AUTOMATED DIAGNOSIS OF LARGE POWER TRANSFORMERS USING ADAPTIVE MODEL-BASED MONITORING

by

Thomas Henry Crowley

Submitted to the Department of
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
In Partial Fulfillment of
the Requirements for the Degrees of

BACHELOR OF SCIENCE

and

MASTER OF SCIENCE IN COMPUTER SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 1990

Signature of Author _____
Department of Electrical Engineering and Computer Science
June 15, 1990

Certified by _____
Ignacio J. Perez-Arriaga
Thesis Supervisor

Certified by _____
Wayne H. Hagman
Thesis Supervisor

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

Acknowledgements

I would like to thank my thesis supervisors, Wayne Hagman and Ignacio Perez-Arriaga, for their unflagging support. I think I will remember most the conversations with Wayne that began "Just one thing...", and ended up lasting three hours and generating a completely new view of the problem. Ignacio's ability to reformulate a problem got me past several roadblocks. Wayne and Ignacio each provided a distinct perspective on this thesis, which grants it a depth of field that would not have been otherwise possible. I thank these two not only for being good advisors, but also for being good friends.

Thanking all the people who helped me and kept me going would read like a mixture of my family tree and laboratory directory, but there are several who deserve special mention. First, there are the administrators, for whom I was always an exception, requiring extra effort and extra aggravation. I thank SharonLeah Brown, Barbara Smith, and Vivian Mizuno of the Laboratory for Electromagnetic and Electronic Systems, and Marilyn Pierce of the EECS Graduate Office. Next, I would like to thank Anthony Callogero and Wayne Ryan for a wide variety of technical support. Mary Jane Close and Kathy McCue gave great assistance in controlling the transformer loading during the experiments. In fact, Mary Jane embodies the concept of "random input". Family and friends provided a haven from the pressures of writing a thesis. My housemates, Karen Covert and Joel Stein, provided a very relaxing environment in which to unwind at the end of the day. I also drew strength from several brothers and sisters who live in the area; I thank David, Cindy, Terry, Betsy, and Chris for their concern and support. Finally, I wish to thank my parents for always pushing, but never pushing too hard.

This work would not have been possible without financial support from a wide variety of sources. The Transformer Monitoring Project was sponsored through the MIT Energy Laboratory Electric Utility Program by: Allegheny Power System, American Electric Power Service Corporation, Bonneville Power Administration, Boston Edison Company, Empire State Electric Energy Research Corporation, New York Power Authority, Northeast Utilities Service Company, Southern California Edison Company, and Tokyo Electric Power Company, Inc.

Additional support for this thesis was given by J.W. Harley, Inc., Bonneville Power Administration, Boston Edison Company, Allegheny Power System, and Syprotec Inc. The Herbert R. Stewart family generously donated funds for the completion of this thesis.

Dedication

To my parents.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The failure of large power transformers is an area of significant cost and concern for electric utilities. In the opening address to the 1973 Conference on Diagnostic Testing of High Voltage Power Apparatus in Service, J.S. Forrest said:[1, p. 4]

> Reliability is probably the most important single problem on the electricity supply system at the present time—not only to give uninterrupted service but also to provide an economic supply. If, for example, a 500 MW generator transformer breaks down it may cost tens of thousands of [British] pounds a day to supply the demand by running a less efficient station—and that sum of money would pay for quite a lot of diagnostic testing to forestall the breakdown.

These words have as much significance today as they did in 1973. In addition to these operating costs, the capital costs associated with repairing or replacing a large power transformer that has suffered a catastrophic failure are considerable; a new transformer may cost as much as $1,000,000[2]. The annual failure rate of power transformers in the 400–500kV range has been estimated as 2%[3]. Even with this low failure rate, the tremendous costs associated with a transformer failure force utilities to purchase spare transformers and install redundant equipment, tying up capital and manpower needed elsewhere.

The ability to identify the existence of incipient failures in in-service transformers before the failures become catastrophic is extremely attractive. In-service monitoring of transformers is not new—however, monitoring has focused almost exclusively on the information revealed through infrequent transformer oil analysis. While this information is revealing, the reliability of this technique is inherently limited by the long sampling intervals. In addition, these tests consume valuable resources that could be used elsewhere. The reliability of this monitoring scheme is leveraged by the addition of sensors to continuously monitor quantities such as oil temperature and dissolved combustible gas content. But these measurements are only compared to static thresholds that do not account for the changing operating conditions of the

transformer. Improvements in modern microprocessor and integrated-circuit technology have enabled an on-line performance monitoring system that is capable of adjusting to changing operating conditions for increased reliability and sensitivity to incipient failure.

This thesis presents an on-line transformer monitoring system that can detect incipient failures. It is also shown that the information revealed by this system is sufficiently rich to enable the diagnosis of incipient failures; furthermore, this diagnostic process can be automated so that the attention of control-room operators can be directed elsewhere.

1.1 Background

The goals of in-service monitoring are accuracy and reliability. These goals can only be achieved through the repeated sensing of *multiple quantities* in conjunction with the recognition of long- and short-term drifts, or *trends*, in the condition of the transformer. Additionally, the uniqueness of every transformer, even among a group of the same basic design, requires a monitoring scheme which is sufficiently sophisticated to learn and interpret the characteristics of a particular transformer—that is, a scheme which *adapts*.

Under electric utility sponsorship, the Laboratory for Electromagnetic and Electronic Systems at MIT established the MIT Transformer Monitoring Project. This work, which was led by James R. Melcher and Chathan M. Cooke, was sponsored through the MIT Energy Laboratory Electric Utility Program. This effort concentrated on the development of basic sensors and on modeling the operation of various transformer subsystems. The culmination of this work was the development of an integrated monitoring system, consisting of hardware and software developed under this project.

This integrated monitoring system based on adaptive models of normal behavior was then applied to the Pilot Transformer Test Facility, a vehicle for advanced transformer monitoring research which was developed as part of the monitoring project. The heart of this facility is a heavily instrumented 50 kVA pole-type transformer. The Pilot Facility can provide information about the thermal, acoustic, electrical, and chemical behavior of this test transformer. While the data acquisition system is flexible enough to accommodate a variety of sampling rates, the work described in this thesis makes use of temperatures and high- and low-side currents and voltages sampled every two minutes; all other data is sampled at ten-minute intervals. The Pilot Facility, from which the vast majority of the data presented in this thesis is drawn, is described in detail in Appendix A. The monitoring system is referred to as the MIT Pilot Transformer Monitoring System.

