge of guaranteeing that the monitoring system itself will remain reliable. In fact, it can often become difficult to distinguish between sensor faults and faults in the target system. Additionally, each of the sensors in an automated monitoring system requires its own source of power and an ability to communicate with a central processing device. At a 1998 NIST-sponsored workshop, these and other issues were cited by many industry experts as some of the principal barriers to the widespread use of automated monitoring systems [26].

Many would argue that sensor-related issues should not prevent the development of monitoring devices with large sensor networks. In particular, they cite the fact that many techniques have been developed to mitigate the effects of sensor faults ([27], [19], [9], [20]). Additionally, they stress that power and communications problems can be addressed through the development of low-power devices and low-cost wireless modules. Although such features are generally beneficial, there are many instances in which they are not sufficient. For

instance, on a Navy warship, security, signal path, and power concerns make it highly unlikely that wireless modules will ever become the primary means of communication. Thus, there is, and will continue to be, a need for millions of feet of expensive and vulnerable cabling[1]. Furthermore, the maintenance of any sensor network will ultimately rest in the hands of a field operator, which means that he alone must be trained to interpret the data that is presented to him by the automated monitoring system. Accordingly, the software designed to detect sensor failures must make it abundantly clear which devices are failing and why. The development of such robust software will be a significant engineering challenge, especially in systems like HVAC that are highly nonlinear and largely affected by quantities such as outdoor temperatures and indoor occupancy. Given the combination of these issues, it is clearly worthwhile to investigate the potential of a simpler monitoring strategy.

1.2 Non-Intrusive Monitoring

The monitoring systems and techniques developed in this thesis are based on a different paradigm than those described in Section 1.1. Specifically, these systems use an approach known as non-intrusive monitoring. In this context, the term non-intrusive refers to the fact that the monitoring device is self-contained, requiring minimal physical access and no reconfiguration of the system to be investigated. Using this approach, the monitoring system can be made to operate with a significantly reduced suite of sensors.

This thesis uses the non-intrusive monitoring approach in order to take aim at several different but, nonetheless, related problems. Ultimately, two different systems are described. The first is a multi-purpose energy monitor and fault detection system. This device, which is known as a non-intrusive energy monitor (NEMO), is installed at a central location in a power distribution network. Its only connections to that network are made via a collection of voltage and current transducers and a set of power-line carrier modems. With this modest combination of sensors, the NEMO is able to track load status and energy consumption.

The other monitoring system described in this thesis uses a single sensor to monitor load activity on an electrical service. This device, which uses only voltage measurements, can be connected at a single electrical outlet. This device is convenient in situations in which it

[1] As a point of reference, consider that each of the Navy's *Arleigh Burke* class destroyers has approximately 1,342,000 feet of cable ([28]).

is undesirable or impractical to hire a trained professional to install the monitoring system. The remainder of this section provides a brief introduction to the two monitoring platforms mentioned previously.

1.2.1 The Non-Intrusive Energy Monitor (NEMO)

As shown in Fig. 1-2, the NEMO consists of two principal subsystems. The first of these is the non-intrusive load monitor (NILM), which is the component that uses electrical measurements to determine both the operating schedule and the "health" of various electrical loads. The other subsystem is a remote sensor monitor that coordinates the operation of remotely located sensing devices that measure quantities that cannot be determined electrically. Ultimately, each of these systems addresses a different set of supervisory monitoring problems. Because of the combined benefits that they could provide to customers, however, it is natural to package them into one unit. Either or both could operate autonomously.

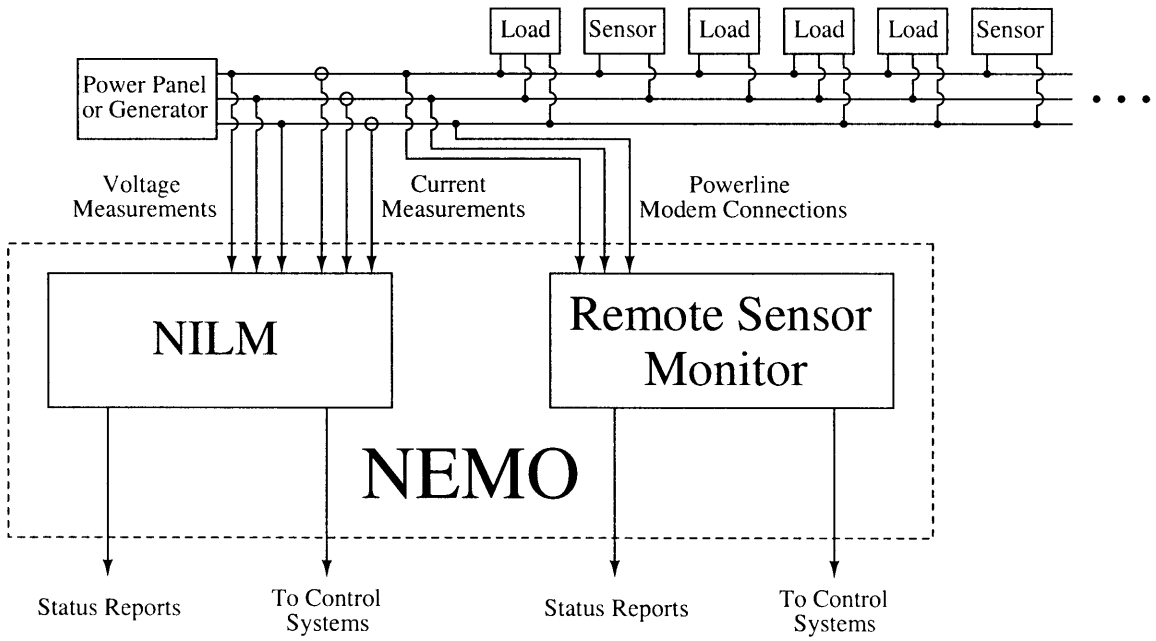


Figure 1-2: Block diagram of a field-ready version of the NEMO. As shown, the device consists of two main sub-systems. The first is the non-intrusive load monitor (NILM) which uses electrical power measurements to determine the status of each of the various loads on the electrical service. The other is a remote sensor monitor that collates and processes measurements from any non-electrical sensors.

The Non-Intrusive Load Monitor (NILM)

Of the two components that comprise the NEMO, the NILM is by far the most mature. In fact, the NILM has existed in several incarnations over the years ([29], [30], [31], [32]) and it has been applied to detect a number of critical faults ([32], [18], [33], [34], [35], [36]). Historically, work in the area of non-intrusive load monitoring can be divided into two categories. The first of these are steady-state approaches that localize and label load events according to their changes in steady-state power. In NILMs based on the other approach, events are detected by analyzing the shape and structure of any transients that occur when a load transitions between two steady states. The differences between these two approaches, as well as the details of the modern implementation of the NILM, will be left to later chapters.

As suggested by Fig. 1-2, a field-ready NILM will return two different streams of information. The first of these is a collection of status reports that are to be examined by a human operator. Such reports contain information indicating the current status of the loads under investigation. These updates can either be sent at regular intervals or immediately following the detection of a fault. In its status reports, the NILM provides both an automated assessment of the state of each target load, as well as a summary of the information used to make that decision.

The other set of outputs that a field-ready NILM will return are commands that can be sent to various controllers. This ability will be critical to the operation of the NILM in high-level supervisory control systems. Ultimately, there are a number of such systems in which the NILM would be useful. For instance, the NILM could act to de-energize any loads that may be approaching a serious failure (i.e. motors that have exceeded their rated run-time) or it could act as power system monitor in a reconfigurable power system ([37]). This thesis will examine the use of the NILM in several supervisory capacities.

The Remote Sensor Monitor

The remote sensor monitor is a new system that was first suggested in [38]. The primary goal of this component is to coordinate the operation of a network of remotely located sensors. As shown in Fig. 1-2, these sensors are scattered throughout the same power distribution network that is being monitored by the NILM. Thus, natural installation points for these devices might be at electrical outlets or light fixtures.

In many instances, a sensor array connected to a building's power distribution network is an alternative that is superior to other options. For instance, existing techniques aimed at measuring indoor air exchange rates are often based on highly expensive and elaborate tracer gas systems ([39], [40]). Consequently, the cost and intrusiveness of the required equipment has made it impossible to deploy these systems on a permanent basis. By incorporating the sensor array into the power distribution network, however, both the cost and intrusion problems can be mitigated. Specifically, if ozone is used as a tracer gas, then it can be easily generated by high voltage DC power supplies driven from the AC line, and its concentration can be sensed by low-cost chemical sensors. Furthermore, assuming that each of these devices is equipped with a power-line modem, their operation can be coordinated by the centrally-located remote sensor monitor.

1.2.2 The Voltage-Based NILM (V-NILM)

The other monitoring platform described in this thesis is an adaptation of the non-intrusive load monitor described above. In this new system, which is described as a voltage-based NILM (V-NILM), access is only provided to the line voltage. In order to track load operation from this single input quantity, th V-NILM exploits the behavior of the power system model presented in Fig. 1-3[2]. In that model, the ports labeled by the node voltages $v_k(t)$ might represent individual electrical outlets. If a load is connected at one of those ports, it will immediately begin to draw current from the source. Since that current must flow through a parasitic source impedance, the voltage measured at the other terminals will be reduced by a small amount.

The phenomenon described above is demonstrated graphically in Fig. 1-4. To generate the waveforms displayed in that Figure, a 1kW heater was connected to the line. Clearly, the resulting load current distorts the measured voltage. Moreover, the resulting change is dependent upon the current, meaning that the voltage contains a scaled and time-shifted replica of it. Because of this, the V-NILM is able to use the voltage to identify the operation of individual loads. The specifics of the identification process are subtle, however, and they will be left to later chapters.

A typical connection for the V-NILM is shown in Figure 1-5. In this case, the monitoring

[2]Several researchers have found this model to be reasonably accurate given certain simplifying assumptions. The basis for those assumptions, as well as a discussion of the validity of the model will be left to later chapters.

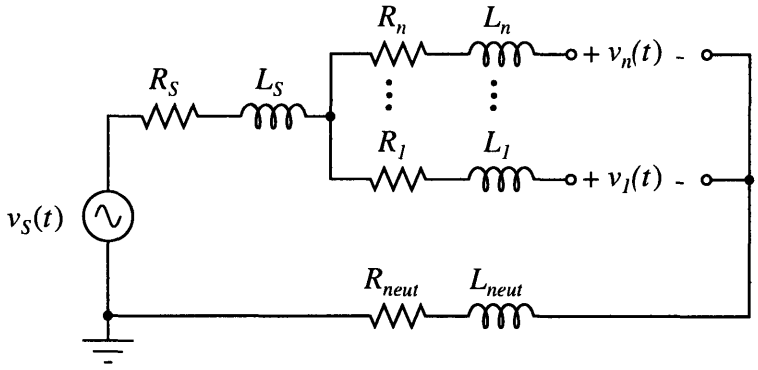


Figure 1-3: A distribution-level model of the electric utility.

device does not have to be installed at a central location; rather, it can be installed remotely, perhaps at an individual service outlet. Moreover, several such devices could be interspersed throughout the network. In fact, certain loads could each have their own built-in V-NILM. If this were to be done, then individual loads could each detect periods of peak demand and respond accordingly. Such a feature would be attractive to utility providers who are interested in leveling peak loads. Additionally, this device could be packaged into a single plug-in unit, making it an attractive tool to use to monitor power usage by residential customers.

1.3 Thesis Contributions

Throughout the course of this project, the goal has been to investigate and to develop several minimally intrusive techniques for performing high-level supervisory monitoring tasks. To that end, a number of different topics have been considered. In this thesis, focus is placed in three distinct areas. First, there is a significant discussion of several techniques for fault detection and diagnosis using electrical measurements. These techniques can be applied in the non-intrusive context in order to track the status of critical loads and systems. Subsequently, there is a discussion of the development of the remote sensor monitor described previously. Finally, there is a presentation of the architecture of the V-NILM.

In the area of fault detection and diagnosis, this thesis presents several model-based methods for tracking the performance of critical loads and systems. Specific emphasis is placed upon models that allow the NILM to monitor the mechanical systems that are serviced by electromechanical actuators. A variety of such models are possible, including

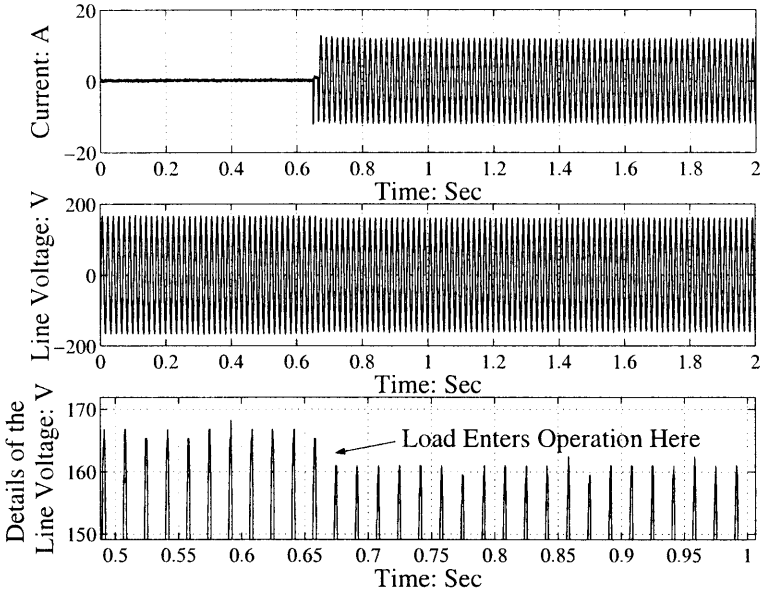


Figure 1-4: Distortion in the measured line voltage as a result of the connection of a 1kW heater. The top trace shows the load current drawn by the heater, the middle trace shows the line voltage applied to the heater, and the bottom trace shows the details near the peak of the line voltage.

those that capture the physical behavior of the system as well as those that treat the system as a "black box." In this thesis, the focus is placed on physical models, and two field-tested examples are presented. The first of these detects leaks in cycling systems, and the other predicts the imminent failure of mechanical couplings in induction machine applications. Examples are demonstrated aboard an active-duty U.S. Coast Guard Cutter.

In addition to the model-based diagnostic methods proposed in this thesis, several other supervisory control features are investigated. In particular, the NILM is shown to be aptly suited for monitoring the "health" of a complex integrated system consisting of many individual actuators. The example system used here is one of the engine rooms aboard *Arleigh Burke* class destroyer.

This document also discusses the development of the remote sensor monitor described previously. In particular, there is a focus on using that monitor to the status of indoor airflows. In this thesis, the airflow system is demonstrated in several test cases, with an eye toward application in naturally ventilated buildings. In those facilities, the monitor could serve to indicate both the speed and the direction of bulk airflow in individual rooms, and

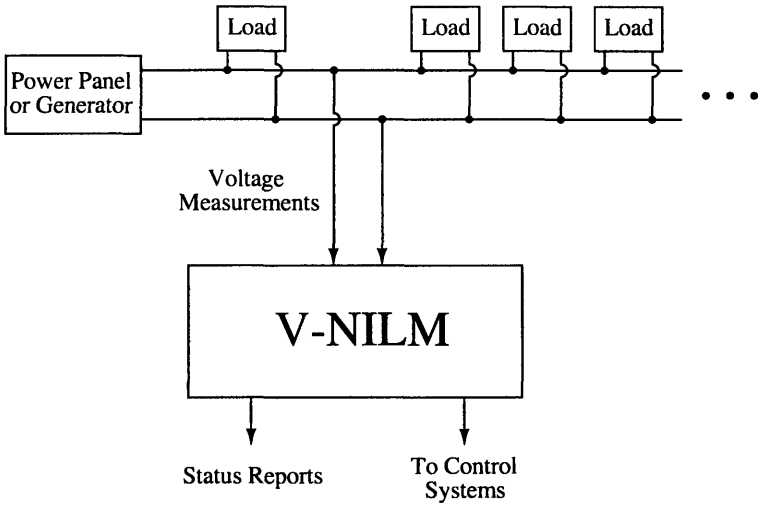


Figure 1-5: A typical connection for a V-NILM in a single-phase distribution system. Note that the device is typically installed downstream of the power panel.

such information could be used to automate the operation of dampers and fans[3]. With the introduction of automatic flow rate control, naturally ventilated buildings could become even more attractive in appropriate locations. Ultimately, this system may also be able to determine indoor air exchange rates and to assist in the validation of the flow models that have been proposed for naturally ventilated facilities [4].

Another major contribution of this thesis lies in the observation that voltage measurements provide enough information to detect the operation of individual electrical loads. The monitoring system that was developed as a result of this observation could make residential load monitoring more attractive to consumers. Additionally, the voltage-based technique proposed in this thesis has significant potential as a surveillance tool.

1.4 Thesis Organization

Chapter 2 of this thesis presents an overview of the non-intrusive load monitor and its prior applications. It begins with a review of the evolution of the NILM, and it culminates with a description of each of the critical subsystems that comprise it. Throughout, there is a

[3]An option in many candidate naturally ventilated facilities is to install fans in order to accommodate for the possibility of extremely warm days.

[4]To date, many of the attempts to validate the models for flow patterns in naturally ventilated buildings have been based on CFD simulations.

focus on those aspects of the NILM that make it particularly well suited to perform fault detection and other high-level supervisory functions.

Chapter 3 presents three example applications of the non-intrusive load monitor described in Chapter 2. The first of these demonstrates the NILM's ability to monitor the health of an integrated system consisting of many individual actuators. The latter two cases present model-based diagnostic methods for cycling systems and mechanical couplings in induction machine applications. The methods and models described in this chapter are all demonstrated in representative field systems.

Chapter 4 describes both the basic framework of the remote sensor monitor, as well as an example application. In this case, the application of interest is an airflow monitoring system that uses ozone as a tracer gas. Part of Chapter 4 is devoted to a description of each of the components of the proposed airflow monitor. Also included are several experimental results that demonstrate the potential capabilities of the proposed airflow monitoring strategy.

Chapter 5 provides a description of the proposed V-NILM system. This chapter begins with an introduction to the fundamental concept upon which this device is based, and it culminates with the presentation of several field examples that demonstrate its potential capabilities. The majority of this chapter is dedicated to a discussion of the spectral envelope preprocessor that is used to extract the required information from the measured line voltage. Finally, Chapter 6 concludes this document with a summary of results and a look at several future applications of each of the three systems that are examined in this thesis.

Chapter 2

Non-Intrusive Load Monitoring

ADVANCED energy monitors that can identify the operation of individual loads can also evaluate the status or "health" of those loads. This ability is extremely useful when monitoring a system that requires many actuators to function in an appropriate sequence. When supplied with appropriate analytical tools, energy monitors can also use electrical data to detect impending or existing faults in individual actuators and in the mechanical systems that they service. Additionally, advanced monitors can detect some of the power quality-related problems that often degrade the performance of large systems.

The purpose of this and the subsequent chapter is to describe several methods by which the non-intrusive load monitor can perform some of the high-level supervisory functions described above. This chapter essentially focuses on the basic framework provided by the modern NILM. In order to place this thesis in the proper historical perspective, this chapter begins with a brief review of NILM devices that have been developed in the past. Subsequently, it describes the basic functionality provided by each of the main components of the modern NILM, with particular emphasis placed upon load identification procedures and non-intrusive diagnostic methods.

2.1 Background

In recent years, numerous studies have been published that indicate that the NILM can successfully identify load operation in many systems ([41], [32], [42], [28], [43], [44]). As indicated in Chapter 1, these schemes tend to fall into one of two broad categories – those that identify loads based on steady-state power consumption and those that perform identification using transient characteristics. This section serves to demonstrate key differences between the two procedures.

2.1.1 Steady-state load identification

NILMs equipped with steady-state identification software classify loads using their steady-state changes in real and reactive power consumption. Typically, this type of identification proceeds in five steps ([45], [42]):

- An edge detector is used to identify changes in steady-state power levels (i.e. on/off events).
- A cluster analysis algorithm sorts on/off events according to their associated changes in real and reactive power. Subsequently, these changes are mapped into a two-dimensional complex power signature space (i.e. a $\Delta P - \Delta Q$ plane). As shown in Fig. 2-1, the signature space technique reduces the potentially complicated aggregate power data to a two-dimensional representation with a pleasing and useful graphical interpretation.
- Positive and negative clusters of similar magnitude are paired or matched (This is especially effective for "two-state" loads that turn-on, consume relatively constant power, and turn-off).
- Unmatched clusters and events are paired or associated with existing or new clusters according to a best likelihood algorithm. This step is known as anomaly resolution.
- Pairs of clusters are associated with known load power consumption levels in order to determine the operating schedule of individual loads. This step typically relies on information gathered during a training or survey phase.

The load identification procedure described above, which has the advantage of intuitive appeal, has been demonstrably successful in certain classes of buildings and, in particular, in residences ([46], [47]). In other settings, however, steady-state approaches have been found to be severely limited. For instance, in medium to large size commercial and industrial facilities with many loads and high rates of event generation, the two-dimensional signature space can become crowded with indistinguishable clusters of events. Furthermore, the nature of the loads in larger facilities is such that it is difficult to define a suitable time scale that yields reliable steady-state measurements ([48]). Given these and other shortcomings, steady-state identification methods are rarely useful in complex industrial or commercial environments.

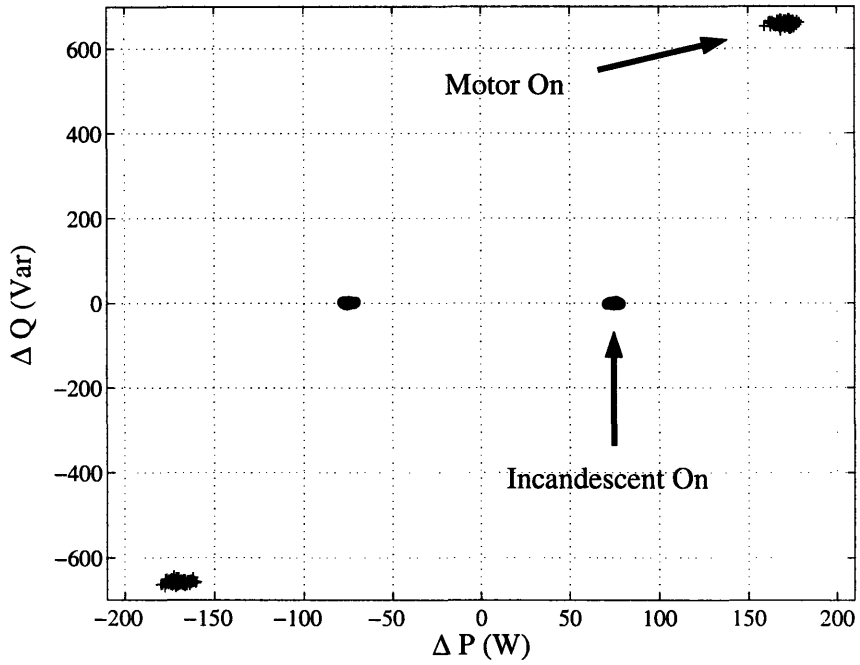


Figure 2-1: A typical signature space as described in [45] and [29]. Turn-on clusters for an incandescent lamp and a single-phase induction machine are labeled; the corresponding turn-off clusters are reflected through the origin.

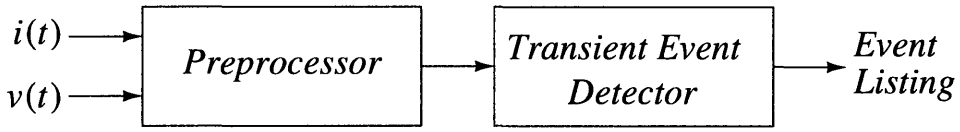


Figure 2-2: The two processing steps required in transient-based NILM devices.

2.1.2 *Transient-based load identification*

Researchers have also considered the use of transient information to distinguish loads. Such an approach, which was proposed in [30], was developed following the observation that the transient behavior of an electrical load is strongly influenced by the physical task that the load performs. As a result, different classes of loads posses unique and repeatedly observable transient profiles that can serve as "fingerprints" indicating the operation of individual loads. For example, the turn-on transients associated with an incandescent lamp and an induction motor are distinctly different, as the physical task of heating the filament of a lamp is not the same as accelerating a rotor.

In transient-based NILM devices, loads are detected using the two step procedure shown in Fig. 2-2. First, a preprocessor, which can be implemented using either analog electronics or digital software routines, computes estimates of the spectral content of a measured current waveform ([30], [49], [32], [50]). Subsequently, the spectral estimates created by the preprocessor are passed to a software module that identifies the operation of individual loads by matching stored templates to transient patterns observed in the preprocessed data stream.

The preprocessor architecture proposed in [30] performs a spectral decomposition of the current, $i(t)$, that is drawn by a collection of loads ([30], [49], [32], [50]). That procedure is based on the fact that a real waveform, $x(t)$, may be described with arbitrary accuracy at time $t \in (t, T]$ by a Fourier series with time-varying, real spectral coefficients $a_k(t)$ and $b_k(t)$:

$$x(t-T+s) = \sum_k a_k(t) \cos\left(k\frac{2\pi}{T}(t-T+s)\right) + \sum_k b_k(t) \sin\left(k\frac{2\pi}{T}(t-T+s)\right). \quad (2.1)$$

In Eq. 2.1, the variable k ranges over the set of non-negative integers, T is a real period of time, and $s \in (0, T]$ ([49], [51]). The coefficients $a_k(t)$ and $b_k(t)$ are calculated using the following formulae ([51]):

$$a_k(t) = \frac{2}{T}\int_0^T x(t-T+s)\cos\left(k\frac{2\pi}{T}(t-T+s)\right)ds \tag{2.2}$$

and

$$b_k(t) = \frac{2}{T}\int_0^T x(t-T+s)\sin\left(k\frac{2\pi}{T}(t-T+s)\right)ds \tag{2.3}$$

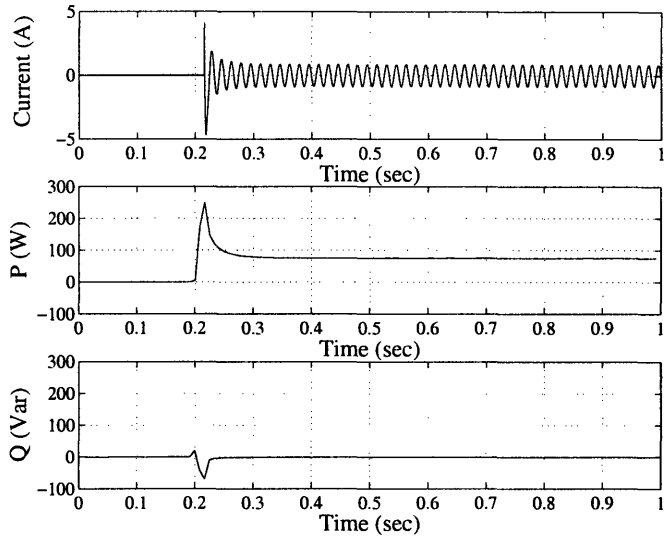
.

In transient-based NILM devices, the preprocessor first synchronizes the basis sinusoids in Eq. 2.1 with the local line voltage, and it then computes estimates of the coefficients $a_k(t)$ and $b_k(t)$ for the current $i(t)$. Historically, several different approaches have been used to estimate the integrals in Eqs. 2.2 and 2.3. In [30] and [49], the calculations were performed using dedicated hardware, and the resulting spectral coefficients were sampled by a data-acquisition card in a digital computer. Given the decreasing cost and increasing capabilities of standard off-the-shelf PCs, the preprocessor was ultimately migrated into a software package that operates on sampled versions of the voltage and current waveforms ([32]). The details of the modern implementation of this software preprocessor will be explained in the next section.

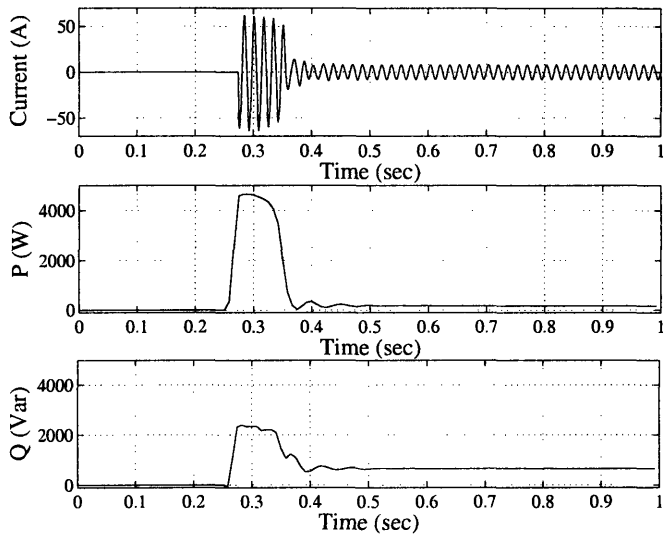
Figure 2-3 shows two different sets of spectral estimates created by the preprocessor described in [50]. In Fig. 2-3a, estimates are shown for an incandescent light bulb, and in Fig. 2-3b, estimates are shown for a single-phase induction machine. Several critical observations can be made by examining these waveforms. For instance, it is immediately obvious that the shape of the transient current and its corresponding spectral envelopes are very different for the two different types of loads. It is these differences that are exploited by the NILM's transient event classifier.

A careful examination of Fig. 2-3 reveals another, more subtle observation. In particular, note that the steady-state values of $a_1(t)$ are proportional to real power, and that the steady-state values of $b_1(t)$ are proportional to reactive power. Given the standard definitions of P and Q, this result is to be expected provided that the cosine function in Eq. 2.2 is in-phase with the observed line voltage.

Once spectral envelope estimates have been created by the preprocessor, they are fed forward to a transient event classifier. The software module used to perform this function searches the incoming data streams for any patterns that match particular templates. As originally implemented in [30], these templates are created during a training phase in which



(a)



(b)

Figure 2-3: Current and P and Q spectral envelope estimates for (a) an incandescent lamp and (b) a single-phase induction motor. In the case of the lamp, P changes rapidly as the filament begins to heat. Once in steady-state, the bulb mainly consumes real power, which is to be expected given that the filament is mainly resistive. In the case of the motor, both P and Q are initially quite large, as a considerable amount of energy is required to accelerate the rotor. Since the machine was unloaded when this data was collected, the power factor is observed to be relatively low once the machine enters steady-state.

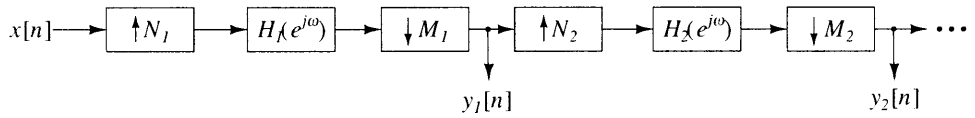


Figure 2-4: Multi-scale decomposition from [30]. Pattern matching is performed on the original $x[n]$ as well as on each of the $y_k[n]$.

transients for each load class of interest are first collected and then classified using relatively high-derivative portions, referred to as v-sections. Since a load transient might consist of several such segments, the NILM identifies events using a two-step process. First, it attempts to match templates to each individual v-section. Subsequently, it determines if a series of matches corresponds to a particular load of interest. Thus, for a NILM to successfully identify a particular transient consisting of three v-sections, the NILM must first fit a separate template to each section and then observe that the matches occurred in the appropriate order. By operating in this manner, it becomes possible for the NILM to detect events that overlap in time[1].

The identification method proposed in [30] also involves a technique whereby each incoming data stream is resampled in order to match transients with similar shapes but different time scales. This procedure is useful, for instance, if the NILM is monitoring several induction machines, each of which has a different rating. The basic data flow in the resampling module is shown in block form in Fig. 2-4. As indicated, the process produces data on several different time scales, and the NILM will search for reference patterns on each. A match will not be declared until each time scale has been examined. Effectively, this process produces a time dilation parameter for each transient pattern.

2.2 The Modern NILM

A system-level block diagram of the modern NILM is shown in Fig. 2-5. In the monitoring scenario considered in that diagram, the NILM is installed in a three-phase AC power system. As such, it samples three voltages and three currents. From those measurements, the preprocessor computes estimates of the spectral content described in Section 2.1.2. Once those estimates have been computed, they are passed to an event classifier that identifies

[1]Note, however, that only a certain amount of overlap is tractable. If, for example, a relatively small load such as lamp turns on at exactly the same instant as an induction machine, the transient behavior of the motor will effectively "swamp out" the transient associated with the lamp.

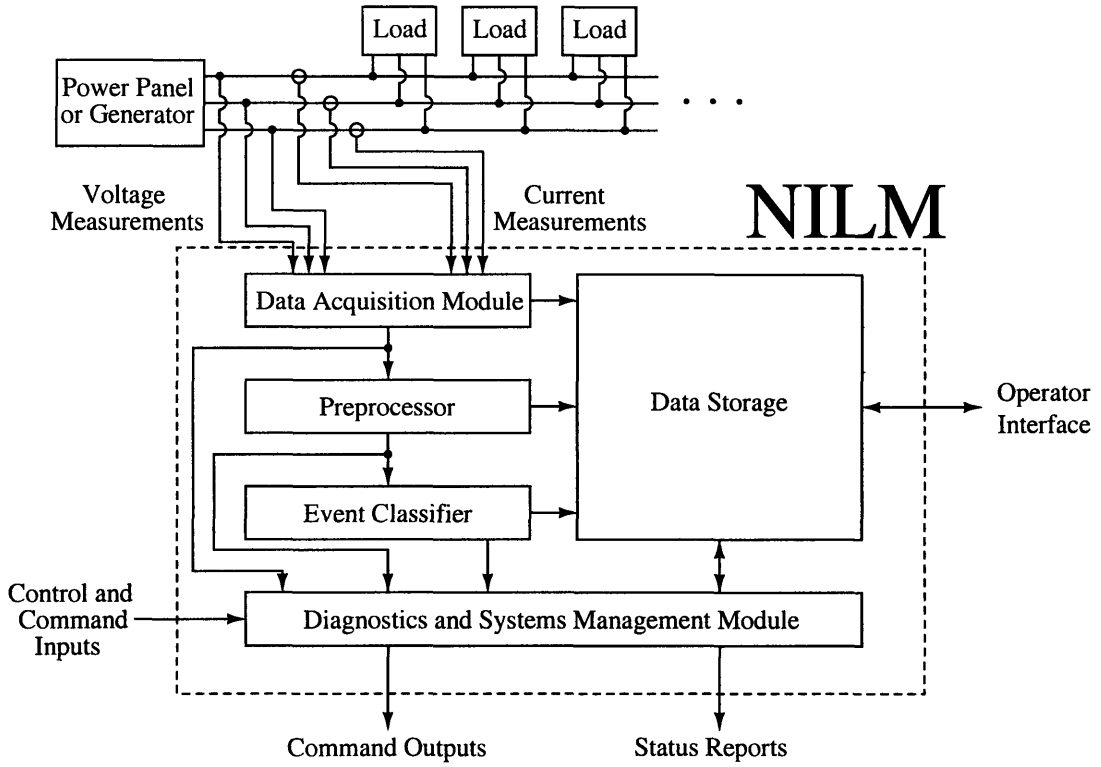


Figure 2-5: System-level block diagram of a modern NILM device.

the operation of each of the individual loads on the monitored electrical service. Each of the three relevant data sets obtained in this process, namely raw waveforms, spectral envelopes, and event listings, are each archived in either a local or remotely located database ([52]).

One of the primary contributions of this work is the development of methods to be used in the NILM's final processing stage. The function of that block, which is known as the diagnostics and systems-management module, is to detect faults in critical loads or systems. Note that as shown in Fig. 2-5, this module does not have to operate in real-time. This section describes the functionality provided by each of the major subsystems shown in Fig. 2-5.

2.2.1 The Preprocessor

Figure 2-6 presents a system-level block diagram of the preprocessor used in the modern NILM. As shown, the two inputs are discrete-time samples of the measured voltage and current waveforms, i.e.

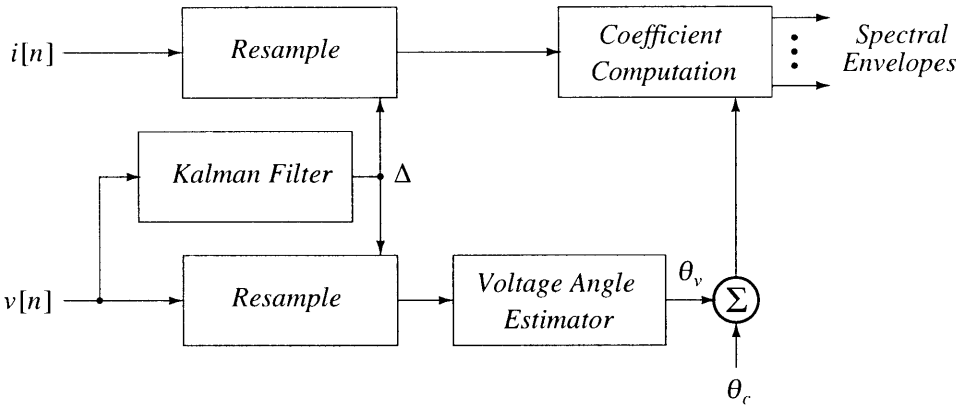


Figure 2-6: System-level block diagram of the spectral envelope preprocessor.

$$v[n] = v(nT) \tag{2.4}$$

and

$$i[n] = i(nT + T_c), \tag{2.5}$$

where T is the sampling period and T_c is a delay resulting from the use of sequential sampling. In the modern NILM, the two waveforms are each sampled at approximately 8kHz ([50]).

The purpose of the preprocessor module is to estimate the spectral envelope coefficients in Eqs. 2.2 and 2.3. Most of the processing shown in Fig. 2-6, however, is a resampling procedure that is required because the sampling clock is not synchronized to the power system frequency. In order to resample the data, a Kalman filter ([53]) estimates the fractional number of samples corresponding to a single period of the input voltage. Once that value has been computed, both data streams are resampled so that they each contain 128 points per period. Following resampling, the FFT is used to estimate the spectral envelope coefficients of the current. Since the size of each period is a power-of-two, this process can be efficiently implemented in software ([54]).

Note that when computing spectral coefficients via the FFT, one does not necessarily obtain estimates of the quantities defined in Eqs. 2.2 and 2.3. In particular, if sampling did not begin at a zero-crossing of the voltage waveform, then the Fourier coefficients will contain an implicit phase error. This error can be corrected, however, by computing the angle θ_v by which the actual analysis window is offset from the zero-crossings of the voltage

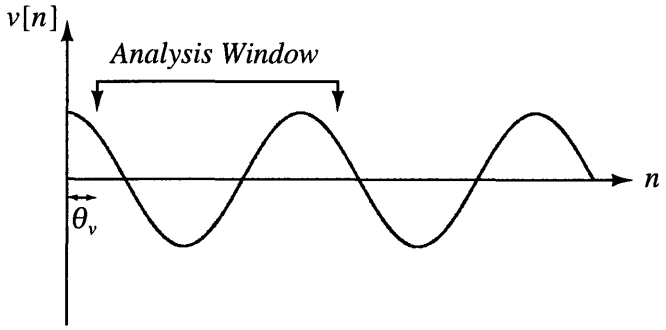


Figure 2-7: Graphical demonstration of the phase-synchronization problem. As shown, the FFT will not necessarily be computed in-phase with the actual line voltage. In this case, the analysis window for the FFT is 45° out-of-phase with the voltage, meaning that $\theta_v = 45^o$.

waveform. The meaning of this offset is demonstrated graphically in Fig. 2-7.

In addition to the phase-synchronization problem, there is also a systematic offset introduced by the use of sequential sampling. This error, which is denoted as θ_c, can be estimated and corrected. Assuming that the discrete-time estimates of the spectral envelope coefficients computed via the FFT are a_k and b_k, then the phase-corrected coefficients are

$$\tilde{a}_k + j\tilde{b}_k = e^{-jk(\theta_v+\theta_c)}\left(a_k + jb_k\right). \tag{2.6}$$

The preprocessor returns estimates of $\tilde{a}_k$ and $\tilde{b}_k$. Complete details of the implementation are presented in [50].

2.2.2 *Event Classifier*

In the modern NILM, the event classifier identifies loads using a combination of both transient and steady-state information. Because of the difficulties associated with steady-state methods, however, the NILM relies heavily on transient profiles. In general, the load-identification software used in this thesis only turns to steady-state information when the transient classifier encounters considerable ambiguity. Similar transient-favored approaches were advocated in both [41] and [42].

The event classification software onboard the NILM proceeds in two broad steps. First, preprocessed data is searched by an event detector. If any events are located, a pattern classifier attempts to match the data surrounding the event to templates known as exem-

plars. Each known transient event has its own exemplar, which consists of one or more sections that identify the given transient. For many loads, the critical sections of interest are those that display significant variation. In the terminology established in [30], these regions are known as v-sections. A positive identification is declared only when the data in each of the preprocessed streams matches the precise time pattern of the sections in the exemplar ([32]).

In this thesis, event detection is performed using a two-step process. First, the spectral envelopes are searched for transients using the change-of-mean detector proposed in [32]. In that method, events are signalled as soon as the absolute value of the difference between the input data and a low-pass filtered version of the input data exceeds a preset threshold. Although this scheme is simple and effective, it does have one shortcoming. Specifically, because the low-pass filter smooths sharp edges, the detector output remains above the threshold for a certain period of time following any sharp changes in the input data. For this reason, the detector is designed to have a "lock out" period during which any new events are ignored. Typically, this is not a problem, since the event detector output is simply used to determine the starting point for the classification process. If multiple events overlap, however, this means that some events may be missed. For example, consider Fig. 2-8, which shows two overlapping pump starts observed aboard the *USCGC Seneca*. In that case, the simple change-of-mean detector located the first pump start, but it missed the second one. In order to correct this problem, a second level of detection was added. This second detector is only invoked once the simple change-of-mean scheme has located an event, and it only searches the data in the "lock out" window. Thus, since the "lock out" window extends 50 samples beyond the location of the event, the second detector searches that 50-point window for any other events. The second detector is the FIR filter $\{1, -1\}$, which is commonly used in many multiscale edge detection applications ([55], [56]). If necessary, a more sophisticated multiscale detector, such as the one described in [55], could be used. Thus far, the procedure described above has been fount to be sufficient in many field applications.

Once a transient event has been detected, a pattern classifier attempts to identify the load by fitting exemplars to the incoming data. The procedure used to classify transient patterns is based on the methods presented in [32]. In that approach, exemplar fitting is conducted using a series of decoupled estimation problems. First, each individual exemplar section is fit to the incoming data. Subsequently, the entire exemplar is fit by determining

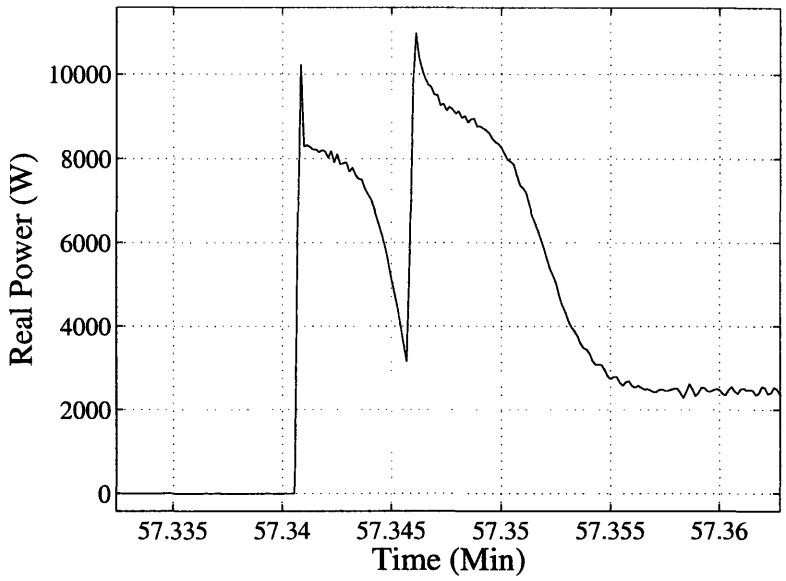


Figure 2-8: Two overlapping pump transients. Note that the simple change-of-mean detector misses the second pump start. The multi-event correction method described in the text, however, detects it successfully.

an appropriate gain constant to apply to all sections. Each of these two steps is described in detail below.

In the first step of the classification process, each exemplar section is fit to the incoming data. To understand this procedure, note that each section is characterized by both an index n_{off} that provides the offset relative to the first section of the exemplar and by a column vector $\mathbf{s}$ that defines the qualitative shape of the section. For example, the data points plotted in Fig. 2-9 are the components of the shape vector for an induction machine template. When classifying transients, it is assumed that the shape vector needs to be scaled and shifted to fit the data appropriately. Thus, if the input data is contained in a vector $\mathbf{d}$, this vector is related to $\mathbf{s}$ via a gain coefficient, α, and an offset parameter, β, i.e.

$$\begin{aligned} \mathbf{d} &= \alpha\mathbf{s} + \beta\mathbf{1} \\ &= [\,\mathbf{s} \quad \mathbf{1}\,] \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \end{aligned} \tag{2.7}$$

When fitting each individual section, it is assumed that there are three parameters to be

selected. Two of these are the parameters α and β in Eq. 2.7. The third is an offset relative to the sample where the event detector signalled an event. Thus, the transient classifier "slides" the shape vector across the incoming data stream and attempts a fit at each location. This "sliding" process begins k samples before the start of the section, and it ends k samples after the section starts. Assuming that an event is detected at sample n_e, the classification problem can be written in the form

$$\begin{aligned} & [\,\mathbf{d}\,[n_e + n_{off} - k] \quad \mathbf{d}\,[n_e + n_{off} - k + 1] \quad \cdots \quad \mathbf{d}\,[n_e + n_{off} + k]\,] \\ &= [\,\mathbf{s} \quad \mathbf{1}\,] \begin{bmatrix} \alpha_{-k} & \alpha_{-k+1} & \cdots & \alpha_k \\ \beta_{-k} & \beta_{-k+1} & \cdots & \beta_k \end{bmatrix} \\ &= [\,\mathbf{s} \quad \mathbf{1}\,]\mathbf{X}, \end{aligned} \tag{2.8}$$

,

where the index n in each of the columns $\mathbf{d}[n]$ on the left-hand side is the first sample in that column, i.e. the column $\mathbf{d}[n_e + n_{off} - k]$ begins at sample $n_e + n_{off} - k$. The estimation problem is solved by choosing the column on the left-hand side that offers the best least-squares fit to the shape vector $\mathbf{s}$. Assuming that we are fitting to the i-th section of an exemplar consisting of N sections, the column vector chosen in this process is denoted as $\mathbf{r}_i$ and the associated parameters are denoted as α_i and β_i ([32]).

The process described above is repeated for each section of a multi-section exemplar. This yields one set of vectors $\mathbf{r}_i$ and two sets of parameters α_i and β_i. As a final step, the entire exemplar is fit to the data by determining a composite gain to apply to all sections. This is done by solving the least-squares problem

$$\begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_N \end{bmatrix} \alpha_{comp} = \begin{bmatrix} \mathbf{r}_1 - \beta_1 \\ \mathbf{r}_2 - \beta_2 \\ \vdots \\ \mathbf{r}_N - \beta_N \end{bmatrix}, \tag{2.9}$$

where $\mathbf{s}_i$ is the i-th shape vector.

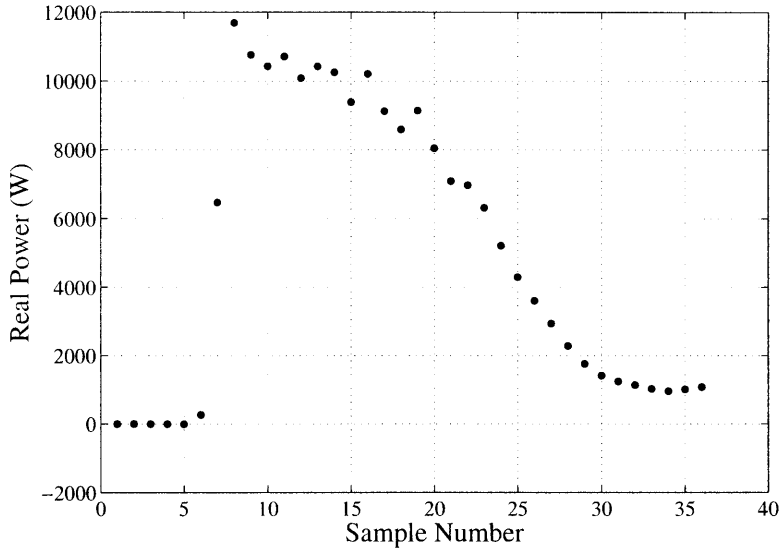


Figure 2-9: Shape vector for an induction motor exemplar. In this case, the exemplar consists of only one section taken from P_1.

2.2.3 *Diagnostics and Systems Management Module*

The diagnostics and systems management module is the component of the NILM that assesses the operational status of the target system and its component loads. This subsystem can perform a number of different functions, including usage trending, power quality analysis, system-level health monitoring, and load diagnostics. In this thesis, the focus is placed upon the latter two functions.

System-Level Health Monitoring

The term system-level health monitoring refers to the ability to assess the overall operational status of a complex integrated system. Essentially, the function of a NILM in this context is to determine whether or not each individual component in the system is operating as desired. The establishment of this capability in air and spacecraft is currently a particularly popular area of research ([57], [58], [24], [59]).

A block diagram showing the typical data flow in a system-level monitoring application is presented in Fig. 2-10. As shown, the systems of interest generally consist of a number of actuators, sensors, and controllers. A typical example system might be a gas turbine engine, which relies on the operation of a number of pumps, mechanical sensors, and solenoid-

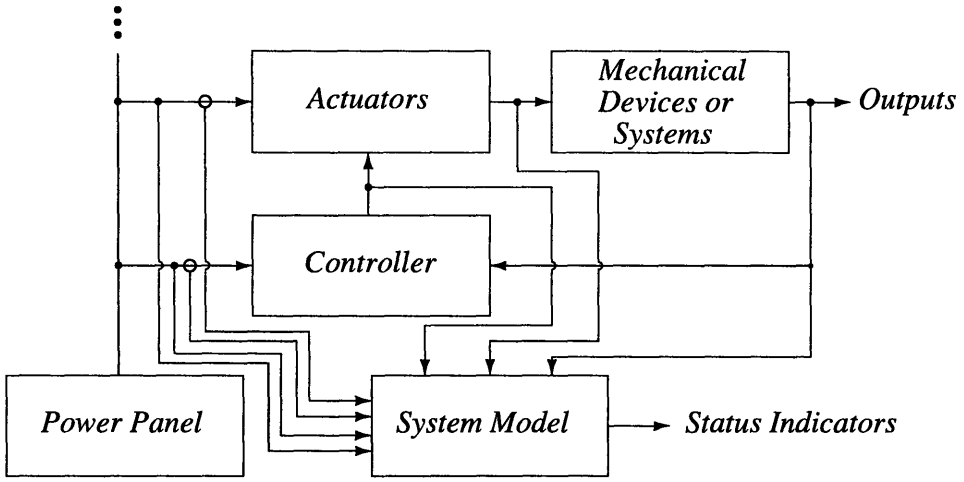


Figure 2-10: Block diagram showing the data flow in a typical system-level health monitoring application.

actuated valves. With the vast number of sensors involved in such a monitoring application, the system model must be designed to tolerate sensor faults ([58]).

When using a NILM to perform system-level health monitoring, the data flow is simplified as shown in Fig. 2-11. In that case, the only measurements are of the voltage and current at a single central location in the power distribution network. In order to assess the status of the complete system, the NILM must be able to use its sparse data set to perform the following operations:

- Identification of the responses of individual actuators to direct commands from the controller
- Identification of the responses of individual actuators to indirect commands from the controller
- Identification of proper actuator sequencing

The meanings and requirements of these three tasks are explained below.