1.2 Thesis Overview

The purpose of this thesis is twofold; it documents the status of the MIT Pilot Transformer Monitoring System as a sensitive detector of incipient transformer failure, and it explores how the system could be extended to diagnose the failures it detects. Chapter 2 reviews the traditional techniques for transformer monitoring, and critiques their strengths and shortcomings. Chapter 3 presents a conceptual overview of the MIT approach to transformer monitoring. Chapter 4 describes an implementation of this monitoring approach, as it is applied to the Pilot Facility, with special emphasis on the mechanism whereby the system adapts to the particular transformer. Chapter 5 discusses how this monitoring system might be integrated with traditional monitoring practice. Chapter 6 presents the data from several experiments simulating one class of transformer failures. Chapter 7 develops the diagnostic process that can be drawn from the information generated by the adaptive transformer monitoring system, and explores how this process could be automated. Chapter 8 summarizes the conclusions that can be drawn from these experiments, and discusses several recommendations for future work.

Chapter 2

Traditional Transformer Monitoring

Device monitoring consists of three phases: *detection*, *diagnosis* and *prognosis*. Detection is the process of screening various measurements of the device for information that may indicate that a newly developing, or incipient, failure is present. Diagnosis entails a detailed analysis of the measurements (possibly requiring additional measurements beyond those used in detection) to determine the likely problem, or to determine that there is, in fact, no problem. Prognosis is the process of extrapolating the likely future course of the device's health based on the information revealed by the detection and diagnosis phases. Due to the underdeveloped state of the art of transformer prognosis, and the limited scope of this thesis, this document is mostly concerned with methods of detection and diagnosis.

Traditional transformer monitoring is based primarily on infrequent oil samples, poorly correlated with temperature and the load history of the transformer. Both detection and diagnosis center on information gathered from these oil samples. To understand the significance of the information gathered from these samples, it is first necessary to understand the degradation processes and failure-triggering stresses present in an oil-filled transformer. The first section of this chapter consists of a brief explanation of the major failure processes, followed by a presentation of traditional techniques for on-line[1] detection and diagnosis of transformer failures. This latter discussion includes some systems which have not, as yet, been widely accepted by the electric utilities, but encompass some interesting advances in transformer monitoring.

[1] *On-line* tests are tests which can be performed while the transformer is energized.

2.1 Failures and Their Causes

The major contributing factors to the eventual failure of the transformer are poor maintenance, incorrect operation, and overloading. S. D. Meyers, *et al.*, state that "transformers do not die of old age, but are killed by neglect![4, p. 119]" The gradual degradation of the transformer is due to a number of chemical reactions that affect the mechanical and dielectric strength of the insulating system. The effects of these reactions on the insulating oil can often be reversed through processing of the oil, yielding oil that can be as good as, or better than, new oils. However, the loss of mechanical or dielectric strength of the solid insulation is often irreversible, short of rebuilding the transformer and replacing the insulation. Thus, the life of the transformer is determined by the life of the cellulosic insulation[4].

Some external events (such as lightning strikes, through faults, or switching surges) can precipitate an instantaneous failure of the transformer. However, transformers are designed to withstand these events to some degree. In the face of a very poor knowledge base concerning the precursors of catastrophic transformer failure, it is believed that these "instantaneous" failures are often preceded by a gradual degradation of the insulating system that renders the transformer more susceptible to failure. Therefore, it is important to understand the degradation processes of the insulating system to monitor all types of failures, even the "instantaneous" ones.

The major field enemies of power transformers are:[4, p. 138]

- Moisture
- Oxygen
- Heat
- Solid contamination
- Gas bubbles (supersaturation or oil decomposition)
- Overcurrent
- Overvoltage (transient or dynamic)
- Short circuit (mechanical forces)

Of these, the primary agents of destruction are moisture, oxygen, and heat. The role of each is presented in Section 2.1.1, Section 2.1.2, and Section 2.1.3, respectively. Finally, Section 2.1.4 describes the origin of the stresses that usually are the immediate causes of transformer failure.

2.1.1 Moisture

Moisture is the insulating system's archenemy. Its main target is the cellulosic insulation, though it plays a role in the degradation of the transformer oil, other solid insulations, and structural components of the transformer.

Water is always present in the transformer. Even with the most conscientious drying of the transformer, a moisture content of 0.3–1.0% of the dry weight of the paper is the best that can be expected. High temperatures are required to drive a greater amount of moisture from the cellulose, temperatures which cause thermal degradation of the cellulose. The moisture present in the paper represents a large quantity of water; when considering the moisture content of the oil, the paper may be viewed as an infinite source or sink for moisture. In addition, moisture can enter the transformer through a leak or as a product of oxidation.

Some moisture is actually necessary to preserve the mechanical properties of the paper; if it becomes too dry, it becomes brittle and loses mechanical strength. This is despite the fact that the electrical properties of the paper improve as the water is driven out of the cellulose. The increased dielectric strength is useless if the transformer fails due to a physical rupture of the paper insulation caused by the mechanical forces arising from vibrations, short circuits, or switching surges.

The cellulose contains the seeds of its own destruction in its glucoside chemical structure. Cellulose is basically long chains of glucose units ($[C_6H_{10}O_5]^*$). Each glucose unit has three hydroxyl (OH) groups: a primary hydroxyl group and two secondary hydroxyl groups. These hydroxyl groups form weak hydrogen bonds with the hydroxyl groups of adjacent cellulose chains to give the entire structure its mechanical strength. Water and other polar compounds are attracted to the hydroxyl groups, interfering with the hydrogen bonds and thus weakening the structure. This attraction of the hydroxyl groups contributes to cellulose's extremely hygroscopic (water-loving) nature.

The secondary hydroxyls also have the unfortunate property that they are liberated relatively easily when the cellulose is heated. The free hydroxyls can form water in a catalyzed reaction. This new water is retained primarily in the cellulose, where it continues to degrade the solid insulation. Water promotes depolymerization, whereby the long chains of glucoside units are broken up into smaller and smaller units. As this chemical breakdown occurs, the insulation can lose mechanical strength, and begin to flow and change shape.

Water is a polar compound with a high permittivity. Because of this, it is attracted to regions of high electric field density. Cellulose's great affinity for water causes it to absorb the water, drying out the insulating oil. The transformer thus stores water in the cellulose in regions of great dielectric stress—exactly where it is most dangerous! It is obvious that keeping the transformer dry is absolutely necessary for prolonging the life of the transformer.

Moisture also attacks the oil; water is a catalyst for the oil oxidation reaction—one of the products of which is water itself. Water interacts with the oxidation decay

products. Newly formed oxidation products are highly water soluble. This fact can be taken advantage of: if water can be extracted from the oil, these contaminants can also be removed before they damage the cellulose. Unfortunately, some of the polar compounds align at the water/oil interface, forming a structure which then attracts more moisture. The large particles which result from this interaction can circulate throughout the oil until they reach the cellulose or a region of intense electric field. Here the particles are deposited, reducing the dielectric strength at that location. The acids which result from oil oxidation are also water-soluble. The water transports the acid throughout the transformer, promoting corrosion even on components that are fully immersed in the oil.