The determination of the response of an individual actuator to a direct controller command is essentially equivalent to the identification of the operation of the actuator. For an off-line device, this simply implies that the NILM must detect and classify the appropriate "fingerprint" sections in the spectral envelopes. In many instances, however, energy is transferred to an actuator via a power converter as shown in Fig. 2-12. Depending upon the

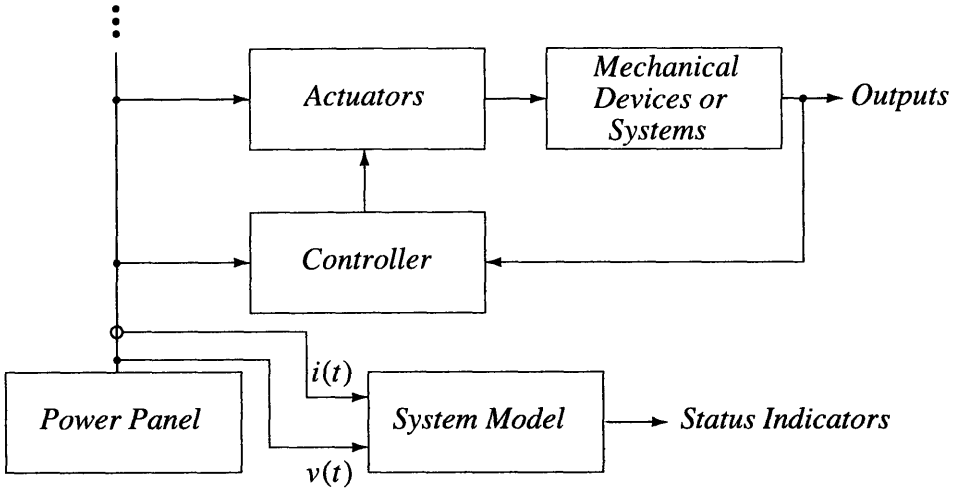


Figure 2-11: Block diagram showing the data flow in a typical non-intrusive implementation of system-level health monitoring.

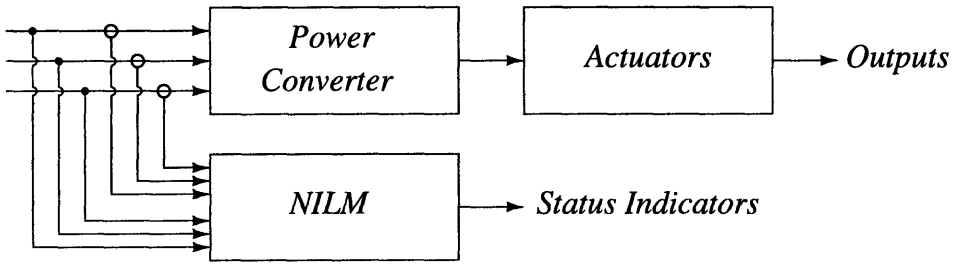


Figure 2-12: In many systems, actuators are driven by power converters. In this arrangement, the NILM detects actuator events using changes in the converter's input current.

specifics of the application, this converter can be as simple as a transformer or as complex as a power electronic load with active waveshaping.

When power electronic circuits drive the actuators in a complex system, the identification process is slightly different. In those cases, the converters usually draw current from the distribution network on a continuous basis. When an actuator is energized, the load on the converter changes, and thus, so too, does its input current. From the NILM's perspective, therefore, the identification of actuator events corresponds to the identification of changes in the current drawn by the converter. Although this subtle difference does not necessarily pose any particular problems for the classification process, it is highlighted because it implies that the transient behavior of interest is actually related to the physics governing the interaction between the converter and the actuator.

System-level health monitoring also requires the NILM to be able to identify the response of an actuator to an indirect command. In this context, indirect commands are those that elicit responses from actuators that they do not directly control. For example, consider a pressure-regulating valve as might be found in the pneumatic starting system for a gas turbine ([60]). Often, these valves are actuated by compressed air supplied from a nearby air receiver. In many cases, this control air is admitted by energizing a separate solenoid-actuated valve ([61]). When the pressure-regulating valve is ordered to throttle the flow in the system, the only command provided by the controller is to the solenoid. In response, however, the load on the compressed-air receiver increases, forcing a change in the operation of the compressor. This latter effect is an example of an indirect command.

Identification of proper actuator sequencing is another critical task performed by the NILM. Specifically, when an operator issues a particular command, a certain set of operations will follow in a specific order. For instance, when an operator starts an LM2500 marine gas turbine[2], the controller issues a series of commands to various valves in the system. First, a normally-closed solenoid-actuated valve in series with the turbine's pneumatic starter is energized, forcing the turbine to spin. Once a particular speed has been reached, a command is transmitted to a fuel shutoff valve, allowing fuel to enter the turbine's combustion chamber. As the turbine approaches its idle speed, the pneumatic starter is no longer needed, so the controller de-energizes the normally-closed valve in series with it ([61], [17]).

When a NILM is monitoring a system such as the turbine described above, it must identify each of the actuator events required to complete the particular operation of interest. Currently, the NILM performs this function by generating a report that lists the time at when each event was detected. An example recorded during the start of an LM2500 is shown below:

```
Timestamp        Event
-----------------------------------
10:21:27 India   Starter Air Regulating Valve OPEN
10:21:34 India   Fuel Shutoff Valve OPEN
10:21:34 India   Ignition Exciters ENERGIZED
10:21:41 India   F/O Service Pump Indicating Flow
10:21:58 India   Starter Air Regulating Valve CLOSED
10:21:58 India   Ignition Exciters DE-ENERGIZED
```

[2]The LM2500 marine gas turbine is a product of General Electric Co.

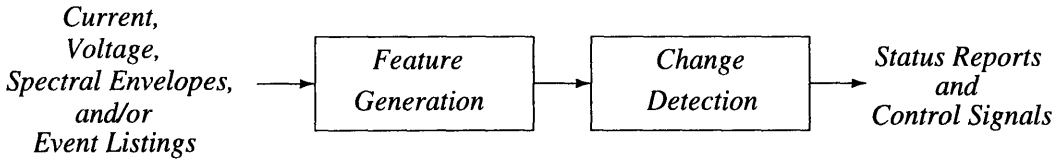


Figure 2-13: General procedure for fault detection and diagnosis using electrical measurements. Typical features of interest include parameter estimates, model residuals, and spectral quantities (i.e. amplitudes and frequencies).

In order to extend the NILM's capabilities, future devices will likely be supplied with a library of finite state machine (FSM) models. These models will describe the proper sequence of events required to perform each desired high-level system operation (e.g., the engine starts described above). In the most basic application of such models, the NILM would be provided with each operator command, and it would try to "match" the detected series of events with the appropriate FSM model. In the case that events did not match, the NILM would issue commands calling for appropriate corrective or emergency operations. A more advanced monitoring approach would be to try to identify high-level operations without knowing anything about the operator's commands. Identification in this case would require an exhaustive search of the NILM's FSM library. An example of the NILM's capabilities in system-level health monitoring is presented in the next chapter.

Load Diagnostics

In this thesis, significant attention is also given to the NILM's ability to detect faults in electromechanical actuators and in the mechanical systems that they service. In general, the NILM'a model-based diagnostic software is implemented according to the procedure presented in Fig. 2-13 ([9]). As shown, the first step is to extract relevant features from the input data. Typical features of interest include parameter estimates, model residuals, and spectral quantities. Once the key feature values have been determined, a change detection module will assess whether or not current trends are indicative of incipient or existing problems.

Diagnostic software used on-board the NILM can accept three different types of input data, namely raw voltages and currents, spectral envelopes, and event listings. The choice of input, of course, depends on the choice of the model for the given application. The bottom line, however, is that any model, regardless of its input source on the NILM, must correlate electrical data with faults observed in the system of interest.

A number of non-intrusive diagnostic routines have been developed and applied by previous researchers. In [32], Shaw demonstrated that a vast of amount of diagnostic information can be obtained from measurements of the transient stator currents drawn by three-phase induction machines. In his ingenious method, non-linear estimation routines are used to determine the machine parameters from measurements of a single phase current. Diagnostics are performed by tracking how the parameters vary over time ([32]).

A number of other non-intrusive diagnostic methods make use of the NILM's collection of event listings. In those instances, the feature generator typically determines how often a particular device or set of devices has been cycling. Under certain conditions, for instances, Luo demonstrated that this technique can be used to detect leaks in ventilation systems ([36]). Subsequently, Armstrong used event listings to detect compressor short-cycling in rooftop air-conditioning units ([18]). In Chapter 3 of this thesis, operating schedule data is used to sense the presence of leaks in vacuum systems and compressed air systems ([33]).

Spectral envelope data is another useful source of diagnostic information. Luo, for instance, demonstrated that spectral envelopes can be used to detect problems in the control systems for large ventilation fans ([36]). Additionally, Armstrong showed that liquid entering a compressor during steady-state operation perturbs the spectral envelope corresponding to the in-phase component of the fundamental current ([18], [34]). In [41], Lee describes the development of periodic disturbances in the spectral envelopes as the result of leaks in the evaporator chamber of a chiller.

Spectral envelopes are often useful when analyzing faults that cause slow variations in either the amplitude or the phase of the measured current. The reason for this can be seen by considering the nature of the spectral coefficients $a_k(t)$ and $b_k(t)$. These quantities, which are defined in Eqs. 2.2 and 2.3, respectively, are computed via the process presented graphically in Fig. 2-14 ([49]). As shown, the current $i(t)$ is first modulated by a set of quadrature sinusoids, producing the signals

$$i_c(t) = \frac{2}{T} i(t) \cos\left(k \frac{2\pi}{T} t\right) \tag{2.10}$$

and

$$i_s(t) = \frac{2}{T} i(t) \sin\left(k \frac{2\pi}{T} t\right). \tag{2.11}$$

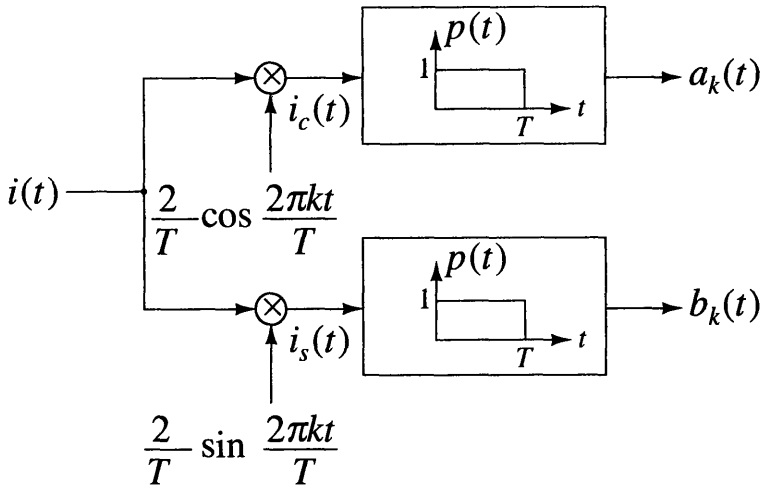


Figure 2-14: Signal flow graph demonstrating the process used to compute the spectral coefficients $a_k(t)$ and $b_k(t)$.

Subsequently, these two outputs are each passed to a filter whose impulse response is of the form

$$p(t) = \begin{cases} 1 & \text{for } 0 \leq t \leq T, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

This filtering operation produces the signals

$$a_k(t) = i_c(t) * p(t) \tag{2.13}$$

and

$$b_k(t) = i_s(t) * p(t) \tag{2.14}$$

where $*$ is the convolution operator. By comparison with Eqs. 2.2 and 2.3, it is clear that these two final output signals are indeed the spectral envelopes ([49]).

In terms of understanding the frequency content contained in the spectral coefficients, Eqs. 2.13 and 2.14 are particularly insightful. For example, the convolution presented in Eq. 2.13 implies that the continuous-time Fourier transform of $a_k(t)$ is given by the relation

$$A_k(f) = I_c(f)P(f), \tag{2.15}$$

where $I_c(f)$ and $P(f)$ are the Fourier transforms of $i_c(t)$ and $p(t)$, respectively ([62]). Assuming that $i(t)$ is a harmonic current with a period T and a slowly varying amplitude and phase, then the mixing operation in Eq. 2.10 suggests that $I_c(f)$ will contain a baseband component that corresponds to the frequency content of the signals that modulate the amplitude and phase of $i(t)$ ([62]). Of course, Eq. 2.10 also implies that $I_c(f)$ will contain mirror images of this baseband content at higher frequencies ([62]). Given the approximately low pass nature of $P(f)$, however, this high frequency content will be attenuated by the filtering operation, and $a_k(t)$ will, in general, be a baseband signal. Similar arguments can be made for $b_k(t)$.

In the event that a fault results in a slow variation of either the amplitude or the phase of any of the harmonic components of the current, then the above presentation suggests that a frequency domain analysis of the spectral envelopes would be particularly useful. To further illustrate this, consider how a NILM would analyze the slow modulations using the current waveform directly. In that case, the NILM would, by necessity, use a fast algorithm such as the FFT to analyze a finite length, windowed series of samples. Windowing, however, reduces resolution and introduces spectral leakage ([54]). The combination of these two effects is particularly problematic in this case, where slow variations in the amplitude and the phase of the current result in closely spaced spectral peaks centered around the harmonics of the fundamental frequency. Given the extreme difference between the sample rate of the current and the frequency spacings of interest, it would be very difficult to resolve the appropriate spectral peaks. Furthermore, given the level of detail required in the frequency spectrum, many DFT samples would need to be generated[3]. Even when using the FFT, the resulting computation would be burdensome and, perhaps, even prohibitive in many systems.

As implied by Eq. 2.15, the spectral envelopes are relatively band-limited. As such, they do not need to be as finely sampled as the current. In practice, the NILM obtains spectral envelopes at roughly 120Hz. With this reduced sampling rate, the resolution and leakage problems are significantly mitigated, making it much easier to analyze the spectral content resulting from slow modulations of the current.

Frequency domain analysis of the spectral envelopes has proven to be highly successful as a non-intrusive diagnostic tool. For instance, several researchers have demonstrated that

[3]Recall that the discrete Fourier transform (DFT) produces samples of the discrete-time Fourier transform (DTFT) of a windowed, finite length sequence ([54]).

it can be used to detect unstable static pressure controllers in ventilation systems ([36], [42]). Additionally, Armstrong used the frequency domain to detect shaft imbalances in fans ([18]). In this thesis, the frequency content in the spectral envelopes is used to diagnose the impending destruction of a coupling via repeated torsional oscillations ([35]).

Chapter 3

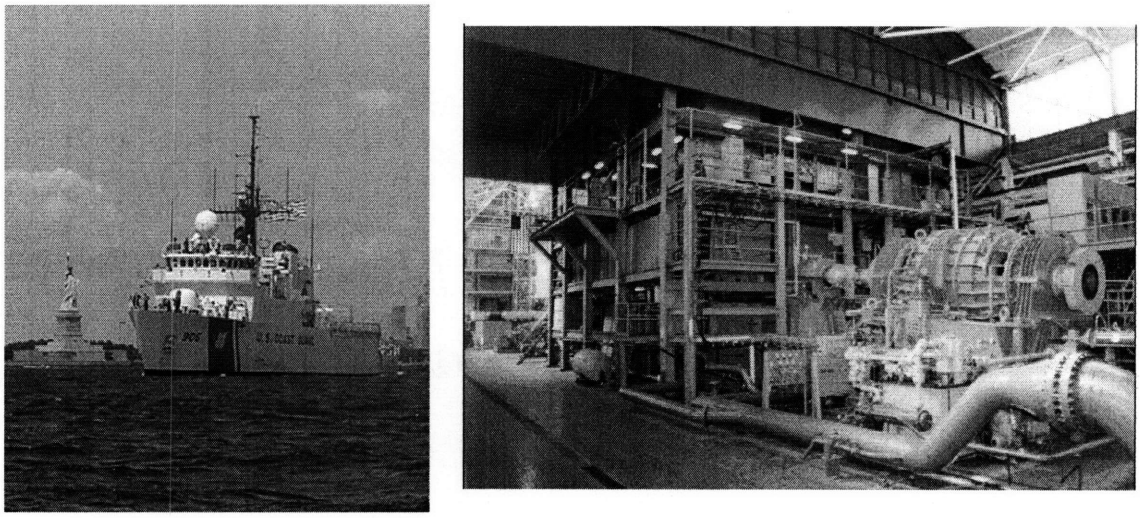
Applications of Non-Intrusive Load Monitoring

THIS chapter presents several example applications of the non-intrusive load monitor. The first two of these demonstrate the NILM's abilities as a platform for model-based diagnostics, while the third example showcases the NILM as a tool for system-level health monitoring. In all three cases, relevant techniques and methods are demonstrated in the field. The test sites, which include an active U.S. Coast Guard Cutter and an active U.S. Navy support facility, are shown in the photographs in Fig. 3-1.

The models and methods presented in this chapter are applicable in many systems. The use of specific examples is not meant to imply that these tools are finely tuned to any one system; rather, field tests are simply used to verify these techniques in representative cases. Moreover, the results presented in this chapter have prompted the U.S. Coast Guard to begin full-scale integration of the NILM into several of the specific systems that are mentioned herein.

3.1 Non-Intrusive Diagnostics for Cycling Systems

There are many systems in which an actuator must periodically operate in order to ensure that a particular mechanical variable remains within a certain fixed range. Consider, for example, compressed-air and vacuum systems. Industrial facilities, marine vessels, and aircraft all commonly use these systems to perform a number of critical functions, including waste collection and valve actuation. To ensure instant availability and to provide for short periods of high demand, these systems typically have an air receiver or vacuum tank that is



(a) (b)

Figure 3-1: The two platforms used to test the non-intrusive diagnostic methods developed in this this thesis. (a) The *USCGC Seneca* and (b) The U.S. Navy's Land-Based Engineering Site (LBES) in Philadelphia, PA.

periodically charged by a compressor or pump. A basic layout for such a system is shown in Fig. 3-2 ([63]). As loads draw from the pressurized reservoir, system pressure decreases. Once a certain low-pressure set point has been reached, the controller transmits a start command to the pump or compressor, causing the device to begin charging the system. Once the pressure has reached the predetermined high set point, the pump motor is either de-energized or the pumps are directed to enter a re-circulation mode. From an electrical perspective, this behavior creates a *cycling system* in which electrical power usage follows a regular cycle of charging and discharging based on a pressure.

The cyclic behavior described above is by no means unique to compressed-air and vacuum systems. There are many other cases in which actuators cycle according to the value of a particular mechanical variable. Many HVAC loads, for instance, cycle in accordance with the value of a certain temperature measurement. Additionally, the pumps in many storage or transfer systems cycle their operation in accordance with the amount of a certain stored quantity such as water or waste.

Although the actuators in cycling systems can suffer from any number of potential problems, one primary concern is the detection of faults that cause the actuators to over-operate. From the perspective of a non-intrusive load monitor, this detection requires

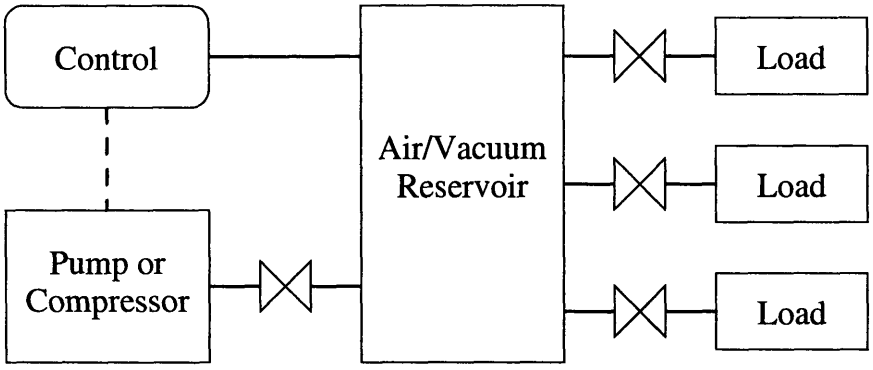


Figure 3-2: Diagram of a typical pneumatic cycling system. Solid lines indicate air paths, and dotted lines indicate electrical connections.

the development of a model that describes how changes in the controlled variable affect the operating schedule of the actuator. To see this, consider Fig. 3-3, which shows both pressure and pump power in a representative cycling system aboard the *USCGC Seneca*. As shown, any time that a system usage event (SUE) removes vacuum from the system, a sharp drop is observed in the measured vacuum pressure. Note that as the number of SUEs increases, the discharge period shortens and the number of pump runs increases. By comparison, the development of a leak, which causes the persistent loss shown in Fig. 3-3, also increases the number of pump runs and decreases the average length of a discharge period. Any attempt at non-intrusive leak detection, therefore, requires the development of a model that makes it possible to distinguish between these two very different effects.

This section examines model-based diagnostic indicators for the cycling systems described above. It begins by introducing a general framework for modeling such systems, and it then employs that framework to model one specific type of cycling system, namely vacuum-assisted waste-disposal systems. Finally, several non-intrusive diagnostic methods are developed and demonstrated aboard an active-duty U.S. Coast Guard Cutter.

3.1.1 Modeling Cycling Systems: A General Framework

In the context of this thesis, the primary reason for modelling cycling systems is to determine how the actuator behaves under both faulty and non-faulty conditions. Once a model has been created, it can then be used develop an analytical procedure that distinguishes between these two very different sets of operating conditions. This section presents a general procedure that can be used to model any specific cycling system.

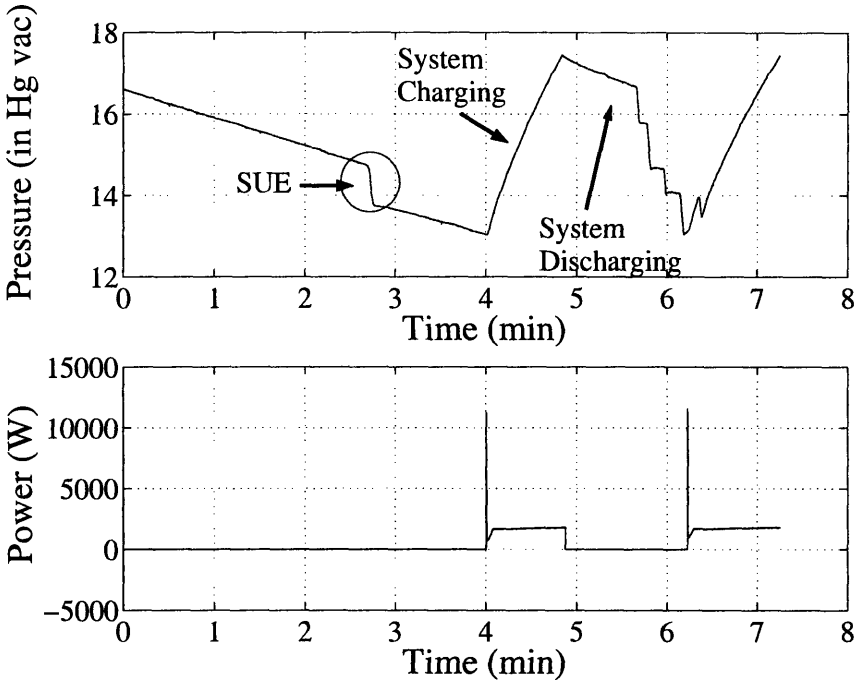


Figure 3-3: Reservoir pressure and pump input power during several charge and discharge cycles in the vacuum-assisted waste-disposal system aboard *Seneca*. Note that the sharp drops in pressure are the direct results of system usage events (SUEs) and that the persistent background loss is the result of a small leak.

To develop a specific model, one must begin by determining how the actuator behaves under non-faulty conditions. In general, this behavior is controlled by the actions of the various elements that either source or sink the variable of interest. In a compressed-air system used to power various pneumatic tools, for example, the actuator operating schedule is determined by the actions of the various tools. By comparison, the cycling of a residential heater is affected by factors such as the quality of insulation, outdoor temperature, and indoor occupancy. As these examples suggest, each of the relevant source and sink elements has its own individual operating schedule as well as its own individual effect on the system. To build a complete model, both of these factors must be considered and addressed.

Specific attention must be provided to the determination of the operating schedule of each individual source or sink. In general, the operation of these elements is under the control of certain external effects. For example, tool usage in a compressed-air system follows operator demand, while occupancy in a heated room varies according to the need or the desire of individuals to be in that room. From the viewpoint of the actuator, these

external driving forces are unknown. As a result, the behavior of any individual source or sink is essentially a stochastic process. For a given element, this process has two potential sources of random variation, namely the time at which the element enters operation and the amount of time that it operates. The best way to model these effects is to turn to algorithmic queueing theory. The next section addresses this issue in some detail for a specific system.

A complete cycling-system model must also describe how individual sources and sinks affect the controlled variable. In general, each element elicits a different dynamic response from the system. Certain elements, such as the tools in a compressed-air system present a very large load and cause a sudden reduction in pressure. Other elements, however, such as the occupancy in a heated room can cause very slow changes. The next section also considers the importance of modeling this effect.

With appropriate models for the behavior of each of the individual sources and sinks, the complete response of the system can be predicted by simulation. If a leak is inserted into the system, the effect will be to add a continuously active sink element. In order to determine the impact of the leak, it can be included into the model and simulated. An analysis of the model's predictions under faulty and non-faulty conditions indicates how a leak can be detected using only the operating schedule of the actuator. In general, the required analysis will differ from one class of systems to the next. The next section addresses this issue as well.

3.1.2 Cycling-System Example: Vacuum-Assisted Waste Disposal

This thesis demonstrates the utility of the modelling process described above by applying it to to an example set of systems. In this case, the examples are vacuum-assisted waste-disposal systems, which are commonly deployed on marine vessels and aircraft. The presentation provided here begins by describing the various pneumatic loads in these systems as well as the process that governs their usage. Subsequently, the full model is simulated and the results are compared to measured data. For convenience, an idealized version of the full model is presented first.

Pneumatic Load Dynamics

In vacuum-assisted waste-disposal systems, there are two relevant types of vacuum reducing events. The first of these are what are termed system usage events, and they result from the

operation of typical system air loads, such as drains. In general, these loads cause sharp, nearly instantaneous drops in vacuum pressure. In addition, there are also leaks, which typically result in a persistent vacuum loss. Assuming that the system can be accurately modeled using lumped-element approximations, the pressure loss due to a leak, which is denoted as $P_{leak}(t)$, is the solution to the first-order differential equation[1]

$$\frac{dP}{dt} = -cP, \tag{3.1}$$

where c is a constant whose value depends upon the parameters of the system (i.e. capacity, etc.). In later sections of this thesis, the solution to Eq. 3.1, as well as several reasonable approximations, will be considered.

Assuming that N usage events have occurred during the current discharge period[2], the reservoir pressure, $P(t)$ can be approximated as

$$P(t) = P_{high} - (\Delta P_1 + \Delta P_2 + \Delta P_3 + \cdots + \Delta P_N) + P_{leak}(t), \tag{3.2}$$

where P_{high} is the high pressure set point, ΔP_k is the pressure removed by the k-th SUE, $P_{leak}(t)$ is a functional description of the leak, and $t = 0$ is defined to be the time at which the current discharge period began. In order for the pumps to energize, the total pressure loss must be sufficient to cause $P(t)$ to fall below the controller's low-pressure set point, P_{low}. Equivalently, the pumps will operate as soon as the total pressure loss, ΔP_{loss}, is greater than the difference between P_{high} and P_{low}. Thus, mathematically, the pump trigger condition can be written as

$$\begin{aligned} P_{high} - P_{low} &\leq \Delta P_{loss} \\ &\leq (\Delta P_1 + \Delta P_2 + \Delta P_3 + \cdots + \Delta P_N) + P_{leak}(t). \end{aligned} \tag{3.3}$$

System Usage Process

The rate at which individual usage events affect the system has a strong impact on the operating schedule of the pumps. In general, however, these events occur at random intervals, as human users typically operate the drains only when needed. As indicated previously,

[1] For convenience, the terms "pressure" and "vacuum pressure" are used interchangeably in this section.
[2] As indicated in Fig. 3-3, the term discharge period refers to the interval when the pump is not operating.

this means that one must propose an appropriate stochastic model. As shown below, a reasonably accurate model can be chosen without performing a great deal of analysis and experimentation.

The determination of the usage process in a cycling system amounts to the selection of an appropriate *queueing* model. These models originated in the early 1900s, when they were investigated by A. K. Erlang for purposes of developing automatic telephone exchanges ([64]). Just as in the case of the telephone problem, cycling systems have a community of users that both request service at random intervals and require service for random lengths of time ([64]). The selection of an appropriate model requires one to consider each of the following ([64]):

- The statistics of the arrival process
- The statistics of the service times
- The number of servers

In the waste-disposal example, the system usage process is approximated as an $M/D/\infty$ queue ([64]). The name of this standard queueing model, which is written in a shorthand form known as Kendall's notation, makes reference to the way in which the model addresses each of the three considerations listed previously ([64]). In this nomenclature, the first character describes the arrival process, the second character describes the service time distribution, and the third character lists the number of servers. In terms of the $M/D/\infty$ queue, this means that the arrival process is Poisson, the service times are deterministic, and the number of servers is infinitely large ([64]).

The use of the $M/D/\infty$ queue to model usage patterns in waste-disposal systems can be justified with relative ease. First, actual field data has shown that the Poisson process reasonably approximates the arrivals in a waste-disposal system. This claim is further justified in the next section. Second, given that usage events have been shown to occur almost instantaneously, one can approximate the distribution for the service times using one deterministic value, namely zero. Third, the claim that the number of servers approaches infinity is effectively true, at least as long as the system has enough toilets and drains to guarantee that patrons will rarely, if ever, have to wait. In most waste-disposal systems, this is likely to be the case.

In the special case that the service times are zero, the combined operating schedule of the individual pneumatic loads is governed exclusively by the statistics of the Poisson arrival process, which is denoted as $N(t)$. In this model, each new arrival increases the value of $N(t)$ to the next largest integer. Assuming that the arrival rate λ does not change as a function of time, this process can be defined as follows. The process starts at $t = 0$, i.e. $N(0) = 0$, and for any times t and s such that $t > s \geq 0$, the increment $N(t) - N(s)$ is independent of $N(\tau)$ for all $\tau \leq s$ ([65]). Moreover, this increment has the following Poisson distribution ([65]):

$$\begin{aligned} \Pr\left[N\left(t\right) - N\left(s\right) = k | N(\tau), \tau \leq s\right] &= \Pr\left[N\left(t\right) - N\left(s\right) = k\right] \\ &= \frac{\left[\lambda\left(t-s\right)\right]^k e^{-\lambda(t-s)}}{k!} \end{aligned} \tag{3.4}$$

Since $N(0) = 0$, Eq. 3.4 implies that

$$\Pr\left[N\left(t\right) = k\right] = \frac{\left(\lambda t\right)^k e^{-\lambda t}}{k!}. \tag{3.5}$$

Physically, Eq. 3.5 expresses the probability that $N(t)$ is equal to a given integer value.

For purposes of modeling the system usage process in a cycling system, it is important to be able to predict more that just the number of arrivals in a given interval. Specifically, it is also necessary to be able to predict the time between individual arrivals. To derive the distribution for this quantity, consider each inter-arrival period to be a random variable T. For a moment, focus on the time to the first arrival. The probability that this event occurs during the interval $[0, t]$ is a quantity known as the cumulative distribution function (CDF) of T. Thus, by definition, this function is

$$F_T(t) = \Pr\left[T \leq t\right]. \tag{3.6}$$

Using both Eq. 3.6 and the total probability law ([66]), Eq. 3.6 can be rewritten as follows

$$\begin{aligned} F_T(t) &= \Pr[T \leq t] \\ &= 1 - \Pr[T > t] \\ &= 1 - \Pr[N(t) = 0] \\ &= 1 - e^{-\lambda t}. \end{aligned} \tag{3.7}$$

From the above result, it is possible to determine the probability distribution function (PDF) for T. By definition, this function, which is denoted as $f_T(t)$, is the derivative of the corresponding CDF, $F_T(t)$. Thus, the PDF is

$$\begin{aligned} f_T(t) &= \frac{dF_T(t)}{dt} \\ &= \lambda e^{-\lambda t}. \end{aligned} \tag{3.8}$$

Although it will not be shown here, the Poisson process is memoryless, meaning that the behavior of the arrivals after any time $t \geq 0$ is itself a Poisson process that is independent of the behavior prior to that time ([66]). Thus, Eq. 3.8 is the PDF for the time between any two arrivals.

Ideal Behavior of the Pump Operating Schedule

In order to predict the operating schedule of the actuator in a cycling system, the usage and load models must be combined. In this case, it is convenient to begin by first considering the following set of idealized operating conditions:

- Each SUE reduces the reservoir pressure by an amount ΔP
- System usage is a homogeneous Poisson process, i.e. λ is not a function of time
- The system usage process resets at the beginning of each discharge period

In this situation, the control relation presented in Eq. 3.2 simplifies as follows

$$\begin{aligned} P(t) &= P_{high} - (\Delta P_1 + \Delta P_2 + \Delta P_3 + \cdots + \Delta P_N) + P_{leak}(t) \\ &= P_{high} - N\Delta P + P_{leak}(t). \end{aligned} \tag{3.9}$$

If no leaks are present in the system, then this result can be reduced even further. With $P_{leak}(t) = 0$,

$$P(t) = P_{high} - N\Delta P. \tag{3.10}$$

Since ordinary usage is the only source of loss in Eq. 3.10, it is clear that the pump will energize as soon as $N\Delta P$ is large enough that $P(t) \leq P_{low}$. Given that each SUE reduces $P(t)$ by the same amount, Eq. 3.10 clearly implies that there is a fixed number of usage events that must transpire before the controller will command the pump to energize. For instance, in the example system used to generate Fig. 3-4, it is clear that the pump will begin to operate immediately following the occurrence of the third SUE. For convenience, we define the variable N_{max}, which represents the maximum number of usage events that can occur during any single discharge period. As implied by Fig. 3-4, the value of N_{max} is simply the smallest integer that guarantees the validity of the inequality

$$P_{low} \geq P_{high} - N\Delta P. \tag{3.11}$$

Solving, it is found that

$$N_{max} = \left\lceil \frac{P_{high} - P_{low}}{\Delta P} \right\rceil. \tag{3.12}$$

where $\lceil x \rceil$ is the ceiling of x.

When operating under the simplified conditions considered here, it is possible to predict the distribution for the time T_p that elapses during any individual discharge period. Given the fact that the number of usage events impacting the system is a fixed quantity, T_p is simply the sum of the N_{max} inter-arrival times, i.e.

$$T_p = T_1 + T_2 + T_3 + \cdots + T_{N_{max}}. \tag{3.13}$$

Because T_p is the sum of a fixed number of random variables, its PDF is given by the relation

$$f_{T_p}(t) = f_1(t) * f_2(t) * f_3(t) * \cdots * f_{N_{max}}(t), \tag{3.14}$$

where $*$ is the convolution operator ([66]). Given the inter-arrival model presented in Eq. 3.8, the PDF for T_p is

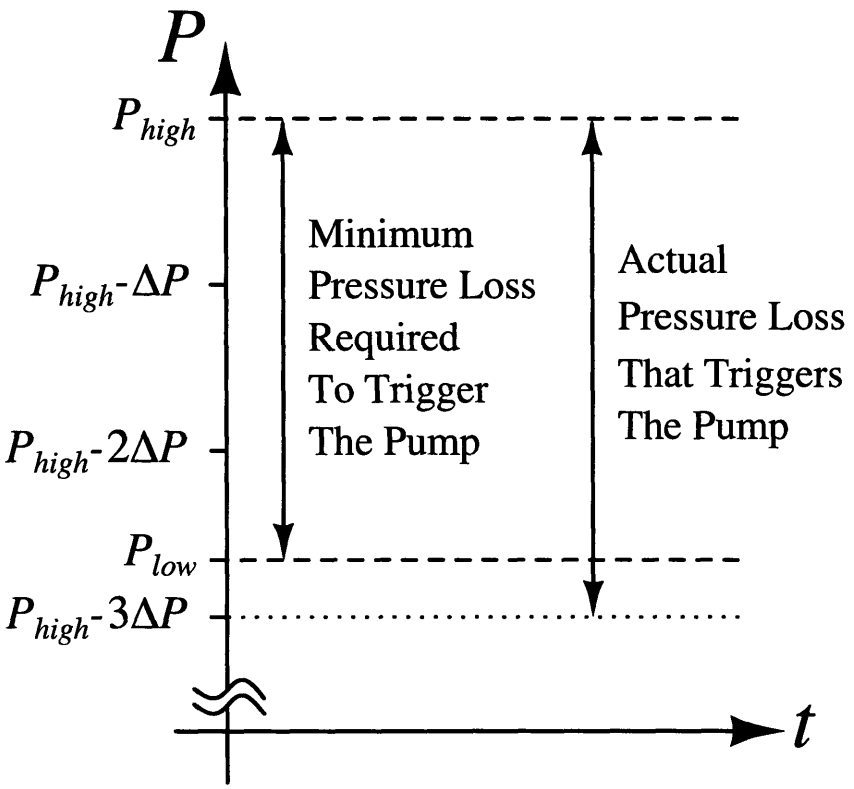
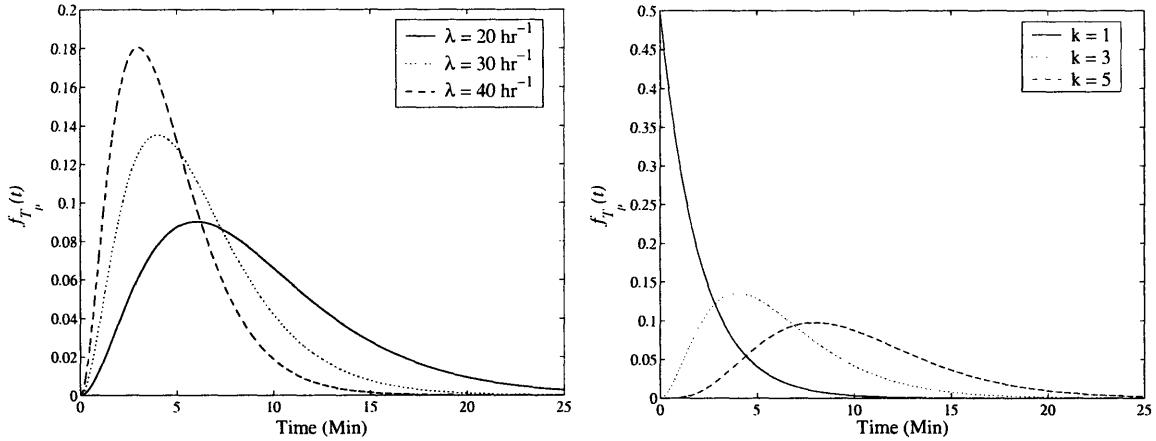


Figure 3-4: Key pressure values in an example cycling system in which $N_{max} = 3$. As shown, the pump will energize as soon as the reservoir pressure drops below P_{low}, which, in this system, can only happen after the arrival of the third SUE.



(a) (b)

Figure 3-5: The Erlang PDF for (a) $k = 3$ and various values of λ and for (b) $\lambda = 30\text{hr}^{-1}$ and various values of k.

$$f_{T_p}(t) = \lambda^{N_{max}} \frac{t^{N_{max}-1} e^{-\lambda t}}{(N_{max}-1)!}. \tag{3.15}$$

In general, this two-parameter distribution is known as the Erlang PDF of order k ([66]). In Eq. 3.15, $k = N_{max}$. Figure 3-5a shows how this distribution behaves for a fixed k and increasing λ; similarly, Fig. 3-5b shows how the distribution behaves for a fixed λ and increasing k. Note that in the special case that $k = 1$, the Erlang PDF reduces to the exponential distribution.

In order to demonstrate the idealized behavior described above, a simulation was designed and executed in Simulink[3] ([35]). In the simulation, the user can set the values of all of the relevant system parameters (i.e. ΔP, P_{high}, P_{low}) as well as the value of the Poisson parameter, λ. Table 3.1 lists the values used to generate the simulated results presented graphically in Fig. 3-6. Also shown in that figure is the frequency distribution predicted by Eq. 3.15.

In order to justify the use of the Poisson model, consider the measured results presented in Fig. 3-7. The data used to generate that histogram was obtained during a one-week study of the vacuum-assisted waste-disposal system aboard the *USCGC Seneca*. Note the qualitative agreement between the measured data and the simulated predictions. Although

[3]The complete details of the simulation are described in Appendix A.

Table 3.1: Parameters used in the ideal system simulation.

Parameter	Value
P_{high}	17.5 in Hg
P_{low}	13.5 in Hg
λ	30 hr^{-1}
ΔP	1.1 in Hg

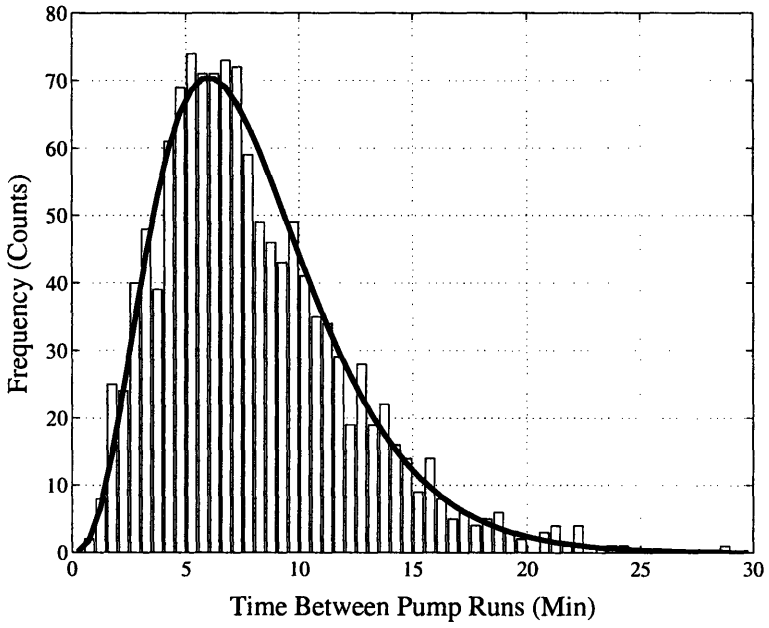


Figure 3-6: Expected (solid line) and simulated (bars) frequency distributions for the discharge time, T_p, under fault-free conditions. The simulation time was one week.

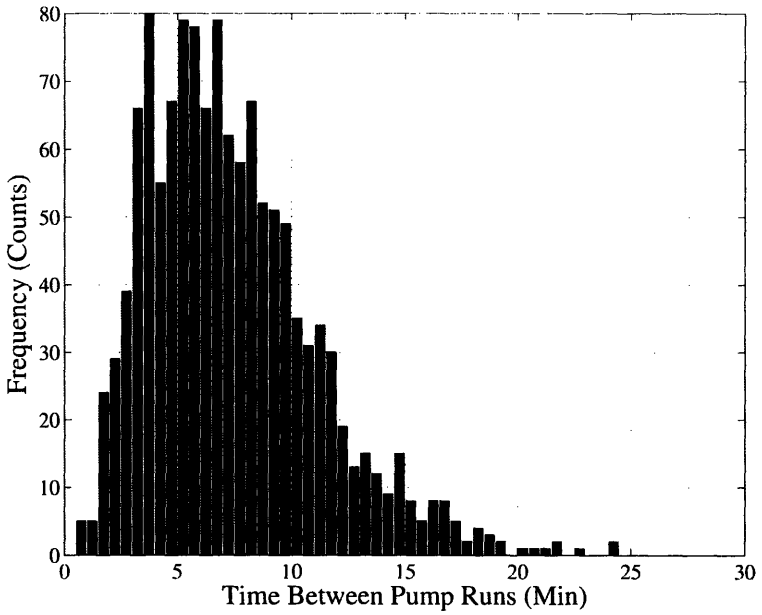


Figure 3-7: Measured frequency distribution obtained when monitoring the vacuum-assisted waste-disposal system aboard the *USCGC Seneca*. These results were gathered during a week of fault-free conditions.

this comparison is not a rigorous proof, it does lend credence to the assertion that the usage process is Poisson.

In the event that the ideal system develops a leak, its behavior will depart from that predicted above. A simple graphical explanation for this departure is presented in Fig. 3-8. As shown, the leak causes the reservoir pressure to decrease continuously. As a result, the number of SUEs required to initiate pump operation becomes dependent upon the time that has elapsed since the beginning of the discharge period. In the example system used to generate Fig. 3-8, for instance, it is clear that the only way to force the pumps to re-energize during the first few minutes of the discharge period is for three usage events to occur in relatively short succession. If that does not happen, however, the leak will continue to reduce the pressure in the reservoir. Eventually, leak-induced loss will be large enough that only 2 SUEs are needed to initiate pump operation. For example, at the time t_0 shown in Fig. 3-8, it is clear that the pressure loss resulting from 2 SUEs will be sufficient to cause the pumps to energize.

With certain simplifying assumptions, a more rigorous foundation can be provided for the leak behavior described above. Specifically, for a lumped element system, the time

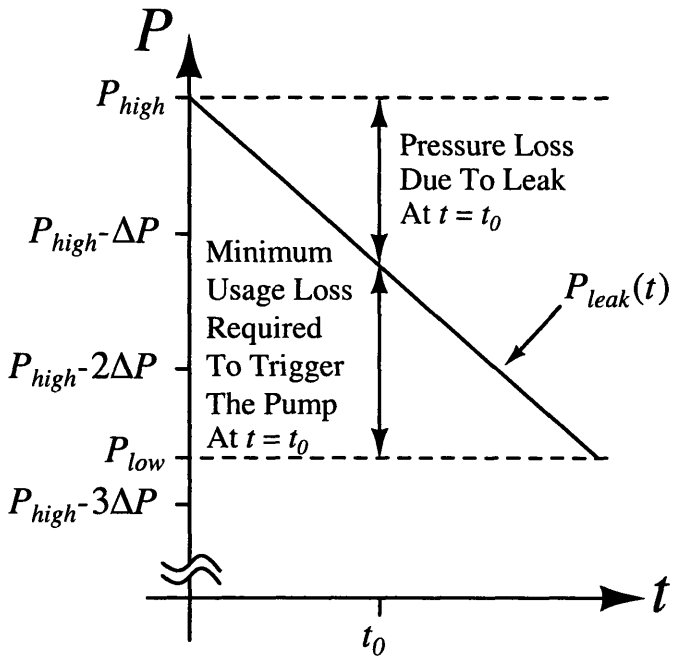


Figure 3-8: Leak-induced pressure loss versus time in an example system with $N_{max} = 3$. Note that the distance between $P_{leak}(t)$ and P_{low} is the amount of usage-induced pressure loss that would be needed in order to cause the pump to energize at time t.

dependence of the leak-induced pressure loss can be described mathematically using the solution to Eq. 3.1. If the controller is programmed so that it keeps the pressure within a relatively narrow range, then the exponential solution to Eq. 3.1 can be represented using the following first-order Taylor series expansion:

$$P_{leak}(t) = P_{high}e^{-ct} \approx P_{high} - \alpha_{leak}t, \tag{3.16}$$

where $\alpha_{leak} = cP_{high}$. For a system that obeys these assumptions, there are exactly N_{max} times at which the required amount of usage-induced pressure loss decreases. The procedure used to identify these times is outlined graphically in Fig. 3-9. As an example, consider what happens if the discharge period lasts until time τ_1. At that exact instant, the leak has reduced the pressure to the point that the pumps will operate if the usage-induced loss is exactly $(N_{max} - 1)\,\Delta P$. Prior to that time, however, only the occurrence of N_{max} SUEs could have forced the system pressure to fall below P_{low}. A similar procedure can be used to determine τ_2, the time at which the required usage loss drops from $(N_{max} - 1)\,\Delta P$ to $(N_{max} - 2)\,\Delta P$. In general, the values of the times τ_k are given by the formula

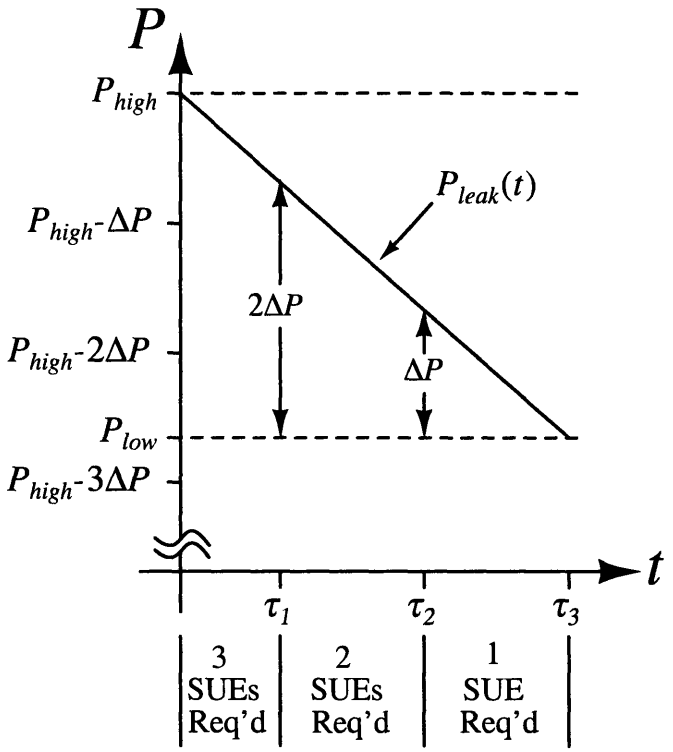


Figure 3-9: Leak-induced pressure loss versus time in an example system with $N_{max} = 3$. Note that at time τ_1, the vertical distance between $P_{leak}(t)$ and P_{low} is exactly $2\Delta P$. As a result, only 2 SUEs are required to trigger the pumps at that time.

$$\tau_k = \frac{P_{high} - P_{low} - (N_{max} - k)\,\Delta P}{\alpha_{leak}} \quad \text{for } k = 1, 2, 3, \ldots, N_{max}. \tag{3.17}$$

Clearly, the behavior described above has an effect on the distribution $f_{T_p}(t)$. In particular, leak conditions cause the PDF to change at the times τ_k. For $0 < t < \tau_1$, the only possible set of events that can elicit pump operation is the arrival of N_{max} SUEs. Thus, the PDF in this region is still given by Eq. 3.15. Recall, however, that the pumps will automatically operate at $t = \tau_1$ if exactly $N_{max} - 1$ SUEs impact the system during the interval $0 < t < \tau_1$. Thus, there is a certain fixed probability that the pumps will operate at that time, and it is equal to the probability that the the system experienced at least $N_{max} - 1$ but not N_{max} SUEs prior to τ_1. To determine the value of this probability, we must employ the Erlang CDF of order k, which is defined as

$$F(t;k,\lambda) = \begin{cases} \frac{\lambda^k}{(k-1)!}\int_0^t x^{k-1}e^{-\lambda x}dx, & t \geq 0, \\ 0, & t < 0. \end{cases} \tag{3.18}$$

For a given value of k, this equation expresses the probability that at least k arrivals occurred prior to time t. Thus, from the above arguments, the probability that the pump runs at time τ_1 can be expressed using the relation

$$F(\tau_1; N_{max}-1, \lambda) - F(\tau_1; N_{max}, \lambda). \tag{3.19}$$

Between the times τ_1 and τ_2, only $N_{max}-1$ SUEs are required to initiate a pump run. The PDF in this region will again be Erlang, but it will be of a reduced order. Specifically, this section of the PDF is the Erlang distribution of order $N_{max}-1$, i.e.

$$f_{T_p}(t) = \lambda^{N_{max}-1}\frac{t^{N_{max}-2}e^{-\lambda t}}{(N_{max}-2)!}, \qquad \text{for } \tau_1 < t < \tau_2. \tag{3.20}$$

In order to generalize the procedure presented above, it is useful to consider the complete CDF of the variable T_p. From the above arguments, it is clear that for the times $0 < t < \tau_1$, the CDF is Erlang with $k = N_{max}$. After $t = \tau_1$, the CDF is still Erlang, but its order is reduced by one. As a result, there is a step change in the overall CDF at $t = \tau_1$, and the height of this change is given by Eq. 3.19. This behavior is illustrated graphically in Fig. 3-10a. As shown, the CDF has similar step changes at each of the τ_k, as the number of required SUEs, and thus the order of the required Erlang, decreases by one. Furthermore, the height of the "jump" at each τ_k corresponds to the probability of observing a pump run at that time.