The damage done by moisture is *permanent*; drying the transformer after the fact can only slow further deterioration. Therefore, keeping the transformer dry must be a high priority. Various conventional methods for monitoring the internal condition of the transformer are discussed later in this chapter. However, the extreme peril posed by moisture makes it important to note here that one test that has traditionally been relied on to reveal the presence of water in the transformer is inappropriate to that purpose. The standard oil dielectric test, ASTM (American Society for Testing and Materials) D-877, can reveal the presence of conductive contaminants (such as rust, dirt, cellulosic fibers, free water, etc.) but *not* the presence of dissolved moisture in the oil. This is because of the cellulose's high affinity for water: the solid insulation does an excellent job of dehydrating the oil. Thus, while there may be a great deal of dissolved moisture in the transformer, only a tiny portion of that moisture will be found in the oil. The moisture content of the insulation can be inferred from that of the oil, provided the two are in equilibrium. But ASTM D-877 does not have the necessary sensitivity to dissolved moisture to enable the moisture content of the system to be deduced[4, p. 316–318].

2.1.2 Oxygen

Second only to moisture in destructive potential, oxygen is a major enemy of the insulating oil. Oxygen is derived from the atmosphere, or from the thermal degradation of cellulose. But even though oxygen may be generated internally from the cellulose, it is desirable to have a low initial oxygen content to retard the aging of the cellulose and oil. However, even in sealed evacuated transformers, the amount of oxygen remaining would, at standard conditions, fill at least 0.25% of the volume of the tank[4, p. 180]. Unstable hydrocarbon impurities combine with the oxygen under the catalytic influence of other materials in the transformer. Pure hydrocarbons are not easily oxidized and can generally be reclaimed from used oil that is otherwise completely unusable.

The problem of oxidation is worsened by the presence of *catalysts* and *accelerators*. Catalysts are materials that increase the rate of reaction without being consumed themselves. The two major catalysts for oxidation are:

- Moisture
- Copper

Moisture is the primary catalyst for oil oxidation. Moisture is always present in the cellulose, and can be driven out of the paper and into the oil by heat. There is such a great quantity of water in the cellulose in comparison to the oil that additional moisture will always be driven out with increased temperature. Moisture can enter the transformer through a leak in the tank, and is also a product of the oxidation reaction itself.

Accelerators are secondary factors that increase oil oxidation. These factors include:[4, p. 182]

- Heat
- Vibration
- Shock loading
- Surge voltages
- High electrical stress

As is true for most chemical reactions, heat is a major accelerator of oxidation. But oxidation is just one of a family of reactions that attack the integrity of the transformer's insulating system. The major role that heat plays in inducing transformer failures is discussed in the following section, Section 2.1.3.

Vibration also speeds up the rate of reaction. The transformer is constantly vibrating in response to the 60 Hz applied power. The major frequency components of this vibration are located at the multiples of 120 Hz.

High electrical stress can increase the amounts of hydrogen and light hydrocarbons produced as a result of oxidation. Electrical stress also promotes the formation of larger particles in the resulting sludge deposits, and can influence the location and shape of these deposits. These deposits can interfere with the heat transfer characteristics of the transformer, they can form electrically conducting "bridges" on surfaces (decreasing the effectiveness of the insulating system through increased heating and increased electrical stress), and they can adversely affect the mechanical strength of the transformer[4, pp. 183–184].

The major stages of the oil deterioration are characterized by the formation of peroxides, acids, alcohols, ketones and sludges. Once a significant amount of peroxides are formed, a chain reaction begins which eventually results in the production of sludge. Sludge is a resinous, partially conductive substance which precipitates out of the oil onto the insulation, the tank, the surfaces of the ventilating ducts, the cooling fins, etc. Even before the sludge has precipitated out to form a heavy, tarry substance, its component decay products have begun to attack the cellulose,

resulting in *shrinkage*. This shrinkage damages the transformer's ability to withstand shock loading; an inability to withstand shock loading can result in movement of the coils and lead to premature failure. Once the sludge has been deposited, the transformer's ability to dissipate heat may be decreased, also leading to damage of the device due to excessive heating.

Natural crude oil contains compounds that interfere with the oxidation process. These compounds are called *inhibitors* or *antioxidants*. Inhibitors are consumed during the service life of the transformer. This means that the rate of oxidation of the oil will increase as the oil ages, an important consideration.

2.1.3 Heat

Excessive temperatures in a transformer are destructive, both because heat is a major accelerator of the chemical degradation due to moisture and oxidation, and because thermal degradation of the cellulose and oil is itself a significant problem (even in the absence of moisture or oxygen). Some incipient faults, such as local short-circuits between laminations of the core, can generate temperatures sufficient to char insulation or even distort the entire core[5, p. 619].

High temperatures can cause cellulosic insulation to shrink and become brittle. This leaves the solid insulation susceptible to failure due to mechanical stress. As mentioned above, thermal degradation of the cellulose yields free water, as well as certain gases. Carbon monoxide (CO) and carbon dioxide (CO_2) are particularly indicative of overheating of the cellulose. If the free water or the gases remain where they are generated, they pose a serious threat. The effect of water on the cellulose is discussed above, but the gases can be every bit as dangerous and, in fact, can pose a more immediate danger to the transformer. The gases may collect in bubbles in regions of high dielectric stress, displacing the oil and lowering the dielectric strength of the insulation. This reduces the transformer's ability to withstand short circuits, impulses, and switching surges; the transformer may actually fail under merely normal dielectric stress, if the dielectric strength of the insulation is reduced enough. At temperatures lower than 150°C, the gases from the thermal decomposition of the cellulose are not significant, but vaporization of the water adsorbed by the cellulose may pose a similar threat. Because of this, the higher the moisture levels in the cellulose, the lower the temperature at which bubble production becomes a problem[6].

General overheating contributes greatly to the aging of the transformer. A rule of thumb used for chemical reactions says that, all other things being equal, the rate of a given reaction will double for each increase of 10°C. (For obvious reasons, this is referred to as the *10°C rule.*) This rule is applied to insulating oil when its temperature exceeds 60°C. For each 10°C increment above 60°C, the expected useful oil life would be half as long: the oil in a transformer operated continuously with an oil temperature of 70°C would be expected to last half as long as an oil

Table 2.1: Typical Transformer Oil Content by Design Period

Year	Gallons/KVA
1915	2
1930	1
1945	1/2
1960	1/3
1975	1/6
1977	1/7
1979	1/9

kept continuously cooler than 60°C. This rule is applied to solid insulation as the *8°C rule*—the life expectancy of the paper is halved for each 8°C rise. In fact, the IEC (International Electrotechnical Commission) believes that in the range 80–140°C a *6°C rule* should be used. In [4], the authors state that "at 70°C even a nitrogen blanket transformer has sufficient oxygen available through breakdown of the cellulosic insulation to cause the **oil** to deteriorate as rapidly as the liquid in a free-breathing transformer![4, p. 583]"

2.1.4 Stress

The degradation processes discussed above greatly affect the ability of the insulating system to withstand the stresses produced by the normal and abnormal operation of a large power transformer. The insulating system of a transformer is subjected to three types of stress:

- Dielectric Stress
- Mechanical Stress
- Thermal Stress

Usually, one, two, or all three of these stresses are the immediate trigger of a catastrophic failure. For example, dissolved moisture in the cellulose may lower the dielectric strength of the solid insulation, but it may be an external event (such as a lightning strike) that raises the dielectric stress to the point of failure. This section briefly discusses the origins of these stresses.

These stresses become ever more significant as new designs try to minimize the amount of material used in the transformer. Traditionally overdesigned, transformers are now having their safety margins trimmed to the bare bone. Table 2.1 shows how the amount of oil per KVA has dropped in this century[4, p. 128].

2.1.4.1 Dielectric Stress

Dielectric stress arises from three major sources. First, there is the voltage stress induced by the normal 60 Hz excitation of the primary winding. Second, dielectric stress is induced by the voltage surges arising from lightning strikes to or near the transmission lines. Switching surges are the third source. While a number of operations can produce switching surges, line switching is the most common cause. A lightning surge is characterized in laboratory studies as a disturbance which reaches its peak in 1.2 μsec and decays to half the peak value in 50 μsec; a switching surge typically takes about 230 μsec to reach its peak and decays to half value in 2000 μsec. In contrast, the normal 60 Hz excitation takes 4167 μsec just to reach its maximum.

Dielectric stress may become concentrated in a particular area due to tracking or coil movement, increasing the vulnerability of the insulation system.

2.1.4.2 Mechanical Stress

The 60 Hz excitation results in a mechanical vibration with harmonics at multiples of 120 Hz. This mechanical stress is present at all times that the transformer is in operation. When the transformer is initially installed, the effects of this vibration on the winding and the core are negated by properly designed, properly adjusted, clamping apparatus. As the transformer ages, this clamping apparatus may be rendered less effective by shrinkage of the solid insulation, loosening of the clamping mechanism, and distortion of the core and coils. The transformer is then susceptible to damage from friction and impact between loose components. At one time, the transformer was given little maintenance attention because it was thought to have no moving parts. It is now known the transformer is in constant motion while energized.

Mechanical stress on the windings and the solid insulation is produced in normal operation through the interaction of load currents flowing in the windings and the magnetic field through which energy is transferred from the primary to the secondary in the transformer. *Through-faults* are events in which excessive current flows through the transformer, usually as a result of a short circuit. Failures due to through-faults are due primarily to the damage produced by mechanical forces in the windings, rather than thermal damage. The intense electromagnetic forces produced by the abnormal current can physically displace the coils, causing an internal short circuit in the transformer. The forces induced in the winding are related to the square of the current; thus, a current twenty times normal will result in forces that are *four hundred* times normal.

2.1.4.3 Thermal Stress

A transformer is an extremely efficient device, yet there are always some losses, which manifest primarily as heat. Certain losses occur whether or not the transformer is loaded. These are called *iron losses* or *no-load losses*, and include core magnetization losses, copper losses in the primary due to the exciting current, and dielectric losses in the insulation. Core losses are the dominant factor, hence the term "iron losses" from the iron (or steel) core. The core losses consist mostly of hysteresis (the power necessary to magnetize the core alternately in one direction, then the other) and circulating and eddy current losses (due to currents induced in the electrically-conductive iron of the core).

Losses which occur due to load are called *copper losses* or *load losses*. Produced by the power lost when the current flows through the resistance of the windings, these losses are also called the I^2R *losses*. The I^2R losses become extremely significant when the transformer is run overloaded for any great length of time. There are also losses generated by small eddy currents in the copper conductors, and losses due to the effect of leakage flux on the tank and other conductive components of the transformer[4, p. 26–27].

An oil-filled transformer relies on the oil to carry damaging heat from the core and the windings to the tank, where it is transmitted to the external environment. The ability of the transformer to dissipate heat can fail or degrade in a number of ways. A cooling duct may become blocked; sludge may interfere with the heat transfer at the oil/surface interface. Pumps for forced convection may fail. Radiating fins on the tank may be damaged. For any of these reasons (among others), normal rates of heat generation will result in higher internal temperatures, increasing the thermal stress on the device.

Other failures can cause excess heat to be generated, resulting in a local hot spot that can damage the insulation or distort components of the transformer. One such failure, called a transposition failure, is when the insulation fails between the individual wires which make up a single conductor in the winding. Abnormal eddy or circulating currents can generate excess heat. Likewise, if the insulation fails between laminations of the core, circulating currents arise. If the insulation fails in such a way as to form a circuit which links a significant portion of the total flux through the core, the heat produced can distort or melt the entire core and char the winding insulation. If a conductor in a bar breaks, the resistance of the winding will increase and excess heat will result. As a final example, if there is a DC component to the AC flux, the core may become saturated during half the cycle, leading to excess heating. Under the influence of solar activity, this can be a serious problem, threatening the entire power system.

2.2 Detection

In this section, present techniques for the detection of transformer failures are presented. The most widely used methods of detection involve the monitoring of various oil properties[7] at intervals ranging from three months to three years. When a problem is suspected, the sampling interval is adjusted accordingly. In addition to the physical, electrical and chemical properties of the transformer oil, other quantities and status indicators (such as temperature or oil level indicators) are sometimes monitored.

2.2.1 Oil Sample Analysis

Most insulating mineral oils used in transformers are complex mixtures of paraffinic and aromatic hydrocarbons. These mixtures act as high-voltage insulation and as a medium for convective cooling of the transformer. In an operating transformer, analysis of the insulating oil can reveal insights into the health of the transformer. Various analytic tools can be used to measure the state of the oil and, indirectly, the state of the solid cellulosic insulation.

Standard D-117 of the American Society for Testing and Materials (ASTM), *Standard Method of Testing and Specifications for Electrical Insulating Oils of Petroleum Origin*[7, Appendix A], lists thirty-three oil properties and fifty-five test methods. Of these, ten tests are suitable for in-service monitoring of transformer performance. The other measures are useful only during the design phase or during acceptance testing of the oil, or are superseded by one of the ten central tests.

The ten tests are:[4, p. 262]

- Dielectric Breakdown Strength (D-877/D-1816)
- Neutralization Number (D-974)
- Interfacial Tension (D-971)
- Color (D-1524)
- Moisture Content (D-1533)
- Specific Gravity (D-1298)
- Visual Examination (D-1524)
- Sediment (D-1698)
- Power Factor (D-924)
- Dissolved Gas Analysis (D-3612)

2.2.1.1 Dielectric Breakdown Strength

The dielectric breakdown voltage of insulating oil at 60 Hz is an important measure of the oil's ability to withstand dielectric stress without failure. These tests measure the voltage at which breakdown occurs between two electrodes under prescribed test conditions. They can reveal the presence of contaminants in the oil, such as free water, dirt, or conducting particles. However, they can *not* reveal the presence of dissolved moisture concentrations below 60% of the saturation level in the oil.