The complete PDF can be derived by differentiating the CDF shown in Fig. 3-10a. In between the τ_k, this is a straightforward operation. At the boundaries, however, the CDF has step changes, meaning that the derivative at those locations is an impulse whose area corresponds to the height of the change[4] ([67]). Thus, the complete PDF for T_p during leak conditions is

[4]Unless a PDF has any impulses, the probability assigned to any single exact experimental value of the corresponding continuous random variable is, by definition, zero. This can be seen mathematically by taking the integral of the PDF over an interval of zero width. Thus, if there is a certain fixed probability associated with a given value of the random variable, the only way to express that is to place an impulse in the PDF at that location ([67]).

$$
\begin{aligned}
f_{T_p}(t) &= f\left(t; N_{max}, \lambda\right)\left[u(t) - u\left(t - \tau_1\right)\right] \\
&\quad + \sum_{i=2}^{N_{max}} f\left(t; N_{max} - i, \lambda\right)\left[u(t - \tau_{i-1}) - u\left(t - \tau_i\right)\right] \\
&\quad + \sum_{i=1}^{N_{max}-1} \left[F\left(t; N_{max} - i, \lambda\right) - F\left(t; N_{max} - i + 1, \lambda\right)\right]\delta\left(t - \tau_i\right) \\
&\quad + \left[1 - F\left(t; 1, \lambda\right)\right]\delta\left(t - \tau_{N_{max}}\right) \qquad (3.21)
\end{aligned}
$$

For demonstration purposes, Fig. 3-10b shows the PDF that corresponds to the CDF presented in Fig. 3-10a.

Figure 3-11 displays a simulated frequency distribution that was obtained by inserting a small leak into the ideal system characterized by the parameters presented in Table 3.1. The actual leak rate in this case is 10 in Hg/hr. Note that the additional probability of observing a pump run at the times τ_k causes several large "pulses" to appear. For comparison purposes, Fig. 3-11 also shows the frequency distribution predicted by Eq. 3.21.

Actual Behavior of the Pump Operating Schedule

In reality, most cycling systems do not obey all of the assumptions used up to this point. Most importantly, the pressure loss ΔP_k due to a single SUE will, in general, be a random quantity. As a result, the number of SUEs required to initiate pump operation during normal conditions is itself a random quantity, which we denote using the discrete random variable η. The value of this variable during any given discharge period will depend upon the amount of pressure removed be each SUE.

Under these conditions, the length of any individual discharge period is

$$T_p = T_1 + T_2 + T_3 + \cdots + T_\eta. \qquad (3.22)$$

Since T_p is now the sum of a random number of random variables, its PDF depends on the distribution for ΔP_k. An approximate distribution for this random variable can be derived by considering the various pneumatic loads in the target system. In the case of the vacuum-assisted waste-disposal system aboard *Famous*-class Coast Guard Cutters, for instance, there are essentially two types of loads, namely toilets and urinals ([68]). As a result, there are roughly two different values for the variable ΔP_k. Moreover, there is

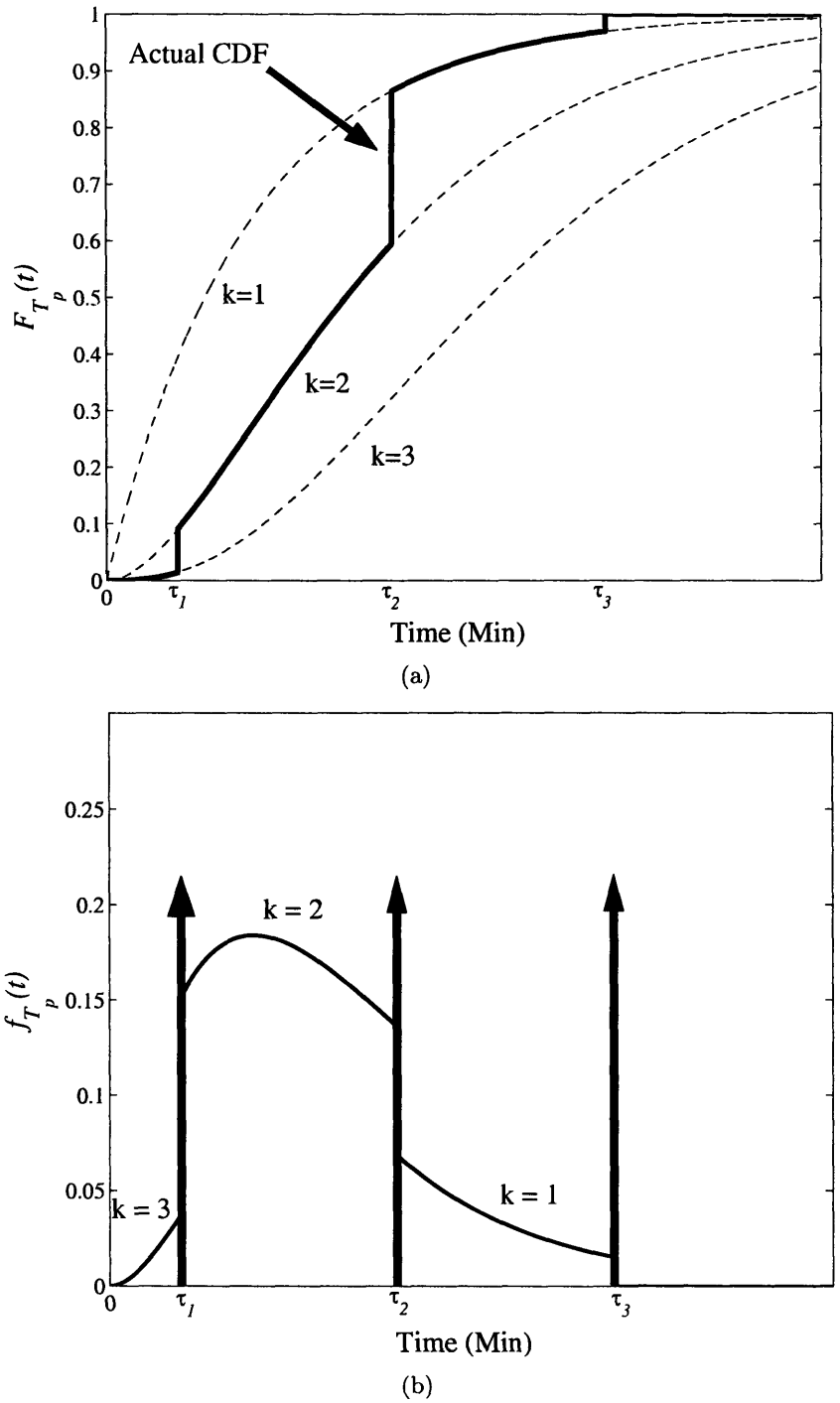


Figure 3-10: Theoretical distributions for T_p when a leak exists in the ideal system. (a) The CDF and (b) The PDF, $f_{T_p}(t)$.

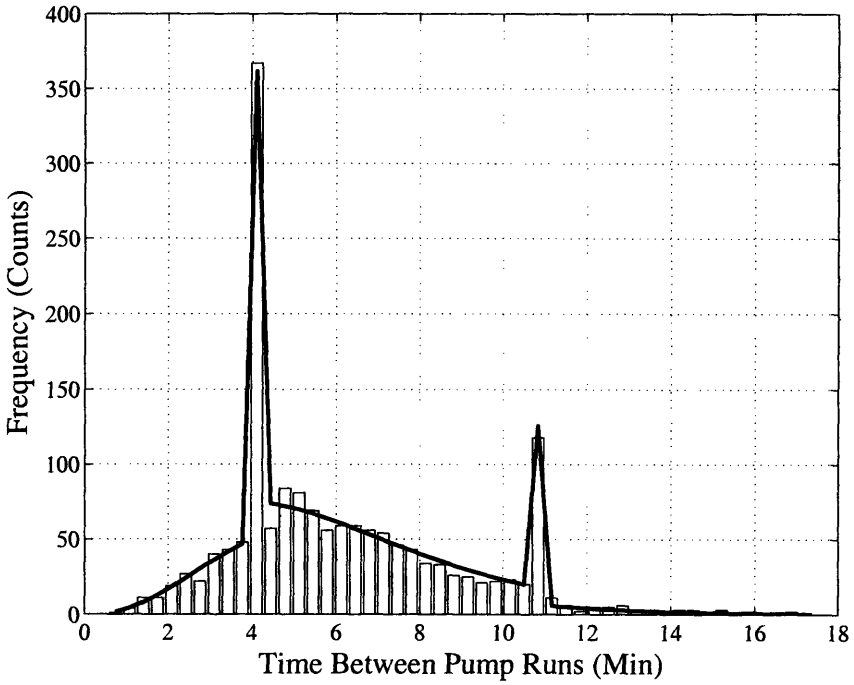


Figure 3-11: Expected (solid line) and simulated (bars) frequency distributions for the discharge time, T_p, in the idealized system with a 10 in Hg/hr leak rate. The simulation time was one week.

some variation in each of these two values. In order to simulate these effects, the following distribution is assumed for ΔP_k:

$$f_{\Delta P}(\Delta p) = \begin{cases} 5, & 0.6 \leq \Delta p \leq 0.7, \\ 5, & 1.3 \leq \Delta p \leq 1.4. \end{cases} \tag{3.23}$$

Another issue in many real cycling systems is that the usage rate, λ, does not remain constant. For instance, it is reasonable to expect that the vacuum-assisted waste-disposal system aboard an operational vessel will experience usage fluctuations throughout the day, as there are simply certain times when the crew is more likely to use the restrooms, i.e. at meal times or at watch changes. These fluctuations, which can be due both to deterministic and random effects, clearly impact the distribution for T_p. An approximate time-domain distribution for λ can be developed without conducting an exhaustive set of experiments. Based on observations of the data obtained aboard the *USCGC Seneca*, for instance, it was discovered that the average usage rate is different during each of the vessel's three watch shifts. Specifically, it is much lower at night when most of the crew is asleep. On that basis, the system simulation was modified to use the following model for $\lambda(t)$:

$$\lambda(t) = \begin{cases} \lambda_1, & 0\text{hr} \leq t < 8\text{hr} \\ \lambda_2, & 8\text{hr} \leq t < 16\text{hr} \\ \lambda_3. & 16\text{hr} \leq t < 24\text{hr} \end{cases} \tag{3.24}$$

where t is in units of hours and $t = 0$ is defined to be at midnight each day. In reality, Eq. 3.24 does not completely reflect all of the random variation in the usage rate, as it would be difficult, if not impossible, to do so. As will be shown, Eq. 3.24 is sufficient for developing powerful diagnostic indicators.

In order to demonstrate the more realistic behavior predicted by the model presented in this section, the ideal system simulation was modified and the simulated results were compared to measurements obtained by non-intrusively monitoring the *Seneca*'s waste-disposal system. The parameter values used in the simulation, which are presented in Table 3.2, were chosen following an analysis of the system aboard the *Seneca*. The histogram presented in the top plot in Fig. 3-12 shows the simulated frequency distribution for T_p over a five-day period; the bottom histogram shows the measured frequency distribution over an equivalent time interval. Note the qualitative agreement between the two histograms.

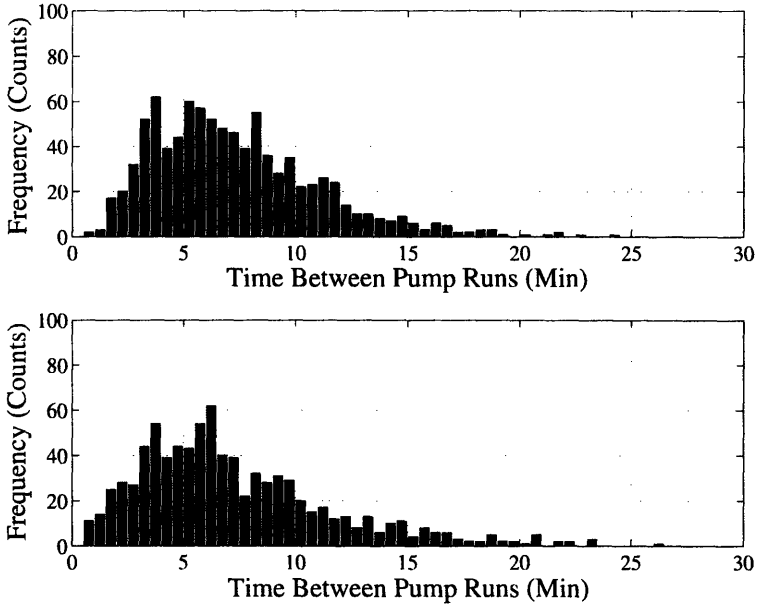


Figure 3-12: Simulated (top graph) and measured (bottom graph) frequency distributions for the discharge time T_p, under normal (i.e. no leak) operating conditions.

Table 3.2: Parameters used in the modified system simulation.

Parameter	Value
P_{high}	17.5 in Hg
P_{low}	13.5 in Hg
λ_1	24 hr^{-1}
λ_2	36 hr^{-1}
λ_3	38 hr^{-1}
ΔP	See Eq. 3.24

When a leak is introduced into the more realistic system, its effect on the distribution is quite different than it was in the ideal case. Specifically, the large pulses shown in Fig. 3-11 will not appear in this situation. There are several reasons for this change. Most importantly, the random variation in ΔP_k will effectively "smear" these pulses over several different bins. To see this, once again consider the ideal system used to generate Fig. 3-11. When a leak is introduced into that system, the pump is guaranteed to enter operation at time τ_1 as long as 2 SUEs occur prior to that time. Consider, however, what would happen if each usage event in that system did not remove the same amount of pressure from the reservoir. In that case, it is possible that exactly 2 SUEs could occur prior to time τ_1 and result in less pressure loss than the amount required to trigger the pump at that time. As a result, the system would continue to discharge for an additional period after τ_1. Similarly, it is also possible that exactly 2 SUEs could occur prior to time τ_1 and result in more pressure loss than the amount required to trigger the pump at that time. As a result, the pump will enter operation slightly earlier. Clearly, this behavior will effectively "spread out" the large pulses observed in the ideal case.

In the modified simulation considered in this section, the assumptions about the leak itself are also changed. Recall, for instance, that in the ideal system, it was assumed that the leak was a linear function of time. As was noted previously, that assumption was based on a first-order Taylor series expansion of the solution to Eq. 3.1. For additional realism, the simulation simply solves Eq. 3.1 directly. This change also causes the pulses observed in the ideal case to "spread" over several different histogram bins.

In order to determine the behavior of the model in the presence of a leak, several sets of both simulated and experimental data were obtained. For instance, Fig. 3-13 shows both simulated and measured frequency distributions for a leak with an initial rate equal to 12 in Hg/hr[5]. For the purpose of this comparison, a leak was manually inserted into the system using a small, adjustable flow meter.

Figure 3-14 presents another set of data. In that case, the experimental results were obtained during a period when the *Seneca*'s waste-disposal system had a faulty check valve that caused a very large leak. Although this type of failure causes a very noticeable change in the measured frequency distribution, it often goes undetected by the system operators for many days or weeks.

[5]In this context, the term "initial leak rate" refers to the value of the leak at $t = 0$, i.e. at the beginning of the discharge period. This distinction is made because the rate is now considered to be a function of time.

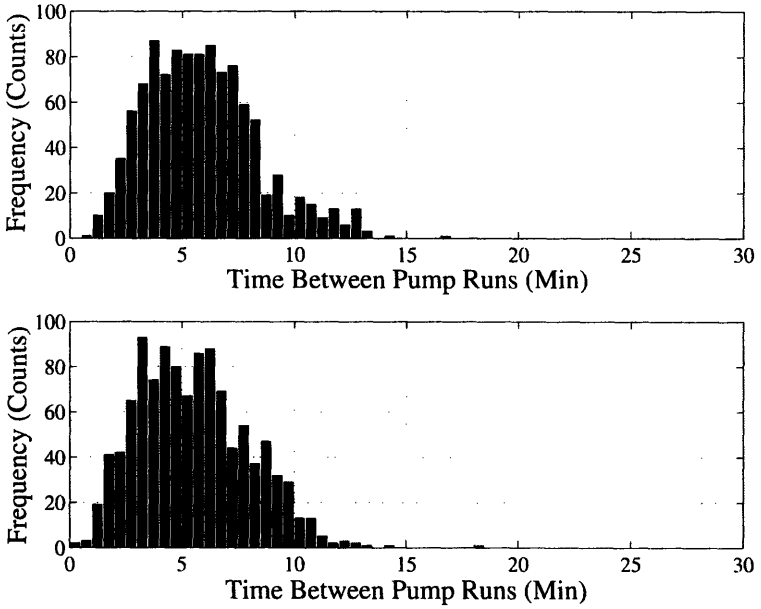


Figure 3-13: Simulated (top graph) and measured (bottom graph) frequency distributions for the discharge time in the presence of a small leak. Both plots contain data collected over a five-day period.

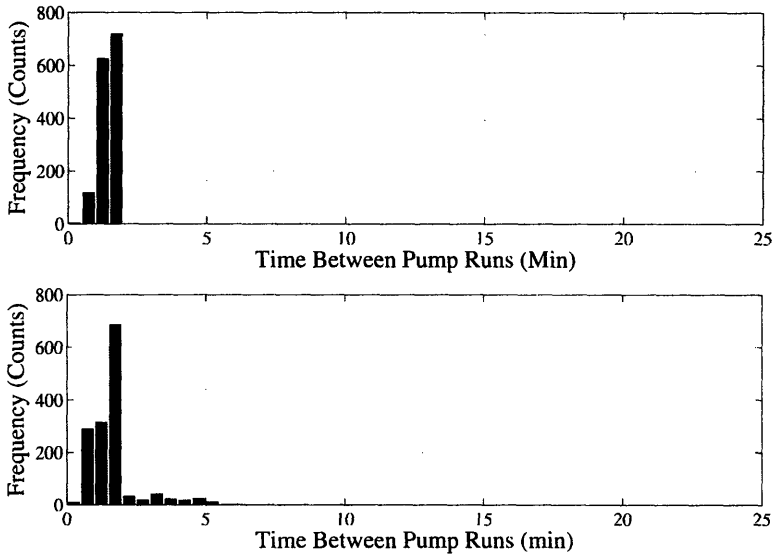


Figure 3-14: Simulated (top graph) and measured (bottom graph) frequency distributions for the discharge time, T_p, in the presence of a large leak. Both plots contain data collected over a five-day period.

As shown in Figs. 3-13 and 3-14, the simulated results obtained under leak conditions qualitatively agree with the real data obtained aboard the *Seneca*. Moreover, the results of the two leak cases considered here demonstrate that leaks do cause noticeable changes in the measured distribution. In the next section, several techniques are presented for detecting these changes.

3.1.3 Diagnostic Indicators for Vacuum-Assisted Waste-Disposal Systems

As suggested by the model developed previously, leaks in vacuum-assisted waste-disposal systems can be detected by analyzing the operating schedule of the system's pumping device. Given that the NILM can determine this operating schedule, it is in an excellent position to perform the required analysis. For this thesis, several non-intrusive leak detection methods have been developed and applied in the field. These methods are presented in this this section.

The diagnostic software currently installed on the NILM is implemented as shown in the flow chart in Fig. 3-15. The basic procedure is rather straightforward. At a pre-specified time each day, the diagnostic software polls the NILM database to obtain all of the on/off times recorded for the relevant pump over the past three days. Subsequently, this data is analyzed using three different techniques, each of which is discussed in detail below. Using the features generated by these three data analysis methods, the NILM decides whether or not a leak is present in the system. As discussed below, the decision-making process is implemented using two different routines. Once the NILM has finally assessed the current state of the system, it generates a status report that is both archived and sent to the system's operator. The remainder of this section presents both the data analysis procedures and the decision-making methods used by the NILM to detect fault conditions in waste-disposal systems.

Data Analysis Methods

As indicated by the model simulations presented previously, the size of a cycling-system leak determines how that leak affects the operating schedule of the system's actuator. In the case of waste-disposal systems, for instance, large leaks cause sudden, drastic changes, whereas small leaks cause more subtle variations. Because of these differences, the analytical techniques needed to detect small leaks are different than those needed to detect large ones.

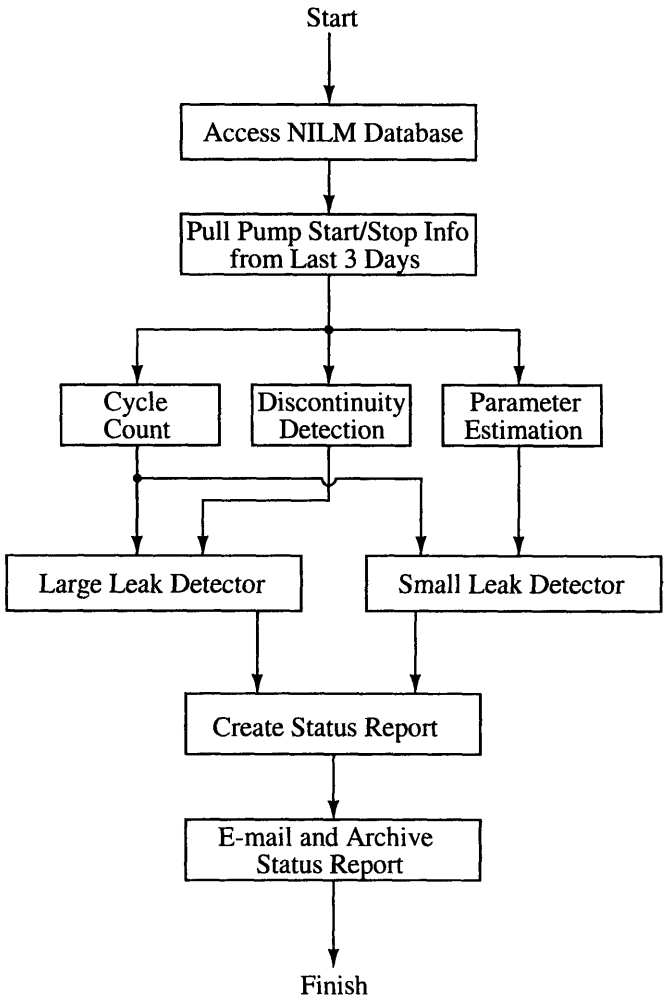


Figure 3-15: Fundamental work flow for the NILM's diagnostic software. The final e-mail is sent to a system operator. In the case of the *Seneca*, the email is sent to the Engineering Officer.

Accordingly, the NILM analyzes the operating schedule data using several different methods. In all, three procedures are employed, and each is discussed in detail below.

Cycle Count

One obvious method that can be used to detect changes in the state of the system is to count how many times the pump has operated during a fixed time interval. This procedure cannot definitively detect leaks, however, as both leaks and increased system usage both affect the frequency of pump operation. Still, this method can be used as a "first-pass" indicator that triggers the use of other, more sophisticated techniques that can determine the difference..

On any given day, the NILM determines two different run count values. The first of these is the number of times that the pump has operated in the past 24 hours. For convenience, the evolution of this one-day run count is viewed as a discrete-time signal $a[n]$, in which each data point reflects the number of runs observed on day n. Additionally, the NILM determines a three-day run count that reflects how many times the pump has operated in the past three days. Again, the daily evolution of this value is also viewed as a discrete-time signal. Thus, on day n, the value of the three-day run count, which is denoted as $b[n]$, is given by the relation

$$b[n] = a[n] + a[n-1] + a[n-2]. \tag{3.25}$$

As will be demonstrated below, the reason for computing two different run count values is that each is useful for detecting a different type of leak. In particular, the one-day count, $a[n]$, is useful for detecting large leaks, and the three-day count, $b[n]$, is useful for detecting small leaks[6]. The application of these two signals will be considered in the section on the NILM's two decision-making routines.

Discontinuity Detection

As indicated by the system model, more sophisticated methods for leak detection should make use of the measured distribution for the discharge time. For instance, in the event that the system develops a leak as large as the one considered in Fig. 3-14, the effect will be the creation of a sharp edge in the measured histogram. From a diagnostic perspective,

[6]Note that $b[n]$ is essentially a low-pass filtered version of $a[n]$.

such features are useful because they make it possible to discern the difference between a leak and a simple change in usage.

The early detection of leaks as large as the one used to generate Fig. 3-14 is desirable, as they cause excessive wear and increased energy consumption. One way to detect such leaks is to search for the corresponding edge in the measured histogram. In order to perform this operation, the histogram, which contains data collected over the last three days, is viewed as a discrete signal $x[n]$. In this context, n is the number of a given histogram bin and $x[n]$ is the value of the measured histogram in that bin. In order to locate sharp edges, $x[n]$ is passed through a five-point median filter and the results are stored in the vector $y[n]$. By definition, such filters determine the median of a given input signal over a sliding window of length $2N + 1$ ([69]). In the case of the five-point median filter (i.e. $2N + 1 = 5$), this implies that ([69])

$$y[n] = \text{median of}\,(x[n-2], \quad x[n-1], \quad x[n], \quad x[n+1], \quad x[n+2])\,. \tag{3.26}$$

The final step in the edge detection process is to compute $d[n] = |y[n] - x[n]|$. This signal is passed to the large leak detection module, which determines if the difference $d[n]$ exceeds a preset threshold at any value of n. The thresholding process is described in the section on the NILM's two decision-making routines[7].

It is important to note that in edge detection problems such as this one, median filtering is often more attractive than linear filtering. In particular, linear filters, which are frequency oriented, will "smooth out" sharp edges such the one shown in Fig. 3-14. By comparison, median filters are geometrically oriented, and they "pass" locally monotonic features such as step edges, while attenuating impulsive noise ([69]).

Parameter Estimation

As suggested by the results presented in Fig. 3-13, small leaks also distort the distribution for the discharge time. Clearly, however, the distortion is not as visually distinct as it is when the system develops a very large leak. In order to detect these small, but potentially crippling fault conditions, other methods must be employed.

To aid in the detection of small leaks, the NILM is provided with a mathematical model

[7]Note that at the ends of the finite length signal $x[n]$, the sliding window in Eq. 3.26 must be padded. Typical methods are to add zeros or to repeat the first and last values as needed. ([69])

for the distribution of the discharge times. Each day, the NILM uses data collected over the previous three days to estimate the parameters of that model. Based on the simulated results and physical reasoning presented in Section 3.2.3, the discharge times are assumed to be gamma distributed. This two parameter distribution, which is a generalization of the ideal PDF presented in Eq. 3.15, is ([70])

$$f_{T_p}(t;k,\lambda) = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{\Gamma(k)} \tag{3.27}$$

where

$$\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt. \tag{3.28}$$

In the NILM's diagnostic software, the values of the parameters k and λ are estimated using the method of maximum likelihood ([71], [72]). Provided that a joint PDF can be written for the observations of the random variable of interest, this approach can be applied to any type of estimation problem ([71]). The fundamental ideas underlying the method can be described as follows. In a general context, suppose that there are n observations $X_1, ..., X_n$ of some random variable X. Furthermore, assume that this variable has a PDF of the form $f(x;\Theta)$, where Θ is a set of parameters that describe the shape, scale, and position of the distribution ([71]). If the complete set of observations is denoted by the vector $\mathbf{X} = (X_1, ..., X_n)$, then the joint PDF of $\mathbf{X}$ is ([71])

$$f(\mathbf{x};\Theta) = f(x_1;\Theta)\, f(x_2;\Theta) \cdots f(\mathbf{x_n};\Theta)\,. \tag{3.29}$$

In the context of probability theory, this joint PDF is known as a *likelihood function* of Θ. The meaning of this name can be understood by considering a particular set of observations denoted as $\mathbf{x}$. If

$$f(\mathbf{x};\Theta_1) > f(\mathbf{x};\Theta_2)\,, \tag{3.30}$$

then the implication is that Θ_1 is a more plausible set of parameters than Θ_2 because Θ_1 ascribes a greater probability to the observations. In estimation by the method of maximum likelihood, this property is exploited. Specifically, we choose the parameter set Θ that is most plausible given the observations, i.e. we choose the set $\hat{\Theta}$ that maximizes the likelihood function ([71]). Stated mathematically,

$$\hat{\Theta} = \arg\max f(\mathbf{x};\Theta). \tag{3.31}$$

In practice, the values of the elements in the parameter set $\hat{\Theta}$ are chosen by solving the set of equations

$$\begin{aligned} \frac{\partial f}{\partial \theta_1} &= 0 \\ &\vdots \\ \frac{\partial f}{\partial \theta_n} &= 0. \end{aligned}$$

Returning to the task at hand, assume that n observations of the gamma-distributed variate T_p are recorded over the NILM's three-day analysis window. In that case, the likelihood function $f(\mathbf{t};k,\lambda)$ is

$$\begin{aligned} f(\mathbf{t};k,\lambda) &= \prod_{i=1}^{n} \frac{\lambda^k t_i^{k-1} e^{-\lambda t_i}}{\Gamma(k)} \\ &= \frac{\lambda^{nk}}{(\Gamma(k))^n} \left(\prod_{i=1}^{n} t_i^{k-1} \right) e^{-\lambda \sum_{i=1}^{i=n} t_i}. \end{aligned} \tag{3.32}$$

To determine appropriate estimates for k and λ, this function must be maximized with respect to both parameters. For implementation purposes, it is much easier to perform this operation using the natural logarithm of the likelihood function[8], i.e.

$$\ln\left(f(\mathbf{t};k,\lambda)\right) = nk\ln(\lambda) - n\ln\left(\Gamma(k)\right) + (k-1)\sum_{i=1}^{n}\ln t_i - \lambda\sum_{i=1}^{n} t_i. \tag{3.33}$$

To maximize Eq. 3.33, we take the partial derivatives:

$$\frac{\partial}{\partial\lambda}\ln\left(f(\mathbf{t};k,\lambda)\right) = \frac{nk}{\lambda} - \sum_{i=1}^{n} t_i = 0 \tag{3.34}$$

$$\Longrightarrow \hat{\lambda} = \frac{k}{\frac{1}{n}\sum_{i=1}^{n} t_i} \tag{3.35}$$

[8]Note that the location of the maximum value of a function f is the same as the location of the maximum value of the function $\ln(f)$ ([72]).

and

$$\frac{\partial}{\partial k}\ln\left(f(\mathbf{t};k,\lambda)\right) = n\ln(\lambda) - n\Psi(k) + \sum_{i=1}^{n}\ln(t_i) = 0 \tag{3.36}$$

$$\Longrightarrow \ln(\hat{k}) - \ln\left(\frac{1}{n}\sum_{i=1}^{n}t_i\right) - \Psi(\hat{k}) + \frac{1}{n}\sum_{i=1}^{n}\ln(t_i) = 0, \tag{3.37}$$

where $\Psi(x)$ is the derivative of the natural logarithm of the Gamma function ([73]), i.e.

$$\Psi(x) = \frac{d}{dx}\ln\Gamma(x). \tag{3.38}$$

Note that Eq. 3.37 was derived by substituting Eq. 3.35 into Eq. 3.36.

To make use of the above results, a two-step procedure is typically employed. First, Eq. 3.37 is solved numerically to obtain $\hat{k}$. Subsequently, that value is used in Eq. 3.35 to compute $\hat{\lambda}$ ([70], [73]). The software used here is the MATLAB function gamfit.m, which employs the above method.

As will be shown in Section 3.2.4, the estimates of both k and λ increase when leaks are inserted into the real and simulated systems. Furthermore, similar trends are not observed when the usage rate increases. It is important to note, however, that in this general context, the parameter λ does not refer to the system's usage rate; rather, it is simply an indicator of the scale of the measured distribution ([70]). When the distribution narrows as it does under leak conditions, the scale changes. As a result, a change in λ is expected.

Note that when the usage rate increases, the only observed change occurs in the estimate of λ. Physically, this makes sense. To see this, recall the idealized system model presented in Section 3.2.2. As shown in that section, the system does tend to follow the idealized model under non-faulty conditions. In that model, a change in the usage rate will not affect k since its value is set by the number of SUEs required to trigger the pumps. Although this explanation is by no means a rigorous proof, it is a likely explanation for the behavior to be demonstrated in Section 3.2.4.

Decision Methods

Once the operating schedule of the pump has been analyzed using the three procedures described above, the NILM must use the generated features to determine whether or not a leak is present in the system. In order to make that determination, two different decision-

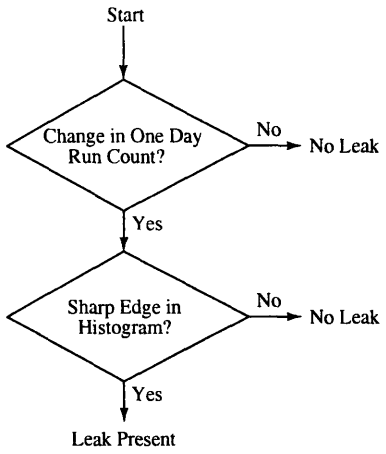


Figure 3-16: Fundamental work flow in the NILM's large leak detection software. The details of the two decision elements are explained in the text.

making routines are used. If a leak is detected by either method, the NILM will alert the system operator in its daily status report.

Large Leak Detector

The first decision-making routine contained in the NILM's diagnostic software package determines if any large leaks are present in the system. As shown in Fig. 3-16, the NILM detects these leaks by proceeding in two steps. First, it determines if there has been a significant change in the number of times that the pump has operated over the past 24 hours. To do so, the NILM checks whether or not the one-day run count value, $a[n]$, exceeds the average recorded over the past three days. Mathematically, this implies that the NILM computes the following:

$$a[n] - \frac{1}{3}\left(a[n-1] + a[n-2] + a[n-3]\right). \tag{3.39}$$

If the corresponding result exceeds a particular threshold, the next step is to check the output of the discontinuity detector (i.e. $d[n] = |y[n] - x[n]|$) to see if the histogram has developed any sharp edges. If the difference $d[n]$ exceeds a threshold at any given histogram bin, the NILM immediately alerts the operator. In addition to an assessment of the state of the system, the status report also includes plots of the measured histogram and the time history of the daily run count.

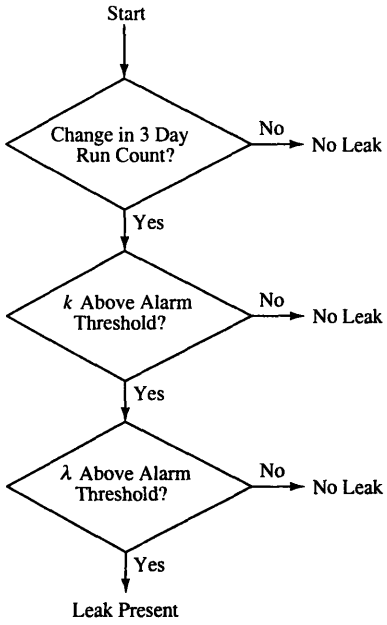


Figure 3-17: Fundamental work flow in the NILM's small leak detection software. The details of the two decision elements are explained in the text.

Small Leak Detector

The second decision-making routine contained in the NILM's diagnostic software package determines if any small leaks are present. In order to detect these leaks, the NILM proceeds as shown in Fig. 3-17. First, it uses the three-day cycle count, $b[n]$, to determine if there has been a change in the number of times that the pump has operated over the past several days. To do so, a simple change-of-mean detector is again employed ([9]). In this case, the detector computes the difference between today's three-day run count value and the average recorded over the past three days, i.e.

$$b[n] - \frac{1}{5}\left(b[n-1] + b[n-2] + b[n-3] + b[n-4] + b[n-5]\right). \tag{3.40}$$

If the corresponding result exceeds a threshold, then the NILM checks whether or not the parameter estimates $\hat{k}$ and $\hat{\lambda}$ exceed their own preset thresholds. If so, the daily status report will indicate that a leak is present in the system. In order to assist the operator, the daily status report always includes the measured histogram as well as the time histories of both the three-day run count and the parameter estimates.

On a day that a change is detected in the number of pump runs, the five-day average

of $b[n]$ is saved to a variable known as *oldAverage*. On subsequent days, the NILM will compute

$$b[n] - oldAverage \tag{3.41}$$

instead of Eq. 3.40. This modification has been found to provide the NILM an improved ability to detect small leaks.

3.1.4 Results

The diagnostic methods discussed above were tested using both simulated and measured data. The test bed for the field experiments was the waste-disposal system aboard the *USCGC Seneca*. Part of that system is shown in the photograph in Fig. 3-18. For full realism, tests were conducted while the ship was at sea.

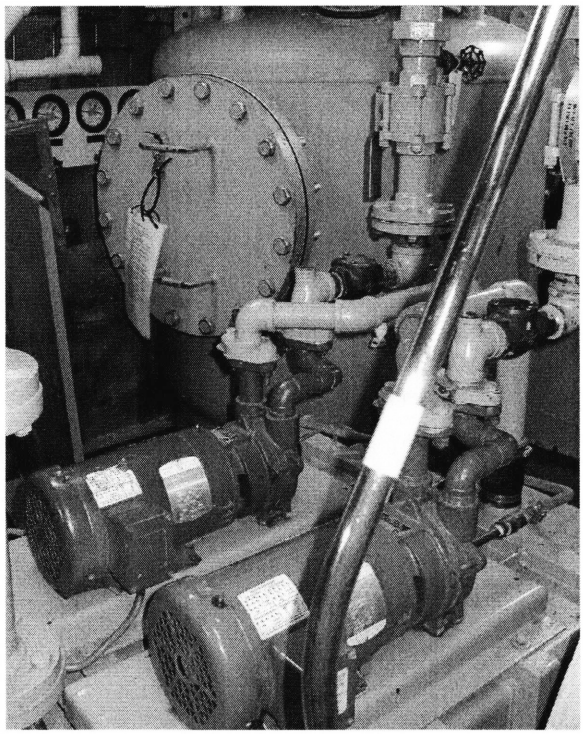


Figure 3-18: The waste-disposal system aboard the *Seneca*. The waste-collection tank is in the background; the vacuum discharge pumps are in the foreground. A NILM is installed on the upstream side of the electrical panel serving both the vacuum pumps and two other pumps.

In order to demonstrate the full capabilities of the non-intrusive leak detection methods, this section presents results computed using four representative data sets. The first two of

these sets were generated using the simulation, and the second two were generated using field measurements. In both the simulated and real cases, there is one data set that is used to test the large leak detector and another that is used to test the small leak detector.

Test Case 1: Large Leak Detector, Simulated Data

To test the large leak indicator, it was applied to a nine-day simulated data set consisting of eight days under non-faulty conditions and one day under faulty conditions. Figure 3-19a shows both the histogram created on day 9 and the corresponding discontinuity detector output. Additionally, Fig. 3-19b displays the daily evolution of the one-day run count. In this case, both $d[n]$ and the change-of-mean detector exceed their thresholds, thus signalling a leak. For reference, the threshold for $d[n]$ is set to 150, and the threshold for the change-of-mean detector is set to 100.

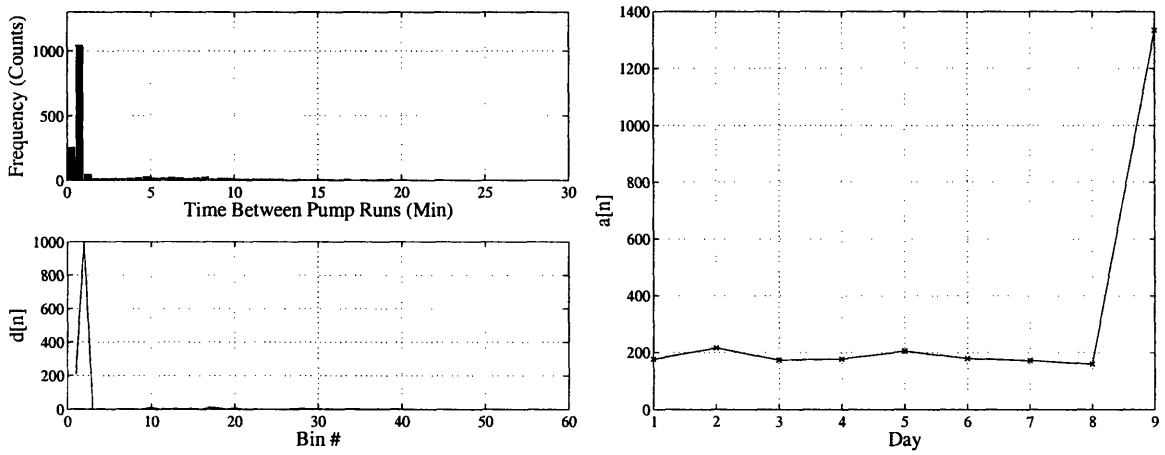


Figure 3-19: (a) Histogram and discontinuity detector output on the first day of a large leak. (b) One-day run count prior to the development of the large leak. These plots were generated using simulated data.

Test Case 2: Small Leak Detector, Simulated Data

To test the small leak detection method, the simulation was used to generate 34 days of synthetic data. For test purposes, this set includes both faulty and non-faulty conditions. In particular, a small leak was inserted into the simulation between days 9 and 13, and the usage rate was elevated between days 22 and 26. The daily evolutions of both the the

three-day cycle count and the parameter estimates are shown in Fig. 3-20. In addition, that figure also includes a binary signal that is set to one on each day that a leak is detected. Note that the NILM does not react during the period of elevated usage, despite the fact that the number of runs is relatively high during that period. For reference, the alarm thresholds for k and λ are set at 5 and 1 Min^{-1}, respectively. The threshold for the change-of-mean detector is 60.

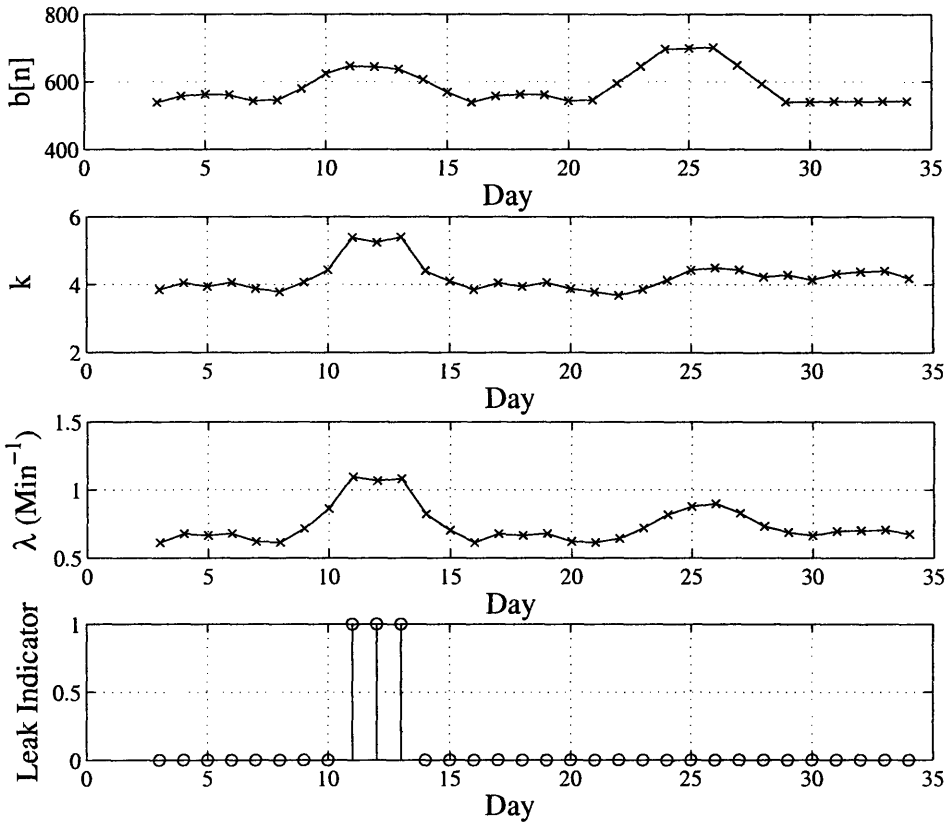


Figure 3-20: Evolutions of the three-day run count, the parameter estimates, and the indicator decision over 34 days of simulated data. Note that a leak was inserted between days 9 and 13, and that the usage rate was elevated between days 22 and 26. The "Leak Indicator" is a binary signal that is set to 1 when a leak is detected.

Test Case 3: Large Leak Detector, Field Data

In the field, the large leak detection scheme has been used to sense actual fault conditions. In particular, the vacuum-assisted waste-disposal system aboard the *Seneca* often experiences failures of the check valves in the suction ports of the vacuum pumps. During preliminary

investigations of the *Seneca*'s system, it was discovered that these failures can cause the pumps to run more than 1000 times in a given day. By comparison, a typical run count is approximately 200. Moreover, these conditions often go unnoticed for several weeks, which can have a deleterious effect on the pump.

Figure 3-21 shows the three plots generated by the NILM's large leak detector on a day that a check valve failed aboard the *Seneca*. Following the detection, the Engineering Officer was alerted, and the valve was replaced. Figure 3-22 shows the state of the faulty valve following its removal from the system. Note the uneven wear on the rubber face.

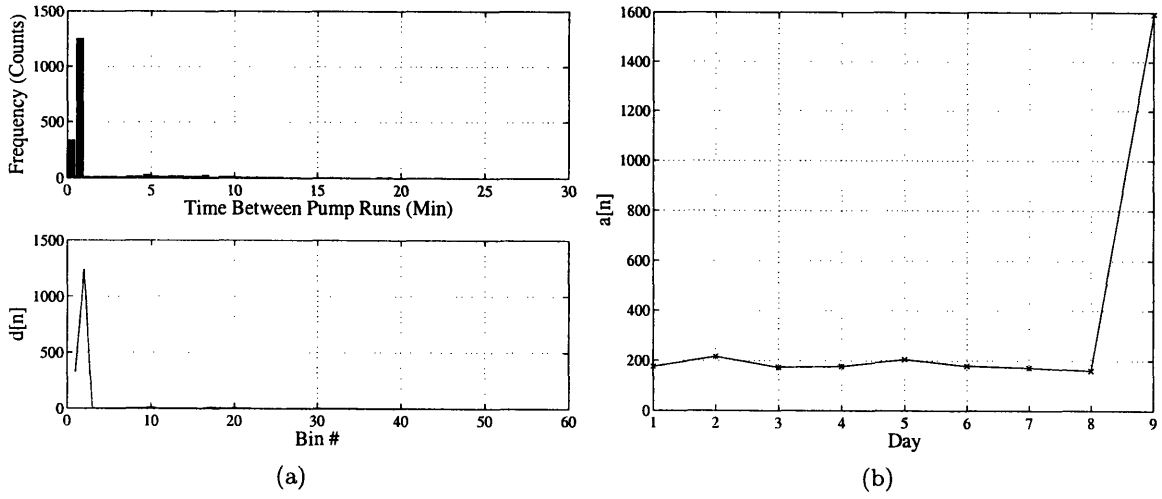


Figure 3-21: (a) Histogram and discontinuity detector output on the first day of a large leak. (b) One-day run count prior to the development of the large leak. These plots were generated using real data.

Test Case 4: Small Leak Detector, Field Data

The small leak detection procedure has also been tested aboard the *Seneca*. To test this scheme, the crew inserted several small leaks into the system using an adjustable flow meter that was provided to them. Once the ship returned to port, an 80-day data set was constructed. That set contains three different leak periods interspersed between periods with no leaks. During the first two sets of leak conditions, the flow meter was set at approximately 30 SCFH, and in the third set, the flow meter was throttled to approximately 60 SCFH. The corresponding three-day run count and parameter estimates are shown in Fig. 3-23. Accompanying those values is a binary signal that is set to one on each day

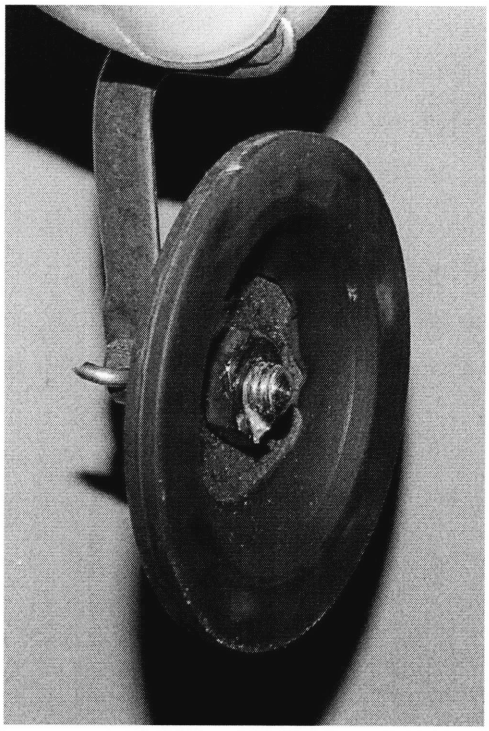


Figure 3-22: Check valve removed from the *Seneca*'s waste-disposal system following the detection of a leak by a NILM. Note the uneven wear on the rubber face.

that a leak is detected. For reference, the alarm thresholds for k and λ are set at 4.2 and 0.7 Min^{-1}, respectively. The threshold for the change-of-mean detector is 60.

It is important to note that the measured data also displays a slight amount of usage variation. In particular, the change-of-mean detector noticed a change in the three-day cycle count on day 8. Since the parameter estimates never exceeded their alarm levels, however, this change was ignored.

3.2 Detection of Coupling Failure in Induction Machine Applications

Numerous real world systems require devices that are mechanically linked to a motor via a coupling or shaft. The failure of such couplings can have a drastic impact, especially in large interconnected systems. Consider, for example, the auxiliary sea water (ASW) system aboard a marine vessel. On many ships, these systems are driven by centrifugal pumps that are linked to a large induction machine via a flexible coupling. Since the ASW system

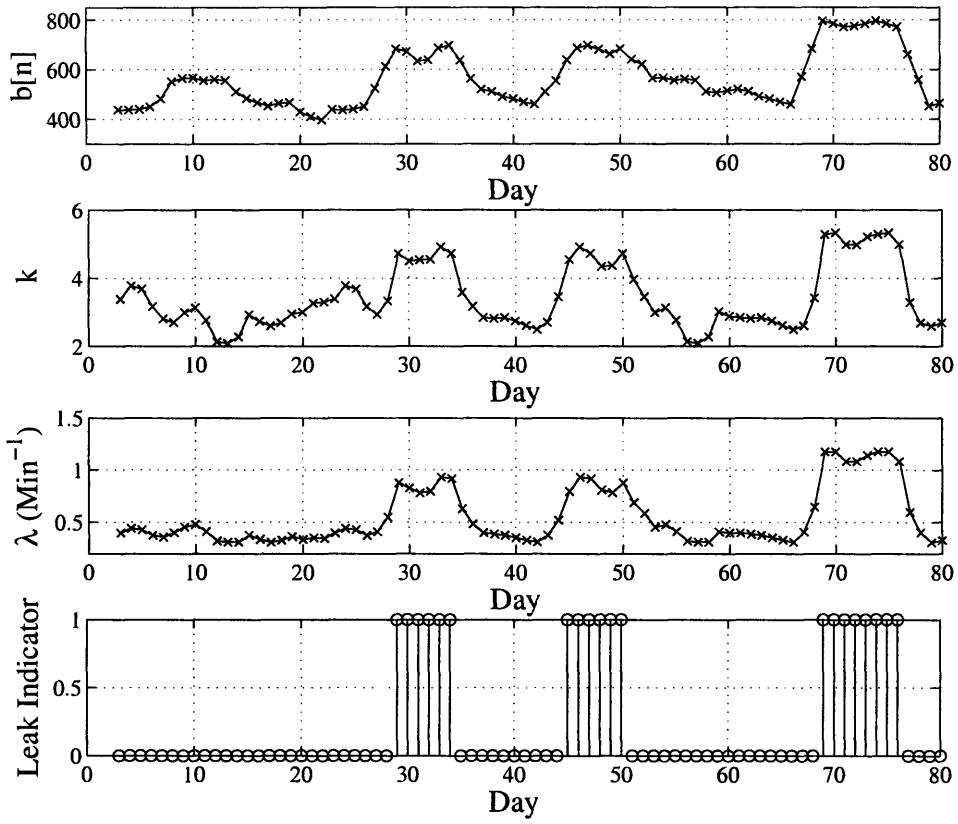


Figure 3-23: Evolutions of the three-day run count, the parameter estimates, and the indicator decision over 80 days of real data. Note that leaks were inserted into the system between days 27 and 34, days 43 and 50, and days 67 and 76. The "Leak Indicator" is a binary signal that is set to 1 when a leak is detected.

supplies cooling water to vital heat loads such as the ship's service generators, failure of this coupling can result in mission-crippling damage ([35]). As shown in this section, a NILM can diagnose such failures before they occur.

The coupling failure described above has been investigated by several researchers who have shown that the oscillatory system formed by the motor and its load can have a natural torsional frequency in the same range as the oscillating components of the starting torque of the machine ([74], [75]). As a result, large torsional torques can build on the coupling, causing serious fatigue and wear. After many starts, this wear may be significant enough to tear the coupling apart.

In this section, a simulation model is used to demonstrate the coupling fatigue phenomenon. As shown, the slow wear manifests itself in measurements of the stator current.

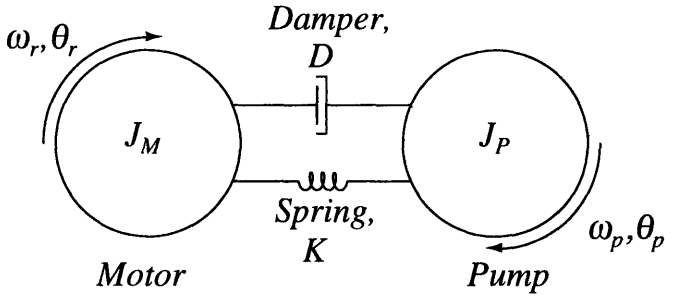


Figure 3-24: Model for the mechanical system. The pump and motor are considered to be separate inertias, and the coupling linking them is a combination of a damper and a spring. The mechanical speeds and displacements of both the pump and the rotor of the machine are shown.

This observation is then used to develop a non-intrusive diagnostic indicator that is demonstrated using field data obtained aboard the *USCGC Seneca*.

3.2.1 Simulation Model

A critical requirement of the simulation developed to study mechanical coupling failure is the inclusion of a model that describes the torsional dynamics of the mechanical system. The model used here, which has been employed by other researchers ([74], [75]), describes an induction machine driving an inertial load through a flexible coupling having a damping coefficient, D, and a spring stiffness factor, K. For purposes of the simulation, the damper and the spring are assumed to have negligible moments of inertia. The model of the mechanical system is shown in Fig. 3-24.

A set of state equations can be developed to describe the relationships between the various components of the mechanical model. These equations are based on the constitutive relations for torsional dampers and springs. In terms of the variables presented in Fig. 3-24, the torque across a torsional damper can be written in the form ([76])

$$\tau = D\left(\theta_r - \theta_p\right), \tag{3.42}$$

and the torque across a rotating spring can be written in the form ([76])

$$\tau = K\left(\omega_r - \omega_p\right). \tag{3.43}$$

Using these relations in combination with the electromagnetic torque, τ_e, the state equations for the mechanical model can written as shown in Eq. 3.44.

$$\frac{\partial}{\partial t}\begin{bmatrix}\omega_r\\ \theta_r\\ \omega_p\\ \theta_p\end{bmatrix}=\begin{bmatrix}-\frac{D}{J_m} & -\frac{K}{J_m} & \frac{D}{J_m} & \frac{K}{J_m}\\ 1 & 0 & 0 & 0\\ \frac{D}{J_p} & \frac{K}{J_p} & -\frac{D}{J_p} & -\frac{K}{J_p}\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}\omega_r\\ \theta_r\\ \omega_p\\ \theta_p\end{bmatrix}+\begin{bmatrix}\frac{\tau_e}{J_m}\\ 0\\ -\frac{\tau_l}{J_p}\\ 0\end{bmatrix} \tag{3.44}$$

In the simulation model, the load torque, τ_l, is assumed to be of the form $\beta\omega_p^2$. This torque models the damping force exerted on the pump impeller by the fluid ([9]).

A complete simulation model was created by combining the mechanical state equations presented above with the electrical state equations of the induction machine. In the synchronously rotating d-q reference frame[9], these equations are[10] ([77])

$$\frac{\partial\lambda_{ds}}{\partial t}=v_{ds}+\omega\lambda_{qs}-r_s i_{ds} \tag{3.45}$$

$$\frac{\partial\lambda_{qs}}{\partial t}=v_{qs}-\omega\lambda_{ds}-r_s i_{qs}, \tag{3.46}$$

$$\frac{\partial\lambda_{dr}}{\partial t}=v_{dr}+(\omega-p\omega_r)\,\lambda_{qr}-r_r i_{dr}, \tag{3.47}$$

and

$$\frac{\partial\lambda_{qr}}{\partial t}=v_{qr}-(\omega-p\omega_r)\,\lambda_{dr}-r_r i_{qr}, \tag{3.48}$$

where

$$\tau_e=\left(\frac{3}{2}\right)\left(\frac{p}{2}\right)(\lambda_{ds}i_{qs}-\lambda_{qs}i_{ds})\,. \tag{3.49}$$

Simulation Results

The full 8th order model described above was simulated in MATLAB using the parameters presented in [77] for a representative 50 hp induction machine. These parameters are presented in Table 3.3.

Based on both physical intuition and an investigation of several failed couplings found in the field, it was decided that the best way to simulate progressive coupling failure is to decrease both D and K from some nominal set of values. To do so, several different sets

[9]The synchronously rotating reference frame is the most useful for dynamic simulations. This frame is rotating at the electrical angular velocity, ω_e ([77]).

[10]A complete description of the nomenclature used here can be found in Appendix C.

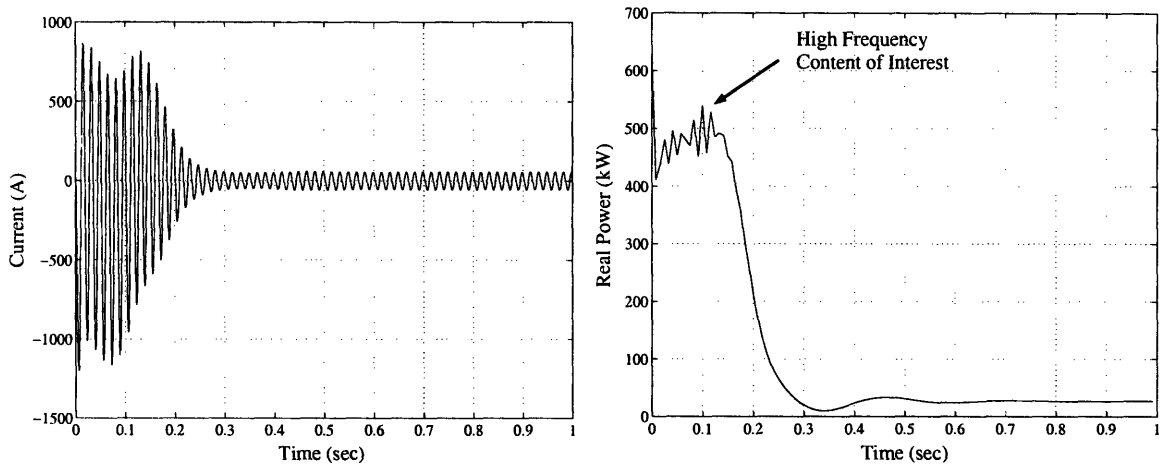
Table 3.3: Parameter values in the induction machine model.

Parameter	Value
r_s	0.087 Ω
X_{ls}	0.302 Ω
X_m	13.08 Ω
X_{lr}	0.302 Ω
r_r	0.228 Ω
J_m	1.662 kg · m^2
J_p	0.3 kg · m^2
p	2
β	4×10^{-3} N · m · s^2/rad^2
ω_e	377 rad/sec
D	Varying (see text)
K	Varying (see text)

of simulations were performed. In the first of these, the value of the damping coefficient was varied while the spring stiffness was held constant. In the second set, the stiffness was reduced while the damping was held constant. Finally, a third set of simulations was conducted in which both variables were varied. The nominal values for D and K were chosen so that the coupling in the simulation would be appropriate for a 50 hp motor driving an inertial load. A manufacturer data sheet was used to select representative values (i.e. $D = 2$ N·m·s/rad and $K = 200$ N·m/rad).

Figure 3-25a shows a simulated stator current waveform obtained with D and K equal to their nominal values. Plotted alongside the current in Fig. 3-25b is the corresponding P_1 spectral envelope. Notice the high frequency oscillations that occur in P_1 during the starting period. After each simulation, the frequency content in this region was analyzed using the discrete Fourier transform (DFT). To perform this analysis, a 16-point segment was extracted for the starting period of each P_1 waveform as shown in Fig. 3-25b. This segment was multiplied by a 16-point Hanning window, and the 128-point DFT of the result was computed according to the equation[11]

[11] The DFT $X[k]$ contains N samples of the discrete-time Fourier transform (DTFT) of the windowed signal



(a) (b)

Figure 3-25: (a) Simulated stator current transient and (b) corresponding P_1 spectral envelope. Note that the current is sampled at 8 kHz and that P_1 is sampled at 120 Hz.

$$X[k] = \sum_{n=0}^{127} x[n]\, w[n]\, e^{j\frac{2\pi kn}{128}}, \tag{3.50}$$

where $x[n]$ is the zero-padded data vector and $w[n]$ is the zero-padded Hanning window defined as

$$w[n] = \begin{cases} 0.5 - 0.5\cos(2\pi n/15), & 0 \le n \le 15, \\ 0, & \text{otherwise.} \end{cases} \tag{3.51}$$

Figure 3-26 shows the interpolated magnitude of the DFT samples obtained for each of the simulations performed with $K = 200$ N·m/rad. Note that the magnitude of the peak observed at approximately 50 Hz increases as the value of the damping coefficient decreases. For these simulations, D was reduced from 2 N·m·s/rad to 0.5 N·m·s/rad in increments of 0.5 N·m·s/rad.

Results similar to those shown in Fig. 3-26 were obtained when D was held constant and K was reduced from its nominal value. Figure 3-27 shows the magnitude of the frequency

$x[n]w[n]$, which is denoted as $X(e^{j\omega})$. These samples are located at the radian frequencies $\omega_k = 2\pi k/128$. Thus, the samples also correspond to the continuous time frequencies $f_k = kf_s/128$, where f_s is the NILM's sampling rate ([54]).

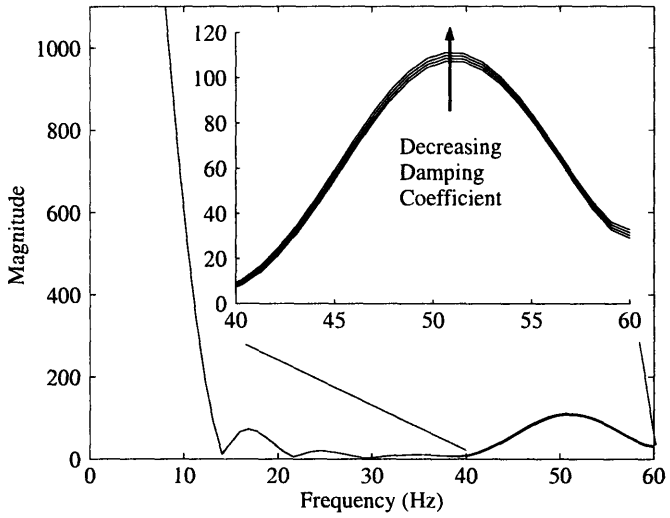


Figure 3-26: Magnitude of the frequency spectrum of the transient portion of P_1 as the damping coefficient decreases. The inset plot shows the details in the area of the spectral peak located in the vicinity of 50 Hz.

spectra obtained when K was varied from 200 N·m/rad to 50 N·m/rad in 50 N·m/rad increments. Physically, these results imply that the amplitude of the high frequency oscillation observed in P_1 increases as the coupling loses stiffness.

In the third and final set of simulations, both D and K were varied. The surface plotted in Fig. 3-28 demonstrates how the magnitude of the 50 Hz spectral peak behaves as a function of these two parameters. As shown, a stiff coupling with a high damping coefficient excites lower amplitude oscillations than a weak coupling with a low damping coefficient.

3.2.2 Diagnostic Indicator

The simulated data presented above suggests that a deteriorating coupling can be detected by trending the behavior of the amplitude of the high frequency oscillations in P_1. To perform the required analysis, the following three step procedure is proposed.

DFT Computation

First, a 16-point segment is extracted from the real power waveform drawn during a motor start. Subsequently, this sequence is mixed with a Hanning window and the DFT, denoted as $X[k]$, is computed using Eq. 3.50. It should be noted that the choice of the window size

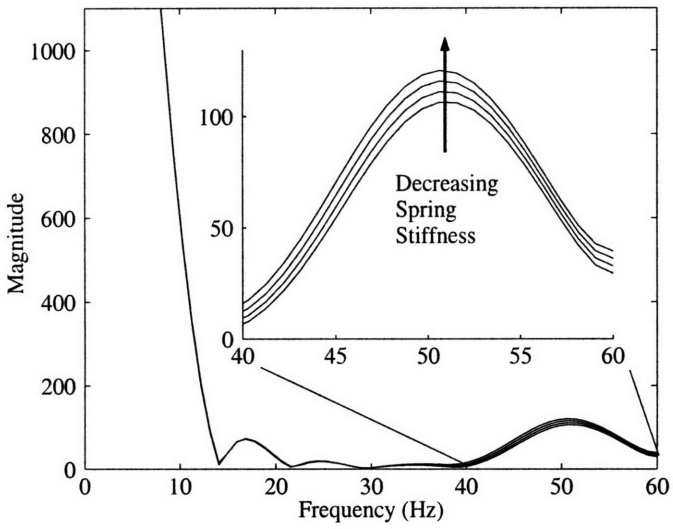


Figure 3-27: Magnitude of the frequency spectrum of the transient portion of P_1 as the spring stiffness decreases. The inset plot shows the details in the area of the spectral peak located in the vicinity of 50 Hz.

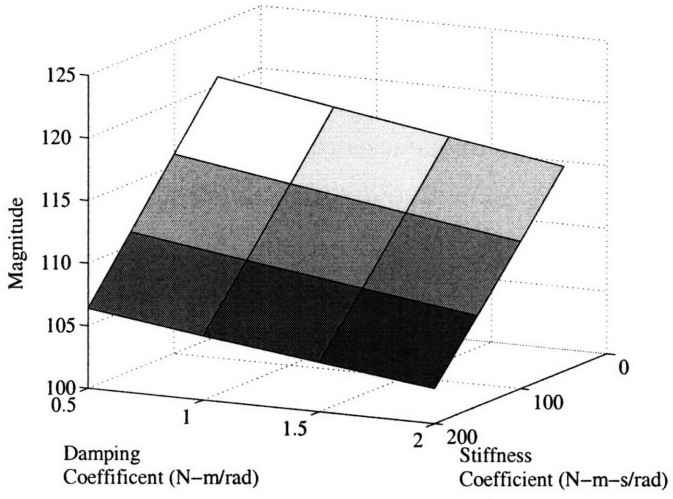


Figure 3-28: Surface plot showing the magnitude of the high frequency spectral peak for several different combinations of D and K.

is critical. The main reason for this is related to the well known inability of the Fourier transform to localize in the time domain ([78]). Specifically, since the oscillations of interest occur only during the beginning of the transient, the selection of a large analysis window will attenuate the spectral peak of interest. Thus, the size of the window is a variable whose value may change slightly from one application to the next. As shown in Section 3.3.3, however, the choice of a 16-point window has been found sufficient in the field for a motor with a rating similar to that used in the simulations.

Peak Detection

Next, the DFT samples spanning the frequency band from 30 Hz to 60 Hz are searched to find the location of the largest spectral peak in that band. In radian frequency, the location of this peak is denoted as ω_{pk}. For convenience, this value can be related to the corresponding continuous time frequency, f_{pk}, via the relation $\omega_{pk} = 2\pi\,(f_{pk}/f_s)$, where f_s is the sampling frequency of the spectral envelopes (i.e. 120 Hz).

Note that from a noise perspective, a more robust procedure for the location of the spectral peak would make use of spectral estimation methods ([71], [79]). In both the simulation and test cases considered thus far, however, the signal-to-noise ratio has been large enough that this was not necessary.

Diagnostic Indicator Computation

Finally, the energy in a band of frequencies centered around ω_{pk} is computed. Assuming that this band spans from $\omega_{pk} - \Delta\omega$ to $\omega_{pk} + \Delta\omega$, the total energy it contains is

$$E = \frac{1}{2\pi} \int_{\omega_{pk}-\Delta\omega}^{\omega_{pk}+\Delta\omega} \left|X\left(e^{\jmath\omega}\right)\right|^2 d\omega, \tag{3.52}$$

where $X\left(e^{\jmath\omega}\right)$ is the discrete-time Fourier transform (DTFT) of the windowed sequence $x[n]w[n]$ ([54]). Since the DFT $X[k]$ contains samples of $X\left(e^{\jmath\omega}\right)$ at the radian frequencies

$$\omega_k = \frac{2\pi k}{128}, \tag{3.53}$$

an estimate is calculated using the DFT samples ([54]). The estimated energy value is the diagnostic feature that is trended over time.

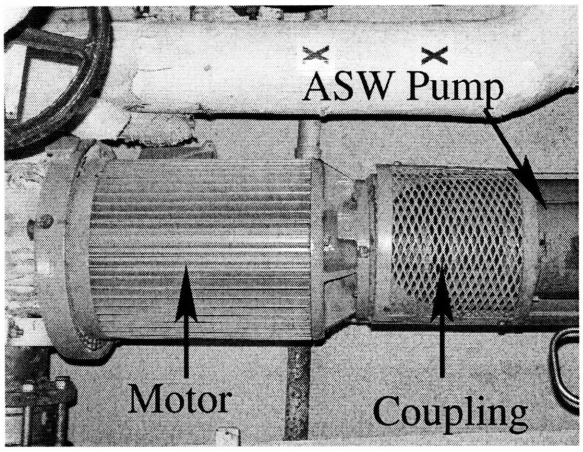


Figure 3-29: One of the two Auxiliary Sea Water pump/motor combinations aboard the *Seneca*.

3.2.3 Field Results

The diagnostic method developed above was tested in the ASW system aboard the *USCGC Seneca*. This system has two 40 hp induction machines that drive centrifugal pumps through flexible couplings. One of the two pump/motor combinations is shown in the photograph in Fig. 3-29. For test purposes, a NILM was installed on the upstream side of the electrical panel servicing both motors.

Figure 3-30a shows the P_1 spectral envelope obtained during a start of one of the two induction machines in the *Seneca*'s ASW system. Just as in the simulated example cases, the measured data exhibits a high frequency oscillation during the transient period. The magnitude of the frequency spectrum of this portion of the time domain waveform is shown in Fig. 3-30b.

Over the life of one of the couplings in the ASW system, the magnitude of the spectral peak located at approximately 44 Hz progressively increases. Figure 3-31 shows its magnitude during five representative starts. The top trace was recorded early in the life of the coupling, the second trace was recorded during mid-life, and each of the final three were recorded during some of the final starts before the end of the life of the coupling. For comparison purposes, Figure 3-32 presents photographs demonstrating the state of the coupling after four of the five starts used to generate Fig. 3-31.

Table 3.4 summarizes the diagnostic indicator values for each of the five starts considered above. Note that this value significantly increases between starts B and C. Currently, the change detector for this system is programmed to trigger an alert as soon as the indicator

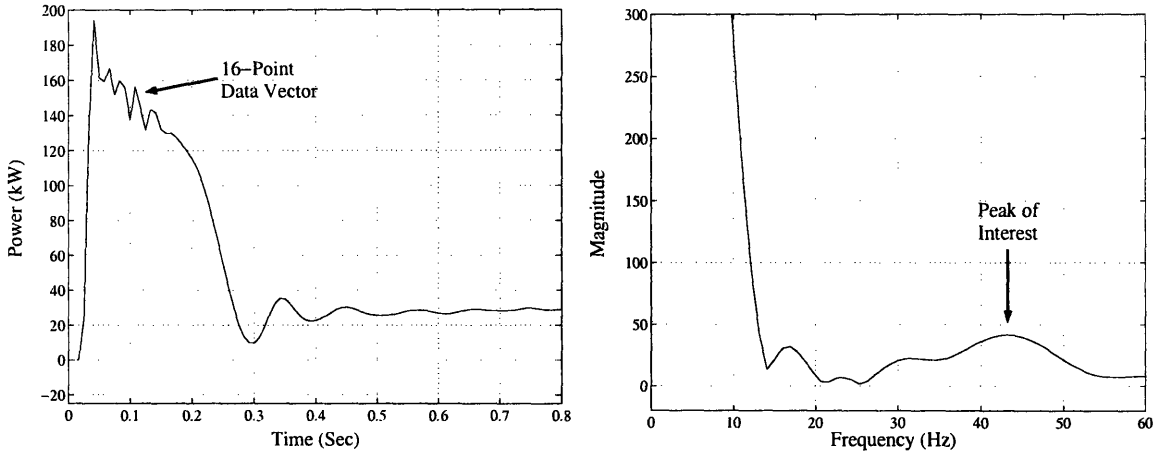


Figure 3-30: (a) Real power drawn during a typical ASW pump start. Note that the relative location of the 16-point data vector shown here was the same for each of the pump starts. (b) Magnitude of the frequency spectrum of the windowed 16-point data vector indicated in (a).

value calculated during any given start exceeds 100. When generating a status report, the NILM also includes a plot showing the time history of the indicator value. This is useful, of course, to the operator because it allows him or her to "tune" the automatic detection threshold as desired. Such flexibility makes it possible to apply this diagnostic indicator in other, similar systems.

Table 3.4: Value of the coupling failure indicator for each of the starts used to form Fig. 3-31.

Start Letter*	Value of Coupling Failure Indicator
A	23.7
B	22.7
C	112.5
D	139.2
E	146.2

* See Fig. 3-31 for reference.

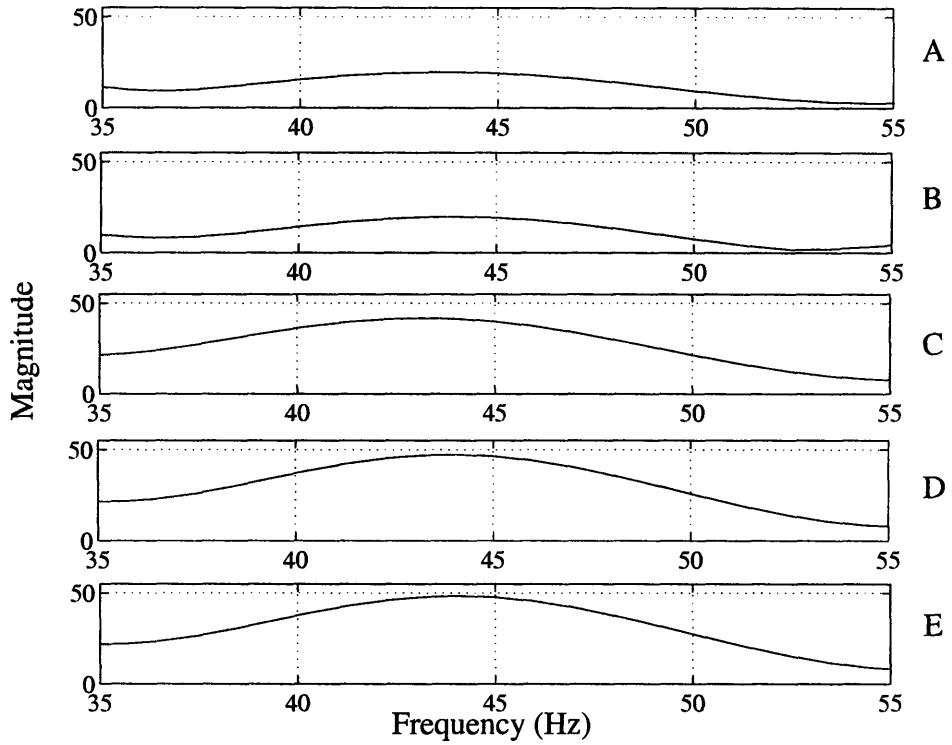


Figure 3-31: Details in the area of the high frequency spectral peak from Fig. 3-30b during several pump starts. The top trace was recorded after one of the first starts with a fresh coupling, and the bottom trace was recorded during the start before the coupling finally failed completely. Each start is designated by a letter written to the right of the corresponding trace.

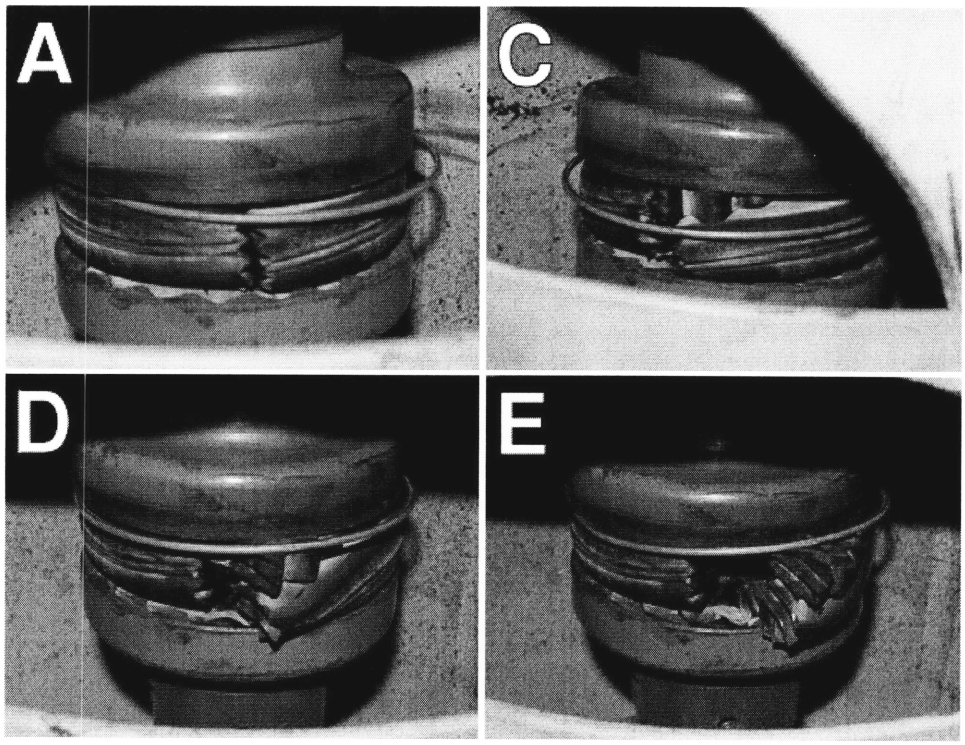


Figure 3-32: Photographs showing the state of the coupling after four of the five starts used to generate Fig. 3-31. The letters in the upper left corner of each photograph correspond to the letters displayed in Fig. 3-31.

3.3 System-Level Health Monitoring in a DDG-51 Engine Room

Many engineered systems consist of large integrated networks of sensors and actuators. Examples include automobiles, aircraft, building ventilation systems, marine engine plants, and industrial assembly lines. In these systems, the interactions between the various sensors and actuators can be extremely complex. Because of this, component failures often trigger time-consuming troubleshooting investigations. In industrial facilities, these investigations can lead to decreased productivity and lost revenue, and in naval vessels, they can leave the ship vulnerable to attack by an enemy.

Electrical monitoring is an effective way to locate faults in complex integrated systems. In order to demonstrate this, several NILMs were installed at the U.S. Navy's Land Based Engineering Site (LBES) in Philadelphia, Pennsylvania. The LBES is a replica of the Number Two Main Engine Room aboard one of the Navy's *Arleigh Burke* class (DDG-51) destroyers. The equipment installed at this facility includes two LM2500 gas turbine propulsion engines (GTMs); three gas turbine generators (GTGs); and auxiliary systems providing fuel oil, lube oil, low-pressure air, and cooling water. For this research, a separate NILM was installed to monitor each of the following loads:

- A low-pressure air compressor (LPAC)
- A fuel-oil service pump
- The controller for each of the GTMs

The next three sections summarize the monitoring activities performed at the LBES. Each individual section focuses specifically on one of the three loads listed above. Methods for system-level health monitoring are considered in Section 3.3.3.

3.3.1 LPAC

In the engine room of an *Arleigh Burke* class destroyer, there are many pneumatically-actuated valves. The pressurized air needed to operate these devices is provided by a low-pressure air header that is charged by two LPACs (see Fig. 3-33). These compressors act to ensure that the header pressure remains between 110 psig and 125 psig. In order to maintain pressure, the LPACs operate using a strategy known as load/unload control ([63]). In this

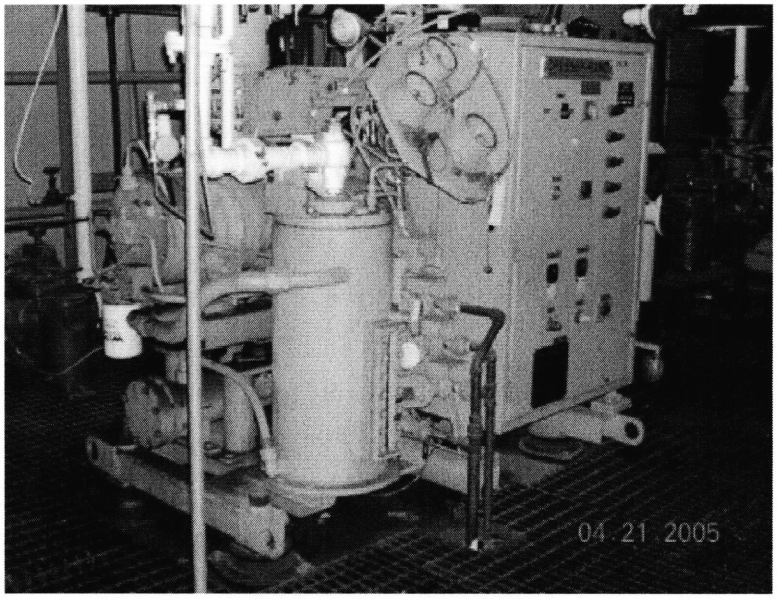


Figure 3-33: One of the low-pressure air compressors installed at the LBES.

approach, the compressor motor is always energized, and the charging process is controlled by periodically redirecting the compressor's output. Figure 3-34 shows the flow diagram for the unloader system used in each LPAC at the LBES. To understand the operation of this system, consider a period when the LPAC (labeled COMPRESSOR in the figure) is charging the header. As soon as the LPAC's discharge pressure reaches 125 psig, the solenoid-actuated unloader valve in the bypass line is opened, thus directing discharge air back to the inlet. Additionally, redirection of the discharge air forces the unloader valve in the inlet line to close ([80]). In this mode, the compressor's motor continues to operate, but no air is directed into the header. Although this type of control is not as energy efficient as an on/off-based approach, it is still used in many systems because it reduces wear on the motor. In systems where compressor cycling is frequent, motor wear is often a primary concern ([63]).

As a result of the control procedure described above, the low-pressure air system in the DDG-51 engine room fits the definition of a cycling system as presented in Section 3.1. Fig. 3-35 displays the real power drawn by the LPAC as a function of time. As shown in that figure, the power drawn by the compressor is considerably higher when it is loaded. For convenience, the period when the compressor is loaded is defined as the "load interval" and the period when the compressor is unloaded is defined as the "unload interval."

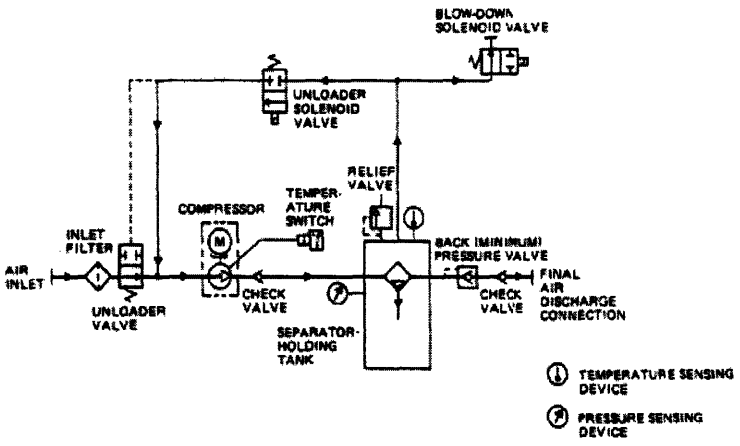


Figure 3-34: Flow diagram of the LPAC unloader system. Taken from [80].

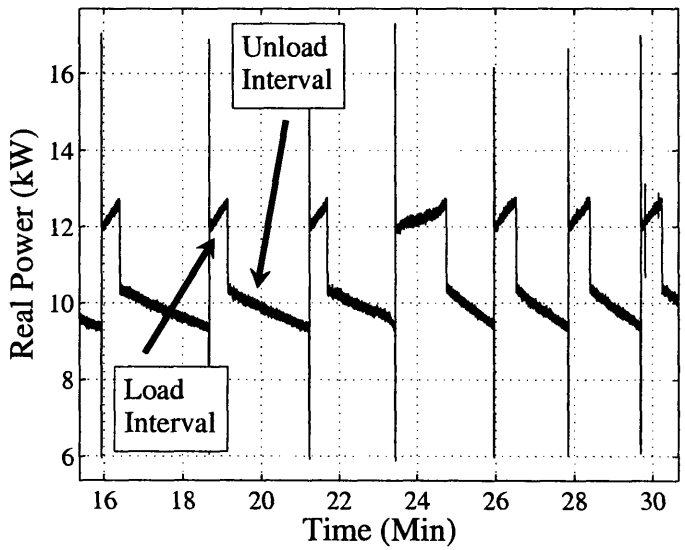


Figure 3-35: Real power drawn by one of the LPACs in the DDG-51 engine room. Note that more power is drawn when the unit is loaded.

In order to appropriately analyze the cyclic energy-usage pattern of each LPAC, it is necessary to consider a model for the low-pressure air system. In this case, the air loads are valves that periodically open to admit compressed air into the various pneumatically-driven actuators. When driving a particular device, the appropriate valves remain open for as long as necessary, which can be for minutes or even hours. Note that this is very different than in the waste-disposal example considered previously, where the drains acted almost instantaneously. Since the air loads operate for such long periods in this system, they effectively act as short-term leaks. Thus, driving an actuator with pressurized air for several minutes is equivalent to inserting a leak into the system for several minutes.

As air demand changes throughout the day, the load on the compressor changes as well. When demand rises, the compressor's load intervals lengthen and its unload intervals shorten. When demand falls, the opposite effects occur. Since such changes happen periodically, a time-based analysis is needed. Figure 3-36 demonstrates one way to analyze the data. At the left in that figure is a plot of the real power drawn by the LPAC before, during, and after the start of one of the GTGs. To understand the behavior shown in the figure, consider the following arguments. When a generator starts, low-pressure air is needed to position several vent dampers and to actuate several valves ([81]). As a result, air demand increases dramatically. Once the generator reaches its operating speed, which takes approximately 60 to 90 seconds, air demand drops, but it remains higher than it was before the start[12]. To analyze these trends, it is useful to generate plots showing how the lengths of the load and unload intervals change as a function of time. Figure 3-36b presents these time histories for the data shown in Fig. 3-36a. Note that before the generator was started, air demand was low, so the load interval was short and the unload interval was relatively long. During the start, which occurred while the compressor was loaded, the length of the corresponding load interval increased dramatically. Once the turbine reached its operating speed, the lengths of the load intervals decreased, but they remained longer than they were before the start. The steady-state change in pneumatic load caused the unload intervals to decrease.

Time histories such as the ones shown in Fig. 3-36b make it possible to perform a number of different diagnostic operations. First, long-term behavior in the load and unload times can be used to detect leaks in the low-pressure air system. A simple leak detector is

[12]The steady-state change occurs because low-pressure air is used to hold open several of the turbine's bleed air valves ([81]).

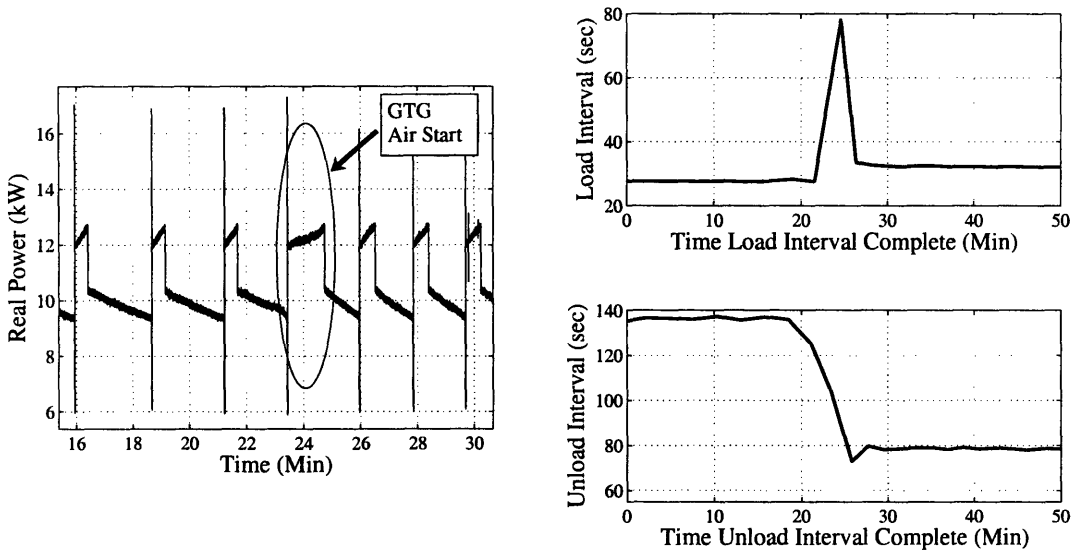


Figure 3-36: Energy usage by one of the LPACs before, during, and after a start of the one of the GTGs. (a) The real power drawn by the compressor. The actual start occurred during the indicated load interval. (b) Time histories of the load and unload intervals. Note that during the start, the length of the load interval increases dramatically.

discussed in Section 3.3.3. Additionally, changes in the load and unload times can be used to determine if the low-pressure air system is responding appropriately to operator commands. The failure to observe expected changes can be useful when attempting to locate faults in the plant.

3.3.2 Fuel-Oil Pump

The fuel-oil system aboard an *Arleigh Burke* class destroyer is designed to provide clean and filtered fuel oil to the GTMs and GTGs ([82]). This system has two pumps that move fuel from a 5000 gallon fuel-service tank into the fuel-supply header. An unloader valve located downstream of the pump discharge acts to maintain the header pressure at 35 psig. If the header pressure changes, this valve is throttled to redirect flow back to the service tank through a recirculation line. Fuel fed to the supply header is heated and filtered before being passed to the GTMs and GTGs ([82]). A photograph of one of the fuel-oil pumps is presented in Fig. 3-37.

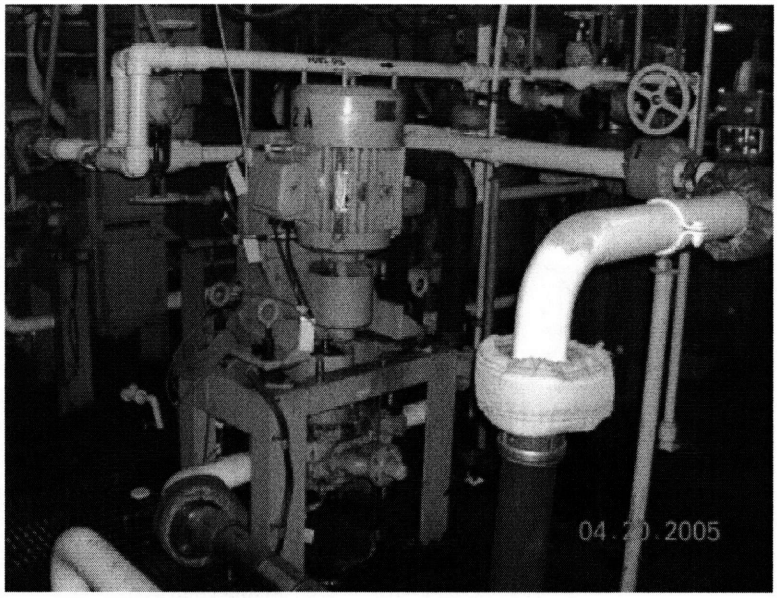


Figure 3-37: One of the fuel-oil pumps installed at the LBES.

The real power drawn by each fuel pump changes in accordance with system fuel demand. For instance, Fig. 3-38 shows the real power drawn during the start of a GTM. Initially, the power draw experiences a sharp drop, as the increase in the load on the fuel system forces the header pressure to fall. This pressure drop triggers the unloader valve to reduce the amount of fuel that is circulated back to the service tank. As this valve responds, the power drawn by the motor slowly rises. Note that the final steady-state power level is slightly lower than it was before the start of the GTM. Using the terminology presented in Section 2.2.3, this is an example of an actuator responding to an indirect command from the controller. As shown in Section 3.3.3, this type of response is useful when attempting to determine if a problem exists in the fuel system.

3.3.3 Engine Controllers

Each GTM aboard an *Arleigh Burke* class destroyer has a control unit known as a universal engine controller (UEC) ([83]). The purpose of this device is to control the actuators that apply start air, fuel, and ignition power. Each UEC contains power electronic circuits that convert three-phase AC power from the ship's distribution system into DC voltages appropriate for individual actuators ([83]). Included among the actuators of interest in this

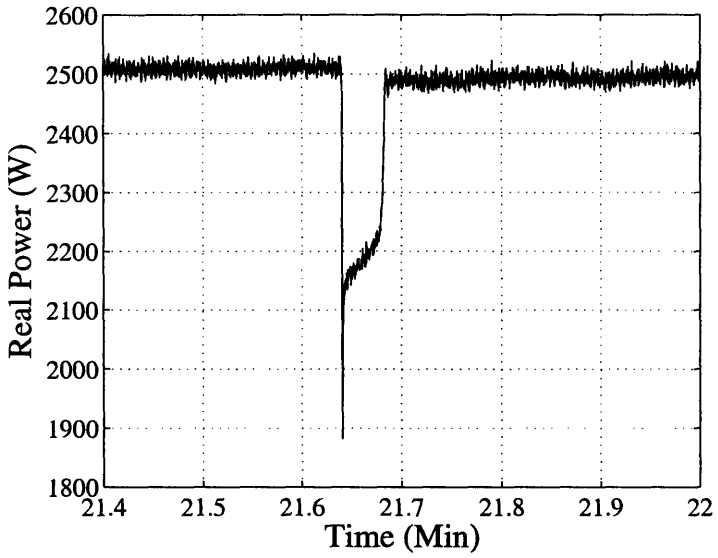


Figure 3-38: Real power drawn by one of the fuel-oil service pumps during the start of a GTM.

thesis are several solenoid-actuated valves such as the starter-air regulating valve and the fuel-oil shutoff valves. Additionally, the UEC contains relay circuits that apply single-phase AC power from the ship's service to other components, including the fuel igniters, the de-icers, and a ventilation fan ([83]). Currently, a NILM is installed on the three-phase input connection to the UEC[13].

A primary function of each UEC is to control the start and stop operations of its associated GTM. During those processes, the UEC monitors the state of the engine and energizes the appropriate actuators in sequence. For example, the UEC performs the following operations when starting a turbine. First, it energizes the normally-closed starter-air regulating valve, which admits compressed air into the turbine's pneumatic starter. This forces the turbine to begin to rotate. As soon the turbine reaches 1200 rpm, it energizes both the fuel igniters and the normally-closed fuel-oil shutoff valves. Once the speed reaches 4500 rpm, which is just below the turbine's self-sustaining speed, the UEC de-energizes both the igniters and the starter-air regulating valve.

During any UEC operation, the NILM is able to monitor the actions of the individual

[13]Note that the components powered via single-phase AC are supplied through a separate single-phase circuit in the distribution network. Thus, the NILM is not currently monitoring the power delivered to these devices.

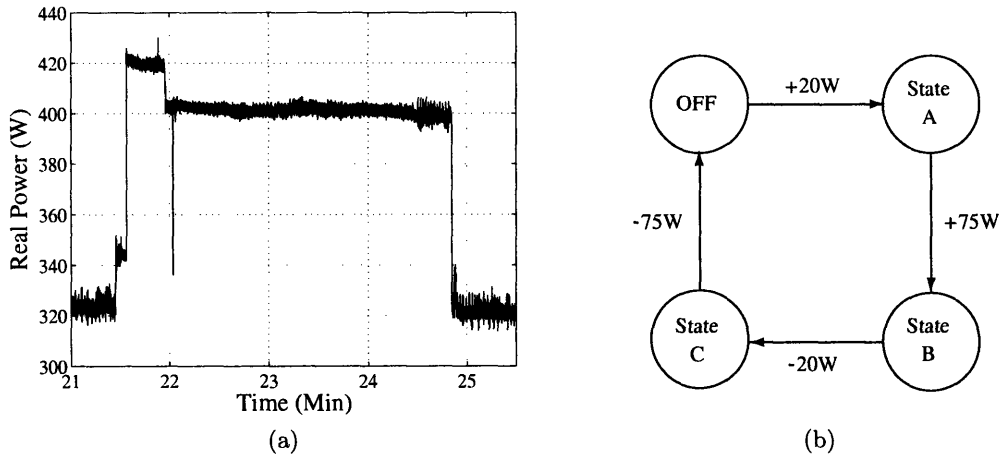


Figure 3-39: UEC power during a GTM start. (a) Real power drawn by the UEC and (b) an FSM model for the level transitions shown in (a). In State A, the starter-air regulating valve is open. In State B, both the starter-air regulating valve and the fuel shutoff valve are open. In State C, only the fuel shutoff valves are open. In the OFF state, all valves are closed. The power level written next to each arc is the amount by which the steady-state power changes during the prescribed state transition.

valves because they affect the power drawn on the UEC's three-phase input bus. Figure 3-39a displays the power drawn by the UEC during a typical start. As shown, the steady-state level changes slightly each time that the UEC issues a new command. As suggested in Chapter 2, these changes can be analyzed using finite-state machine (FSM) models. Figure 3-39b shows the FSM model for the start procedure described above. To understand that model, consider the accompanying set of of power trace. Before the operator issues a start command, the accompanying turbine is at rest, and the system is in the OFF state. As shown, the UEC draws about 325 W during this period. When the start sequence begins, the starter-air regulating valve is energized. causing an approximately 20 W increase in the power drawn by the UEC. This change places the system into State A. Once the speed reaches 1200 rpm, the fuel-oil shutoff valves are energized, forcing the power level to rise again. This 75 W increase signals a transition into State B. At 4500 rpm, start air is no longer needed, so the UEC closes the starter-air regulating valve, thus causing the power to drop by about 20 W. This change places the system into State C. Operation continues in this state until the operator issues a shutdown command. At that point, the fuel shutoff valves close, and the UEC power returns to its initial level.

With a set of FSM models, the NILM can identify each of the major operations in the

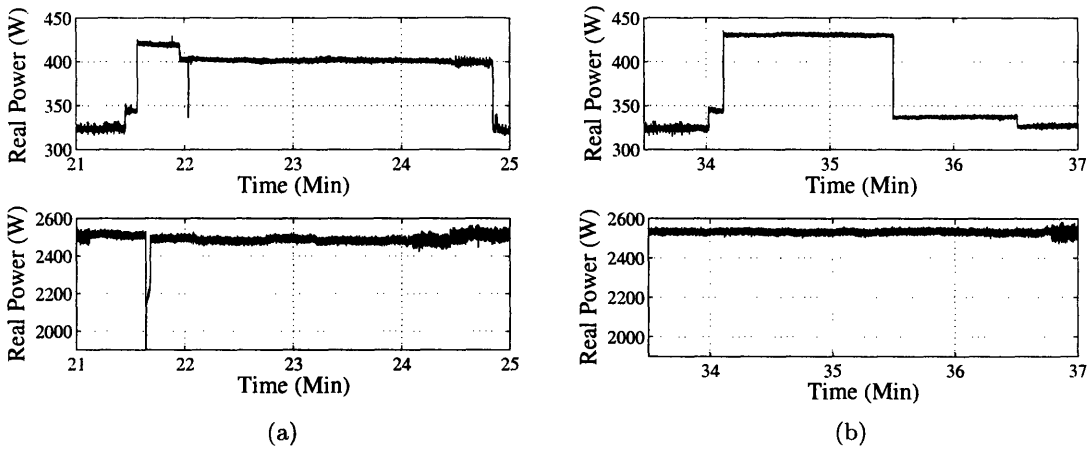


Figure 3-40: UEC power and fuel pump power during (a) a normal start and (b) a failed start. The failure was due to a fault in the fuel system.

plant. Once a certain set of commands has been classified, it is possible to determine if the system responded appropriately. This procedure is particularly useful when attempting to find the source of a fault. As an example, consider Fig. 3-40, which shows both the UEC power and the fuel-pump power during two different GTM starts. During a normal, non-faulty start (e.g. Fig.3-40a), note that the fuel pump responds as soon as the fuel shutoff valves open to admit fuel into the turbine. By comparison, Fig. 3-40b demonstrates what happened during a start when the fuel valves failed to respond. As shown, the system initially transitioned into State A and then into State B. At that point, the fuel shutoff valves were commanded to open, but no corresponding change occurred in the fuel pump power. This particular failure was the result of a fault in the fuel system. A likely fault indicator would first use the UEC state transitions to determine that the start failed[14], and it would then examine the accompanying actuator responses. Similar procedures could be implemented for each engine plant operation.

The NILM can also use its ability to identify controller commands as a tool to help locate leaks in the low-pressure air system. For instance, one way to detect leaks is to try to correlate controller commands with long-term trends in the load on the compressor.

[14]The UEC power trace shown in Fig. 3-40b possesses two characteristics that indicate a fault. First, the system remains in State B until exactly 90 seconds have passed since the beginning of the start sequence. If the starter-air regulating valve is still open at that time, the UEC knows that the turbine failed to reach its self-sustaining speed. It then transitions into a fail-safe mode for exactly 60 seconds. During that state, the turbine is spun with start air, but no fuel is provided to the engine ([61]).

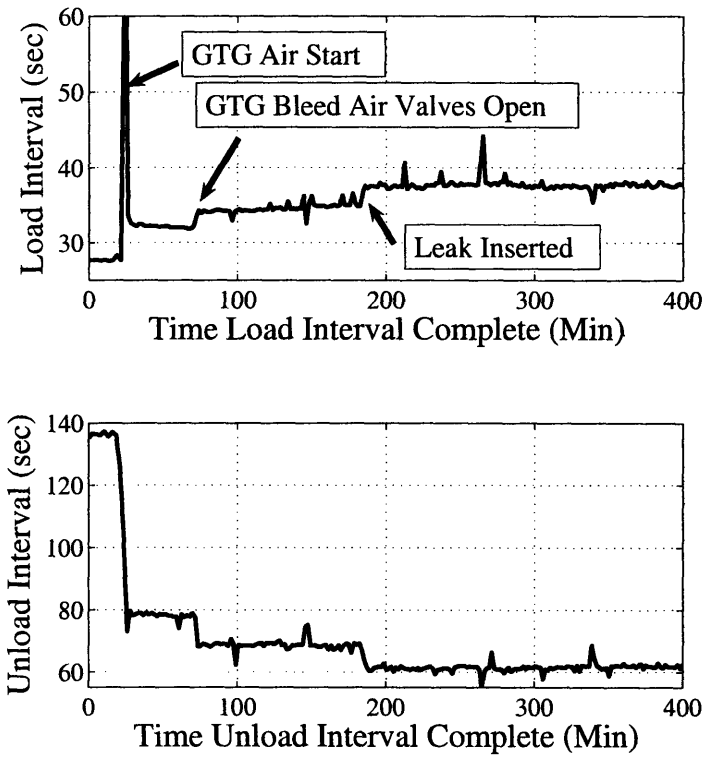


Figure 3-41: Time histories of the LPAC's load and unload intervals over a nearly seven-hour window. The top trace is annotated in order to indicate the physical event corresponding to each long-term shift.

As a specific example, consider Fig. 3-41, which shows the evolution of the lengths of the load and unload intervals over a nearly seven-hour window. During that period, the load on the compressor experienced three long-term shifts. As indicated in the figure, two of those three changes were the result of controller-issued commands. The third shift, however, occurred because a small leak was inserted into the system using a flow meter. With no accompanying controller command, the NILM would consider this change to be a leak. In order to detect leaks that build slowly over time, multiscale techniques could be used to augment this process.

Chapter 4

Monitoring Indoor Airflow

ACCORDING to recent estimates, ventilation is one of the largest components of overall energy consumption in modern and retrofit buildings ([39]). As energy costs continue to rise, building engineers have focused intense efforts on creating methods to increase the efficiency of ventilation systems. Several methods have been proposed, including the use of natural ventilation in suitable climates ([84]), ventilation system balancing ([39]), and the reduction of uncontrolled air infiltration ([85]). In order to realize the complete potential of these methods, however, it is necessary to obtain real-time information about airflow patterns and ventilation efficiency. While techniques have been developed to monitor indoor airflow using tracer gases such as SF_6, the required equipment is both prohibitively expensive and invasive ([86], [40]).