2.2.1.2 Neutralization Number

The neutralization number of an oil sample is the number of milligrams of potassium hydroxide (KOH) that is necessary to neutralize a one gram sample of oil. This test measures the acidity (due to oil oxidation) of the oil, and is a measure of the level of deterioration of the oil.

2.2.1.3 Interfacial Tension

Interfacial tension is the molecular attractive force between dissimilar molecules at an interface. In this test, the tension at an oil/water interface is measured. Two methods for measuring this force entail the measurement of the amount of force needed to pull a platinum wire away from such an interface or the determination of the largest droplet of water that can be supported by the oil.

This test is a sensitive mechanism for detecting the presence of soluble polar contaminants and the products of oxidation. These oil contaminants are hydrophilic; the attraction across the interface between the oil and the water interferes with the mutual attraction among the water molecules at the boundary. This interaction results in a decrease in the strength of the film at the interface between the two materials. Thus, this reduction in strength indicates the presence of the contaminants.

Interfacial tension has a definite relationship to the neutralization number of an oil sample. An increase in the neutralization number should be followed by a decrease in the interfacial tension. This correlation makes it possible to achieve a reliable measure of these two quantities; if the relationship fails to hold, the measurements are suspect. These two quantities have been related directly to sludging. The result is a classification system for oil quality, seen in Table 2.2[4, p. 289].

2.2.1.4 Color

The color of transformer oil is measured through comparison with a series of tinted glass standards, calibrated from zero to eight. An increase in the color number is an indication that the oil has deteriorated or become contaminated.

Table 2.2: Transformer Oil Classifications

NN	IFT	IFT/NN	Color	Diagnosis
0.00–0.10	30.0–45.0	300–1500	Pale yellow	Good oil
0.05–0.10	27.1–29.9	271–600	Yellow	Proposition A oil
0.11–0.15	24.0–27.0	160–318	Bright yellow	Marginal oil
0.16–0.40	18.0–23.9	45–159	Amber	Bad oil
0.41–0.65	14.0–17.9	22–44	Brown	Very bad oil
0.66–1.50	9.0–13.9	6–21	Dark brown	Extremely bad oil
1.51+			Black	Oil in disastrous condition

2.2.1.5 Moisture Content

Moisture content of the oil is measured by the Karl Fischer titration technique. This method is based on the reduction of iodine by water and sulfur dioxide. Another method involves the liberation of moisture from the oil sample through heat, vacuum, and agitation. The water is then collected in a moisture-absorbing trap containing P_2O_5. The increase in weight of the trap is reveals the amount of collected water.

2.2.1.6 Specific Gravity

Specific gravity is the ratio of the mass of a given volume of liquid weighed in vacuum at 15.6°C (50°F) to an equal volume of pure water, also measured at 15.6°C. If the specific gravity of an oil sample is greater than unity, then contamination may be indicated.

2.2.1.7 Visual Examination

The oil may also be visually examined for increased cloudiness, particles of insulation, metal corrosion products or other undesirable suspended materials in the oil. Visual inspection is sometimes used to screen oil samples in the field. If no deterioration is evident, then the sample is not sent to a central laboratory for complete quantitative evaluation. Good oil should be clear and sparkling. Cloudiness indicates moisture, carbon, and/or sludge. If carbon is present, arcing or partial discharge should be suspected and a dissolved gas analysis should be conducted. An unusual smell should also be investigated further.

2.2.1.8 Sediment

In this test, the oil sample is centrifuged to separate the oil and the sediment. The oil is decanted and saved. The sediment is dried, weighed, burned, and weighed

again. The difference between the two weights is organic material; the material that remains is inorganic, such as rust.

The decanted portion of the sample contains soluble sludge. The sludge is precipitated out by diluting the sample with n-pentane, a hydrocarbon. This occurs because the sludge is not soluble in n-pentane. The amount of sludge which precipitates out is an indication of the condition of the oil.

2.2.1.9 Power Factor

Power factor is defined as $\cos\phi$, where ϕ is the phase angle between the applied voltage and the resulting current. It is equivalent to the ratio of power dissipation in watts to the product of voltage and current in volt-amperes. The power factor is a measure of the dielectric losses in the oil, and thus the amount of heat dissipated. This test involves applying a 60 Hz voltage to an oil sample in a three-terminal or guarded-electrode test cell. The power factor is measured using a bridge circuit, according to the instructions appropriate to the bridge.

The power factor is sensitive to the presence of polar compounds in the oil, but it is not able to distinguish between different polar compounds. Because of this, the test is primarily used as a negative test to determine if the oil has *not* changed. If this test indicates the presence of polar compounds, other tests become necessary. The power factor test may detect sludges, but some types of sludges have very little effect on the power factor. If the power factor at 100°C is more than seven to ten times the measure at 25°C, then a soluble contaminant other than water is indicated.

2.2.1.10 Dissolved Gas Analysis

Gas is extracted from the oil sample using vacuum extraction. The volume of oil is measured at atmospheric pressure. The gas sample is then analyzed using gas chromatography. High levels of gases which result from the deterioration of the oil or the paper are a strong indication of incipient failure. The relative levels of gas production can be used to distinguish between different types of faults (see Section 2.3).

2.2.1.11 Baseline Oil Properties

The following tests of transformer oil are sometimes used in the detection of failures in large power transformers:

- Qualitative infrared absorption (D-2144)
- Refractive index and specific optical dispersion (D-1218/D-1807)
- Resistivity (D-1169)

- Oxidation inhibitor content (D-1473)
- Peroxide number (D-1563)

Qualitative Infrared Absorption The infrared absorption spectrum is recorded from 2.5 to 15 μm (4000 to 667 cm^{-1}.) The spectrum of transformer oil indicates the general chemical composition of the sample. Because of the complexity of the mixture of components which form electrical insulating oil, the spectrum is not sharply defined. Thus, it is difficult to quantitatively evaluate the composition of the oil based on the infrared absorption. Instead, the spectrum is compared to a baseline spectrum. This test is suitable for quickly determining whether the sample has the same composition as previous samples.

Refractive Index and Specific Optical Dispersion The *refractive index* of a substance is the ratio of the velocity of light in air to its velocity in the substance. The refractive index is an indication of the nature and amount of contaminants held in solution.

Specific optical dispersion is the difference between the refractive indexes of light of two different wave lengths, both indices measured at the same temperature. This difference is divided by the specific gravity measured at this same temperature.

Specific optical dispersion is a quick index as to the amount of unsaturated compounds present in the transformer oil. Since the dispersion values of paraffinic and naphthenic compounds are nearly equal and mostly independent of molecular weight and structural differences, values above 97 are directly related to the amount of aromatic compounds in the oil.

Resistivity High resistivity indicates a low concentration of free ions and ion-forming particles. This normally means the oil is free of conductive contaminants.

Oxidation Inhibitor Content During service, levels of natural or added oxidation inhibitors become depleted. Thus, while this test is not an indicator of incipient transformer failure, it can be used in estimating the remaining service life of the oil and recognizing the increased threat of oil oxidation.

Peroxide Number The peroxide number indicates the quantity of oxidizing constituents in the oil. It does this by measuring those compounds that oxidize potassium iodide (KI).

2.2.2 Temperature

Traditional transformer temperature monitoring is limited to the measurement of a single temperature, top oil temperature. From this temperature, the "hot spot"

temperature is estimated by adding 15°C.

Normal design criteria for transformers allow for a 65°C average rise of top oil temperature over ambient temperature. Thus, operated in an ambient temperature of 30°C at full rated load, the transformer will have a top oil temperature of 95°C and an estimated hot spot temperature of 110°C.

Since the manufacturers claim that their products can be operated continuously at 110°C hot spot temperature, many utilities choose to adopt this as their operating criteria. Emergency loading guides even allow for higher hot spot temperatures, although only for a short period of time. But, as discussed above, elevated temperatures have a severe negative effect on the longevity of the insulating system. A proposed ANSI/IEEE Loading Guide says that a fully-loaded, 65°C winding-rise transformer in a 30°C environment has a normal life expectancy of only 7.5 years. Lowering the hot spot temperature by 25°C increases the estimate to more than 40 years[4, p. 587]!

The top oil temperature is monitored by many utilities to ensure that the estimated hot spot temperature does not exceed their adopted guidelines. If the top oil temperature is below the upper limit, its value is ignored.

2.2.3 Partial Discharges

A *Partial discharge*, also called *corona*, is a transient electric discharge which only partially bridges the insulation gap between two conductors, as opposed to a breakdown, which completely bridges the gap. These discharges are important because they are indicative of a loss of dielectric strength of the insulation or an increase in electric stress, such as might occur if two conductors are moved closer together by through-fault forces. Partial discharges can be detected either electromagnetically, acoustically, or chemically (i.e., through dissolved gas analysis, see Section 2.3)[8,9].

If the partial discharges persist over a long period, or are particularly energetic, then the dissolved gas analysis is likely to detect the incipient problem. But because of the information that partial discharges impart concerning the insulation's dielectric strength, it is desirable to get this data as early as possible. Continuous discharge in the solid insulation may quickly be followed by a complete failure of the insulation. For this reason, electromagnetic and acoustic techniques for the detection of discharges have received a great deal of attention. With sufficient information, it may be possible to compute the energy of the discharge, and thus estimate the ensuing damage[10].

While it is possible to detect excessive partial discharges, it is very difficult to determine their location, either their electrical location or their physical location. In fact, one source deemed it unlikely that a system for the automatic location of partial discharges will be developed[11].

2.2.4 Other Detection Methods

2.2.4.1 Gas Relay

The Buchholz relay was introduced in 1928. This device detects the passage of gas bubbles between the transformer and the conservator tank. A drawback is that it can only be installed on conservator-type transformers.

The relay is an oil-tight compartment with two internal elements that are connected to mercury switches. The elements are buoyant, and float in the oil. An incipient fault may generate small bubbles that then get trapped in the housing of the relay, displacing the oil. As the oil level falls, the top floating element descends and, when enough oil is displaced, the circuit is tripped.

A serious fault will generate gas much more violently. This will deflect the bottom element, and again a circuit is tripped[5, p. 611]. Because of the remoteness of the relay from the actual point of failure, it is impossible for the device to provide *early* warning of a failure. Dissolved gas analysis provides a more direct way of determining the gas production of a failure.

2.2.4.2 Hydran

All types of failures that involve gassing produce significant levels of hydrogen. For this reason, there has been a great deal of attention directed to the continuous monitoring of dissolved hydrogen levels.

Syprotec's Hydran 201R® sensor is one that gives a continuous indication of the level of hydrogen dissolved in the oil. Additionally, the Hydran is partially sensitive to three other gases besides hydrogen (carbon monoxide, ethylene, and acetylene). The sensor produces a reading that is a composite of the concentrations of the four gases. The reading reflects 100% of the hydrogen concentration, 15% of the carbon monoxide, 8% of the ethylene, and 1% of the acetylene.

2.2.4.3 Westinghouse TMS

The Westinghouse Transformer Monitoring System (TMS) is a continuous on-line monitor of transformer performance[12,13]. TMS continuously monitors the following inputs: top oil temperature, cabinet temperature, winding current, relative corona, gas in oil, and status inputs (oil level, oil flow, pump bearings, and cooling equipment contactors).

The system indicates a problem exists if:

- Top oil temperature exceeds 120°C.
- Hot spot temperature exceeds 150°C. Hot spot temperature is the sum of the top oil temperature and the estimated temperature rise based on load current.
- Control cabinet temperature exceeds 70°C.

Table 2.3: TMS Transformer Performance Classification

Action	Top Oil Temperature (°C)	Calc. Hot Spot (°C)	Relative Corona (%)	H_2 in oil (ppm)	Change of H_2 in oil (ppm/day)	Oil Level	Oil Flow	Cooling Bank	Pump Bearing
Normal Periodic Reporting	90	+ 105	+ 10	+ 300 200*	+ 25	+ On	On	+ On	+ Off
							Off	+ Off	
Increase Data Acquisition by Host Computer	90	or 105	or 10	or 300 200*	or 25				
Contact Utility: more extensive analysis required	100	or 140	or 20	or 500 400*	or 50				
						Off			
							On	+ Off	
							On	+ On	
									On
Contact Utility: recommend reduce load	105	or 150	or 30	or 1000 700*	or 75				
			25	+ 800 600*					
			25		+ 50				
Contact Utility: recommend take transformer off-line	110	or 170	or 50	or 1500 1000*	or 100				
			30	+ 750 900*					
			30		+ 75				

(* units with gas space)

- Any analog channel is outside of operator-specified bounds.
- Oil level drops too low.
- Non-contaminating oil pump bearing wear indicator is activated.
- Cooler group contactor is energized, but no oil flow in the coolant circulation system is indicated.
- Analog-to-digital converter fails self-test.
- Data is copied during a system reset or power restore.

While TMS does not try to perform a diagnosis of an indicated problem, it does attempt to classify the performance of the transformer into five classes, with five associated actions: normal operation, increased data acquisition frequency, contact utility—further study required, contact utility—recommend reduced load, and contact utility—recommend shutdown. Table 2.3 summarizes this classification process.