In order to meet the need for a low-cost, minimally-intrusive airflow monitoring system, a technique was developed that uses ozone as a tracer gas. In the current system, high-voltage DC power supplies generate ozone via corona discharge. Several of these ozone-generating devices as well as their corresponding ozone detectors are deployed at different points in a building. Accompanying each individual generator or detector unit is a power-line carrier (PLC) modem that communicates with the remote sensor monitor (RSM) described in Chapter 1. When an airflow test is conducted, the RSM issues appropriate commands to each generator and detector. At the end of an experiment, ozone generation ceases and the detectors transfer their concentration measurements to the RSM for analysis.

This chapter presents a detailed analysis of the complete airflow monitoring system. Section 4.1 describes the overall system layout. That section also discusses both the corona discharge process used to generate ozone and the chemical sensing devices used to detect it. Section 4.2 presents an in-depth analysis of the high-voltage power electronics that comprise

the ozone generator, and Section 4.3 includes a similar presentation for the ozone detector circuitry. Subsequently, Section 4.4 explains the methods used to coordinate the operation of the complete system. Finally, the chapter concludes in Section 4.5 by describing the details of the airflow measurement process and by presenting the results of several example measurements.

Because of the health risks posed by exposure to high levels of ozone[1], significant effort was made to generate and to detect trace concentrations. Thus, much of the work on this airflow monitoring system was devoted to circuit design and implementation. Accordingly, this chapter primarily focuses on the circuits used to generate and to detect ozone.

4.1 System Overview

The main challenge in creating a low-cost airflow monitoring system is the selection of a gas that can be easily generated and reliably detected at trace levels. Ozone (i.e. O_3) was chosen as the tracer gas in this system because it meets both of those requirements. This section provides an overview of the complete system.

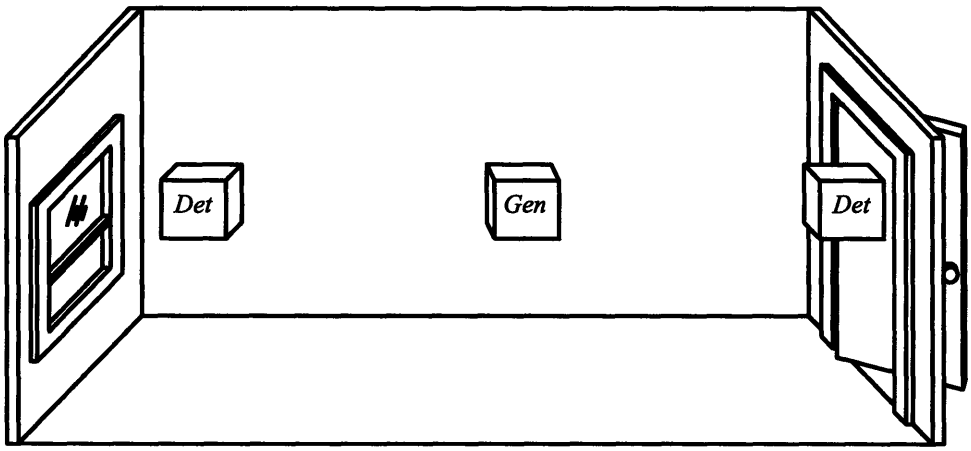
4.1.1 System Layout

A typical experimental arrangement for our system is shown in Fig. 4-1a. Using that configuration, it is possible to determine both the speed and the direction of air movement at various points in a room. Each unit depicted in Fig. 4-1a is shown in only one of its two states. In order to maximize flexibility, each station is actually comprised of both a generator and a detector as shown in Fig. 4-1b. Using power-line communications, the remote sensor monitor commands each station into one of its two states prior to the start of any given measurement. Those commands are generated by control software that routinely reads user-provided configuration files. In this way, the system can be easily and automatically reconfigured to provide different velocity and direction measurements.

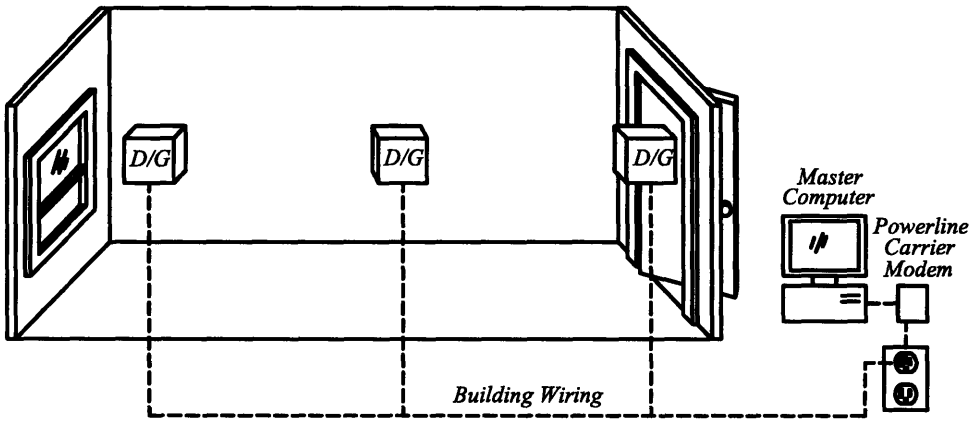
4.1.2 Ozone Generation

The ozone used in this system is generated during the process of corona discharge at a point electrode. Corona is a term describing the gaseous discharge that occurs when air breaks

[1]Appendix B addresses the health risks posed by ozone, and it also discusses how this system avoids exacerbating them.



(a)



(b)

Figure 4-1: The basic system layout for the airflow monitor. (a) A typical experimental arrangement for determining both air speed and direction. (b) Diagram of the complete system showing that each unit is actually comprised of both a generator and a detector. A centrally-located master computer communicates with each of the remote stations using the power distribution network.

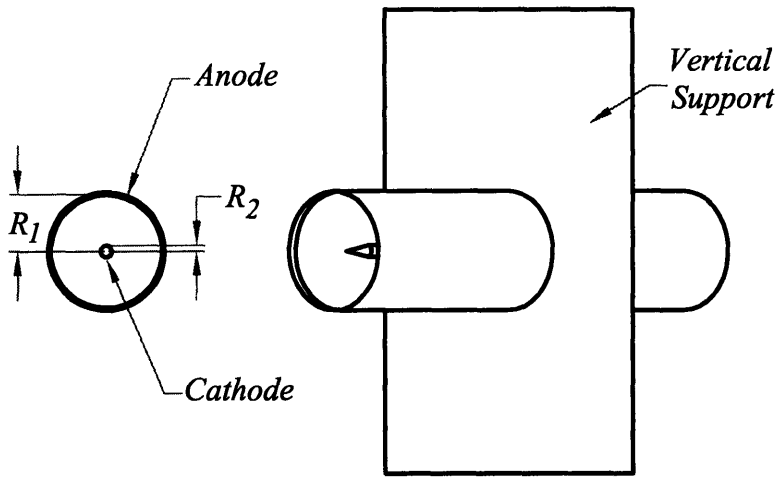


Figure 4-2: Head-on and side view of the electrode arrangement used in this system. The inner radius of the outer conductor is 8 mm, and the radius of the inner conductor is 0.8 mm. With such a large spacing between the two electrodes, corona discharge occurs at the tip of the inner conductor. Note that the side view also shows the plexiglass sheet used to support the outer cylinder.

down under the stress of a non-uniform electric field ([87], [88]). Because the breakdown process is driven by non-uniform fields, corona discharge is a highly complex phenomenon ([88]). Although an exhaustive analysis is beyond the scope of this work, it is useful to discuss the basic physics of the process since they influence the design of both the electrodes and the high-voltage power supply.

The electrode geometry used in the ozone generators consists of a central cathode raised to highly negative potential and a coaxial cylindrical anode that is held at ground. As shown in Fig. 4-2, the physical spacing of the two electrodes is fairly large relative to the actual size of the inner conductor. When a large potential difference is imposed upon electrodes in this arrangement, breakdown will tend to be confined to the high-field regions near the electrode surfaces ([87]). At atmospheric pressure, the breakdown will localize even further, occurring mainly near the tip of the inner electrode. A manifestation of this process is the emission of light near the tip of the inner conductor ([87]). This arrangement can be analyzed in detail by making the simplifying assumption that the inner cathode is a point conductor and that the outer anode is a plane ([87], [89]).

As a result of the corona discharge, a series of chemical reactions are triggered in the space between the electrodes. One of the by-products of these reactions is ozone. Details of the reaction dynamics for the point-to-plane geometry can be found in several references ([90], [89]).

The critical component of the ozone generator is the power supply that imposes a potential difference of several thousand volts upon the two electrodes. In this system, a parallel resonant DC/DC converter generates the required voltage from a 12 V DC input. The details of the power supply design are presented in Section 4.2.

4.1.3 Ozone Detection

The ozone concentration measurements required in this system are performed using low-cost, metal-oxide chemiresistors. In these devices, a metal-oxide surface is typically raised to a temperature several hundred degrees above the ambient. At those temperatures, the chemisorption of gases such as ozone can take place at the surface ([91], [92]). In the process of chemisorption, a chemical reaction occurs between the surface and the gas being adsorbed ([93]). In this particular case, researchers believe that ozone oxidizes the surface, removing an electron and producing both a gas-phase oxygen molecule and a surface oxygen ion, i.e. O^- ([92]). This reaction is summarized in Eq. 4.1.

$$O_3(g) + e^- \rightarrow O^-(s) + O_2(g) \tag{4.1}$$

By removing free electrons, this reaction decreases the electrical conductivity of the oxide. Thus, the resistance of the sensor increases in proportion to ambient gas concentration ([92], [91]).

In the current system,the ozone sensor is a tin-oxide (i.e. SnO_2) device that is heated to 430^o C by applying a voltage to a heating resistor located below the oxide surface ([94]). This device, which is sold in a discrete package by MicroChemical Systems, is the heart of the ozone detector ([94]). To monitor gas concentration, the detector measures the resistance of the sensor using a lock-in sensing scheme. Using that technique, errors due to low frequency phenomena are reduced, making it possible to detect very small changes in resistance. The interface circuitry used in the detector is described at great length in Section 4.3.

4.2 The Ozone Generator

In the current system, trace amounts of ozone are generated using the high-voltage output of a parallel resonant DC/DC converter. This type of supply is particularly useful in high-voltage applications since the leakage inductance and parasitic winding capacitance of the

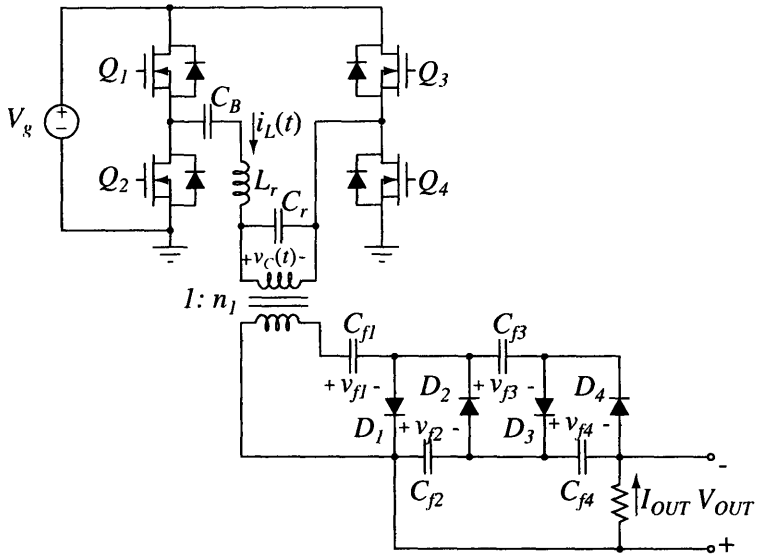


Figure 4-3: A full bridge implementation of the parallel resonant converter. Note that the average value of the voltage across C_{f1} is a constant denoted as V and that the average value across each of the other 3 filter capacitors is $2V$.

step-up transformer can be incorporated into the resonant tank network ([95]). This section presents a discussion of the converter design, and it also describes the control circuitry used to synchronize the operation of the power supplies with the rest of the system.

4.2.1 Parallel Resonant Converter Design

The ozone-generating power supply used in initial experiments has the topology shown in Fig. 4-3 and is designed to the specifications summarized in Table 4.1. This circuit consists of a resonant inverter, a step-up transformer, and a series of voltage-multiplying rectifiers. During operation, the full-bridge inverter drives the tank network with a square wave of voltage, causing the tank to ring in a piecewise sinusoidal manner. A DC output voltage is produced by transforming, rectifying, and filtering the voltage across the resonant capacitor. In order to control the magnitude of the output voltage, the switching frequency is varied about the natural frequency of the tank ([95]). Note that the use of a voltage-quadrupling rectifier allows the voltage ratings of the output capacitors and diodes to be significantly lower than the ratings required in a standard parallel resonant converter designed to the same set of specifications ([95]).

Table 4.1: Specifications for the high-voltage power supply

Quantity	Specified Value
V_{out}	5800V
I_{out}	3.2mA
V_{in}	12V

Analysis

Before designing the power supply used in this system, the intended topology was first analyzed in order to determine the ideal converter waveforms and to develop several useful steady-state design relations. Of particular interest in this analysis was the explicit determination of the effect of the voltage multipliers. Similar analyses for converters with standard full or half bridge rectifiers can be found in the literature ([96]).

In steady state, the parallel resonant converter can operate in any one of several modes, depending upon the order in which the semiconductor devices switch ([96], [97]). The analysis begins by considering operation over a single cycle in a mode known as mode 1 ([96]). In that mode, the switches Q_1, Q_4, D_1, and D_3 are all in operation at the beginning of each cycle, i.e. at $t = 0$. Neglecting the forward voltages of the diodes and assuming that the filter capacitors are large enough that ripple can be neglected, the voltage across the resonant capacitor, $v_C(t)$, will be $-V/n_1$, where V is the average value of the voltage across C_{f1}. Accordingly, the voltage across L_r at time $t = 0$ is held at the constant value $V_g + V/n_1$, thus forcing the tank current, $i_L(t)$, to rise linearly. Since the voltage across the resonant capacitor is assumed to be constant, no current flows through it. Thus, $i_L(t)$ has a magnitude that is equal to that of the reflected secondary current. Furthermore, a KCL analysis shows that $i_L(0) < 0$.

During the period when the tank current is negative, charge is supplied to the capacitors C_{f1} and C_{f3}. As soon as $i_L(t)$ crosses zero, that charging process ceases because the diodes D_1 and D_3 block the flow of positive tank current. After that time, which is defined as T_{a1}, the rectifier no longer imposes a voltage across C_r, allowing the tank to begin to ring.

The ringing of the tank network does not last indefinitely. Once the voltage $v_C(t)$ reaches the value V/n_1, the diodes D_2 and D_4 turn on, thus clamping $v_C(t)$. At that time, which we define as T_{a2}, the voltage imposed on the inductor also becomes fixed, and its

value is $V_g - V/n_1$. Accordingly, $i_L(t)$ assumes a linear shape. This subinterval is defined to end when the FETs switch at time $T_{sw}/2$. By continuing the analysis of the circuit beyond that time, one finds that over the remainder of the cycle $v_C(t) = -v_C\,(t - T_{sw}/2)$ and $i_L(t) = -i_L\,(t - T_{sw}/2)$.

The results of the analysis performed above can be used to derive a set of equations that describe the ideal converter waveforms. When considering conduction losses in the diodes, those state equations are

$$v_C(t) = \begin{cases} -\dfrac{V + V_{D2}}{n_1}, & 0 < t < T_{a1}, \quad (4.2a) \\ V_g - \left[V_g + \dfrac{V + V_{D2}}{n_1}\right] \cdot \cos\left[\omega_o\,(t - T_{a1})\right], & T_{a1} < t < T_{a2}, \quad (4.2b) \\ \dfrac{V + V_{D2}}{n_1}, & T_{a2} < t < \frac{T_{sw}}{2}, \quad (4.2c) \end{cases}$$

and

$$i_L(t) = \begin{cases} i_L(0)+ \left[V_g + 2V_{D1} + \dfrac{V + V_{D2}}{n_1}\right] \dfrac{t}{L_r}, & 0 < t < T_{a1}, \quad (4.3a) \\ \dfrac{1}{R_o}\left[V_g + \dfrac{V + V_{D2}}{n_1}\right] \cdot \sin\left[\omega_o\,(t - T_{a1})\right], & T_{a1} < t < T_{a2}, \quad (4.3b) \\ i_L(T_{a2})+ \left[V_g - \dfrac{V + V_{D2}}{n_1}\right] \cdot \dfrac{t - T_{a2}}{L_r}, & T_{a2} < t < \frac{T_{sw}}{2}, \quad (4.3c) \end{cases}$$

where V_{D1} is the forward voltage drop of the FET body diodes and V_{D2} is the forward voltage drop of the rectifier diodes. Values for the unknowns T_{a1}, T_{a2}, $i_L(0)$, and $i_L(T_{a2})$ are found by matching boundary conditions. The accuracy of these equations is demonstrated in the next subsection.

Using the results of the above analysis, it is possible to derive a closed-form equation that expresses how the normalized output current, J, depends on both the normalized output voltage, M, and the normalized switching frequency, F. As shown below, this equation is

extremely useful in a design context because it allows the designer to examine the output characteristics of the converter without needing to perform numerous iterations ([96]). In mathematical terms M, J, and F are defined as follows:

$$M = \frac{V_{OUT}}{nV_g}, \tag{4.4}$$

$$J = \frac{nI_{OUT}}{V_g/R_o}, \tag{4.5}$$

and

$$F = \frac{f_{sw}}{f_o}, \tag{4.6}$$

where V_g is the converter's input voltage; V_{OUT} and I_{OUT} are its output voltage and current, respectively; $n = 4n_1$ is the combined voltage gain of the transformer and voltage quadrupler; f_{sw} is the switching frequency; and f_o and R_o are the resonant frequency and characteristic impedance of the tank, respectively. An equation relating J, M, and F can be derived in this case by writing an expression for the average value of the current flowing through C_{f4}. During that process, one discovers that the zero average current restriction for the capacitor forces the following relationship to hold:

$$|I_{OUT}| = \frac{1}{2n_1 T_{sw}} \left(\int_{T_{a2}}^{T_{sw}/2} i_L(t)dt + \int_{T_{sw}/2}^{T_{sw}/2+T_{a1}} i_L(t)dt \right). \tag{4.7}$$

After evaluating Eq. 4.7 and normalizing the result, we find that

$$\begin{aligned} J \;=\; & \left\{ M_4 \left[M_2 (\gamma - \alpha)^2 /2 + M_1 (\gamma - \alpha) \sin\alpha \right] \right. \\ & \left. - \; (M_1 \sin\alpha)^2 /2 \right\} / (\gamma M_3) . \end{aligned} \tag{4.8}$$

In Eq. 4.8, we have used the following definitions, which are consistent with those found in the literature ([96]):

$$\gamma = \pi/F, \tag{4.9}$$

$$M_1 = 1 + M + V_{D2}/(n_1 V_g), \tag{4.10}$$

$$M_2 = 1 - M - V_{D2}/(n_1 V_g), \tag{4.11}$$

$$M_3 = 2 + 2V_{D1}/V_g, \tag{4.12}$$

$$M_4 = 1 + M + 2V_{D1}/V_g + V_{D2}/(n_1 V_g), \tag{4.13}$$

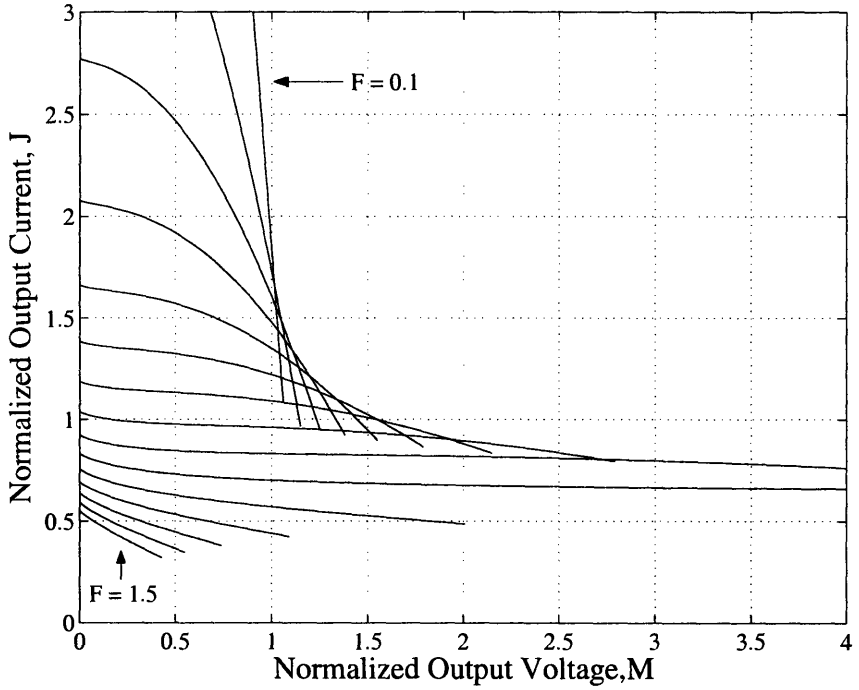


Figure 4-4: Mode 1 output characteristics of the parallel resonant converter for various normalized switching frequencies, F. For the curves shown here, the value of F varies from 0.1 to 1.5 in increments of 0.1. Note that the J-axis intercepts of the various curves decrease as F increases.

$$\alpha = \cos^{-1}\left(M_2/M_1\right). \tag{4.14}$$

Fig. 4-4 displays several curves generated using Eq. 4.8. Note that because Eq. 4.8 was derived under the assumption that the converter operates in mode 1, the curves plotted in Fig. 4-4 end abruptly at the point (M, J) that corresponds to a mode boundary. Beyond that boundary, the semiconductor devices begin to switch in a different order, and the results of the analysis performed here no longer hold. An analysis of the converter's operation in its two other modes is not included here because only mode 1 was considered during the design process.

Design

The design of parallel resonant converters used in high-voltage applications such as this one must begin by considering both the magnitude of the parasitics of the required step-up transformer and their corresponding effect on the choice of the operating point. In order to understand the significance of the winding capacitance and the leakage inductance in such

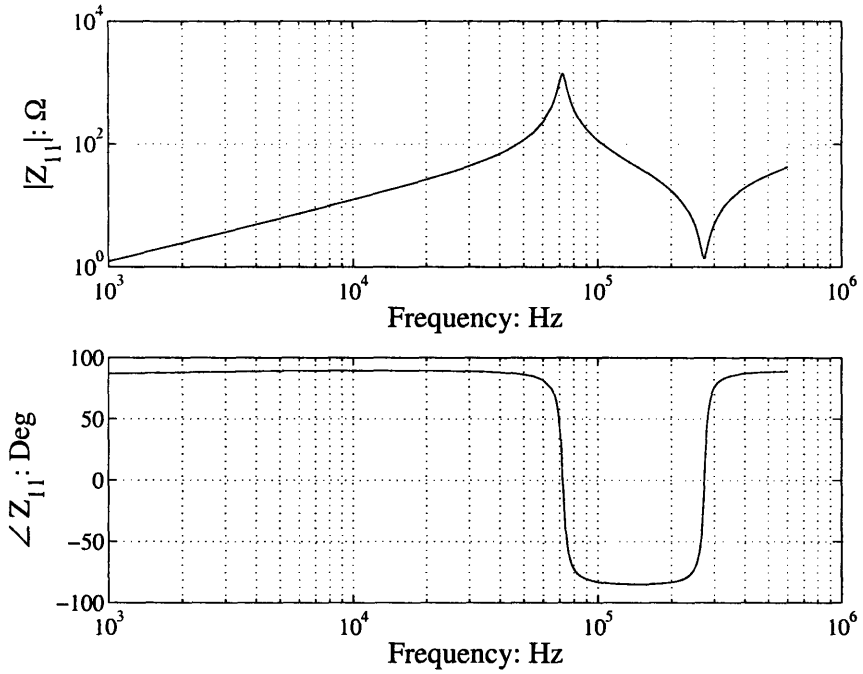


Figure 4-5: Impedance measured looking into the primary of the step-up transformer with the secondary open-circuited.

transformers, consider Fig. 4-5, which shows a plot of the impedance that is measured looking into the primary winding of an example transformer with its secondary open-circuited. From that plot, it is clear that the device exhibits both a parallel and a series resonance in the frequency range considered. One model that explains the observed behavior is shown in Fig. 4-6. According to that model, the impedance below the parallel resonance is dominated by the magnetizing inductance and the impedance beyond the series resonance is dominated by the leakage. With the measured inductance values presented in Table 4.2, the measured resonance locations suggest a capacitance on the order of 26 nF. As predicted by the model in Fig. 4-6, this value corresponds closely to the reflected value of the measured secondary winding capacitance. Note that parasitics on this order could prove troublesome in a comparable switched-mode converter ([95]).

The winding parasitics in this application have a direct impact on the selection of the operating point. Specifically, both M and J must be chosen so that the required values of L_r and C_r are at least as large as the leakage inductance and reflected winding capacitance, respectively. For instance, for any given value of C_r, only a limited subset of all of the possible combinations of M and J can be achieved. To illustrate this, consider the value of

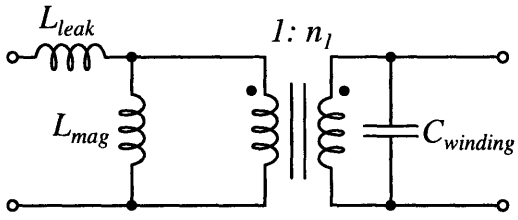


Figure 4-6: Lumped-element model for a high-voltage step-up transformer.

Table 4.2: Summary of transformer parameters

Parameter	Measured Value
L_{mag}	183.2μH
L_{leak}	14.1μH
$C_{winding}$	14.2pF
n	42.8

the tank capacitance referred to the secondary, which can be written in the following form:

$$C_r' = \frac{1}{n_1^2 \omega_o R_o}. \tag{4.15}$$

Using Eqs. 4.4, 4.5, and 4.6, that relationship can be re-formulated as

$$C_r' = \left(\frac{16 I_{OUT}}{2\pi f_{sw} V_{OUT}}\right)\left(\frac{MF}{J}\right). \tag{4.16}$$

Fig. 4-7 displays several curves generated from Eq. 4.16. Note that for large values of J and small values of M, the required secondary capacitance becomes smaller than the typical winding capacitance. Thus, in order to accommodate this parasitic successfully, M should be chosen to be appropriately large and, correspondingly, J should be chosen to be appropriately small. A similar analysis found in [95] shows that the same result holds for the leakage inductance.

In accordance with the above considerations, the converter was designed so that it operates at values of M and J for which C_r is at least $n_1^2 C_{winding}$ and L_r is at least L_{leak}. Thus, to begin the design process, we first selected a reasonably large value for M (i.e. 2.82) and a reasonably small value for J (i.e. 0.815). According to Eq. 4.8 this choice requires a normalized switching frequency F that is equal to 0.84. Furthermore, with M

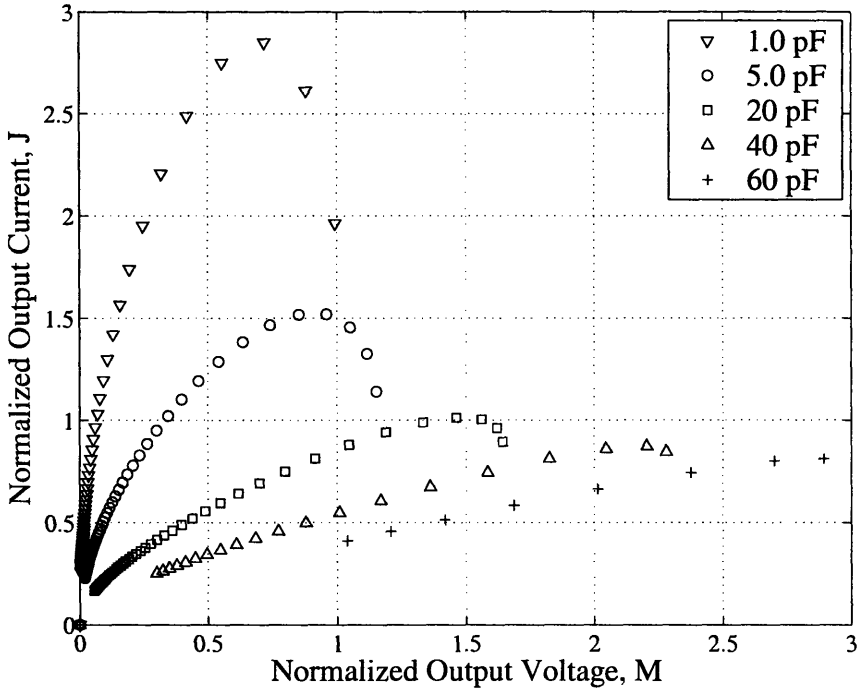


Figure 4-7: Points on several curves generated from Eq. 4.16. Each curve was drawn using one of the capacitance values listed in the legend. Note that these curves are drawn assuming operation in mode 1, and note that diode conduction losses are neglected in this plot since they have a rather negligible effect on the result.

chosen to be 2.82, the specifications stipulate that the turns ratio must be 42.8. Given that requirement, we wound the transformer considered in the earlier example, which has a total secondary capacitance of approximately 14.2 pF. As desired, that value is less than the amount required, which is found from Eq. 4.16 to be 57.1 pF. Accordingly, an additional 43 pF was placed in parallel with the secondary winding.

Before choosing values for the two remaining parameters, which are L_r and f_{sw}, some further consideration was required. Specifically, in order to achieve the desired J, the characteristic impedance of the tank must be 17.7 Ω. With the given tank capacitance, this requirement dictates the use of a 32.8 μH inductor. Because that value is greater than the transformer's leakage, additional inductance was added in series with the primary. In order to determine the exact value, it was first necessary to revisit the transformer model shown in Fig. 4-6. An analysis of that circuit shows that the effective inductance near the series resonance is actually the parallel combination of L_{mag} and L_{leak}. Thus, if an external inductance L_{ext} is added in series with the primary, the effective tank inductance in the vicinity of the series resonance becomes $(L_{leak} + L_{ext}) \| L_{mag}$. In order to achieve the required characteristic impedance, approximately 26 μH was added. With that amount of additional inductance, the series resonance is located at approximately 86 kHz. Accordingly, the converter must be switched at approximately 72.25 kHz.

Fig. 4-8 displays the measured secondary voltage and tank current for the actual power supply. Note that the waveforms demonstrate behavior similar to that predicted by Eqs. 4.2 and 4.3. In particular, note that the voltage clamps at approximately 1450 V and that the current peaks at approximately 2.5 A. Given the design values, Eq. 4.3 predicts that the peak current should be 2.59 A, which is very close to the measured value. These waveforms were obtained when switching the converter at 71 kHz. At that frequency, the output voltage is approximately 5800 V (i.e. 4×1450 V).

4.2.2 Control

As will be explained in Section 4.5, ozone is generated by creating a series of high-voltage pulses. Typically, a single pulse lasts several minutes, but its length depends upon the nature of the given test. Because of the inherent need for flexibility, the power supplies must be able to receive commands from the centrally-located remote sensor monitor (RSM). This functionality is implemented by controlling the resonant converter with a PIC microcontroller.

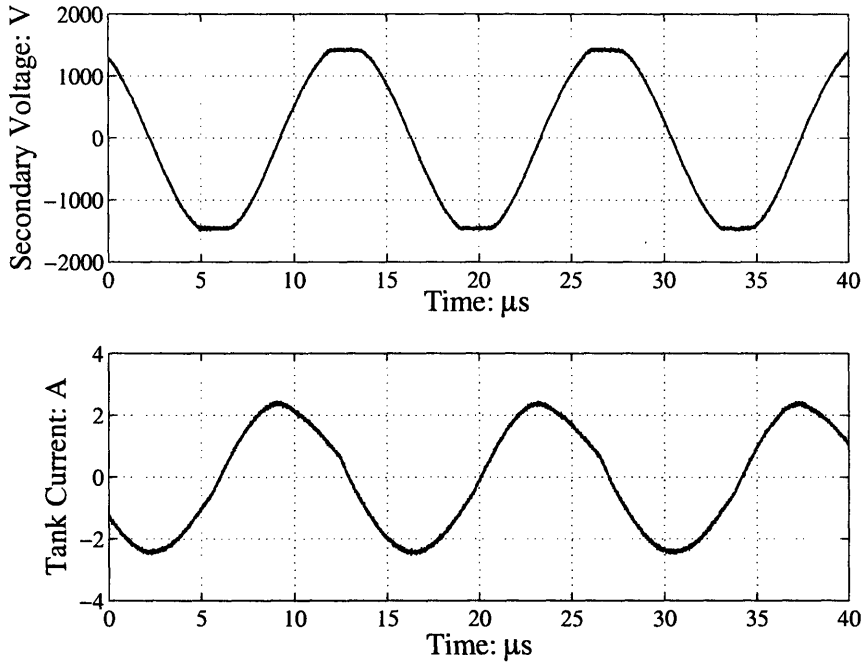


Figure 4-8: Secondary voltage and tank inductor current in the converter prototype. Note that the output voltage is 4 times the peak value of the secondary voltage.

In order to change the operation of a power supply, the RSM first creates a control program for the PIC and then uploads it to the PC at the remote station. Once uploaded, this new code is transferred to the PIC using an RS-232 data link. In order to be able to accept new code on a regular basis, the PIC is programmed with bootloader software. Because all of the generators and detectors in the system must act in synchrony, the bootloader software does not shift execution to the new control program until the PIC has received a start command from the PC. A partial schematic showing the relevant connections to the PIC is contained in Fig. 4-9.

4.3 The Ozone Detector

The ozone detector measures sensor resistance, and thus gas concentration, using a lock-in sensing scheme. By employing this measurement technique, it is possible to detect small resistance changes that would otherwise be "swamped" by high levels of $1/f$ noise. This useful strategy is applied in many similar applications requiring sensitive measurements of faint signals ([98], [99], [100], [101]).

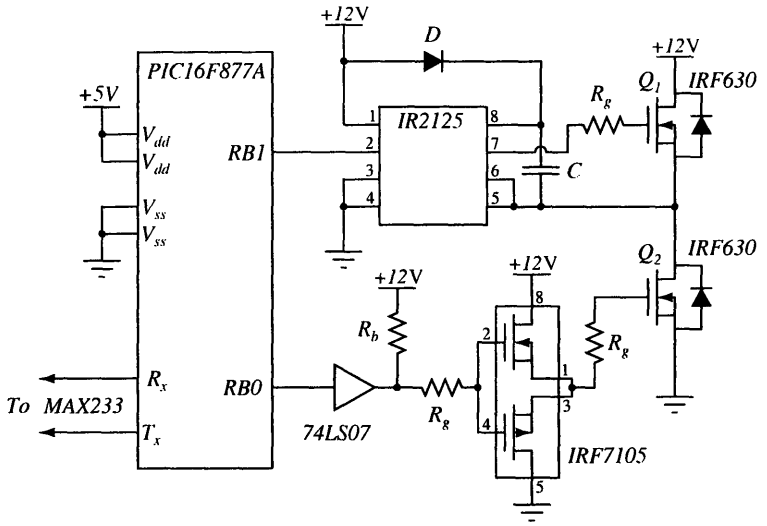


Figure 4-9: Gate drive circuitry for one leg of the full bridge inverter. Note that the FETs in the other leg are controlled in a similar manner. Additionally, note that the IRF630 FETs in the inverter could be replaced with devices with lower current ratings.

This section presents a complete analysis of the ozone detector, and it also details several key design issues. It begins with a discussion of the factors that affect both the choice of the measurement circuit and the choice the measurement strategy. Subsequently, it presents a description of the circuit used to amplify the low-level input signal, and it explains the discrete-time operations that process the amplifier output. Finally, this section concludes with an analysis of the noise properties of the complete ozone detector.

4.3.1 *Interface Circuit Design*

As in any instrumentation problem, the properties of the sensing device impose certain constraints upon both the design of the measurement circuit and the choice of the measurement procedure. When detecting small changes in resistance, there are a number of factors that significantly impact the measurement ([99], [101]). In the case of chemical sensors, the most significant of these are excess noise, self-heating, and cross-sensitivity ([101]). When designing the required electronics, each of these issues must be given just consideration.

The starting point for the design process was the characterization of the intrinsic behavior of the sensor in an open environment. According to manufacturer data, the sensor resistance varies linearly with concentration over a range from approximately 10 kΩ to about 150 kΩ ([94]). From experimental measurements conducted in several example rooms, it

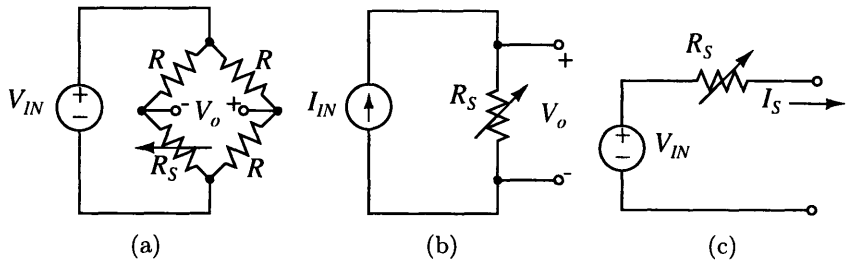


Figure 4-10: Three circuit topologies commonly considered when measuring the unknown resistance R_S. (a) The Wheatstone bridge circuit has an output voltage, V_O, that is proportional to the difference between R_S and R. (b) The current clamp circuit fixes the current through R_S so that the resulting voltage, V_O, is proportional to R_S. (c) In the voltage clamp circuit, the voltage across R_S is fixed, and the resulting current, I_S, is proportional to the conductance $1/R_S$.

has been found that the resistance of a typical sensor is nominally around 30 kΩ. This value can vary, however, from sensor to sensor and as a function of time and ambient conditions. From tests conducted with 15 sensors placed in several different environments, it has been found that the resistance typically varies between 20 kΩ and 50 kΩ, which approximately corresponds to ozone concentrations between 14 ppb and 35 ppb. Most importantly, since the background ozone concentration varies as a function of weather conditions and ventilation patterns ([102]), the sensor resistance can vary throughout the 20 kΩ to 50 kΩ range during the course of an individual test.

The observed behavior of the sensor has a significant impact on the choice of the measurement circuit. The three basic topologies considered are shown in Fig. 4-10 ([103]). The most desirable approach is to place the sensor in the Wheatstone bridge network depicted in Fig. 4-10a. Because the output of that circuit is directly dependent upon the change in sensor resistance, it can be used to obtain a high level of precision[2]. As noted, however, the background ozone concentration, and thus the sensor resistance, can change significantly during the course of a single experiment. In order to address this practical concern without sacrificing resolution, it is necessary to introduce a method whereby either the range of the bridge or the gain of the sensing electronics can be automatically adjusted.

Given the above considerations, it was clear that our measurement system required either an adjustable bridge or one of the other two interface networks depicted in Fig. 4-10.

[2]Actually, the output of the Wheatstone bridge depicted in Fig. 4-10a has a slightly non-linear dependence upon the change in the value of R_s. The output can be linearized, however, through the use of feedback techniques ([104]).

Ultimately, we decided to use the topology shown in Fig. 4-10c. Using that approach, a voltage is applied to the sensor and the resulting current is measured with a circuit that is based on the chopper-stabilized integrator topology presented in ([105]). When measuring the sensor current with that integrator, the need for range adjustment is eliminated for the currents considered here, as the integrating amplifier has a dynamic range that is several orders of magnitude larger than that of most standard instrumentation amplifiers. The reasons for this claim are addressed at length in the next subsection. As illustrated in Section 4.3.4, this rather complex amplifier achieves superb performance when interfaced with the ozone sensor, as it provides a tremendous amount of gain while simultaneously adding a negligible amount of noise to the measurement. As shown later, however, the intrinsic noise of the sensor is rather large, and thus it tends to dominate the noise of any amplifier used to process the sensor's output signal. Accordingly, a simpler, lower-gain topology could be used without significantly affecting the noise floor. For instance, rather precise measurements can be obtained with the circuit shown in Fig. 4-10b when using an amplifier with a moderate voltage gain and a high-resolution A/D converter. That method may be considered in the future. The details of the chopper-stabilized integrator topology are presented at length in Sections 4.3.2, 4.3.3, and 4.3.4.

Having chosen an interface network, the next step of the design process was to chose a measurement procedure. As stated previously, lock-in sensing is used in this application, and its choice is motivated mainly by the presence of low-frequency excess noise. In this and in other tin oxide sensing elements, this noise has a $1/f$ spectral distribution that results from the trapping of carriers in the surface layer ([106]). In order to overcome the effects of this troublesome error source, the sensor is excited at 1 kHz, which according to manufacturer specifications is beyond its $1/f$ noise corner ([107]). In this way the measurement is limited by the sensor's broadband random noise sources, which have a spectral density that is significantly lower than that of its $1/f$ noise. Because of the sensor's elevated operating temperature, its broadband noise is dominated by thermal components ([107]). This issue is discussed in greater detail in Section 4.3.4.

One final issue to address during the design phase is the matter of self-heating. This phenomenon is a rise in resistance that results when passing current through a resistor ([99]). If this effect is not properly considered, then it can easily become the limiting factor of the measurement process ([99], [101]). In order to eliminate the effect of this error source, the sensor manufacturer recommends operating at an average power dissipation level that is less

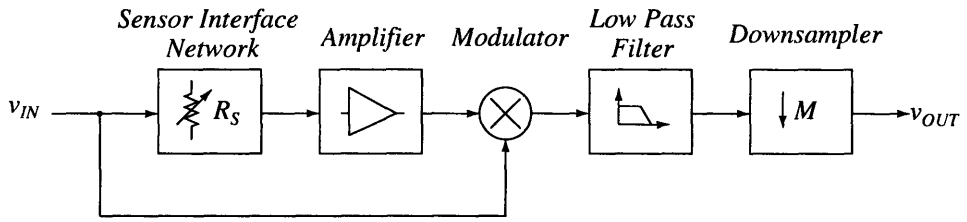


Figure 4-11: Block diagram of the ozone detector for the case in which the measurement circuit is driven by the AC voltage v_{IN}. The output of the sensor interface network is amplified and subsequently demodulated in-phase with the original input voltage. The decimated output signal is related to the low-frequency variations in sensor conductance.

than 2 μW ([107]). In order to do so, the amplitude of the AC drive voltage was reduced to 150 mV.

Given the above arguments, the complete ozone detector was designed as shown in Fig. 4-11. In that system the AC output from the sensor network is passed to an amplifier. As shown in the next section, that amplifier consists of an analog integrator followed by a discrete-time differentiator. The output of that block is thus a discrete-time signal that is proportional to its input. Following amplification, the signal is demodulated using the same waveform that was originally applied to the sensor. Finally, the demodulated signal is filtered and downsampled so that the ultimate output is proportional only to the low-frequency variations in sensor conductance. Note that the successful completion of the lock-in sensing scheme only requires the band-limiting provided by the low-pass filter operation and does not require downsampling. The reasons for decreasing the data rate are explained in greater detail in the next section.

4.3.2 The "Hourglass" Integrator

The amplifier used in the ozone detector is depicted in block form in Fig. 4-12. The first stage is a chopper-stabilized integrator designed using the ingenious topology presented in [105]. Prior to integration, the input signal is modulated so that the integrator remains in its linear regime. Following conversion into the digital domain, the integrator output is unwrapped and passed to a software-based differentiator that recovers the amplified input signal. The lock-in sensing process is then completed by multiplying the amplifier's output signal by a sampled and scaled version of the square wave that was originally applied to the sensor. In order to band-limit the measurement and to ease computational and storage

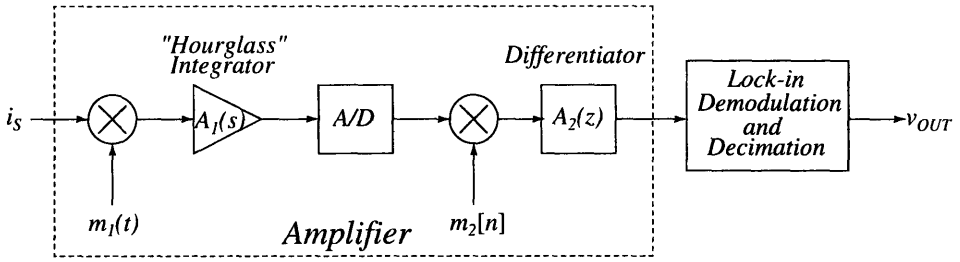


Figure 4-12: Block diagram of the complete low-noise amplifier. The signal i_s is the current through the sensor. The signal $m_1(t)$ modulates the input current in order to ensure that the integrator remains in its linear regime. The integrator output is converted into the digital domain where it is unwrapped and differentiated, returning a signal that is proportional to the original sensor current. Following the processing conducted by the amplifier, the lock-in sensing measurement is completed by performing the demodulation, filtering, and downsampling operations shown in Fig. 4-11

burdens, a series of decimators follows the lock-in demodulation step.

Basic Operation of the "Hourglass" Integrator

The essential building block of the chopper-stabilized integrator is the floating gyrator circuit shown in Fig. 4-13. Gyrators are a class of circuits that are used to map the characteristics of one node to another ([108]). The gyrator network shown in Fig. 4-13 is comprised of two cross-coupled operational amplifiers. The internal negative feedback forces each of the input nodes to be at ground potential. When the control signal Φ_1 is asserted, the current i_s flows into the upper feedback capacitor. Assuming that negligible current is drawn by the operational amplifiers' inverting input terminals, the signal current i_s will flow through the lower feedback capacitor with the opposite polarity. Additionally, the internal negative feedback will force the two output voltages to be centered around ground. Thus, v_{OUT_1} and v_{OUT_2} are given by Eqs. 4.17 and 4.18.

$$v_{OUT_1} = \int \frac{i_s}{C_F} dt \tag{4.17}$$

$$v_{OUT_2} = -\int \frac{i_s}{C_F} dt \tag{4.18}$$

When the signal Φ_2 is asserted, the current through each of the two feedback capacitors reverses polarity. From the above analysis, it is clear that the input characteristics of this

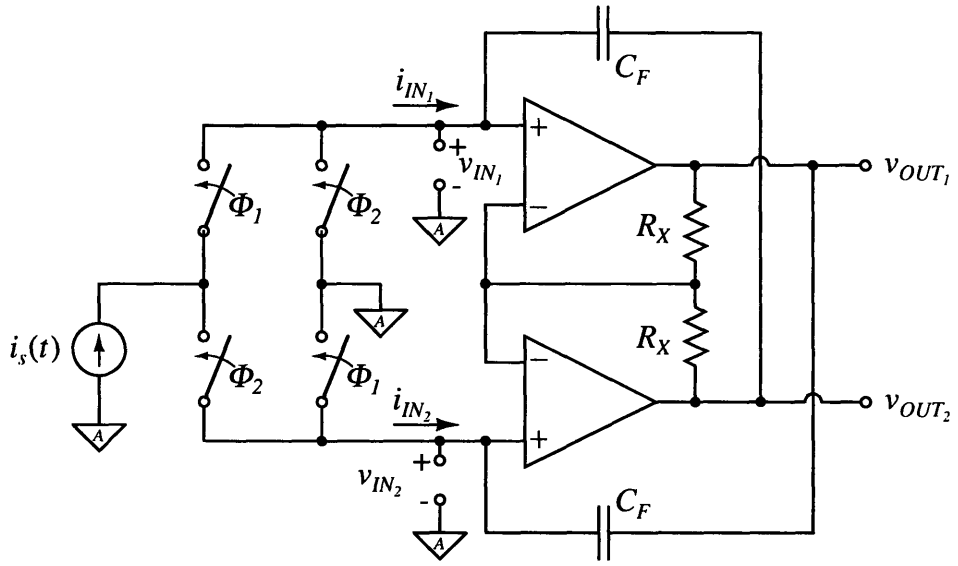


Figure 4-13: The floating gyrator topology used to construct the hourglass integrator. The letter "A" that appears in the ground symbols in this circuit indicate the use of analog ground. Because the printed circuit board has both an analog and a digital ground plane, all of the schematics presented in this section use notation that is intended to indicate the proper reference plane.

circuit can be summarized using the following two relations:

$$v_{IN_1} = v_{IN_2} \tag{4.19}$$

$$i_{IN_1} = -i_{IN_2} \tag{4.20}$$

In order to control the gyrator circuit, the signals Φ_1 and Φ_2 are generated by a feedback network that monitors the two integrator output voltages. Any time that one of the outputs exceeds a threshold that is set at approximately 4.75 V, the control signals reverse polarity and the signal current is passed to the other input. In this way, the input current is integrated indefinitely. This circuit is referred to as an "hourglass" integrator, a name that makes reference to the fact that the alternate charging and discharging of the two feedback capacitors is analogous to the movement of sand from one half of an hourglass to the other ([105]). Fig. 4-14 shows the output of one of the integrators in the presence of a fixed DC input current.

For the hourglass integrator to operate successfully, several auxiliary circuits must accompany the basic gyrator. Fig. 4-15 shows a complete block diagram of the hourglass integrator. As noted previously, a global feedback network is included to force the circuit to

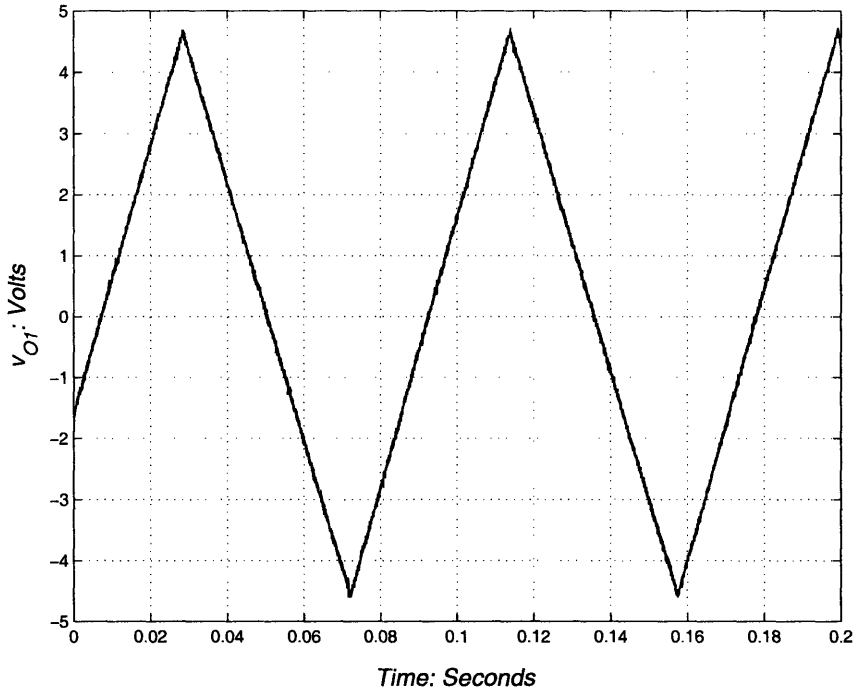


Figure 4-14: Output of one of the integrators in response to a DC input current.

operate in its linear regime. Additionally, the hourglass requires an offset-nulling network, as any low-frequency offset voltage at the amplifier inputs will affect the rate of integration. Each of the components of the integrator circuit are discussed below.

Input Network and Gyrator Interface

Figure 4-16 presents a schematic of the basic interface circuit. As shown, the critical components are the circuitry required to create the drive waveform and the switching network required to chopper-stabilize the amplifier. In this system, the AC input voltage is synthesized using two low-noise voltage references and a pair of analog switches. The control signal for those switches is provided by the data-acquisition card and is synchronized with the A/D clock. Both of the reference voltages are provided by resistively-biased LM199 Zener diodes. These devices were chosen because of their low temperature coefficients and their low $1/f$ noise contributions ([104]). The magnitude of the output voltage of both networks is reduced to the desired value of 150 mV by the inclusion of the resistive divider formed by R_{B2} and R_{B3}.

Because the low-frequency noise generated in the voltage reference networks is returned

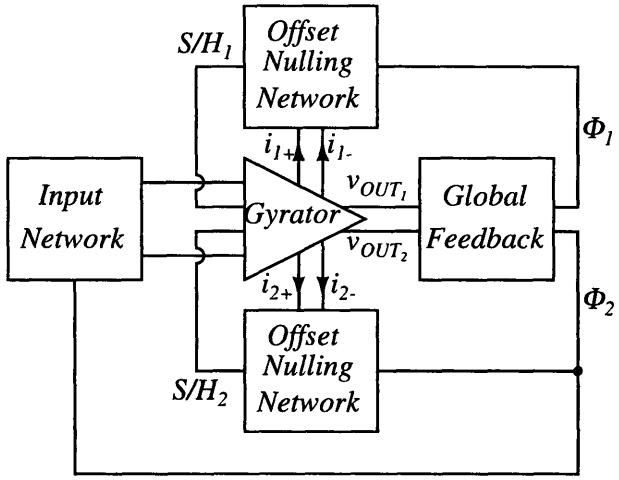


Figure 4-15: Block diagram of the hourglass integrator. The input network is comprised of the sensor and its bias circuit as well as the full bridge input switches. The two integrator outputs are monitored by the global feedback network, which provides control signals to the input switches as well as to the offset-nulling circuits. Each of the offset-nulling networks monitors the offset voltage of one of the two operational amplifiers in the gyrator. The currents i_{n+} and i_{n-} are fed to the balance inputs of the op amp in order to reduce their offsets.

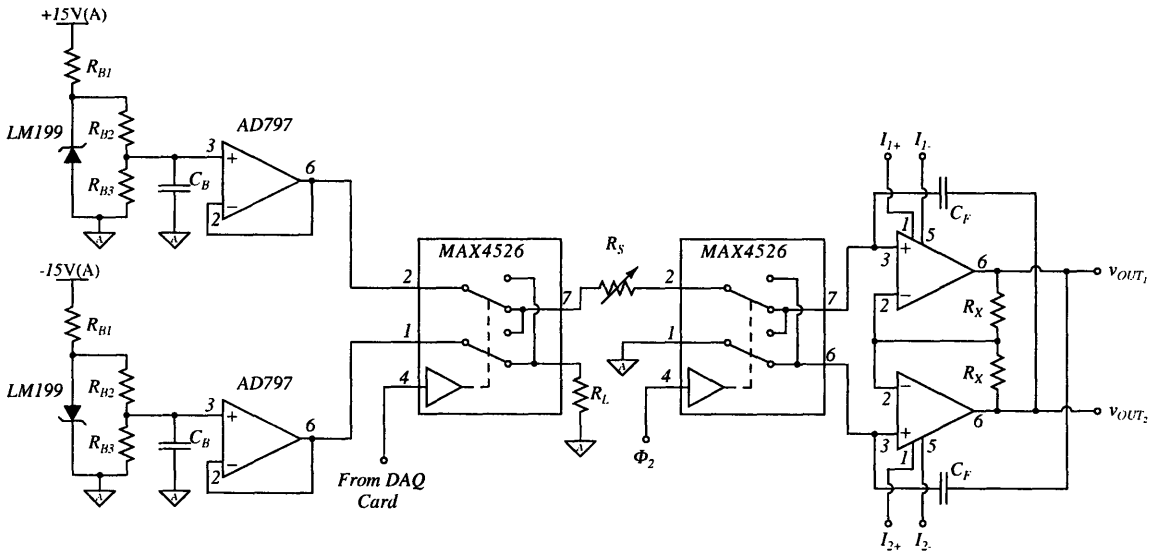


Figure 4-16: Input network for the "hourglass" integrator.

to baseband by the lock-in demodulation step, it is important to minimize the low-frequency noise present at the output of those two circuits. Given the extremely low value of the dynamic impedance of the LM199, both of the reference networks have an output-referred voltage noise with a power spectral density (PSD) of the form

$$\begin{aligned} S_{V_{ref}}(f) &= \frac{\overline{e^2_{n,V_{Ref}}}}{\Delta f} \\ &= \left(\frac{R_{B3}}{R_{B2}+R_{B3}}\right)^2 \left|\frac{1}{1+j\omega\tau}\right|^2 \left(\frac{\overline{e^2_{nd}}}{\Delta f} + \frac{\overline{e^2_{nR_2}}}{\Delta f}\right) \\ &+ \left(\frac{R_{B2}}{R_{B2}+R_{B3}}\right)^2 \left|\frac{1}{1+j\omega\tau}\right|^2 \frac{\overline{e^2_{nR_3}}}{\Delta f} \\ &+ \frac{\overline{e^2_n}}{\Delta f}, \end{aligned} \tag{4.21}$$

where e_n is the input-referred noise voltage of the op amp; e_{nd} is output noise of the Zener diode; e_{nR_2} and e_{nR_3} are the noise voltages of R_{B2} and R_{B3}, respectively; and τ is the time constant of the input filter, i.e. $\tau = C_B\,(R_{B2} \| R_{B3})$. In order to minimize the low-frequency noise terms in Eq. 4.21, we used extremely low-noise AD797 op amps and low noise index thin-film chip resistors.

The only major note related to the input switching network is that of the charge injection that occurs when the integrator changes state. According to the specifications for the MAX4526, it is typical for approximately 2 pC of charge to be injected forward into the integrator input terminals during a transition ([109]). Given the relatively large value of the feedback capacitors (i.e. 100 pF), a charge packet of that size has a minimal effect on the voltage across them. Charge injection can be effectively removed in this case, however, by exploiting the specifics of the circuit topology. The technique used here is based on the fact that the amount of charge injected from an analog switch depends upon the signal voltage ([110]). Thus, charge injection can be significantly reduced by biasing the signal paths of the MAX4526 to a voltage at which very little charge injection is observed. This biasing scheme is relatively simple to achieve through an asymmetric adjustment of the power supply voltages ([105]). As shown in the next section, the state changes of the integrator are synchronized to the A/D clock, making it possible to reduce any transients that still remain in the output waveform as a result of charge injection ([105]).

Global Feedback Circuitry

The basic functionality of the global feedback network needed to control the operation of the chopper-stabilized integrator is provided by the circuit shown in Fig. 4-17. In that circuit, each of the two gyrator output voltages is compared to a positive reference voltage using an LM311. Hysteresis is essential in this system, as multiple transitions at the output of the comparator could force the gyrator circuits to saturate. In order to achieve a sufficient amount of hysteresis, positive feedback is applied to one of the LM311's offset-adjust terminals ([111]). This technique has been found to outperform conventional positive feedback in this application.

As shown in Fig. 4-17, both of the comparator outputs are fed forward to a D flip-flop clocked by the signal that controls the rate of the analog-to-digital conversion process. The two flip-flop outputs are input to an OR gate that toggles the final state register. Whenever one of the gyrator outputs exceeds its preset threshold, the signals Φ_1 and Φ_2 both change state, triggering events in three different subsystems. First, the input switches change their position so that the signal current is passed to the other op amp. Additionally, one of the offset-nulling networks transitions from sample mode to hold mode, while its companion performs the complementary operation. Finally, the software that unwraps the integrator output waveforms reverses the polarity of the chopping signal $m_2[n]$.

Although the circuit shown in Fig. 4-17 provides the basic functionality required by the global feedback network, one addition must be made if the integrator is to operate successfully in the presence of a purely AC input. Consider the series of events that transpires if the input waveform reverses its polarity shortly after the integrator changes state. As an example, assume that v_{out_1} has just exceeded its upper threshold, thus forcing the output of the upper comparator in Fig. 4-17 to drop low. Accordingly, the integrator will change state on the next clock edge. Until that edge arrives, however, v_{out_1} will continue to increase. When the clock finally transitions, the signal current will pass to the other input terminal, thus forcing v_{out_1} to begin to decrease. A problem arises in this situation, however, if the polarity of the input waveform transitions on the same clock edge. If it does, the result will be to cause the two op amp outputs to begin to increase again. Unfortunately, the output of the upper comparator will still be low, so the integrator outputs will continue to diverge until the op amps saturate.

In order to avoid the problematic situation described above, the feedback circuit was

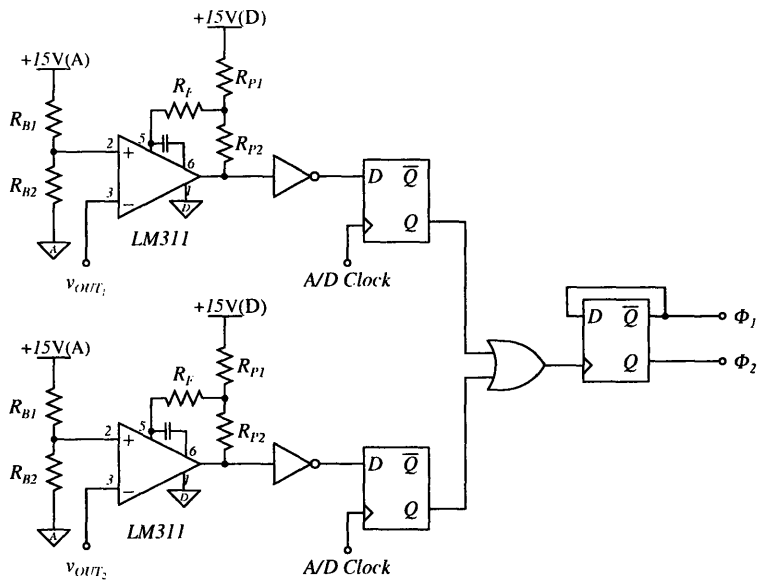


Figure 4-17: Schematic showing the basic functionality required in order to control the hourglass integrator. Note the use of both digital and analog ground referencing in this circuit.

modified as shown in Fig. 4-18. In that implementation there is a second set of comparators with threshold voltages that are set slightly higher than those of the original comparators. The thresholds for the new comparators are set such that transitions only occur when the integrator outputs begin to diverge. If that situation arises, the appropriate comparator output drops low. Accordingly, a short negative pulse is generated at the output of the one shot. Since the diverging output condition forces the output of the OR gate labeled OR_1 to be stuck in a high state, the short negative pulse from the one shot causes a similar pulse to appear at the AND gate output. As a result, the state of the integrator changes and operation returns to normal. In order to guarantee that the outputs never exceed the limits of the A/D converter, care must be taken when setting the thresholds of the two sets of comparators. The integrator has never saturated during a testing period when using this scheme.

Offset-Nulling Circuitry

Although the use of lock-in sensing greatly reduces the effects of low-frequency offsets, the difference between the offset voltages of the two op amps in the gyrator circuit can still be problematic. This difference causes the charging currents in the two feedback capacitors to have different values in response to the same input current. For this reason, the offset

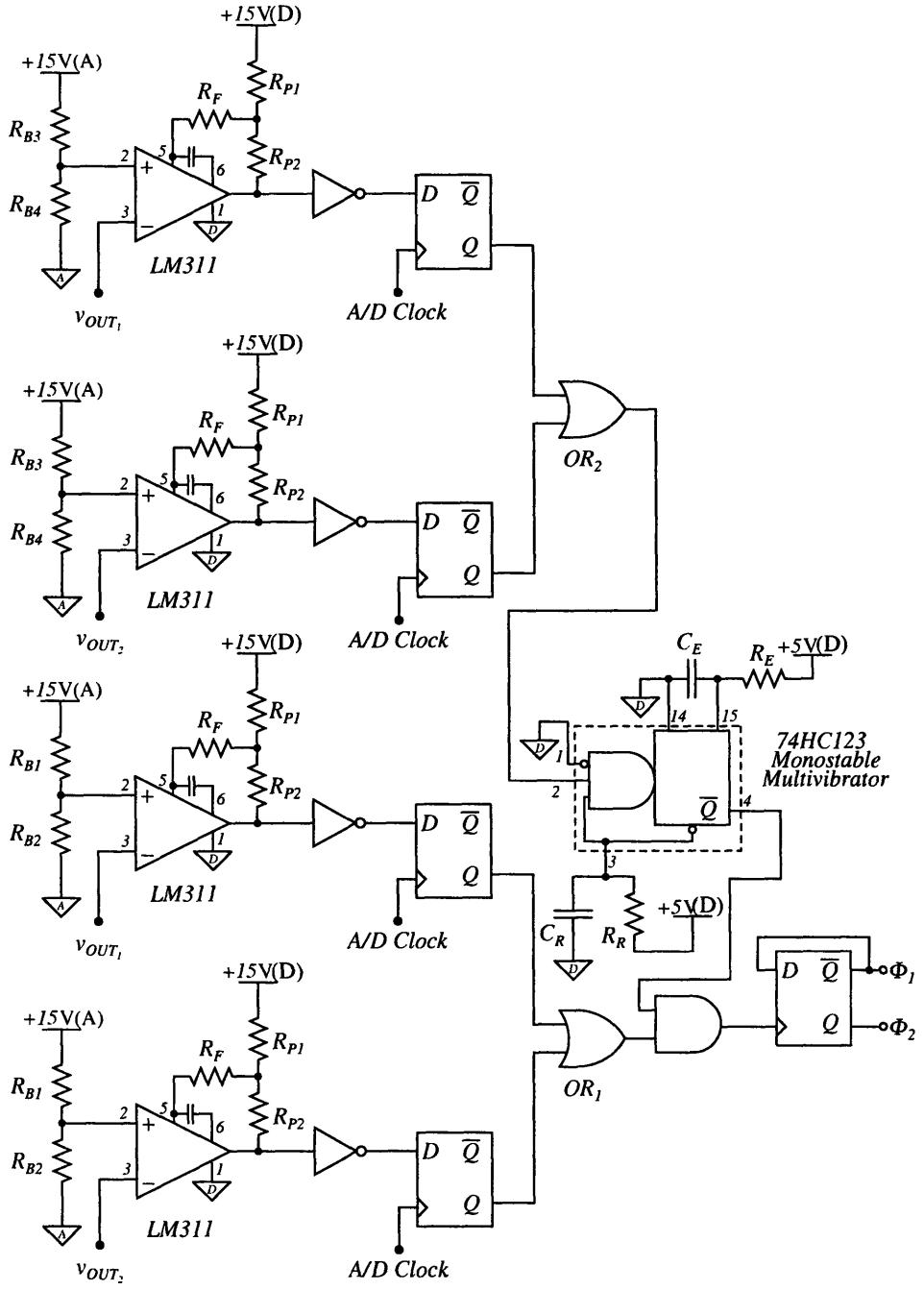


Figure 4-18: The complete feedback circuit used to control the operation of the hourglass integrator. Note the addition of the one-shot and the AND gate at the clock input of the state register. Any time that the integrator outputs begin to diverge, the output of one of the additional comparators will transition from high to low, creating a short negative pulse at the output of the one shot.

voltages of the two op amps must be reduced.

The offset-nulling approach used here exploits the fact that the offset voltage of each op amp can be sampled without perturbing the signal inputs ([105]). To see how this works, consider the circuit shown in Fig. 4-13 during the period when Φ_1 is asserted. Throughout that interval, the non-inverting input of the lower op amp is connected to analog ground. As a result, the voltage at the inverting terminal is equal to the op amp's offset voltage. By sampling the voltage at that node, it is thus possible to null the op amp's offset. When the integrator changes state so that Φ_2 is asserted, the nulling voltage for the lower op amp is held and a similar sampling process begins for the other op amp.

The circuitry used to implement offset nulling is shown in Fig. 4-19. During the sampling period, one of the analog switches in the MAX301 connects the inverting terminal of the desired op amp to the input of the nulling amplifier, which is a ground referenced integrator. In order to ensure that the nulling amplifier contributes a negligible offset, it is formed using a chopper-stabilized LTC1052, which is specified to possess a maximum offset voltage of 5 μV ([112]). Given the low leakage of the LTC1052 and the use of 1 μF integrating capacitors, the drift of the nulling circuit is negligible over typical hold intervals. In order to eliminate parasitic offset effects due to dissimilar contacts, the nulling amplifier has symmetric connections to its two input terminals. The output of the nulling amplifier feeds a differential pair that is biased using a $50\mu A$ current source.

4.3.3 Demodulation and Differentiation

In order to complete the chopper-stabilized amplifier, the integrator output must be demodulated and differentiated. As shown below, digital techniques are well suited for performing these operations because of the relatively limited bandwidth requirements imposed in this application.

Implementing the derivative operation in the discrete-time domain typically amounts to designing an FIR filter whose frequency response matches closely to that of an ideal differentiator over the frequency range of interest. Although window-based methods are commonly used for this purpose ([113], [54]), the limited bandwidth required in this application allows the derivative to be approximated using a simple first-difference filter. For a sampling frequency f_s that filter is defined as ([113])

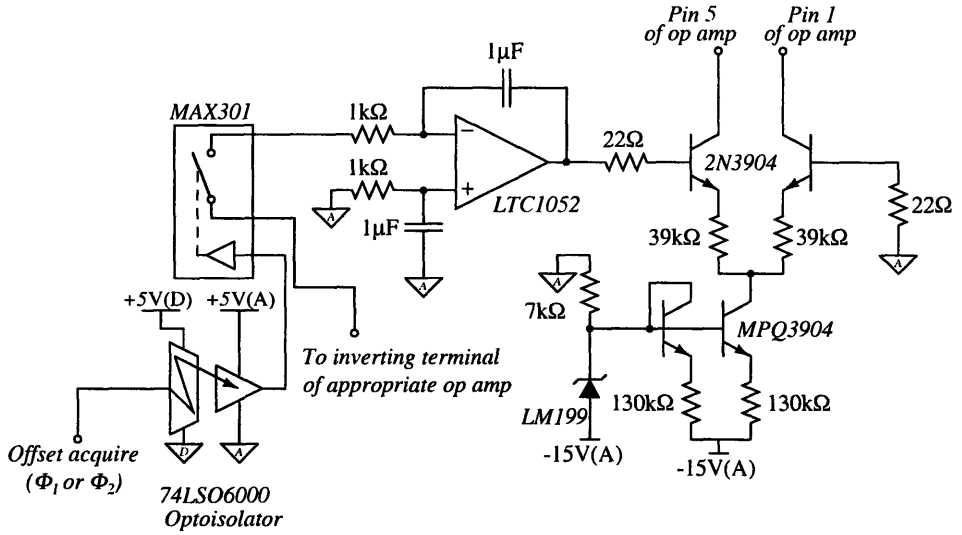


Figure 4-19: Offset-nulling circuitry for one of the operational amplifiers in the gyrator circuit. The optoisolator is used to separate the analog and digital ground planes. Both op amps have similar nulling circuits.

$$y[n] = f_s\left(x[n] - x[n-1]\right). \tag{4.22}$$

Fig. 4-20 compares the magnitude of the frequency response of the first-difference filter to the magnitude of the response of the ideal differentiator. In order to limit the detrimental effects that result from the spectral weighting of high-frequency noise, the differentiator response is appropriately band-limited. Note that over the frequency range of interest, the magnitude of the transfer function of the first-difference filter, which is

$$|H\left(e^{j\omega}\right)| = 2f_s|\sin\left(\frac{\omega}{2}\right)|, \tag{4.23}$$

is nearly identical to the magnitude of the frequency response of the ideal filter. In order to implement the required band-limiting, a low-pass FIR filter follows the first-difference operation ([105]). The design of that filter is discussed below.

The particularly attractive feature of the first-difference approximation is that it greatly simplifies the unwrapping process. With this filter, the integrator output can be unwrapped simply by alternating the sign of the two filter coefficients each time the integrator changes state. A higher-order digital differentiator would require a far more complicated demodulation process and would provide little or no performance enhancement ([105]).

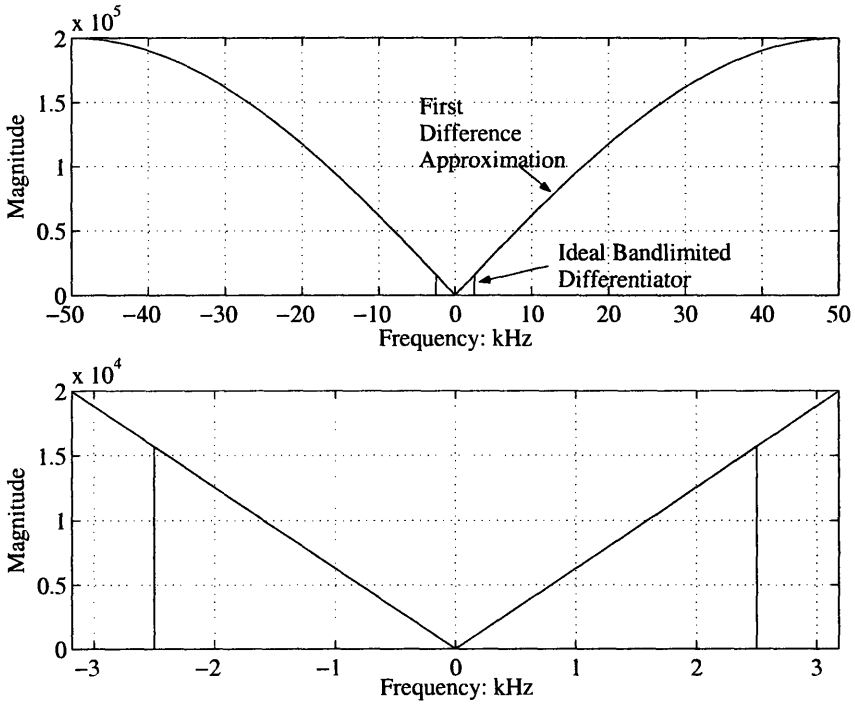


Figure 4-20: The magnitude of the frequency response of both an ideal band-limited differentiator and a simple first difference filter. The details over the band of interest are shown in the bottom trace. Note that the ideal response is limited to the passband of the first anti-aliasing filter shown in Fig. 4-21.

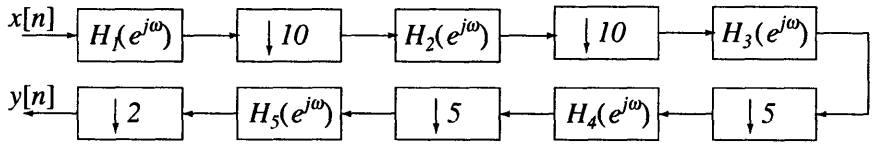


Figure 4-21: Five decimators follow the first difference filter. With the given downsampling ratios, the output signal $y[n]$ is sampled at 20 Hz.

Once the integrator output has been unwrapped and passed through the first-difference filter, it is demodulated in-phase with a sampled and scaled version of the AC waveform applied to the sensor network. In this way, the low-frequency variations in resistance are returned to baseband and the problematic flicker noise is transposed to a higher frequency. Following demodulation, the output signal is filtered and downsampled. These operations are described below.

In order to sample the integrator output with sufficient temporal resolution, the A/D converter's sampling clock is set to 100 kHz. Although the specifics of the hourglass dictate the use of this data rate, it is far greater than what is needed to monitor changes in airborne ozone concentration. Additionally, since testing periods can last for minutes or even hours, this oversampling operation creates a tremendous storage burden. In order to alleviate this problem, the differentiator output is heavily decimated, reducing its data rate to a frequency that preserves all of the relevant concentration information. The anti-aliasing filter required prior to downsampling also provides the band-limiting needed by both the differentiator and by the lock-in sensing process. Performing this filtering operation in only one step would be computationally taxing, however, as the required filter length is quite large. For this reason, the output signal is decimated in several stages as shown in Fig. 4-21. Prior to each downsampling by a factor of M, the signal is passed through an Mth-band FIR filter which has an impulse response satisfying the constraints presented in Eq. 4.24 ([114]).

$$h(Mn) = \begin{cases} 1/M, & \text{for } n = 0 \\ 0, & \text{for } n \neq 0 \end{cases} \tag{4.24}$$

By exploiting the zero coefficients in these filters, it is possible to reduce the overall computational burden by a factor of $1/M$ in comparison to a general symmetric FIR filter of the same length ([114]). The Mth-band filters employed here were designed using the

eigenfilter approach, a method that easily allows the user to incorporate time-domain constraints such as those in Eq. 4.24 ([115]). Eigenfilters are optimal in the sense that that they minimize the weighted mean-squared error between the desired response, $D\left(e^{\jmath\omega}\right)$, and the actual amplitude response, $H\left(e^{\jmath\omega}\right)$. This amounts to designing a filter that minimizes the objective function

$$E_{LS} = \int_R W\left(e^{\jmath\omega}\right)\left[D\left(e^{\jmath\omega}\right) - H\left(e^{\jmath\omega}\right)\right]^2 d\omega, \tag{4.25}$$

where $W\left(e^{\jmath\omega}\right)$ is a non-negative weighting function and R is the region $0 \leq \omega \leq \pi$ with the exception of the transition band ([115]).

With the decimation process implemented as shown in Fig. 4-21, five filters are required, and each filter is comprised of 49 coefficients. By comparison, if the signal were to be downsampled in a single step, an Mth-band filter would need well over 1000 points, even with relaxed specifications.

4.3.4 Amplifier Gain and Noise Analysis

In lock-in sensing applications, it is imperative that the amplifier has sufficient gain and a negligible noise contribution at the drive frequency. In this final section on the ozone detector, we demonstrate that the chopper-stabilized integrator meets both of these requirements, and we briefly describe the ability of the offset-nulling circuitry to reduce the low-frequency noise of the gyrator op amps.

An expression for the overall gain of the complete amplifier is obtained by considering the combined effects of the hardware-based integrator and the software-based differentiator. Recall that the amplifier input signal is the sensor current. Thus, from Eqs. 4.17 and 4.18, we can write the magnitude of the gain from one of the integrator input terminals to one of its output terminals as

$$\left|\frac{V_o}{I_s}\right| = \frac{1}{\omega C_F}. \tag{4.26}$$

Since the two integrator outputs are mirror images of each other, they can be sampled differentially, thus increasing the gain by a factor of two. Following conversion into the discrete-time domain, the integrator output is passed to the differentiator. The effect of this operation is to weight the transfer function by $\jmath\omega$. Thus, the overall gain of the amplifier is $2/C_F$.

In order to attain some perspective on the amount of gain achievable with an hourglass integrator, consider an example implementation. In this application, the feedback capacitors are 100 pF. With these components, the overall transimpedance gain is 2×10^{10} V/A. In the current configuration, the sensor is excited using a 300 mV_{pp} square wave at 1 kHz. At that level, both self-heating and amplifier slew are avoided, and the amplifier can detect resistance changes as small as 1 Ω throughout the 20 kΩ to 50 kΩ range.

The other key performance metric for a lock-in amplifier is that its noise contribution be negligible at the drive frequency. Thus, the input-referred noise current of the amplifier at 1 kHz should be dominated by the thermal noise current of the sensor. An expression for that current can be obtained by considering the schematic shown in Fig. 4-22, which displays the basic gyrator circuit with all of the relevant noise sources in place. In that model, the input-referred noise voltage of the op amps is represented by the sources e_{n1} and e_{n2}. Given the use of op amps with JFET inputs, these voltages contain both a white noise component and a $1/f$ component ([116]). In addition to the amplifier noise sources, the model also includes the noise of the sensor, e_{n,R_s}, and the noise generated by the quantization process, $e_{n,quant}$. For completeness, the model should also include the noise currents present at the op amp input terminals. These sources are omitted, however, since they contribute a negligible amount of noise at 1 kHz[3].

Recalling that the integrator-differentiator combination has an overall transimpedance gain of $2/C_F$, its input-referred noise current has a PSD of the form

$$\begin{aligned} S_{IN}(f) = &\left(4\pi^2 f^2 C_F^2 + \frac{1}{R_s^2}\right)\left(\frac{\overline{e_{n,1}^2}}{\Delta f} + \frac{\overline{e_{n,2}^2}}{\Delta f}\right) \\ &+ \frac{1}{R_s^2}\frac{\overline{e_{n,R_s}^2}}{\Delta f} \\ &+ \left(\pi^2 f^2 C_F^2\right)\left(\frac{\overline{e_{n,quant}^2}}{\Delta f}\right). \end{aligned} \tag{4.27}$$

In order to determine the overall noise contribution at 1 kHz, each of the terms in Eq. 4.27 must be considered in turn. Since we are operating beyond the $1/f$ noise corner of the

[3]Input-referred noise currents in JFETs are extremely small at low frequencies, but they rise at higher frequencies as the thermal noise of the channel is capacitively coupled to the gate ([116]). In the devices used here, (i.e. the OPA627) this current is specified to be approximately 2.5 $fA/\sqrt{Hz}$ at 1 kHz ([117]). At that level, these currents are several orders of magnitude less than any other noise source.

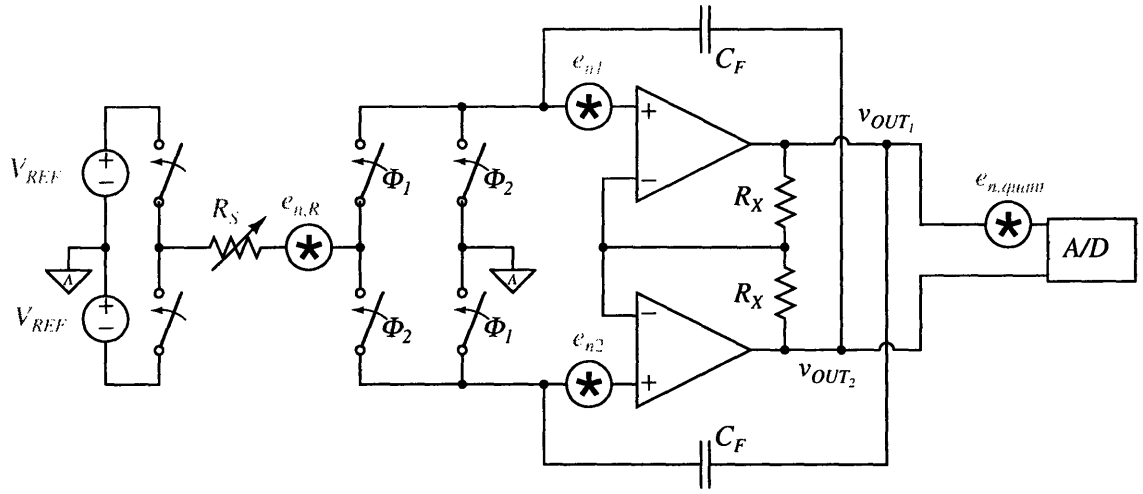


Figure 4-22: Gyrator circuitry showing the sensor noise source and all of the relevant amplifier noise sources. This model is used to compare the input-referred noise of the amplifier to the noise of the sensor.

sensor, its contribution is due primarily to its thermal noise, which has a PSD of the form

$$S_{th}(f) = \frac{\overline{e_{n,Th}^2}}{\Delta f} = 4kTR_s, \tag{4.28}$$

where k is Boltzmann's constant and T is absolute temperature ([104]). At the sensor's nominal resistance of 30 kΩ and operating temperature of 430° C, it contributes 1.3×10^{-24} A^2/Hz. The next set of sources to consider are the input-referred noise voltages of the two gyrator op amps. Since 1 kHz is just beyond the specified $1/f$ noise corner of the OPA627 operational amplifiers used here, they provide a minor amount of residual flicker noise ([117]). Regardless, the overall voltage noise of both op amps is extremely low, with each contributing a spot noise of 5.2 $nV/\sqrt{Hz}$ ([117]). When interfaced with the 30 kΩ sensor and 100 pF feedback capacitors, the op amps add a combined input-referred noise current of 6×10^{-26} A^2/Hz at 1 kHz. The final noise source to consider is $e_{n,quant}$. Assuming that the uncertainties arising from the quantization procedure can be modeled as a white noise process, they contribute a voltage noise having a power spectral density of the form

$$S_{quant}(f) = \frac{\overline{e_{n,quant}^2}}{\Delta f} = \frac{\left(\frac{\Delta V}{2^N}\right)^2}{12 f_s}, \tag{4.29}$$

where ΔV is the full-scale range of the converter, N is the total number of bits, and f_s is

Table 4.3: Summary of noise sources for $R_s = 30\text{k}\Omega$

Noise Source	Term in Eq. 4.27	Spot Noise at 1kHz
Amplifier Noise Terms		
$e_{n,1}$	$\left(4\pi^2 f^2 C_F^2 + \frac{1}{R_s^2}\right) \frac{\overline{e_{n,1}^2}}{\Delta f}$	3×10^{-26} A^2/Hz
$e_{n,2}$	$\left(4\pi^2 f^2 C_F^2 + \frac{1}{R_s^2}\right) \frac{\overline{e_{n,1}^2}}{\Delta f}$	3×10^{-26} A^2/Hz
$e_{n,quant}$	$\left(\pi^2 f^2 C_F^2\right) \frac{\overline{e_{n,quant}^2}}{\Delta f}$	7.7×10^{-27} A^2/Hz
Sensor Noise		
e_{n,R_s}	$\frac{1}{R_s^2} \frac{\overline{e_{n,R_s}^2}}{\Delta f}$	1.3×10^{-24} A^2/Hz

the sampling frequency ([118]). Since the global feedback circuitry limits the excursions of the individual integrator outputs to approximately ± 4.75 V, a reasonable conversion range is 20 V[4]. With a 16-bit converter sampling at 100 kHz, the resulting spot noise at 1 kHz is approximately 280 nV/$\sqrt{\text{Hz}}$. When referred to the input, quantization will result in a noise current that rises with increasing frequency. This is not problematic, however, since we are only interested in a very narrow bandwidth centered around 1 kHz. At that frequency, quantization adds an input-referred noise current of approximately 7.7×10^{-27} A^2/Hz.

Table 4.3 summarizes the theoretical noise contribution from each the relevant sources described above. Clearly, the total noise is dominated by that of the sensor, demonstrating that the chopper-stabilized integrator can provide significant gain without adding significant noise.

As mentioned previously, the low-frequency noise of the hourglass integrator is impacted by the offset-nulling circuitry. Although the details of this reduction mechanism are reasonably complex, the basic principle is rather intuitive. Essentially, the process is similar to that encountered in continuous-time auto-zeroed amplifiers ([119], [120], [121], [122]). In those devices, the offset-nulling circuitry periodically samples the voltage at the amplifier's input node. When it does so, it also samples the amplifier's input-referred noise voltage. For noise sources that are much slower than the auto-zeroing rate, subsequent samples will be

[4]Recall that the integrator outputs are sampled differentially.

highly correlated. From the perspective of the offset-nulling circuit, therefore, low-frequency noise is indistinguishable from DC offsets, and the circuit will null them in the same manner that it attempts to null offsets ([119], [123]). The effect of this noise reduction is not to remove the $1/f$ component completely; rather, the auto-zeroing flattens the noise PSD for frequencies within the bandwidth of the offset-nulling circuitry ([123]). Because lock-in sensing is used in this application, the noise reduction provided by the the offset-nulling process is of little significance. Nonetheless, it is described here because one can imagine other applications of the hourglass integrator that do not require lock-in sensing.

4.4 Supervisory Control

Intelligent supervisory control is required in order to coordinate the operation of each of the individual components in the airflow monitoring system. This is achieved using the remote sensor monitor (RSM). This section describes the functions performed by the RSM, and it also explains how the remote stations are programmed to respond.

Before any measurements begin, the RSM prepares a series of command files to send to each unit in the system. These files are created by software that gathers information from the system operator about upcoming tests. Two different sets of files are generated from this information. The first of these is a series of configuration files, with one file created for each station. The file for a specific unit contains both the start time for the test and a command indicating the mode in which the unit is to operate. Additionally, the RSM's software package also creates new control software for each of the units that are to function as generators. These new control programs are executable files that allow the PIC microcontroller to create the desired series of high-voltage pulses. Several minutes prior to the start of any test, the appropriate files are transferred to the remote stations over the power line.

At the beginning of each minute, a daemon running on each of the remote PCs checks to see if any new configuration files have been uploaded. If so, the daemon executes a PERL script that parses the new configuration file. If the mode of operation is determined to be generation, then the software on the remote PC looks for the new control program and uploads it to the PIC using an RS-232 data link. Once that transfer is complete, both the computer and the PIC wait until the appointed start time. When that time arrives, the

computer sends a command to the PIC instructing it to begin executing its new control software. If it is desired to stop the test prior to the pre-specified time, the computer can send an emergency stop sequence to the PIC instructing it to immediately cease ozone generation. Otherwise, the PIC will continue to generate ozone until it has performed the pre-determined number of ozone pulses.

For those units which operate in detection mode, the control procedure is similar. At the scheduled start time, these units begin to sample, process, and store the detector output signal. As soon as the current time matches the stop time provided in the configuration file, the control daemon on the detector computers stops data collection and transfers the resulting data file to the RSM.

4.5 *Measuring Bulk Airflow*

The primary goal of this system is to provide the information needed to enhance the performance of energy-saving ventilation techniques. For instance, the air speed and direction information provided by this system could be used to control airflow rates in naturally-ventilated buildings. This section describes how ozone is used to determine air speed and direction, and it also presents several example measurements.

4.5.1 Airflow in Indoor Spaces

As stated in Section 4.1, this system uses experimental arrangements similar to the one shown in Fig. 4-1a. The use of that configuration is predicated on the fact that convection is the the dominant transport mechanism for ozone molecules. The reasons for that stipulation are described in this section.

Consider a tracer gas that is present in a space and has a time- and position-dependent concentration denoted as $C(\mathbf{x}, t)$. In a turbulent flow, $C(\mathbf{x}, t)$ is viewed as a random signal that is the sum of a slowly-varying mean component, $\overline{C}$, and a high-frequency AC term, C' ([40]), i.e.

$$C(\mathbf{x}, t) = \overline{C} + C'. \tag{4.30}$$

Using the conservation of mass, a transport equation can be written for $\overline{C}$. This relationship is

$$\frac{\partial \overline{C}}{\partial t} + \nabla \cdot \left(\overline{\mathbf{u}}\overline{C}\right) = \nabla \cdot \left[\left(D_{mol} + D_{turb}\right) \nabla \overline{C}\right], \tag{4.31}$$

where $\overline{\mathbf{u}}$ is the mean velocity vector and D_{mol} and D_{turb} are the molecular and turbulent diffusivities, respectively ([39]). Note that turbulence must be included in this model because flows in most rooms are, indeed, turbulent ([40]). In fact, it is generally true that turbulent diffusion dominates molecular diffusion for all but nearly stagnant flows ([40]).

In order to understand the effects of Eq. 4.31. consider several limiting cases. First, if an impulse of ozone is injected into a large open space with no diffusion, then the impulse will simply move through the space at a speed equal to the average velocity of the air. Second, if an impulse of ozone is injected into a large open room with no convection, then the effect will be to create an expanding cloud of gas. For intermediate cases, the result will be a moving cloud that expands as it moves ([40]). For this system to be able to successfully estimate air speeds, it must be true that convective transport is dominant.

In order to address the diffusion issue, initial experiments were performed using a range of air speeds in an environment with a forced, uni-directional flow. For the experiments discussed in Section 4.5.3, three different air speeds were considered. In the low-speed tests, the average velocity measured at two different test locations was approximately 0.04 m/s. That value was chosen on the basis of several different considerations. First, the minimum detectable signal for most commercial velocity transducers is approximately 0.01 m/s, meaning that it would be difficult to compare our results to any other measurements if we were to consider velocities below that magnitude ([39]). Additionally, speeds on the level of 0.04 m/s are similar to those typically observed in many indoor facilities ([40]). Finally, an approximate analysis indicates that convective transport dominates at that speed. That analysis proceeds as follows. In a room with a characteristic length L, the diffusion time scale, t_D, is defined as

$$t_D = \frac{L^2 Sc_t}{D_{turb}} \tag{4.32}$$

and the convection time scale, t_C, is defined as

$$t_C = \frac{L}{U}, \tag{4.33}$$

where Sc_t is a dimensionless quantity known as the turbulent Schmidt number and U is a characteristic velocity in the room ([40]). A zeroth-order approximation for the turbulent diffusivity can be obtained using the equation ([124])

$$D_{turb} = \frac{0.03874UL}{Sc_t}. \tag{4.34}$$

For the test room considered here, which has a characteristic length of 10 ft and a characteristic velocity of 0.04 m/s, the above approximations yield a convective time scale of 76.2 s and diffusion time scale of 1967 s. Given the factor of 26 difference between t_D and t_C, it is reasonable to expect that diffusion will have some effect, but that it will be minimal in comparison to that of convection[5] ([125]). For the experiments conducted with velocities larger than 0.04 m/s, the same results hold.

4.5.2 Determining Direction and Estimating Velocity

The experimental procedures used during initial tests were designed to determine both the direction and speed of air movement in a single room. In the case of direction measurements, one or more ozone pulses were created by a specific generator and concentration was monitored by several detectors. Assuming that convective transport dominates, a larger change in concentration was expected at those detectors located downstream of the generator.

In order to obtain speed estimates, a time-of-flight procedure was developed. In that approach, measured concentration data was used to calculate the time-of-flight between a generator and a given detector. In this context, time-of-flight is defined to be the amount of time that elapses between the start of a high-voltage pulse and the time at which a corresponding change is observed at the detector. In the absence of sensor dynamics and both molecular and turbulent diffusion, this technique can be used to generate extremely accurate speed estimates. Unfortunately, both diffusion and sensor non-idealities impact the measured ozone concentration, so the time-of-flight measurements contain some additional error. Despite these shortcomings, the time-of-flight values are still quite useful. For instance, as shown in the next subsection, the measured time-of-flight at a given detector increases as the local air speed decreases. Additionally, different detectors display different time-of-flight numbers depending upon their position and the local air speed. Thus, these measurements can be used to create a qualitative map of the airflow in an indoor space.

[5]Better estimates of the effects of diffusion and convection can be obtained using computational fluid dynamics software that allows one to introduce specific transport equations for several key quantities related to the turbulence of the flow field ([40]). Published results demonstrate, however, that the zeroth-order model produces reasonable results ([124]).

Table 4.4: Summary of initial experiments

Test Set #	Flow Direction	Supply Velocity	Air Exchange Rate (ACH)	Arrangement
1	Left to Right	High	22.6	A
	Left to Right	Med	15.1	A
	Left to Right	Low	8.8	A
2	Right to Left	High	22.6	B
	Right to Left	Med	15.1	B
	Right to Left	Low	8.8	B
3	Left to Right	High	22.6	B
	Left to Right	Med	15.1	B
	Left to Right	Low	8.8	B

4.5.3 Example Measurements

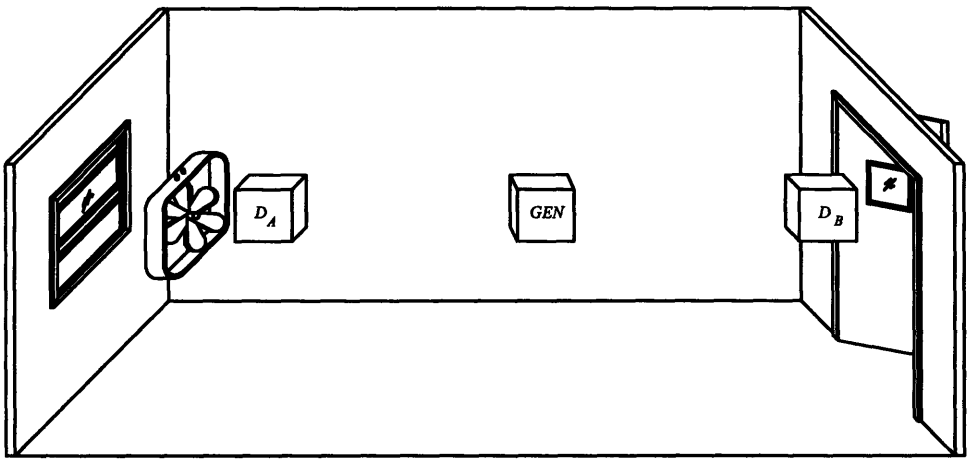
Initial testing of this system has been performed in rooms where convection is the dominant transport mechanism. Furthermore, the experimental conditions have been designed to be relatively similar to those that might be encountered in a naturally-ventilated facility. For example, several initial sets of experiments were performed in a small office with an open window at one end and an open doorway opposite[6] (Fig. 4-23). The speed and direction were controlled using a fan that was placed near the window. This configuration was used to simulate scenarios that are typically encountered in rooms with wind-induced cross-ventilation[7]. Table 4.4 summarizes the various experimental conditions used during initial tests. The air exchange rates listed there were estimated using the average velocity measured over the area of the open doorway[8].

The purpose of the first set of experiments listed in Table 4.4 was to test the ability of

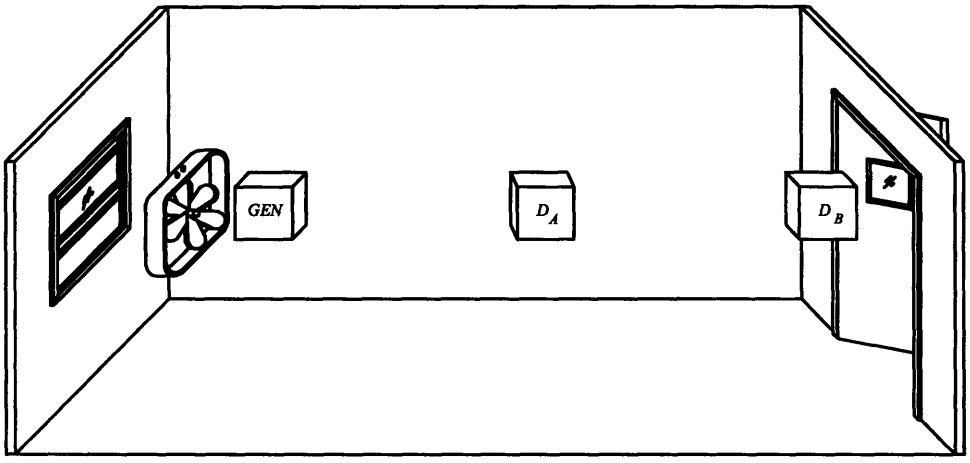
[6]The dimensions of the test room are 10 ft × 12 ft × 16 ft

[7]Since a fan was used, the flow is clearly not driven by wind-induced pressure gradients. The fan was needed for purposes of experimental control. Regardless, the observed flow patterns are rather similar ([39]).

[8]The air exchange rate, N, in a room or sealed chamber is defined as $N = Q/V$, where V is the volume of the room and Q is the volumetric flow rate ([39]). The value for Q was estimated in this case by measuring the average velocity across the area of the doorway. Air exchange rate is typically specified in units of air changes per hour, or simply ACH ([39]).



(a)



(b)

Figure 4-23: The two experimental arrangements used for the tests described in this section. In configuration (a) the generator is placed between two different detectors. In configuration (b) the positions of the generator and detector D_A are switched. In both arrangements shown here, the generator and detectors are placed along a line, with 6 ft between neighboring stations. During each experiment the fan speed and direction are varied in order to control flow.

Table 4.5: Summary of results from test set 1

Test Set #	Supply Velocity	Ave. Resistance Change (Ω)	
		D_A	D_B
1	High	N/A	126
	Med	23	398
	Low	68	533

the system to make direction measurements. During these tests, we used the configuration shown in Fig. 4-23a, with air moving from left to right. By design, convective transport should dominate in these experiments, meaning that the majority of the ozone should be carried into the vicinity of the downstream detector. At each fan speed, ozone was generated using a series of ten high-voltage pulses, with each pulse lasting approximately 14 minutes. Fig. 4-24 displays the detector responses resulting from one of the high-voltage pulses created during the low-speed experiments. Clearly, a short time after the voltage pulse begins, a change in ozone concentration is observed at the downstream detector. As expected, a far less significant change is observed at its upstream counterpart.

For each pulse, an average change in sensor resistance was calculated using a simple change-of-mean scheme. In this approach, the data was first processed using a one-minute moving-average filter. A change was declared to begin whenever the filter output exceeded its previous value by a preset threshold and then maintained that difference over the next 60 seconds. Once a change was observed, we computed the average value of the sensor resistance over the ten minute interval preceding the change and we calculated a similar average for the ten minute interval following it. The difference between those values was calculated, and the results were averaged over the course of ten pulses. Table 4.5 summarizes the results obtained for each of the three different fan speeds listed in Table 4.4. Note that the changes observed at the upstream detector were consistently smaller than those measured downstream. During the high-speed experiments, the change was not even large enough to trigger the event-detection scheme.

The second set of experiments was also intended to test the ability to measure direction. The experimental arrangement used in these experiments is shown in Fig. 4-23b, with air moving from right to left. Since the air speed was chosen so that convection would again be the dominant transport mechanism, it was expected that little or no change would be

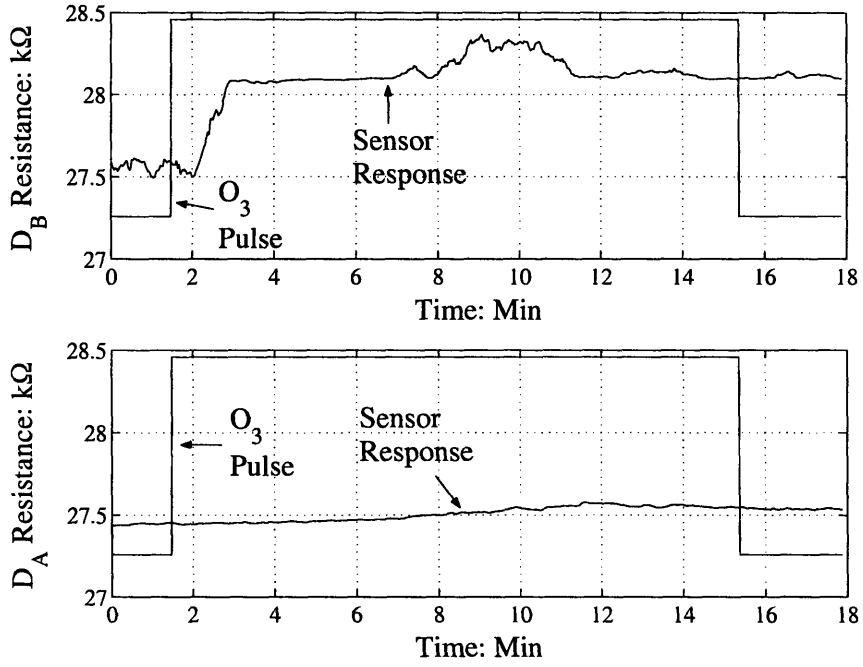


Figure 4-24: The responses of two detectors when ozone is injected between them. Detector D_B is located downstream of the generator, and detector D_A is located upstream. During this test, the fan was set to its low speed.

observed at either detector in response to an ozone pulse. In this case, ozone was again created by energizing the high-voltage electrode with a series pulses lasting approximately 14 minutes each. As expected, no appreciable change was noticed at either detector in the case of the high-speed experiments. At low speed, a slight increase was noticed at the detector nearest the generator, but the change was only 11 Ω.

A third set of experiments was designed to determine if the system can also estimate speed. The configuration used during these tests is the one shown in Fig. 4-23b, with air moving from left to right so that ozone molecules are carried into the vicinity of both detectors. With this design it is actually possible to test two important functions. The first of these is the ability of the system to measure changes in speed at any one location, and the other is the ability of the system to detect differences between the speeds at two different points in the same room.

The protocol for the third set of experiments was similar to that used during the first two sets. Once again, a series of ten high-voltage pulses was created by the generator, and each pulse lasted for approximately ten minutes. At the end of each test, time-of-flight values were computed for each of the ten pulses.

Table 4.6 summarizes the results of the third set of experiments. An estimated velocity value was computed for each of the ten voltage pulses performed at each fan speed using the measured time-of-flight and the known distance between the detector and the generator. The estimates presented in Table 4.6 are the averages of the ten values obtained at each speed. Several trends are clearly evident in these results. First, the estimated values consistently decrease as the actual air speed decreases. Additionally, each of the estimates matches closely to the actual velocity measured by a hot-wire anemometer.

Table 4.6: Summary of results from test set 3

Supply Velocity	**Velocity at D_A (m/s)**		**Velocity at D_B (m/s)**	
	Estimated	Actual*	Estimated	Actual*
High	0.82	0.86	0.62	0.60
Med	0.18	0.20	0.12	0.13
Low	0.033	0.04	0.034	0.04

* The hot-wire anemometer used for comparison purposes has an accuracy of 0.01 m/s.