2.3 Diagnosis

The primary tool for power transformer diagnosis is dissolved gas analysis. Dissolved gas analysis has proven to be a powerful tool for diagnosing incipient trans-

former failures. The analysis is based on the fact that many important failure types generate gases in predictable ways. The gases produced in the greatest quantities are: hydrogen (H_2), nitrogen (N_2), carbon monoxide (CO), carbon dioxide (CO_2), oxygen (O_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2).

Dissolved gas analysis can distinguish between different types of failure because each type of failure tends to produce the various gases at different rates. Several different methods have been proposed for interpreting the dissolved gas data:[4, p. 347]

- Detected gases
- Key gas
- Key components
- Dörnenburg ratios
- Amount of key gases
- Total combustible gases
- Rogers ratios
- IEC ratios
- Combustible concentration limits
- 90% norms concentration

Each method takes advantage of the "failure fingerprint" in a slightly different way. The two most widely used methods are the Key Gas and Rogers Ratios methods.

2.3.1 Rogers Ratios

The Rogers Ratios method is a technique for interpreting the results of a dissolved gas analysis. The Rogers method is refinement of the Dörnenburg Ratios method, one of the earliest methods of interpreting the dissolved gas concentrations. Two ratios of dissolved gas concentrations (methane/hydrogen and acetylene/ethylene) were used in the Dörnenburg Ratios method. The Rogers method uses four gas concentration ratios (methane/hydrogen, ethane/methane, ethylene/ethane, and acetylene/ethylene). The classification method for these ratios is summarized in Table 2.4 and Table 2.5[14].

A variation on the Rogers method is presented by DiGiorgio, *et al.*, in [15]. Figure 2.1 shows their graphical representation of the Rogers method. This variation is notable not only because it efficiently conveys the encoding and diagnosis table of

Table 2.4: Rogers Ratio Code

Ratio	Range	Code
CH_4/H_2	≤ 0.1	5
	$> 0.1, < 1.0$	0
	$\geq 1.0, < 3.0$	1
	≥ 3.0	2
C_2H_6/CH_4	< 1.0	0
	≥ 1.0	1
C_2H_4/C_2H_6	< 1.0	0
	$\geq 1.0, < 3.0$	1
	≥ 3.0	2
C_2H_2/C_2H_4	< 0.5	0
	$\geq 0.5, < 3.0$	1
	≥ 3.0	2

Table 2.5: Rogers Fault Diagnosis Table

$\frac{CH_4}{H_2}$	$\frac{C_2H_6}{CH_4}$	$\frac{C_2H_4}{C_2H_6}$	$\frac{C_2H_2}{C_2H_4}$	Diagnosis
0	0	0	0	Normal deterioration
5	0	0	0	Partial discharge
1/2	0	0	0	Slight overheating (< 150°C)
1/2	1	0	0	Overheating (150 - 200°C)
0	1	0	0	Overheating (200 - 300°C)
0	0	1	0	General conductor overheating
1	0	1	0	Winding circulating currents
1	0	2	0	Core and tank circulating currents, overheated joints
0	0	0	1	Flashover without power follow through
0	0	1/2	1/2	Arc with power follow through
0	0	2	2	Continuous sparking to floating potential
5	0	0	1/2	Partial discharge with tracking (note CO)

Table 2.6: Key Gas Diagnostic Method

Key Gas	Failure
Acetylene	Arcing in oil
Hydrogen	Partial discharge in oil
Ethylene	Overheated oil
Carbon Monoxide	Overheated cellulose

Table 2.4 and Table 2.5, but adds lower limits on the gas levels. These lower limits were lacking from the Rogers method, allowing the application of the method by naive users to gas levels that were not significant. This new version also takes the levels of carbon monoxide and carbon dioxide into consideration, solving a major shortcoming of the original Rogers Ratios method.

2.3.2 Key Gas

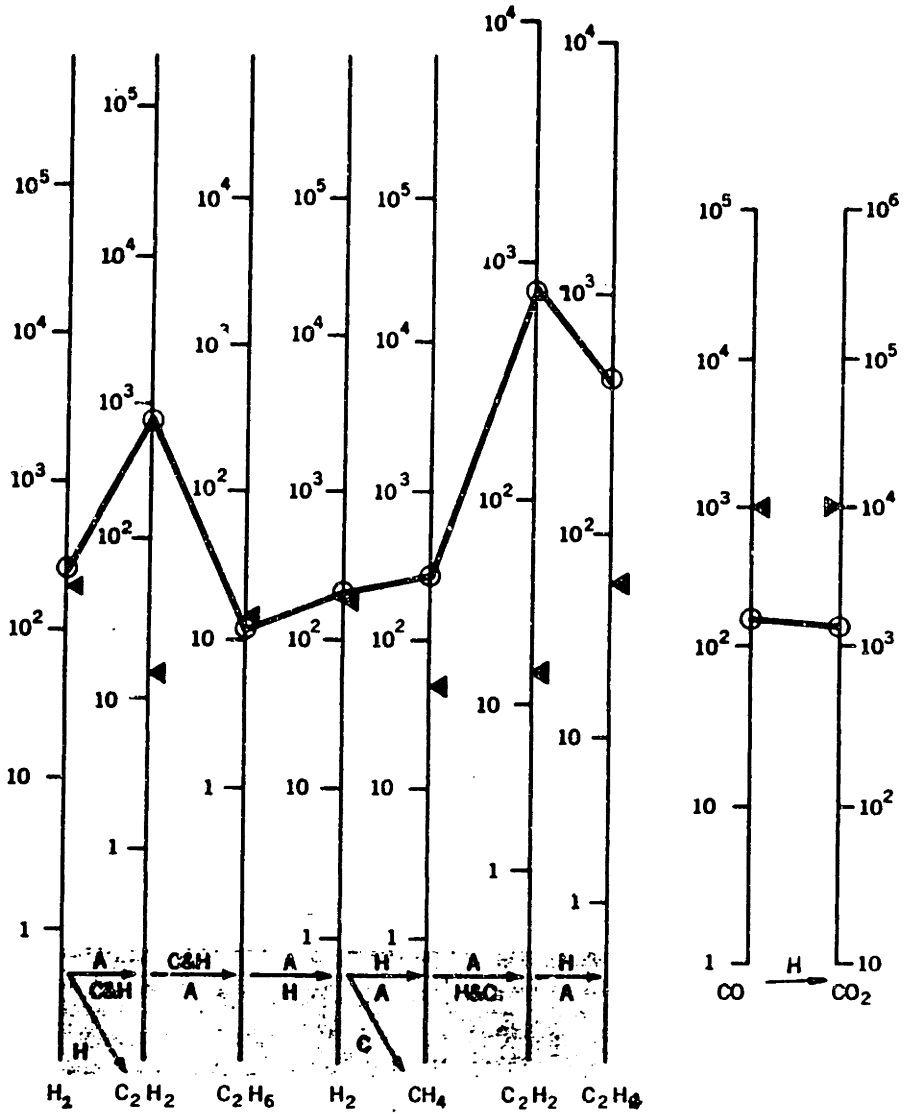
The Key Gas method involves looking for a single significant gas level that is associated with a particular type of failure. Table 2.6 lists the key gases and the associated failure[4, pp. 349–350]. This method is qualitative; the user is asked to decide on the significance of particular distributions without hard guidelines. This is a both a strength and a weakness: the Key Gas method is more flexible in the face of multiple failures than the Rogers Ratios method and small changes in concentrations do not drastically affect the diagnosis, but it is also more difficult to define the Key Gas method algorithmically.