Chapter 5

Non-Intrusive Load Monitoring Based on Voltage Distortion

Since the 1970s, there has been considerable discussion over the potential of demand-side energy management (DSM). Originally, this idea was proposed in an effort to reduce the need for new generating capability by increasing efficiency at the point-of-load ([126]). Over the years, however, many other potential uses have been considered, including load levelling and the implementation of liberalized market schemes ([127]). Recently, there has also been discussion about how demand-side management can help to prevent problems related to the age of the power grid ([127]).

In order to implement demand-side management strategies, energy monitoring systems such as the NILM were developed ([30]). Because of its relative simplicity, the NILM is very attractive to many energy consumers, particularly those in large industrial and commercial facilities. Unfortunately, however, the NILM is still not attractive to many small consumers, as its installation requires the services of a trained electrical professional. In order to deliver the benefits of the NILM to these consumers, a new approach is needed.

This chapter presents a monitoring strategy that further reduces the number of sensors required by the NILM. In particular, this system operates using voltage measurements obtained at a single electrical outlet, thus making monitoring as simple as "plugging in" an appliance. This device is able to function because the transient currents drawn by different classes of electrical loads distort the voltage waveform in ways that a computer can be trained to recognize. Because it uses only voltage measurements, this device is referred to as a voltage-based NILM (V-NILM).

This chapter begins by introducing the theoretical background upon which the new

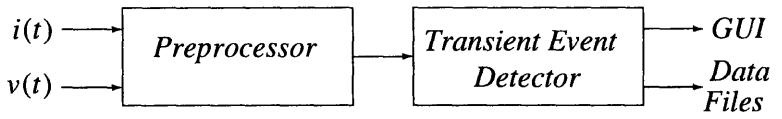


Figure 5-1: System-level block diagram of the front-end of a standard NILM. The system described in this chapter uses a similar overall architecture, but its only input is the voltage measured at a single electric service outlet.

system is based. Subsequently, it describes the preprocessor that computes estimates of the spectral content of the measured voltage waveform. The capability of this preprocessor is demonstrated using several field measurements. Finally, this chapter concludes by summarizing the results and by describing several areas of future research.

5.1 Spectral Envelope Estimation

Recall that in standard approaches to non-intrusive monitoring, loads are detected using the two-step procedure shown in Fig. 5-1. In the first step, a preprocessor computes estimates of the short-time spectral content of a measured current waveform ([49], [32], [50]). Subsequently, the spectral estimates created by the preprocessor are passed to a software module that identifies the operation of individual loads by matching stored templates to transient patterns observed in the preprocessed data stream ([32]). In the V-NILM, the procedure is similar. The only difference is that the preprocessor performs a spectral decomposition of the measured line voltage.

The preprocessors found in both the standard NILM and the V-NILM determine the time-varying coefficients of the Fourier series presented in Eq. 2.1, which is repeated here for convenience:

$$x\left(t-T+s\right)=\sum_{k}a_k(t)\cos(k\frac{2\pi}{T}(t-T+s))+\sum_{k}b_k(t)\sin(k\frac{2\pi}{T}(t-T+s)).$$

In the traditional NILM, which has access to both voltage and current, the coefficients $a_k(t)$ and $b_k(t)$ are estimated using a two-step process. First, the basis sinusoids in Eq. 2.1 are synchronized to the local utility voltage. Subsequently, the NILM approximates $a_k(t)$ and $b_k(t)$ by computing estimates of the integrals presented in Eq. 2.2 and 2.3. Recall that those

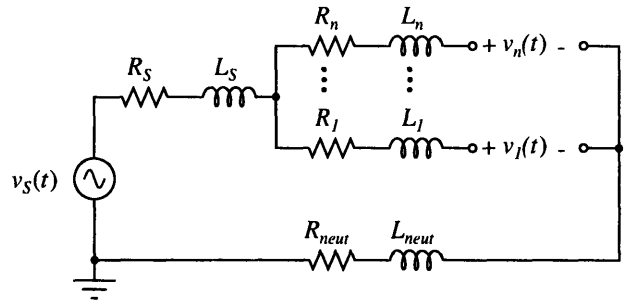


Figure 5-2: A distribution-level model of the electric utility.

integrals are

$$a_k(t) = \frac{2}{T}\int_0^T x(t-T+s)\cos(k\frac{2\pi}{T}(t-T+s))ds$$

and

$$b_k(t) = \frac{2}{T}\int_0^T x(t-T+s)\sin(k\frac{2\pi}{T}(t-T+s))ds.$$

In the V-NILM, the computation of the a_k and the b_k proceeds in a similar manner. Instead of analyzing a current, however, the V-NILM performs a decomposition of the line voltage measured at a single electrical outlet. This approach is based on the utility model presented in Fig. 5-2. In that model, single-phase loads draw current from a source $v_s(t)$ through lumped parmaters R and L that model the effects of physical cabling, protection devices, and other impedances. Physically, v_s might corresponds to the voltage on the bus bars of a breaker cabinet, meaning that it may be shaped by power-factor correction capacitors, upstream voltage distortion, and the aggregate current drawn by the loads. The terminal voltages $v_k(t)$ represent the line-to-neutral voltages at individual service outlets. For completeness, this model includes impedances in each branch circuit as well as impedances in the return path[1].

In the V-NILM, the waveform to be analyzed is one of the voltages $v_k(t)$ shown in Fig. 5-2. Whenever a load is either connected to or disconnected from any of the terminals shown in that figure, the voltage at the measurement node, which is denoted as $v_m(t)$, will

[1]Numerous researchers have found the model in Fig. 5-2 to be reasonably accurate provided that one considers a frequency range consisting of several low order harmonics of the fundamental power system frequency ([128], [10], [129], [130]). In fact, the authors of [128] and [10] used this model to estimate the impedance of the local distribution network for frequencies extending to about the 15^{th} harmonic of 60 Hz.

change as a result of the harmonic currents drawn through the parasitic impedances. For example, Fig. 5-3 shows the line voltage measured at one outlet in the laboratory both before and after the connection of a vacuum cleaner at another outlet[2]. Note that when the vacuum's motor is accelerating from rest, the magnitude of the current is rather large, and, correspondingly, the magnitude of the voltage is significantly reduced. Furthermore, as the motor approaches steady-state, the magnitude of the current slowly decreases, forcing the amplitude of the voltage to rise slowly. As this example demonstrates, voltage distortion is related to load current, and, thus, it is also related to the physics of the load. The purpose of the preprocessor in this application is to measure the relatively small changes in the amplitude and phase of the voltage waveform. The details of the voltage-based preprocessor are discussed in the next section.

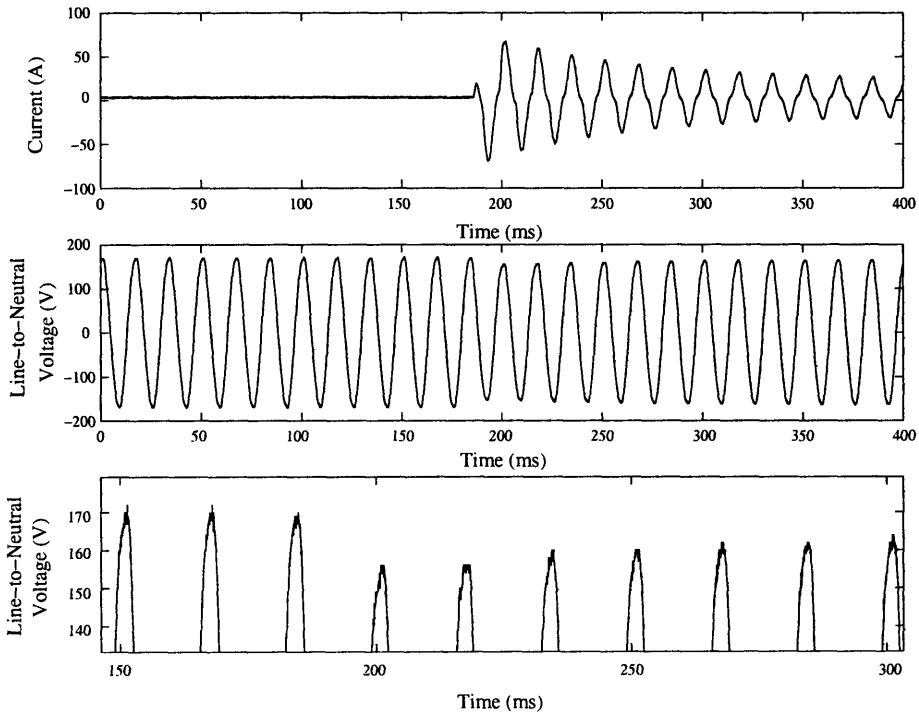


Figure 5-3: Distortion in the measured line voltage as a result of the connection of a 4 hp vacuum cleaner. The top trace shows the load current drawn by the vacuum (for reference), the middle trace shows the line voltage at a nearby outlet, and the bottom trace shows the details near the peaks of the line voltage.

[2]The vacuum cleaner in question, which is rated at 4 hp, has a universal motor ([131]).

5.2 Spectral Envelope Preprocessor Implementation

In the preprocessor prototype, a set of spectral coefficients denoted as $a_{k,AN}(t)$ and $b_{k,AN}(t)$ are computed using the synchronous detection scheme outlined in block diagram form in Fig. 5-4. As shown, there are currently four channels of analog monitoring circuitry. These four channels compute estimates of the in-phase and quadrature voltage distortion at both the fundamental frequency and its third harmonic[3]. The final analog signal on each channel is sampled using a Tektronix TDS3014B oscilloscope, and subsequent discrete-time processing is performed in MATLAB. All of the relevant operations are introduced in this section. Circuit schematics and other specific implementation details are presented in Appendix D.

The variable $v_m(t)$ in Fig. 5-4 represents an analog signal that is a scaled and isolated version of the line-to-neutral voltage at the measurement point. This scaled and isolated waveform is produced by a Hall-effect transducer not shown in the figure. This voltage is approximately a sinusoid at the fundamental line frequency, and it is arbitrarily assumed to be a cosine wave. Note that the assumptions about the phase and frequency are for convenience and do not impact the generality of the techniques or results presented here. Looking across the top of Fig. 5-4, the first channel in the analog preprocessor computes an estimate of the envelope of the in-phase voltage distortion. To perform that computation, the measured voltage is multiplied by the output of a sinusoidal oscillator that is synchronized to the line voltage using a slow phase-locked loop. The output of the multiplier circuit is fed to a second order Butterworth low pass filter with a cutoff frequency of 20 Hz. This filter is designed to pass only the slowly-varying baseband component of the multiplier output, which is an estimate of the integral in Eq. 2.2 for $k = 1$ and $x(t) = v_m(t)$. The filter output is level-shifted, amplified, and then sampled at 10 kHz. Once quantized, the signal is averaged using a 167-point FIR filter of the form

$$d[n] = \frac{1}{167}\{1,\ 1,\ 1, \cdots\ 1\}. \tag{5.1}$$

Note that this operation, which significantly reduces any remaining content at the second harmonic of the line frequency, is a scaled numeric approximation of the integral in Eq. 2.2. Since the output of the FIR filter is band-limited well below the sampling frequency, it is slightly decimated.

[3]In this section, the coefficients $a_{k,AN}(t)$ and $b_{k,AN}(t)$ are termed in-phase and quadrature voltage distortion, respectively.

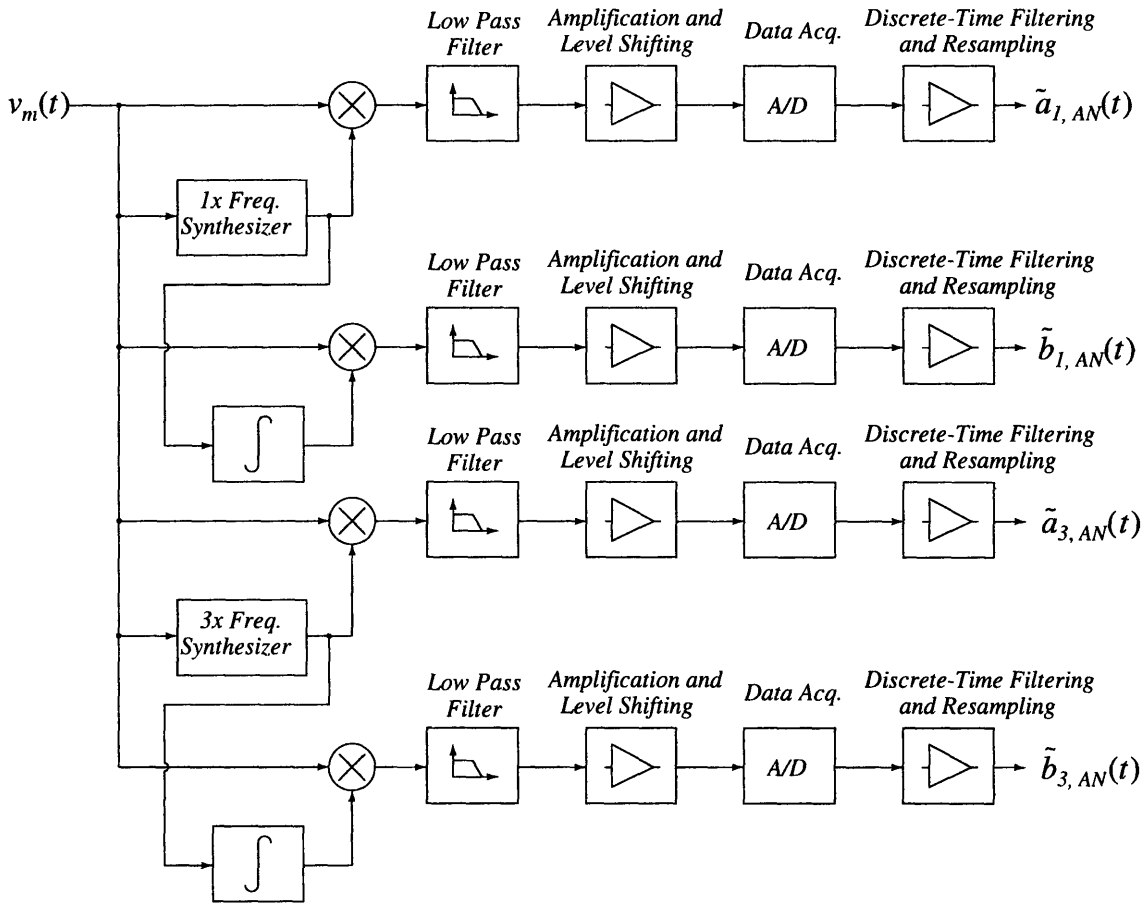


Figure 5-4: Block diagram of the spectral envelope preprocessor used in the current system. The frequency synthesizers are constructed using phase-locked loops. The amplification and level-shifting circuitry following the low pass filters is required in order ensure that the output signals are within the range of the A/D converter in the data acquisition system. Note that the two integrators in this system are used to shift the appropriate reference waveforms by 90^o.

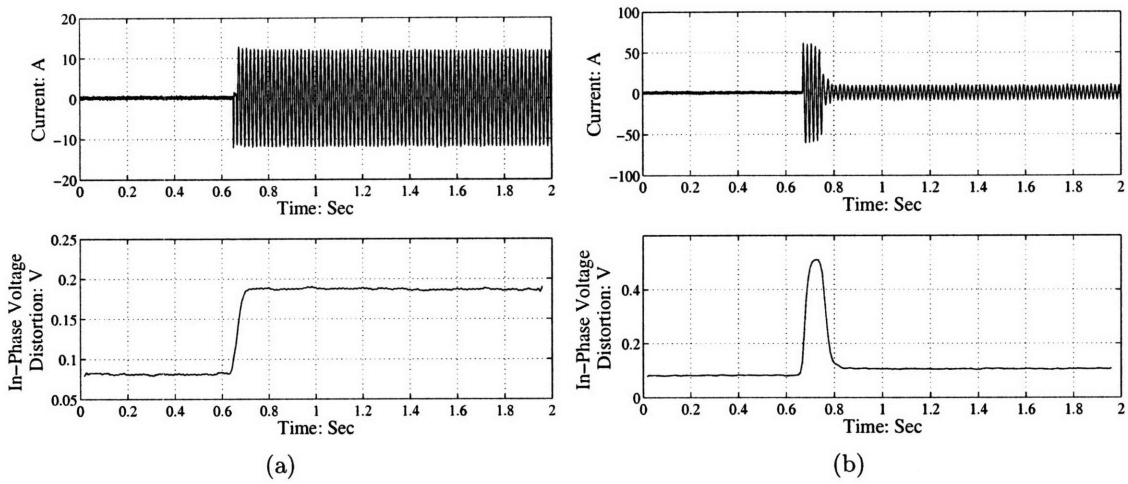


Figure 5-5: (a) The measured load current (for reference) and the measured estimate of $a_{1,AN}(t)$ during the start of a 1 kW heater. (b) The measured load current (for reference) and the measured estimate of $a_{1,AN}(t)$ during the start of an unloaded, 1/3 hp, single-phase induction motor. Note that the estimates of $a_{1,AN}(t)$ are labeled as in-phase voltage distortion.

For a number of different loads, estimates of the spectral coefficient $a_{1,AN}(t)$ have been found to be reliable and repeatable indicators of load operation. For instance, Fig. 5-5a shows the behavior observed during the start of a 1 kW heater.[4] Since this load is well approximated as a pure resistance, the real power that it draws is approximately a step function of time. This is the behavior shown in Fig. 5-5a. A similar trend is observed in Fig. 5-5b for an unloaded, 1/3 hp single phase induction machine. In this case, $a_{1,AN}(t)$ is quite large during the transient period when the rotor is accelerating, and the steady state change is quite small. This behavior is expected, as an unloaded induction machine draws minimal real power during steady operation ([131]).

The procedure used to estimate $b_{1,AN}(t)$ is similar to that used to estimate $a_{1,AN}(t)$. The only difference is that the oscillator voltage is integrated in order to shift it by 90°. There is an important caveat to be noted when estimating $b_{1,AN}(t)$ in this way. Specifically, since the phase-locked frequency synthesizer must eventually synchronize to the new zero crossings

[4]The scales on the y-axes of all of the spectral estimate plots shown from Fig. 5-5a onward correspond to the scales of the voltages measured at the outputs of the analog measurement circuits (i.e. at the inputs to the A/D converters). In order to determine the exact amount of line distortion, each of the plots shown here would need to be appropriately scaled. Additionally, the plots of the spectral estimates are labeled as in-phase or quadrature voltage distortion. In this case, the a_k are considered to be measures of the in-phase voltage distortion, and the b_k are considered to be measures of quadrature voltage distortion.

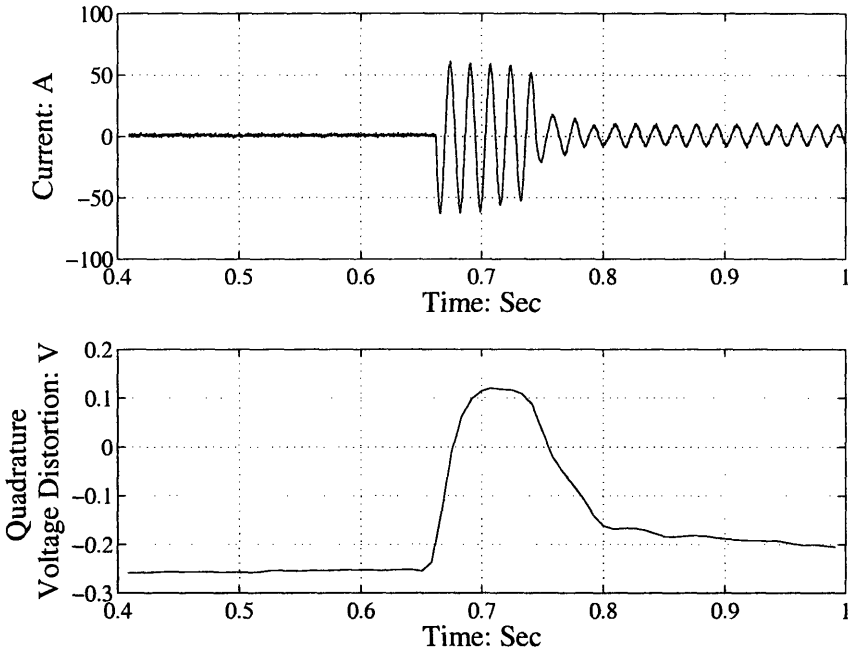


Figure 5-6: Line current (for reference) and the estimate of $b_{1,AN}(t)$ during the start of an unloaded, 1/3 hp, single-phase induction motor. Note that the estimate of $b_{1,AN}(t)$ initially shows a very large change, but that the difference ultimately begins to decrease.

of the voltage waveform, the overall steady state change in the estimated value of $b_{1,AN}(t)$ should be zero. Thus, the estimate that is calculated by the prototype preprocessor is only valid for a short time following a transient event. This is reasonable, however, given that the goal of our system is to identify the operation of individual loads from their transient behavior.

Figure 5-6 displays the estimate of $b_{1,AN}(t)$ during the start of an unloaded, 1/3 hp, single-phase induction motor. As expected, the initial estimate is quite reasonable. As the motor enters steady state conditions, however, the estimate begins to decay. The amount of time that the estimate remains valid is largely a function of the dynamics of the phase-locked loop in the frequency synthesizer. The difference between the estimated value of $b_{1,AN}(t)$ immediately before the motor start and the estimated value immediately following the end of the transient in-rush period is larger than the corresponding change observed in $a_{1,AN}(t)$ during the start of the same unloaded motor (see Fig. 5-5b). This result is logical, as an unloaded induction motor mainly draws reactive power.

In addition to the load information contained in the spectral coefficients of the line-

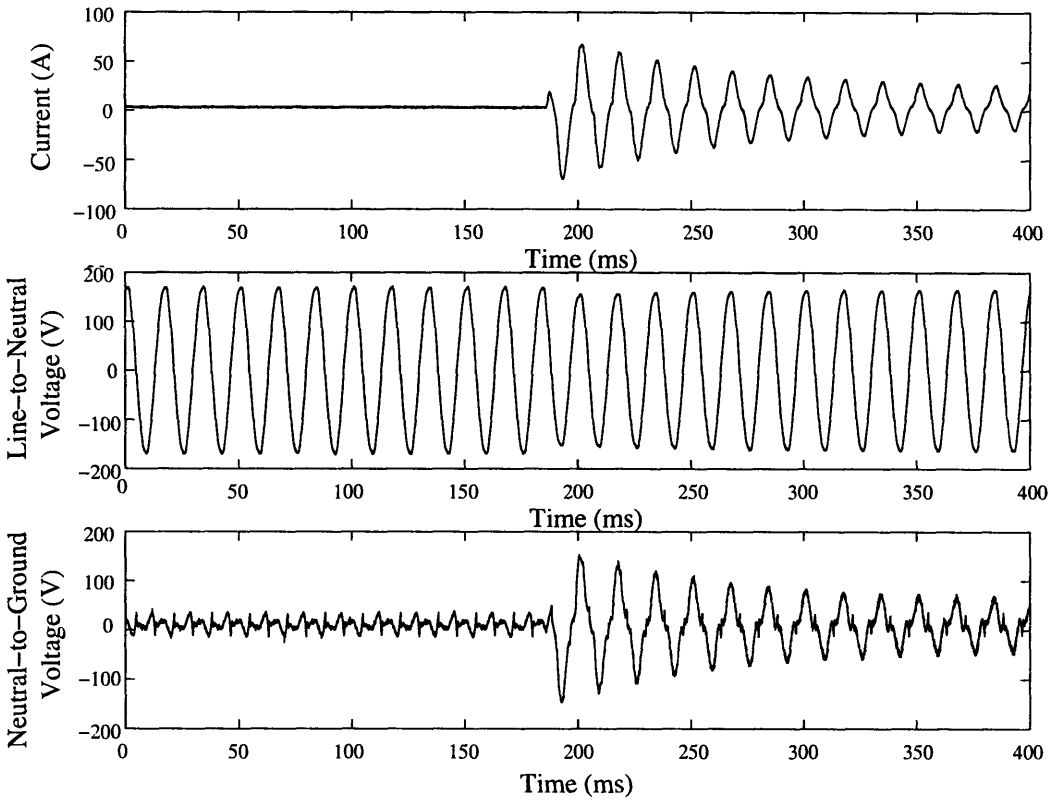


Figure 5-7: Distortion in the neutral-to-ground voltage as a result of the connection of a 4 hp vacuum cleaner. The top trace shows the load current drawn by the vacuum (for reference), the middle trace shows the line voltage at a nearby outlet, and the bottom trace shows the neutral-to-ground voltage at the same measurement point.

to-neutral voltage, there is also information contained in the spectral coefficients of the neutral-to-ground voltage. As shown in Fig. 5-7, the neutral-to-ground voltage at the measurement point, which is denoted as $v_{NG}(t)$, also contains a component that is related to the currents drawn by individual loads. In order to assist in the identification process, the preprocessor also computes estimates of the spectral content of this waveform. In the prototype, the spectral coefficients of $v_{NG}(t)$ are estimated using a procedure similar to that used to estimate the spectral content in the the line-to-neutral voltage. Fig. 5-8 shows a block diagram. Note that the oscillators in this system are phase-locked to the measured line-to-neutral voltage. Circuit schematics and other specific implementation details are presented in Appendix D.

For a number of different loads, the estimated spectral coefficients of the neutral-to-

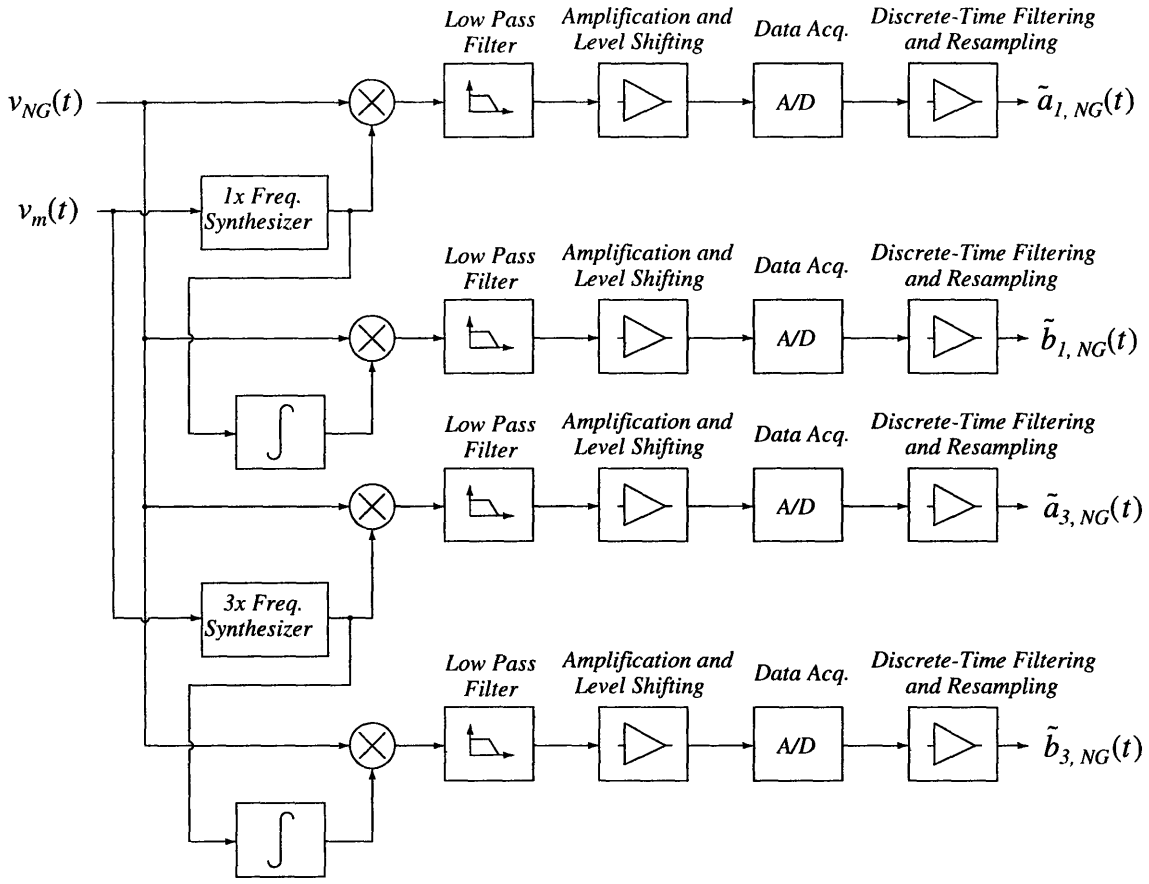


Figure 5-8: Block diagram of the spectral envelope preprocessor used to estimate the spectral coefficients of the neutral-to-ground voltage. The voltage $v_m(t)$ is the measured line-to-neutral voltage.

ground voltage are also reliable and repeatable indicators of load operation. For example, Fig. 5-9 shows estimates of $a_{1,NG}(t)$ and $a_{1,AN}(t)$ as computed during three consecutive incandescent light bulb starts. During a single start, the magnitudes of both estimates display a significant initial rise, as a considerable amount of current is drawn by the bulb. As the filament inside the bulb heats, the peaks of the estimates $\hat{a}_{1,NG}(t)$ and $\hat{a}_{1,AN}(t)$ decrease as expected.

There are several potential advantages to be derived by measuring the spectral content of both the line-to-neutral voltage and the neutral-to-ground voltage. For instance, experimental measurements performed in two different locations have shown that in steady state, the typical amplitude of the line-to-neutral voltage is on the order of several volts (see Fig. 5-7 for an example). By comparison, the fundamental component of the voltage

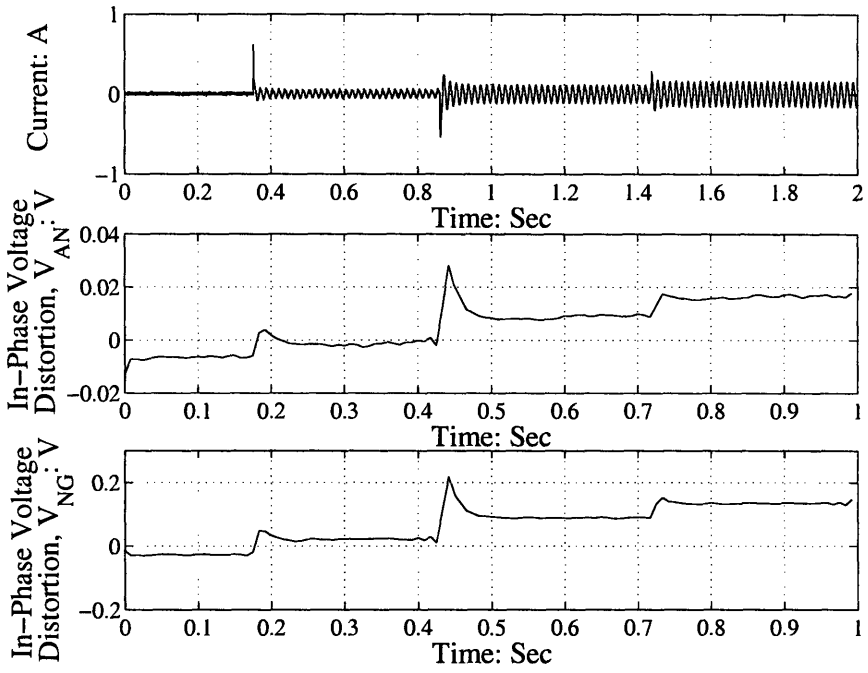


Figure 5-9: The measured current and estimated spectral coefficients $a_{1,AN}(t)$ and $a_{1,NG}(t)$ as a three-way incandescent lamp progresses through its various states.

measured between the phase conductor and the neutral wire has a nominal RMS value of 120 V. As a result, it has been found to be somewhat easier to detect changes in the neutral-to-ground voltage, especially at the higher order harmonic frequencies. As an example, consider Fig. 5-10, which displays estimates of the spectral coefficients $a_{1,NG}(t)$ and $a_{3,NG}(t)$ during the starting period of a rapid-start fluorescent lamp. In particular, note that the estimate of $a_{3,NG}(t)$ has a shape that is remarkably similar to that of the spectral envelope of the third harmonic current (see [32] for an example).

The neutral-to-ground voltage also helps to localize the identification problem. For example, homes in the United States are required to have their neutral conductor tied to earth ground at the utility service entrance ([132]). As a result, the connection of a load inside the monitored home causes the neutral-to-ground voltage to increase and the line-to-neutral voltage to decrease. By comparison, if a load is connected in a nearby home fed by the same distribution transformer, then both voltages will decrease. During limited field testing of the prototype system, the use of the neutral-to-ground voltage has been found to aid in the load identification process.

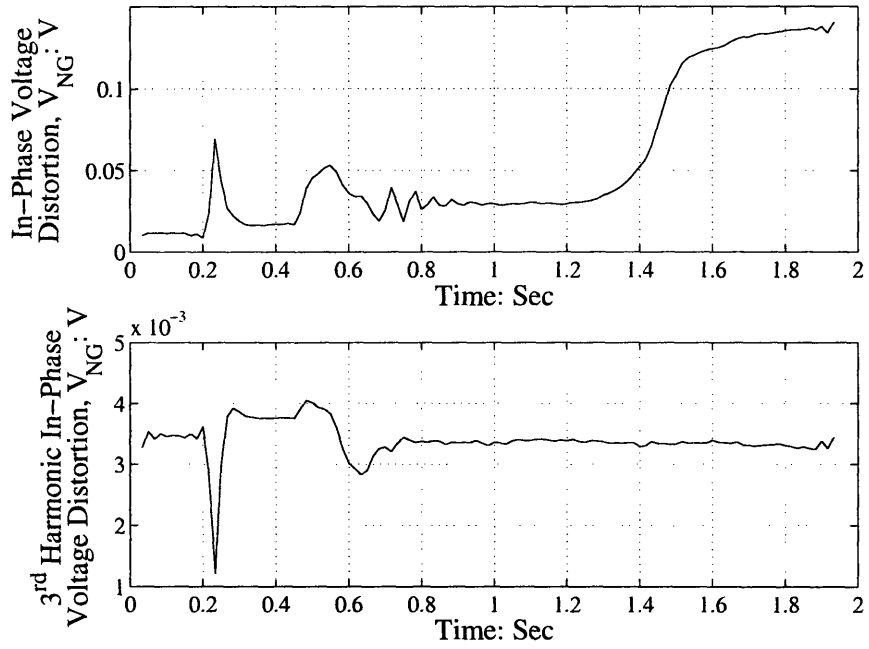


Figure 5-10: Estimates of $a_{1,NG}(t)$ and $a_{3,NG}(t)$ during the starting period of a rapid-start fluorescent lamp.

5.3 *Load Identification*

Once spectral envelope estimates have been created by the preprocessor, they are passed to a transient classifier that identifies the operation of individual loads. The identification procedure used here is essentially the same as the one used in the standard NILM. The only difference is that the load classifier in the V-NILM uses two sets of spectral envelopes, namely those generated from the line-to-neutral voltage and those generated from the neutral-to-ground voltage. For a description of the identification process, the reader is referred to Section 2.2.2.

Fig. 5-11 demonstrates the success of the load identification procedure employed in the V-NILM. The upper trace in that figure is the estimate of the spectral coefficient $a_{1,AN}(t)$, and the bottom trace is the estimate of the spectral coefficient $a_{1,NG}(t)$. As indicated by the exemplar fits, the transient event detector correctly classified both observed events as incandescent lamp starts. During testing, many other loads have been successfully identified, including induction motors, fluorescent lamps, and personal computers.

The prototype system described here can track the behavior of loads whose operation

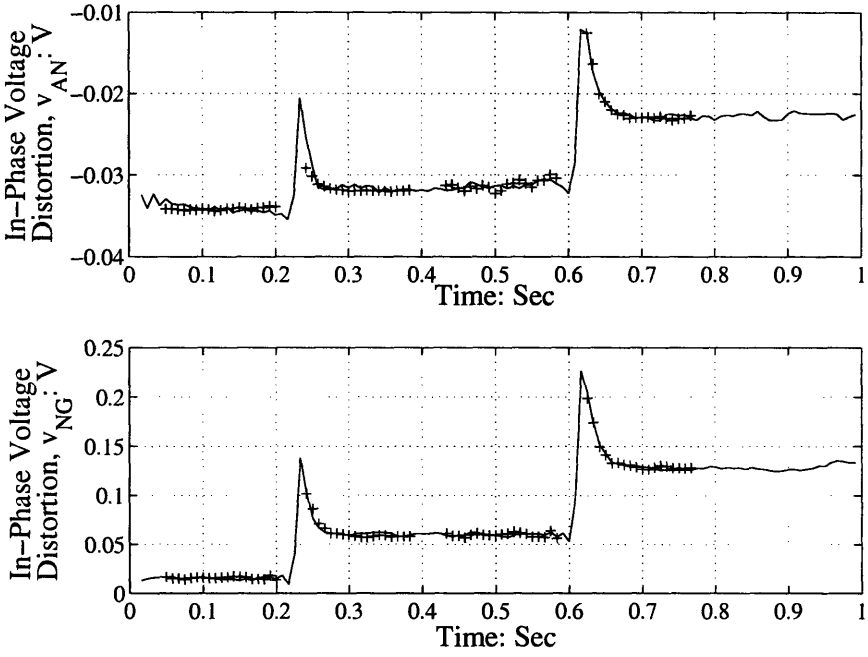


Figure 5-11: Demonstration of the transient event detection software for two consecutive incandescent light bulb starts. The upper trace is the estimate of $a_{1,AN}(t)$ and the bottom trace is the estimate of $a_{1,NG}(t)$ over the same time interval. Overlaid atop both traces are the exemplars that were fit to each observed event. Note that two exemplars are shown in each plot, and that each exemplar consists of two regions of support (i.e. there is one region of points before each start and another region of points after each start).

is more sophisticated than that of the simple on-off devices described thus far. Specifically, the prototype system has shown an ability to be able to track the operation of loads whose behavior follows a finite-state machine (FSM) model. Fig. 5-12 shows state diagrams for two example FSM-type loads that have been tracked using this system ([29]). As an example, consider Fig. 5-9, which displays estimates of $a_{1,AN}(t)$ and $a_{1,NG}(t)$ as a three-way lamp transitions from its off state to its high power state. In order to track such loads automatically, future incarnations of this system will include software similar to that described in [29].

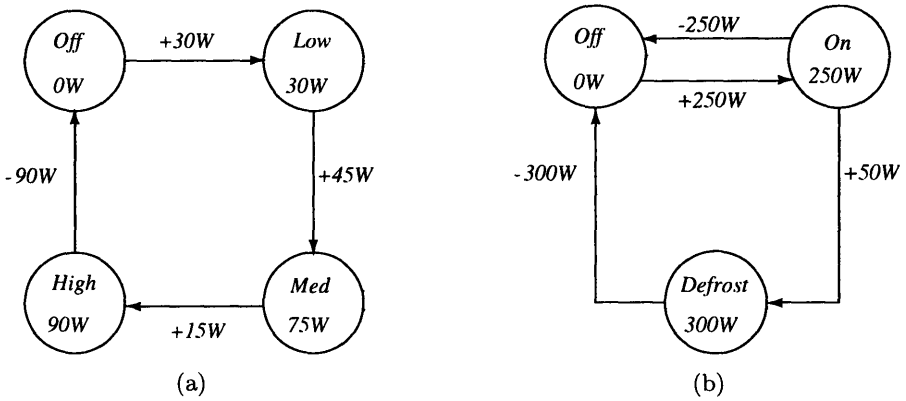


Figure 5-12: Two finite-state appliance models: (a) a three-way incandescent lamp and (b) a refrigerator with a defrost state. The power levels written inside each state correspond to the nominal power drawn by the load during steady operation in that state. The power levels written next to each arc correspond to the amount by which the steady state power drawn by the load must change in order to make the prescribed state transition.

5.4 Summary

This chapter has demonstrated a new monitoring system that uses voltage measurements to track the operation of each load on an aggregate power service. The system described in this chapter opens the door to easy, inexpensive, and accurate load monitoring. Installation of the monitor is as easy as "plugging in" an appliance. A variety of demand-side energy management features could be implemented through the use of this device.

The operational examples included in this chapter are a subset taken from the many tests that have been performed to date. The prototype system balances analog and digital processing in order to provide an inexpensive tool to detect small load transients. Current

results indicate that the prototype system can identify nearly all household loads. Future work on this system will be aimed at determining whether or not the analog preprocessor can be migrated into the digital domain.

Chapter 6
Conclusions

6.1 Summary

This thesis proposes several techniques and methods for increasing the efficiency and reliability of engineered systems. The specific contributions include several field-tested non-intrusive diagnostic methods, a minimally invasive airflow monitoring system, and a single-sensor approach for monitoring load activity on an electrical service.

A common theme uniting the many topics in this thesis is the use of strategies that minimize sensor count and overall physical intrusion. These methods indicate that small, minimally intrusive sensor networks can effectively monitor many large and complex systems. This claim is supported using numerous field examples.

Aspirations of generality notwithstanding, many of the methods described in this thesis are ready for almost immediate deployment in the field. In particular, several of the non-intrusive diagnostic procedures described in Chapter 3 have already been used successfully by the crew aboard the *USCGC Seneca*. Other applications, such as the V-NILM, are nearing a field-ready status.

6.2 Future Work

The results presented in this thesis demonstrate that these methods are generally useful in many different environments. Several exciting opportunities exist to extend the capabilities of each technique described in the previous chapters. Some of the avenues for future work are discussed below.

6.2.1 Non-Intrusive Diagnostics

This thesis demonstrates the vast potential of non-intrusive diagnostic methods. Future work in this area should follow several parallel paths. First, effort should be made to extend the methods presented in this thesis. In particular, the cycling system modelling technique could be applied in many systems, especially those related to heating, ventilation, cooling, and refrigeration. Also, strong efforts should be made to explore new non-intrusive techniques. Fan and pump-driven systems, for example, could strongly benefit from the development of non-intrusive monitoring methods.

6.2.2 System-Level Health Monitoring

When a process requires many components to function in sequence, it is often difficult to locate the source of a fault. The system-level monitoring methods discussed in this thesis can address that problem. The examples presented here demonstrate that the NILM is particularly well suited for use in these applications. Many systems could benefit from having a NILM monitor the status of component loads. Examples include aircraft, automobiles, satellites, building ventilation systems, and manufacturing plants.

Future applications in complex integrated systems will likely place the NILM in an automated supervisory role. In this capacity, the NILM would receive the commands issued by the operator, and it would then monitor the ensuing operations. If a fault was detected, the NILM would either alert the operator or automatically take corrective action. This interactive ability would make the NILM an extremely powerful tool, especially in industrial plants where down-time can not be tolerated.

6.2.3 Airflow Monitoring

The results presented in Chapter 4 demonstrate that it is possible to monitor the speed and direction of airflow in rooms with air exchange rates as low as approximately 8.8 ACH. Future work on the airflow monitoring system should focus on several topics. First, there is a need to determine if ozone can be used to measure speeds in rooms with air exchange rates even lower than 8.8 ACH. Additionally, it is also important to determine if the ozone-based system can be used to estimate air exchange rates directly. Currently, these rates are measured using inert gases such as SF_6. Since ozone is a highly reactive species, it is considerably more difficult to perform the same type of experiments. Further work is

needed in order to determine if ozone can be used in this capacity.

6.2.4 V-NILM

Future work on the V-NILM should focus on both development and applications. On the development side, effort should be made to determine how much of the processing must remain in the analog domain. It is likely that most of the operations in the V-NILM could be moved into software. On the application side, it would be wise to install V-NILMs in a few different homes for several weeks or months. This type of field study would help to determine how to employ the V-NILM as a tool for DSM and load diagnostics.

Bibliography

[1] S. Lawson, "Fancy features don't sell cell phones," *PC World*, vol. 24, no. 5, 31 May 2006.

[2] "Consumer Reports: gas prices having big impact," *CNN.com*, 24 May 2006. [Online]. Available: http://money.cnn.com/2006/05/22/Autos/cr_fuel_concerns

[3] E. H. Mathews, C. P. Botha, D. C. Arndt, and A. Malan, "HVAC control strategies to enhance comfort and minimise energy usage," *Energy and Buildings*, vol. 33, no. 8, pp. 853–863, Oct. 2001.

[4] M. DiUlio, C. Savage, and E. Schneider, "Taking the integrated condition assessment system to the year 2010," in *Proc. 13th International Ship Control Systems Symposium*, Orlando, FL, Apr. 2003.

[5] J. A. Crossman, H. Guo, Y. L. Murphey, and J. Cardillo, "Automotive signal fault diagnostics-part I: Signal fault analysis, signal segmentation, feature extraction and quasi-optimal feature selection," *IEEE Trans. on Vehicular Technology*, vol. 52, no. 4, pp. 1063–1075, July 2003.

[6] Y. L. Murphey, J. A. Crossman, Z. Chen, and J. Cardillo, "Automotive fault diagnosis-part II: A distributed agent diagnostic system," *IEEE Trans. on Vehicular Technology*, vol. 52, no. 4, pp. 1076–1098, July 2003.

[7] H. E. Troost, T. Butcher, and A. M. Rudin, "Results of the flame quality monitor field test program," in *Proc. of the 2002 National Oil Heat Research Alliance Technology Symposium*, R. J. McDonald, Ed., Providence, RI, 20 Aug. 2002, pp. 135–147.

[8] R. L. Lopushansky, "All optical shipboard sensing system," in *Proc. 45th International Instrumentation Symposium*, Albuquerque, NM, 2 May 1999.

[9] R. Isermann, *Fault-Diagnosis Systems.* Berlin: Springer-Verlag, 2006.

[10] S.R. Shaw, C.R. Laughman, S.B. Leeb, and R.F. Lepard, "A power quality prediction system," *IEEE Trans. on Industrial Electronics*, vol. 47, no. 3, pp. 511–517, June 2000.

[11] D. Filbert, "Fault diagnosis in nonlinear electromechanical systems by continuous-time parameter estimation," *ISA Transactions*, vol. 24, no. 3, pp. 23–27, 1985.

[12] K. Cho, J. H. Lang, and S. Umans, "Detection of broken rotor bars in induction motors using parameter and state estimation," *IEEE Trans. on Industrial Applications*, vol. 28, no. 1, pp. 702–708, May 1992.

[13] U. Kiencke, "Diagnosis of automotive systems," in *Proc. IFAC Symposium of Fault Detection, Supervision, and Safety for Technical Processes (SAFE-PROCESS).* Hull, UK: Pergamon Press, Aug. 1997.

[14] G. Rizzoni, A. Soliman, and K. Passino, "A survey of automotive diagnostic equipment and procedures," in *Proc. International Congress and Exposition.* Detroit, MI: SAE, 1993.

[15] A. Wolfram, D. Fussel, T. Brune, and R. Iserman, "Component-based multi-model approach for fault detection and diagnosis of a centrifugal pump," in *Proc. American Control Conference (ACC)*, Arlington, VA, 2001.

[16] C. Kallesoe, V. Cocquempot, and R. Izadi-Zamanabadi, "Model based fault detection in a centrifugal pump application," *IEEE Trans. on Control Systems Technology*, vol. 14, no. 2, pp. 204–215, Mar. 2006.

[17] D. Groghan, "Gas Turbines," in *Marine Engineering*, R. Harrington, Ed. Jersey City, NJ: Society of Naval Architechts and Marine Engineers, 1992, pp. 146–183.

[18] P. R. Armstrong, "Model identification with application to building control and fault detection," Ph.D. dissertation, Massachusetts Institute of Technology, June 2004.

[19] S. Wang and F. Xiao, "AHU sensor fault diagnosis using principal component analysis method," *Energy and Buildings*, vol. 36, pp. 147–160, 2004.

[20] S. W. Wang and J. B. Wang, "Law-based sensor fault diagnosis and validation for building air-conditioning systems," *International Journal of HVAC&R Research*, vol. 5, no. 4, pp. 353–380, 1999.

[21] H. Li, "A decoupling-based unified fault detection and diagnosis approach for packaged air conditioners," Ph.D. dissertation, Purdue University, Aug. 2004.

[22] J. Melody, T. Basar, W. Perkins, and P. Voulgaris, "Parameter estimation for inflight detection of aircraft icing," in *Proc. 14th IFAC World Congress*, Beijing, China, 1991, pp. 295–300.

[23] J. Musgrave, T.-H. Guo, E. Wong, and A. Duyar, "Real-time accomodation of actuator faults on a reusable rocket engine," *IEEE Trans. on Control Systems Technology*, vol. 5, no. 1, pp. 100–109, 1997.

[24] S. Budalakoti, A. Srivastava, and R. Akella, "Discovering atypical flights in sequences of discrete flight parameters," in *Proc. 2006 IEEE Aerospace Conference*, 4 Mar. 2006.

[25] "Technology for the United States Navy and Marine Corps, 2000-2035 - Becoming a 21st century force, Volume 4: Human Resources," Naval Studies Board Panel on Human Resources, 1997, enter text here.

[26] R. Bartholomew, "ATP position paper on condition-based maintenance," National Institute of Standards and Technology, 17 Nov. 1998. [Online]. Available: http://www.atp.nist.gov/files/cbm_wp1.pdf

[27] E. Y. Chow and A. S. Willsky, "Analytic redundacny and the design of robust failure detection systems," *IEEE Trans. Automat. Contr.*, vol. 29, no. 7, pp. 603–614, July 1984.

[28] W. Greene, J. Ramsey, R. Cox, and T. DeNucci, "Non-intrusive monitoring for condition-based maintenance," in *Proc. ASNE Reconfiguration and Survivability Symosium*, Jacksonville, FL, 16 Feb. 2005.

[29] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[30] S. Leeb, "A conjoint pattern recognition approach to nonintrusive load monitoring," Ph.D. dissertation, Massachusetts Institute of Technology, Feb. 1993.

[31] U. A. Khan, S. B. Leeb, and M. C. Lee, "A multiprocessor for transient event detection," *IEEE Trans. on Power Delivery*, vol. 12, no. 1, pp. 51–60, Jan. 1997.

[32] S. R. Shaw, "System identification techniques and modeling for non-intrusive load diagnostics," Ph.D. dissertation, Massachusetts Institute of Technology, Feb. 2000.

[33] R. W. Cox, J. P. Mosman, T. D. McKay, S. B. Leeb, and T. J. McCoy, "Diagnostic indicators for shipboard cycling systems," in *Proc. ASNE Day 2006*, Arlington, VA, 19 June 2006.

[34] P. R. Armstrong, C. R. Laughman, S. B. Leeb, and L. K. Norford, "Detection of rooftop cooling unit faults based on electrical measurements," *HVAC+R Research Journal*, vol. 12, no. 1, pp. 151–175, Jan. 2006.

[35] T. DeNucci, R. Cox, S. B. Leeb, J. Paris, T. McCoy, and W. C. Greene, "Diagnostic indicators for shipboard systems using non-intrusive load monitoring," in *Proc. 1st IEEE Electric Ship Technologies Symposium (ESTS)*, Philadelphia, PA, 25 July 2005.

[36] D. Luo, "Detection and diagnosis of faults and energy monitoring of HVAC systems with least-intrusive power analysis," Ph.D. dissertation, Massachusetts Institute of Technology, Feb. 2001.

[37] K. L. Butler, N. D. R. Sarma, C. Whitcomb, H. D. Carmo, and H. Zhang, "Shipboard systems deploy automated protection," *IEEE Computer Applications in Power*, vol. 11, pp. 31–36, Apr. 1998.

[38] R. W. Cox, S. B. Leeb, and L. K. Norford, "A minimally intrusive, low cost system for determining indoor air flow patterns," in *9th IEEE Workshop on Computers in Power Electronics (COMPEL)*, Urbana, Illinois, 15 Aug. 2004.

[39] H. Awbi, *Ventilation of Buildings.* Spon Press, 2003.

[40] D. Etheridge and M. Sandberg, *Building Ventilation: Theory and Measurement.* John Wiley & Sons, 1996.

[41] K. D. Lee, "Electric load information system based on non-intrusive power monitoring," Ph.D. dissertation, Massachusetts Institute of Technology, June 2003.

[42] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power signature analysis," *IEEE Power and Energy Magazine*, pp. 55–63, Mar. 2003.

[43] T. Zhu, S. R. Shaw, and S. B. Leeb, "Transient recognition control for hybrid fuel cell systems," *IEEE Transactions on Energy Conversion*, vol. 21, no. 4, pp. 195–201, Mar. 2006.

[44] S. R. Shaw, D. Evangelista, S. B. Leeb, and C. R. Laughman, "Non-intrusive load monitoring and identification in an automotive environement," in *Proc. of ELECTRIMACS 1999*, Lisbon, Portugal, Sept. 1999, pp. 199–204.

[45] S. Drenker and A. Kader, "Nonintrusive monitoring of electric loads," *IEEE Computer Applications in Power*, vol. 12, no. 4, pp. 47–51, Oct. 1999.

[46] L. Carmichael, J. Kelly, and H. Dominguez, "The beta test of a non-intrusive load monitoring system," in *1996 DA/DSM Proceedings*, Vienna, Austria, 1996.

[47] G. W. Hart, "Residential energy monitoring and computerized surveillance via utility power flows," *IEEE Technology and Society*, pp. 12–16, June 1989.

[48] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy and Buildings*, vol. 24, pp. 51–64, 1996.

[49] S.B. Leeb, S.R. Shaw, and J.L. Kirtley, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Trans. on Power Delivery*, vol. 10, no. 3, pp. 1200–1210, July 1995.

[50] S.R. Shaw and C.R. Laughman, "A Kalman-filter spectral envelope preprocessor," submitted for publication in *IEEE Trans. Instrum. Meas*, Feb. 26. 2004.

[51] S.R. Sanders, J.M. Noworolski, X.Z. Liu, and G.C. Verghese, "Generalized averaging method for power conversion circuits," *IEEE Trans. on Power Electronics*, vol. 6, no. 2, pp. 251–259, Apr. 1991.

[52] J. Paris, "A framework for non-intrusive load monitoring and diagnostics," M. Eng. thesis, Massachusetts Institute of Technology, Cambridge, MA, Feb. 2006.

[53] A. Gelb, Ed., *Applied Optimal Estimation.* Cambridge, MA: The MIT Press, 1999.

[54] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[55] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, July 1992.

[56] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, Mar. 1992.

[57] F. Figueroa, R. Holland, J. Schmalzel, and D. Duncavage, "Integrated system health management (ISHM): Systematic capability implementation," in *Proc. 2006 IEEE Sensors Applications Symposium*, Houston, TX, 7 Feb. 2006, pp. 202–206.

[58] J. Schmalzel, F. Figueroa, J. Morris, S. Mandayam, and R. Polikar, "An architecture for intelligent systems based on smart sensors," *IEEE Trans. on Instrumentation and Measurement*, vol. 54, no. 4, pp. 1612–1616, Aug. 2005.

[59] M. Feather and L. Markosian, "Emerging technologies for V&V of ISHM software for space exploration," in *Proc. 2006 IEEE Aerospace Conference*, 4 Mar. 2006.

[60] H. Cassee, "Piping Systems," in *Marine Engineering*, R. Harrington, Ed. Jersey City, NJ: The Society of Naval Architechts and Marine Engineers, 1992, pp. 782–845.

[61] *Technical Manual for LM2500 Propulsion Gas Turbine Module*, Naval Sea Systems Command, 26 May 2000, S9234-AD-MMO-010/LM2500.

[62] W. M. Siebert, *Circuits, Signals, and Systems.* Cambridge, MA: MIT Press, 1986.

[63] Talbott. E.M., *Compressed Air Systems: A Guidebook on Energy and Cost Savings.* Atlanta, GA: Fairmont Press, 1986.

[64] H. C. Tijms, *A First Course in Stochastic Models.* New York: John Wiley and Sons, 2003.

[65] A. S. Willsky, G. W. Wornell, and J. H. Shapiro, "Stochastic Processes, Detection and Estimation," MIT Course 6.432 Course Notes, Feb. 2001.

[66] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability.* Belmont, MA: Athena Scientific, 2002.

[67] A. W. Drake, *Fundamentals of Applied Probability Theory.* McGraw-Hill, 1967.

[68] *Envirovac vacuum sewage collection system, United States Coast Guard, U. S. C. G. 270' B Class WMEC*, Envirovac, Inc., Apr. 1983.

[69] W. C. Karl, S. B. Leeb, L. A. Jones, J. L. Kirtley, and G. C. Verghese, "Applications of rank-based filters in power electronics," *IEEE Transactions on Power Electronics*, vol. 7, no. 3, pp. 437–443, July 1992.

[70] G. J. Hahn and S. S. Shapiro, *Statistical Models in Engineering.* New York: John Wiley and Sons, 1967.

[71] M. B. Priestley, *Spectral Analysis and Time Series.* San Diego, CA: Elsevier, 1981.

[72] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences.* Boston, MA: McGraw-Hill, 1992.

[73] J. A. Nachlas, *Reliability Engineering: Probabilistic Models and Maintenance Methods.* Boca Raton, FL: CRC Press, 2005.

[74] A. A. Shaltout, "Analysis of torsional torques in starting of large squirrel cage induction motors," *IEEE Trans. on Energy Conversion*, vol. 9, no. 1, pp. 135–142, Mar. 1994.

[75] C. B. Mayer, "Torsional vibration problems and analyses of cement industry drives," *IEEE Trans. on Industry Applications*, vol. IA-17, no. 1, pp. 81–89, Jan. 1981.

[76] H. H. Woodson and J. R. Melcher, *Electromechanical Dynamics - Part I: Discrete Systems.* New York: John Wiley and Sons, 1968.

[77] P. C. Krause, O. Wasynczuk, and S. D. Sudhoff, *Analysis of Electric Machinery.* McGraw Hill, 1986.

[78] G. Strang and T. Nguyen, *Wavelets and Filter Banks.* Wellesley, MA: Wellesley-Cambridge Press, 1997.

[79] K. D. Hurst and T. G. Habetler, "Sensorless speed measurments using current harmonic spectral estimation in induction machine drives," *IEEE Trans. on Power Electronics*, vol. 11, no. 1, pp. 66–73, Jan. 1996.

[80] *Technical Manual for Air Compressor, Low Pressure, Oil Free*, Naval Sea Systems Command, 1 Aug. 1989, S6220-DG-MMA-010/88663.

[81] L. Skarbeck, Personal communications, LBES Test Engineer, 24 Apr. 2006.

[82] *USS ROOSEVELT (DDG-80) Damage Control and Engineering Handbook*, United States Government Printing Office, 25 Nov. 2000, S6263-B7-MMA-010.

[83] *Technical Manual for Universal Engine Controller*, Naval Sea Systems Command, 25 Nov. 2002, S6263-B7-MMA-010.

[84] G. Tan, "Study of Natural Ventilation Design by Integrating the Multizone Model with CFD Simulation," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2005.

[85] M. Sohn and M. Small, "Parameter estimation of unknown air exchange rates and effective mixing volumes from tracer gas measurements for complex multi-zone indoor air models," *Building and Environment*, vol. 34, pp. 293–303, 1999.

[86] C. A. Roulet and L. Vandaele, "Airflow patterns within buildings measurement techniques," AIVC, Tech. Rep., 1991.

[87] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases.* New York: John Wiley and Sons, 1939.

[88] H. M. Schneider, "A Review of the Basic Properties of Corna Discharges," Electric Power Systems Laboratory, Massachusetts Institute of Technology, Tech. Rep. 34, 7 Dec. 1971.

[89] H. Kawamoto, "Ozone generation in corona discharge at pin electrode of electrophotographic charger," *J. Imaging Sci. Technol.*, vol. 44, no. 5, pp. 452–456, Sept. 2000.

[90] G. Naidis, "Modeling of plasma chemical processes in stable corona discharges at thin wires," *J. Phys. D: Appl. Phys.*, vol. 25, pp. 477–480, 1992.

[91] D. Wilson, S. Hoyt, J. Janata, K. Booksh, and L. Obando, "Chemical sensors for portable, handheld field instruments," *IEEE Sensors Journal*, vol. 1, no. 4, pp. 256–274, Dec. 2001.

[92] T. Becker, L. Tomasi, C. Bosch-v.Braunmuhl, G. Muller, G. Sberveglieri, G. Fagli, and E. Comini, "Ozone detection using low-power-consumption metal-oxide gas sensors," *Sensors and Actuators A*, vol. 74, pp. 229–232, 1999.

[93] F. Daniels and R. A. Alberty, *Physical Chemistry*, 4th ed. New York: John Wiley and Sons, 1975.

[94] *MiCS-2610 O_3 Gas Sensor Preliminary Data Sheet*, MicroChemical Systems, 2003.

[95] S. Johnson, A. Witulski, and R. Erickson, "Comparison of resonant topologies in high-voltage DC applications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 3, pp. 263–274, May 1988.

[96] S. Johnson and R. Erickson, "Steady-state analysis and design of the parallel resonant converter," *IEEE Trans. Power Electron.*, vol. 3, no. 1, pp. 93–104, Jan. 1988.

[97] R. Erickson and D. Maksimovic, *Fundamentals of Power Electronics*, 2nd ed. Norwell, MA: Kluwer Academic Publishers, 2001.

[98] M. Tavakoli and R. Sarpeshkar, "An offset-cancelling low-noise lock-in architecture for capacitive sensing," *IEEE Journal of Solid State Circuits*, vol. 38, no. 2, pp. 244–253, June 2003.

[99] S. Senturia, *Microsystem Design.* Boston, MA: Kluwer Academic Publishers, 2001.

[100] C. Kolle and P. O'Leary, "Low-cost, high-precision measurement system for capacitive sensors," *Meas. Sci. Technol.*, vol. 9, pp. 510–517, 1998.

[101] P. Harris, W. Arnold, A. Andrews, and A. Partridge, "Resistance characteristics of conducting polymer films used in gas sensors," *Sensor and Actuators B*, vol. 42, pp. 177–184, 1997.

[102] C. Weschler, H. Shields, and D. Naik, "Indoor ozone exposures," *Journal of the Air Pollution Control Association*, vol. 39, pp. 1562–1568, 1989.

[103] H. R. Taylor, *Data Acquisition for Sensor Systems.* New York: Chapman and Hall, 1997.

[104] C. Motchenbacher and J. Connelly, *Low-Noise Electronic System Design.* New York: John Wiley and Sons, 1993.

[105] T. A. Denison, "The Development of a Nanoscale Coulter Counter for Rapid Genetic Sequence Recognition," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, July 2000.

[106] C. M. Van Vliet, "Responsivity and noise in illustrative solid-state chemical sensors," *Sensor and Actuators B*, vol. 24-25, pp. 6–16, 1995.

[107] S. Trautwieler, Personal communications, 2004.

[108] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits.* Cambridge, UK: Cambridge University Press, 1998.

[109] *MAX4526/MAX4527, Phase-reversal analog switches*, Maxim, 1996.

[110] G. Wegmann, E. Vittoz, and F. Rahali, "Charge injection in analog MOS switches," *IEEE Journal of Solid State Circuits*, vol. SC-22, no. 6, pp. 1091–1097, Dec. 1987.

[111] *Circuit Techniques for Avoiding Oscillations in Comparator Applications (Linear Brief 39)*, National Semiconductor, Jan. 1978.

[112] *LTC1052, Zero-Drift Operational Amplifier*, Linear Technology, 1985.

[113] R. Hamming, *Digital Filters*, 3rd ed. Mineola, NY: Dover Publications, Inc., 1989.

[114] F. Mintzer, "On half-band, third-band, and n-th band FIR filter and their design," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-30, no. 5, pp. 734–738, Oct. 1982.

[115] P. P. Vaidyanathan and T. Q. Nguyen, "Eigenfilters: A new approach to least-squares FIR filter design and applications including nyquist filters," *IEEE Trans. on Circ. and Sys.*, vol. CAS-34, no. 1, pp. 11–23, Jan. 1987.

[116] A. van der Ziel, "Noise in solid state devices and lasers," *Proc. IEEE*, vol. 58, no. 8, pp. 1178–1206, Aug. 1970.

[117] *OPA627/OPA637, Precision High-Speed Difet Operational Amplifiers*, Burr-Brown, 1995.

[118] R. M. Gray, "Quantization noise spectra," *IEEE Trans. on Information Theory*, vol. 36, no. 6, pp. 1220–1244, Nov. 1990.

[119] C. Enz and G. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: auto-zeroing, correlated double sampling, and chopper stabilization," *Proceedings of the IEEE*, vol. 84, no. 11, pp. 1584–1614, 1996.

[120] *LTC1050, Precision chopper stabilized operational amplifier with internal capacitors*, Linear Technology, 1991.

[121] M. Coln, "Chopper stabilization of MOS operational amplifiers using feed-forward techniques," *IEEE Journal of Solid-State Circuits*, vol. SC-16, no. 6, pp. 745–748, Dec. 1981.

[122] Bakker. A., K. Thiele, and J. Huijsing, "A CMOS nested-chopper instrumentation amplifier with 100-nV offset," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, pp. 1877–1883, Dec. 2000.

[123] I. G. Finvers, "Noise analysis of a continuous-time auto-zeroed amplifier," *IEEE Trans. Circ. and Sys.-II: Analog and Digital Signal Processing*, vol. 43, no. 12, pp. 791–800, Dec. 1996.

[124] Q. Chen and W. Xu, "A zero-equation turbulence model for indoor airflow simulation," *Energy and Buildings*, vol. 28, pp. 137–144, 1996.

[125] O. Levenspiel, *Chemical Reaction Engineering.* New York: John Wiley and Sons, Inc., 1962.

[126] "Electric utility demand-side management 1999: executive summary," Energy Information Agency, 1999. [Online]. Available: http://www.eia.doe.gov/cneaf/electricity/dsm99/dsm_sum99.html

[127] Anne Bengston, Ed., *IEA-DSM 2005 Annual Report: Implementing Agreement on Demand-Side Managment Technolgies and Programmes*, 2005. [Online]. Available: http://dsm.iea.org/NewDSM/Work/AnnualReports/IEADSMAnnualReport2005.pdf

[128] A. de Oliveira, J.C. de Oliveira, J.W. Resende, and M.S. Miskulin, "Practical approaches for AC system harmonic impedance measurements," *IEEE Trans. on Power Delivery*, vol. 6, no. 4, pp. 1721–1726, Oct. 1991.

[129] W.C. Beattie and S.R. Matthews, "Impedance measurement on distribution networks," in *Proc. 29th Universities Power Engineering Conf.*, Sept. 1994, pp. 117–120.

[130] A. Oury, R. Bergeron, and A. Laperrière, "Source impedances of the Canadian distribution systems (residential and industrial)," in *Proc. 14th International Conf. and Expo. on Electrity Distribution*, vol. 2, June 1997, pp. 34/1–34/8.

[131] A. Fitzgerald, C. Kingsley, and S. Umans, *Electric Machinery.* Boston, MA: McGraw-Hill, 1983.

[132] J. Garland, *National Electrical Code Reference Book.* Englewood Cliffs, NJ: Prentice-Hall, 1988.

[133] M. Boeniger, "Use of ozone generating devices to improve indoor air quality," *American Industrial Hygiene Assoc. Journal*, vol. 56, pp. 590–598, 1995.

Appendix A

Cycling System Simulation Software

A.1 The Simulink Model

The cycling system simulation presented in this thesis uses the Simulink model shown in Fig. A-1. In that model, the user can set the following parameters:

- **Po** - The high-pressure set point.
- **Plow** - The low-pressure set point.
- **Plower** - The back-up low-pressure set point, i.e. the point at which the second pump energizes.
- **LRP** - Persistent leak rate. This is meant to simulate a systematic, built-in leak.
- **LR** - Leak rate. This simulates a leak resulting from fault.

To execute the simulation, the user must specify the values of each of the parameters listed above. Additionally, the user must also specify the size of the flush drops (i.e. the ΔP_k) in the embedded function fcn2.m and the pump rates in the embedded function fcn3.m. These two codes are presented in Section A.3.

A.2 Executing the Simulation Software

To conduct a simulation, the user must perform the following operations:

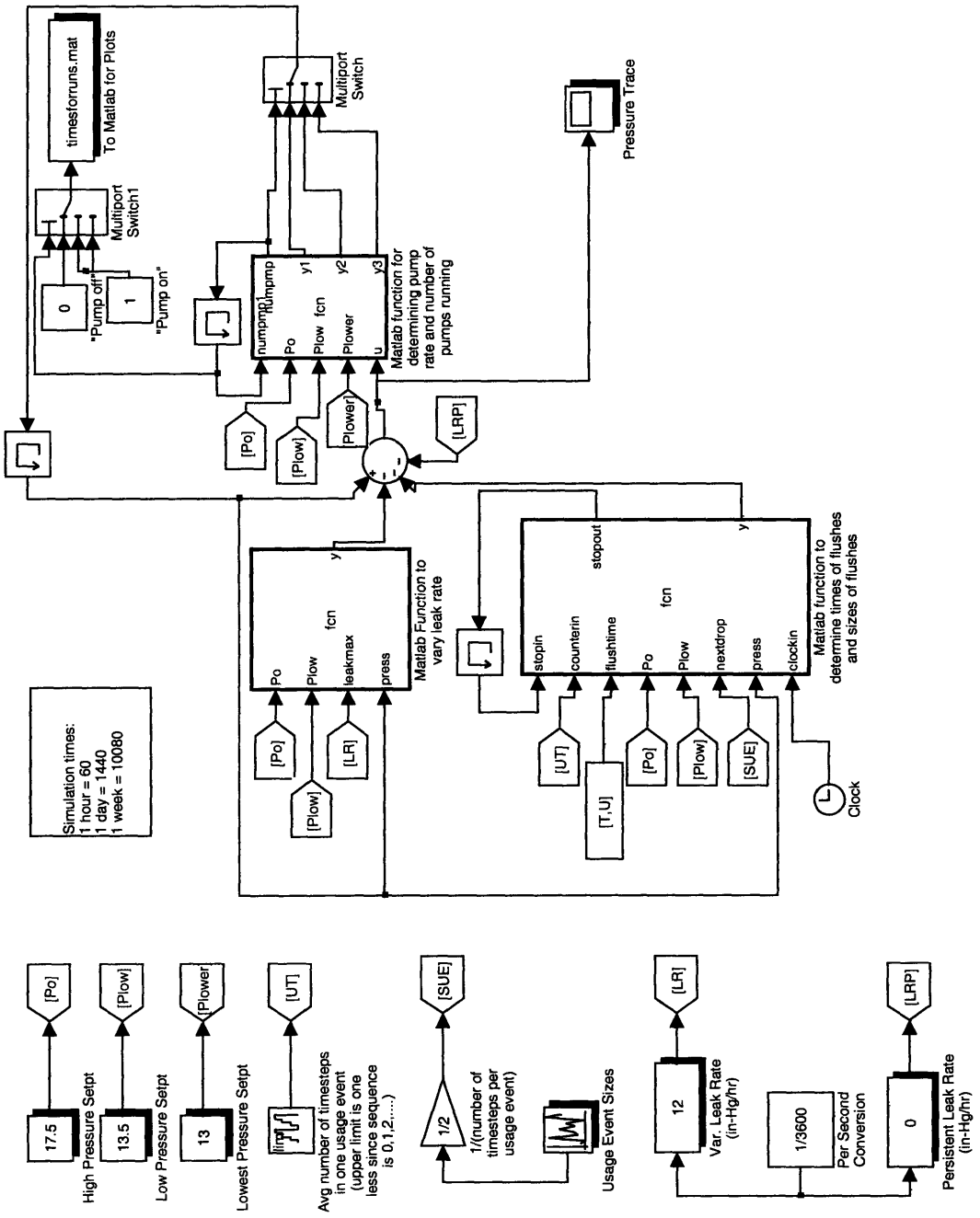


Figure A-1: The Simulink model used to simulate the vacuum-assisted waste disposal system aboard the *USCGC Seneca*. The blocks labled "fcn" are embedded functions, and the codes for each of these are presented in Section A.3.1.

- Execute sim_prep.m from the MATLAB command line. This script generates flush times according to the exponential distribution ([72]). The user is asked to set the following:
  - Simulation time
  - A λ value for each of the 8 hour periods discussed in Chapter 3.
  - A percent variation in each λ value
  - The name of the output file, which will contain the times between pump runs.
- Set the values of each of the parameters in the model file.
- Run the simulation.
- Execute sim_post.m from the MATLAB command line. This script will generate a histogram showing the distribution of the times between pump runs. These times are saved in a separate file for later analysis.

A.3 Simulation Code

Two different sets of code are used in the cycling system simulation. First, there are three embedded functions that are used directly in the Simulink model. The blocks containing these functions are labeled "fcn" in Fig. A-1. These functions are presented in Section A.3.1. Additionally, the simulation requires the user to execute sim_prep.m and sim_post.m. These scripts are presented in Section A.3.2.

A.3.1 Embedded Functions

fcn1.m

```
function y = fcn(Po,leakmax,press)

% Implements dP/dt = -k*P
leakmod = leakmax/Po*press

% Output
y = leakmod;
```

fcn2.m

```
function [stopout,y] = fcn(stopin,counterin,flushtime,Po,Plow,nextdrop,press,clockin)
% This block places a flush into the system
% The code compares the sim time with the times generated in the
% sewage_model_rev1_prep.m file and inserts a flush where dictated by the T and U
% vectors.

counterin1 = double(counterin);

% To account for predominance of two flush sizes.  If 100*clock in is even, the first
% factor is used and the other is used if it is odd.
if rem(ceil(100*clockin),2) == 0
  nextdropfactor = 1;
else
  nextdropfactor = 0.75;
end

% To account for flushes occuring over multiple time steps.
if flushtime == 1
  y=nextdrop*nextdropfactor;
  stopout=counterin1;

elseif stopin ~= -1
  y=nextdrop*nextdropfactor;
  stopout=stopin;
  if counterin1 == stopin
   y=0;
   stopout=-1;
  end
else
  y=0;
  stopout=-1;
end
```

fcn3.m

```
function [numpmp,y1,y2,y3] = fcn(numpmp1,Po,Plow,Plower,u)
% This block determines the output of the pumps, if any
```

```
% Definition of pumping rates
pump1rate = 4.5; %in-Hg/min
pump2rate = 4.7; %in-Hg/min

% These are the pumping rates
y1 = u;
y2 = u + pump1rate/60;
y3 = u + (pump1rate+pump2rate)/60;

% Check to see if the pumps should be turned off or how many should be on
if u >= Po
  numpmp = 0;
elseif u <= Plow && u > Plower
  if numpmp1 == 0
     numpmp = 1;
  else
     numpmp = numpmp1;
  end
elseif u <= Plower
  if numpmp1 ~= 2
     numpmp = 2;
  else
     numpmp = numpmp1;
  end
else
 numpmp = numpmp1;
end
```

A.3.2 Pre- and Post-Processing Scripts

sim_prep.m

```
%This m-file is the prep for setting up to run the
%sewage_model_rev1 Simulink model.

clear;
% First get the required inputs
T=input('What is the simulation time you intend to run (in minutes)?');
lambda_w=input('What is the lambda value for the workday (in flushes/hour)?');
lambda_e=input('What is the lambda value for the evening (in flushes/hour)?');
```

```
lambda_n=input('What is the lambda value for the night (in flushes/hour)?');
perc_var=input('What is the percent variation for the lambda values (in %)?');
perc_var=perc_var/100;
filename = 'test';%input('What is the file name to save this run under?','s');

% Now set it up to run for a third of the time with each lambda
flush_times=[;];
t = 0;
rand('state',sum(100*clock));%97531
while t <= T/3
  t = t - 60*(log(rand) / lambda_w/(1+perc_var*rand(1))(1-perc_var*rand(1)));
  flush_times=[flush_times,[t;1]];
end
while t > T/3 && t <=2*T/3
  t = t - 60*(log(rand) / lambda_e/(1+perc_var*rand(1))/(1-perc_var*rand(1)));
  flush_times=[flush_times,[t;1]]
end
while t > 2*T/3 && t <= T
  t = t - 60*(log(rand) / lambda_n/(1+perc_var*rand(1))/(1-perc_var*rand(1)));
  flush_times=[flush_times,[t;1]];
end
save('flush_times'); %for use if want to compare different timesteps

% Now account for any errors that will occur if the flush times are too
% close together
timestep=1/60;
j=2;
count=0;
while j <= length(flush_times)-1
  if flush_times(1,j)-flush_times(1,j-1) < timestep;
     flush_times(1,j) = flush_times(1,j)+timestep;
     count = count + 1;
  end
  if flush_times(1,j+1)-flush_times(1,j) < 0
     flush_times(1,j+1) = flush_times(1,j+1)+timestep;
  end
  j=j+1;
end
number_of_times_moved=count

%Now put this in a format that Simulink can understand and use
i=2;
times=[];
```

```
while i <= 2*length(flush_times)
  times(:,i-1) = flush_times(:,i/2);
  times(:,i) = [(flush_times(1,i/2)+timestep);0];
  i=i+2;
end

T = times(1,:)';
U = times(2,:)';

total_number_of_flushes = length(times)/2 - 1
```

sim_post.m

```
% This will read "timesforruns" and create the histogram
load('timesforruns.mat');


timeson = [];
timesoff = [];
timediff = [];
lnth = length(timesforruns(1,:));

% First find where the 1's and 0's change.  1 means pump on. 0 means pump off.
for i = 2:lnth
  if timesforruns(2,i-1) == 0 && timesforruns(2,i) == 1
    timeson = [timeson,timesforruns(1,i)];
  elseif timesforruns(2,i-1) == 1 && timesforruns(2,i) == 0
    timesoff = [timesoff, timesforruns(1,i)];
  end
end

if length(timeson) == length(timesoff)
  timediff = [timeson(1) timeson(2:end)-timesoff(1:end-1)];
else
  timediff = [timeson(1) timeson(2:end)-timesoff(1:end)];
end

% First, pullout the outliers and count them
j=1;
outlier_count = 0;
```

```
ln_timediff = length(timediff);
while j <= ln_timediff
  if timediff(j)>=25
    timediff(j)=[];
    outlier_count = outlier_count+1;
    ln_timediff = length(timediff);
  else
    j=j+1;
  end
end
outlier_count

% Now do the histogram
[N,X] = hist(timediff,[0.25:0.5:25]);
N2 = medfilt1(N,7);
figure(1); clf;
bar(X,N);
hold on;
grid on;
xlabel('Time (min)');
ylabel('Counts');
set(gca,'FontName','Times','FontSize',14);
xl = get(gca,'Xlabel');
yl = get(gca,'Ylabel');
set(xl,'FontSize',18,'FontName','Times');
set(yl,'FontSize',18,'FontName','Times');
%plot(X,N2,'r','Linewidth',2);
hold off;

total_number_of_runs=sum(N)

gfit=gamfit(timediff);
k = gfit(1)
lambda = 1/gfit(2)
```

Appendix B

Health Concerns Related to Ozone

Ozone, which is an extremely powerful oxidizing agent, is known to cause several health problems in humans. According to recent research, acute exposure to ozone concentrations in excess of 120 ppb can cause respiratory irritation, decreased lung function, and headaches ([133]). As a result, several federal agencies have established health standards intended to limit human exposure to ozone. For example, the Occupational Safety and Health Administration (OSHA) has established that workers should not be exposed to ozone levels greater than 100 ppb averaged over an 8-hour day ([133]).

Given the above health risks, the airflow monitoring system was designed to introduce ozone in concentrations far below federally-recommended standards. As a demonstration of the relatively low levels of ozone produced by this system, consider the two detector output waveforms shown in Fig. B-1. Both of the responses shown in that Figure were recorded at the same location using the same sensor. The upper trace shows the response during a two hour period when ozone was injected into the room using a series of high voltage pulses, and the bottom trace shows the response during a similar two hour window when no experiments were conducted. Note that the magnitude of the sensor resistance is quite similar in both cases. In fact, the average value during the two hour period when ozone is injected is approximately 27.3 kΩ, which is only slightly larger than the 27 kΩ average recorded when the high voltage power supply was de-energized. For the particular sensor used in these experiments, this change in resistance corresponds to a change in concentration that is less than 1 ppb. Furthermore, measured concentrations at various points in the room show similar behavior. Clearly, this evidence suggests that our system does not produce dangerous amounts of ozone during testing periods.

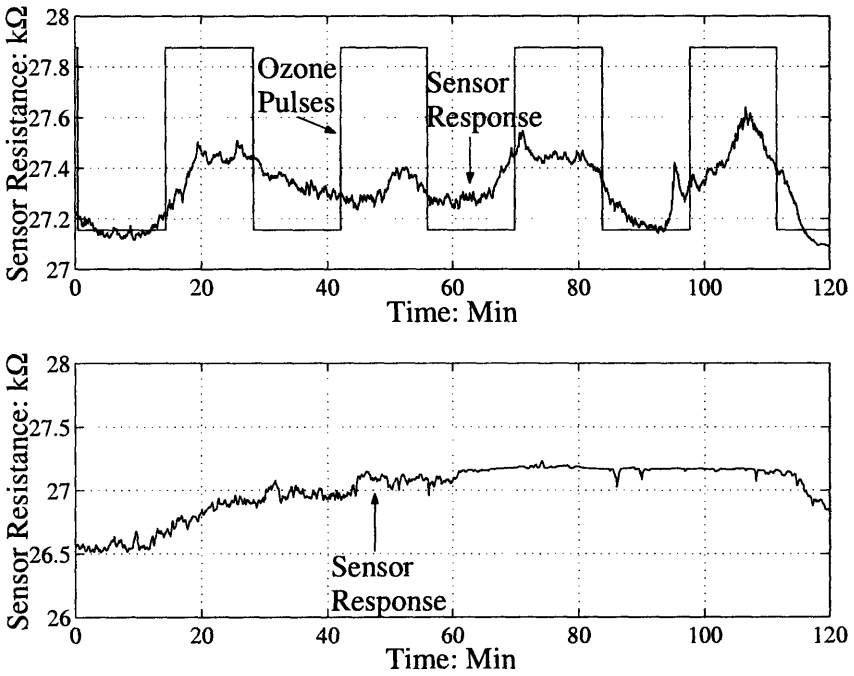


Figure B-1: Response of the same sensor over two different two hour time intervals. The upper trace shows the response when ozone is periodically injected into the room, and the bottom trace shows the response when no ozone is generated.

Appendix C

Nomenclature Used in Induction Machine Simulations

λ_{ds}	Direct-axis stator flux
λ_{qs}	Quadrature-axis stator flux
λ_{dr}	Direct-axis rotor flux
λ_{qr}	Quadrature-axis rotor flux
r_s	Stator resistance
r_r	Rotor resistance
X_m	Magnetizing reactance
X_{ls}	Stator leakage reactance
X_{lr}	Rotor leakage reactance
i_{ds}	Direct-axis stator current
i_{qs}	Quadrature-axis stator current
i_{dr}	Direct-axis rotor current
i_{qr}	Quadrature-axis rotor current
v_{ds}	Direct-axis stator voltage
v_{qs}	Quadrature-axis stator voltage
v_{dr}	Direct-axis rotor voltage
v_{qr}	Quadrature-axis rotor voltage
p	Number of pole pairs
ω	Frame speed
ω_r	Rotor mechanical speed

ω_p	Pump mechanical speed
θ_r	Rotor mechanical angle
θ_p	Pump mechanical angle
D	Damping coefficient
K	Spring stiffness coefficient
J_m	Moment of inertia of rotor
J_p	Moment of inertia of pump
τ_e	Electromagnetic torque
τ_l	Load torque
β	Viscous damping coefficient

Appendix D

V-NILM Implementation Details

This appendix presents schematics of each of the key components of the V-NILM. The following circuits are included:

- The line-to-neutral interface circuit (Fig. D-1).
- The 1× frequency synthesizer (Fig. D-2).
- The 3× frequency synthesizer (Fig. D-3).
- The line-to-neutral spectral envelope preprocessor. (Fig. D-4).
- The neutral-to-ground spectral envelope preprocessor. (Fig. D-6).

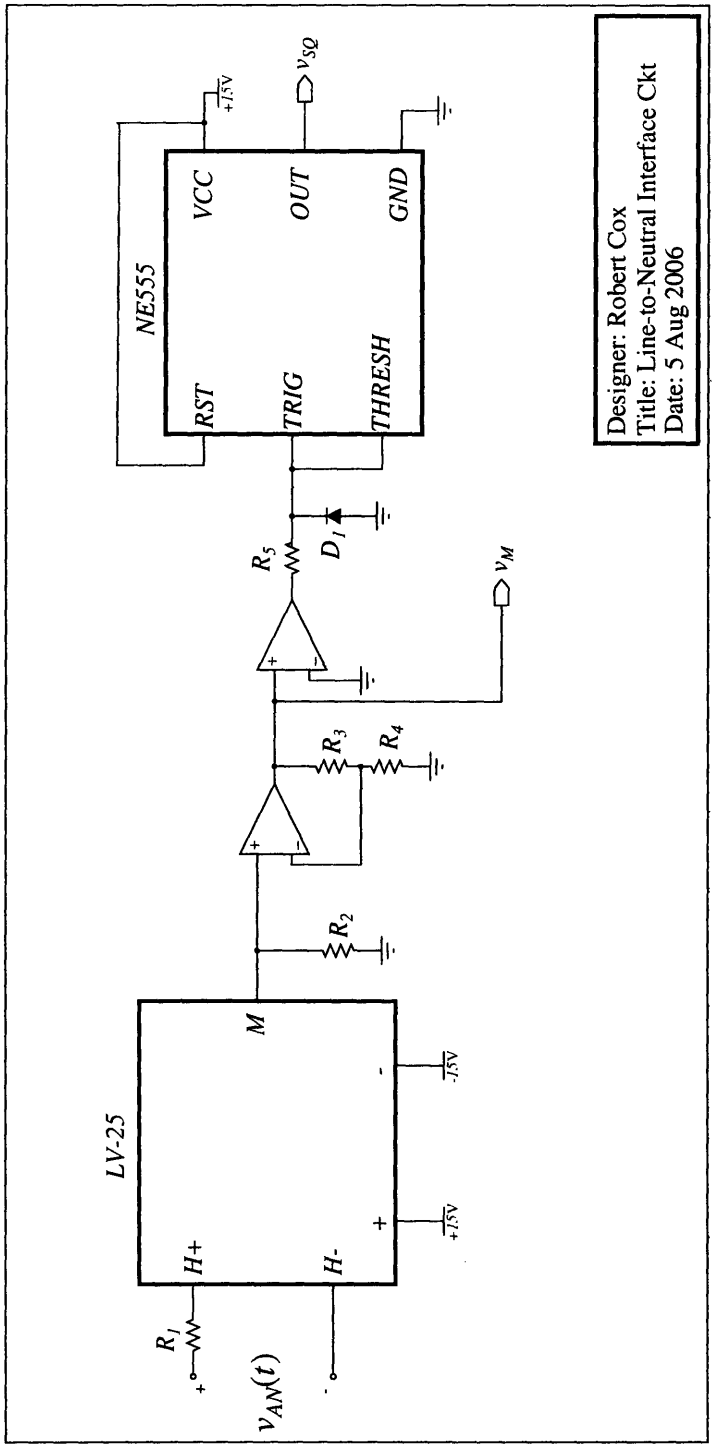


Figure D-1: Schematic of the line-to-neutral interface circuit. All of the op amps are OP-07s and all supply voltages are ±15 V unless stated otherwise.

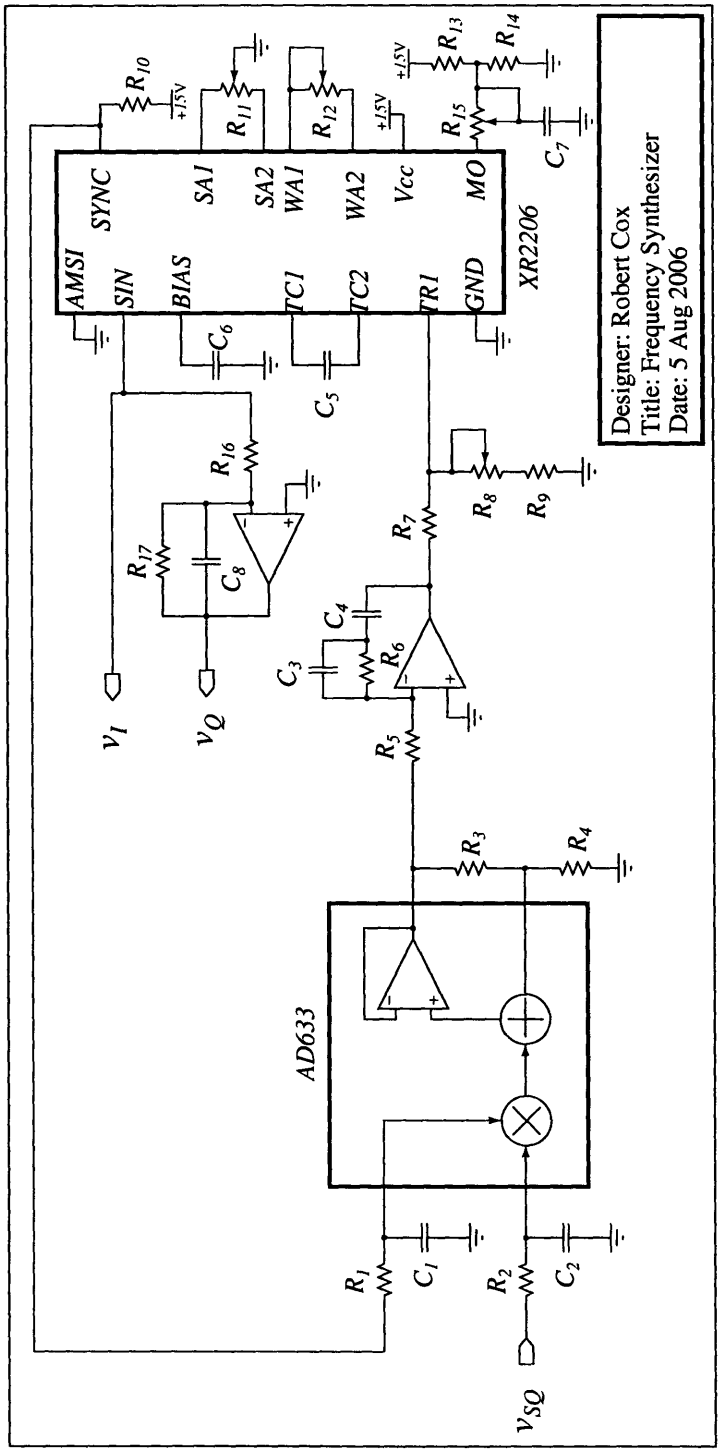


Figure D-2: Schematic of the 1× frequency synthesizer. All of the op amps are OP-07s and all supply voltages are ±15 V unless stated otherwise.

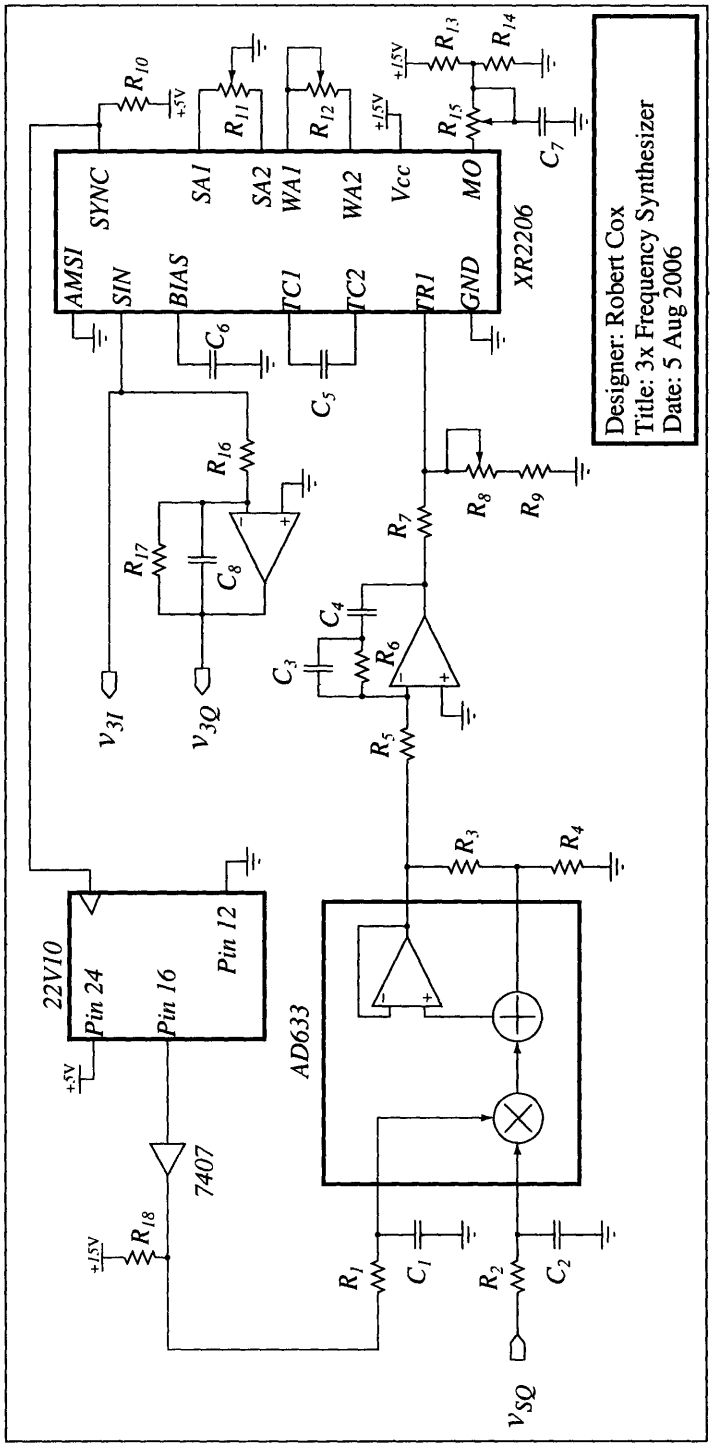


Figure D-3: Schematic of the 3× frequency synthesizer. All of the op amps are OP-07s and all supply voltages are ±15 V unless stated otherwise.

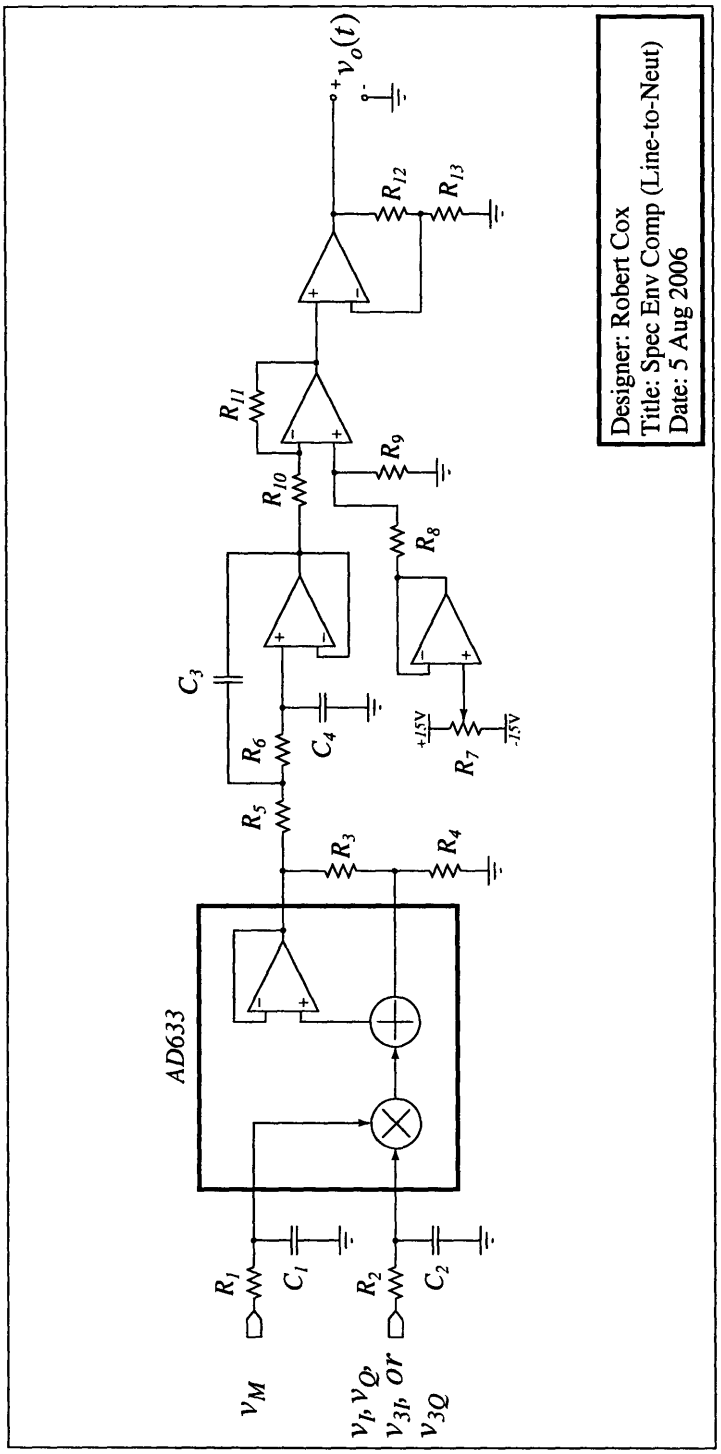


Figure D-4: Schematic of one of the spectral envelope preprocessor circuits. The input to this circuit is proportional to the line-to-neutral voltage. All of the op amps are OP-07s and all supply voltages are $+/-15$ V unless stated otherwise.

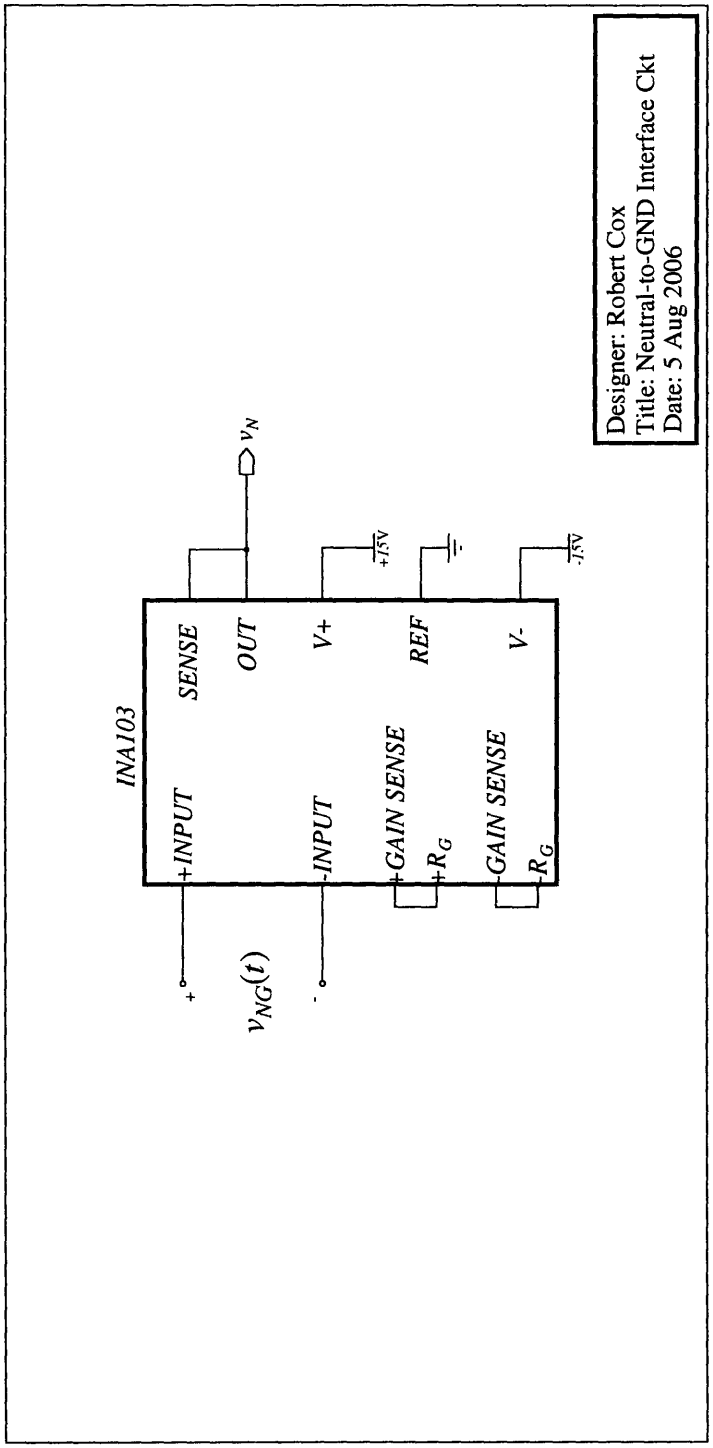


Figure D-5: The interface circuit for the neutral-to-ground voltage. All of the op amps are OP-07s and all supply voltages are $+/-15$ V unless stated otherwise.

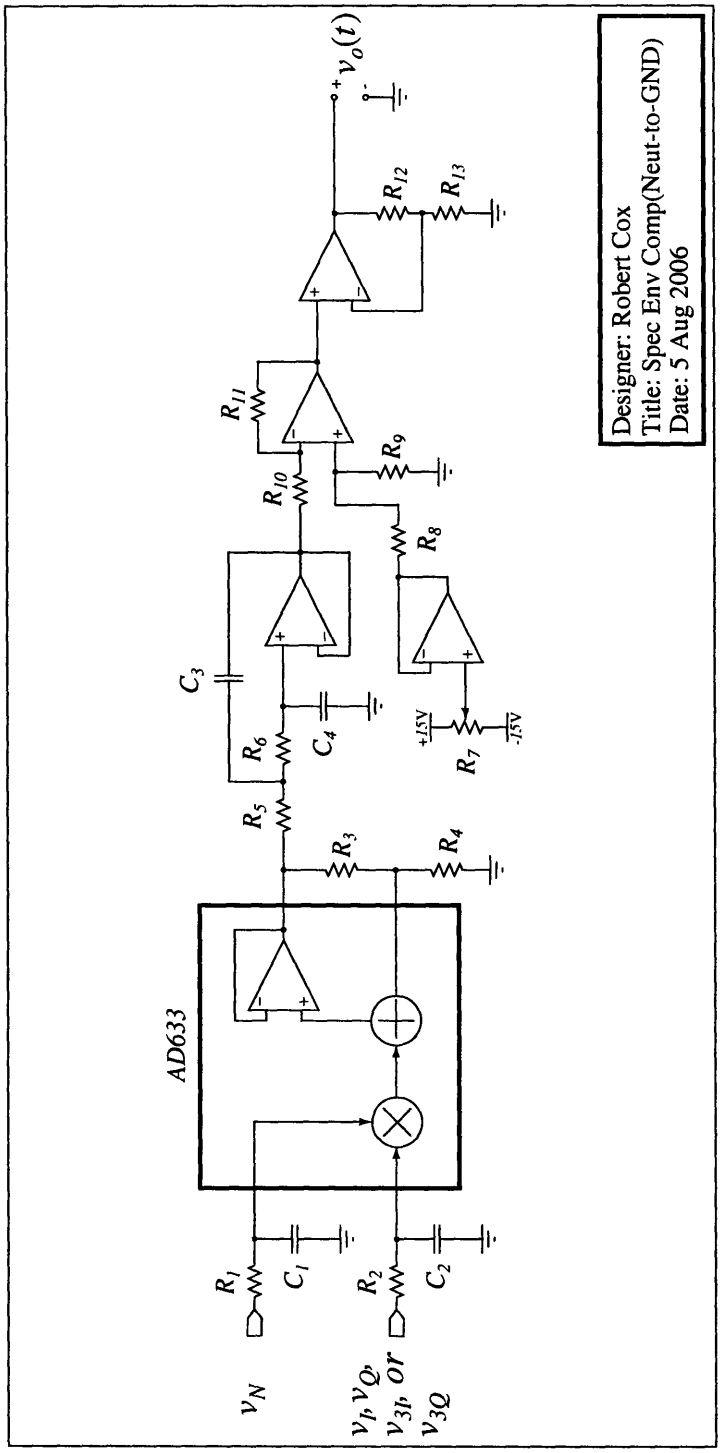


Figure D-6: Schematic of one of the spectral envelope preprocessor circuits. The input to this circuit is proportional to the neutral-to-ground voltage. All of the op amps are OP-07s and all supply voltages are $+/-15$ V unless stated otherwise.