2.3.3 Advanced Analysis

2.3.3.1 TOGA

The Transformer Oil Gas Analyst system (TOGA) is not intended to be used to generate comprehensive diagnoses, but rather to screen out transformers that seem to be in good condition[16]. In this way, the transformer expert does not need to review their conditions. TOGA is an expert system that checks for gas concentrations outside of threshold levels and looks at relative concentrations of these gases. Apparently this is mostly an implementation of one of the standard dissolved gas analysis codes (Rogers or IEC[2]). TOGA also analyzes reports of dielectric strength, power factors at ambient and elevated temperatures, acidity, and interfacial tension to evaluate the condition of the oil.

[2]The IEC Ratios method is a variation of the Rogers Ratios method which only makes use of three of the four ratios employed by the Rogers method. This method was developed for the International Electrotechnical Commission.



Fault causes (arcing A, corona C, heating H) are determined by measuring slope of connecting lines after gas extraction. Shaded area defines diagnostic for various slopes.

Figure 2.1: Graphical Presentation of Rogers Ratios Method

A database has been integrated into the system to give the transformer expert access to trending data. Work is being pursued to allow the expert system access to the database to enable reasoning about trends and additional factors, such as age and manufacturer of the transformer[17].

2.3.3.2 Exformer

Exformer is an expert system implementation of the IEC code for dissolved gas analysis[18]. An interesting aspect of Exformer is the application of fuzzy logic to the problem of dissolved gas analysis. This produces the desirable property that tiny changes in gas concentrations do not cause drastic changes in diagnosis.

2.4 Shortcomings of Traditional Techniques

The traditional techniques for the detection and diagnosis of transformer failures described above are widely used, familiar to the electric utilities, and have prevented numerous failures at an enormous cost savings. Unfortunately, they are also inadequate. A typical annual failure rate for 400–500kV transformers has been estimated as 2%[3]. Due to the central role of the power transformer in the electric power network, the capital costs and opportunity costs of a catastrophic failure provide a strong economic incentive for improving this failure rate[2]. In this section, some particular shortcomings of existing transformer monitoring techniques are discussed.

The use of oil samples for the on-line analysis of transformer performance allows the direct measurement of the state of the oil and indirect measurement of the state of the paper. Because of the harsh internal environment of large power transformers, more direct information as to the state of the cellulose is difficult and expensive to obtain. Most utilities perform an oil analysis annually; some progressive utilities do an analysis more frequently, semi-annually or even quarter-annually.

Compared to the time scales at which failures may occur, this sampling frequency is relatively low. Each analysis, however, represents a sizable investment of a technician's time, to sample and perform the analysis, and a transformer engineer's time, to interpret the results of the analysis. These resources limit the frequency at which the units can be tested. The frequency is low in that a failure may progress from undetectability to catastrophe between samples, but it is not so low that these tests are not a valuable and *vital* part of the maintenance routine. There is certainly sufficient evidence to support the claim that these tests can avert catastrophic failure. However, it must be remembered that much of the damage incurred by a transformer is permanent; tardy detection of the problem can only slow further damage or, in the worst case, allow the graceful shutdown of a transformer which must be completely rebuilt or discarded. These are worthy achievements when compared to the consequences of letting a failure proceed undetected, but

there is great deal of room for improvement. It is, therefore, extremely important to detect failures as early in their evolution as possible, to prevent the shortening of the transformer's life.

Oil samples are not the only vehicle for failure detection in transformers, of course. Many transformers are equipped with temperature sensors and combustible gas sensors (such as the Hydran 201R) which can be sampled almost continuously—at least when compared to a three-month sampling interval. These sensors can help avert catastrophic failure, but they are not sensitive to incipient failures. This is not an inherent property of the sensors, but rather a consequence of the way in which they are used: as raw inputs to set-point threshold detectors.

Threshold detectors compare each sensor reading to a pre-determined set-point; when this threshold is violated, a flag is raised. The result may be an operator warning, or even a trip of the transformer. Multiple thresholds may be implemented to tailor the response to the severity of the detected fault. But hard-and-fast thresholds are not very sensitive to small changes in the behavior of the transformer. Consider temperature monitoring: a temperature is chosen over which the transformer is considered as running hot. Any temperature below this threshold is considered normal. This ignores under what load the transformer is operating or what the ambient temperature is. Thus, if the transformer runs at the same top oil temperature while fully loaded in a hot environment and, later, while lightly loaded in a cool environment, no discrepancy will be noted, no alarm will be tripped. This is despite the fact that, in a normally operating transformer, the top oil temperature is almost entirely driven by the load and the ambient temperature.

The detection of trends is already part of traditional transformer monitoring, albeit on a very simple level. A gas concentration rise taking place over a number of years may be ignored, while the same rise occurring in a single year may be a cause for great concern. In this example, a trend (the rate of gassing) was used to distinguish between two conditions: normal aging and potential failure. Continuous monitoring opens the doors on a much larger statistical population from which to draw trending information. Because of the greater statistical significance available, smaller changes in the trends can be detected and used to increase the transformer engineer's understanding of the internal condition of the unit.

The discussions above may be summarized as three basic shortcomings in traditional methods of detection and diagnosis:[19,7]

- Dependence on infrequent oil samples.
- Reliance on hard thresholds for detection.
- Inadequate support for trend analysis.

2.5 Summary of Traditional Methods

In this chapter, the major failure processes of oil-filled transformers were presented. The primary roles of moisture, oxygen, and heat in the degradation of the insulating system were chronicled. The stresses which drive this deterioration and, ultimately, trigger catastrophic failure were also discussed.

The traditional methods of failure detection were discussed in the next section. The central role of oil sample analysis for detection was made obvious.

The importance of oil sample analysis was confirmed in the section concerning diagnosis of transformer failures. Diagnoses are at present mainly based on the results of the dissolved gas analysis.

Three basic shortcomings of this body of techniques were presented: infrequent sampling, static thresholds, and inadequate trend analysis. A novel approach to transformer monitoring, addressing these shortcomings and allowing the incorporation of superior detection, diagnosis, and prognosis capabilities, is presented in the following chapters.

Chapter 3

MIT Transformer Monitoring Approach

In Chapter 1, the prevention of catastrophic failures of large power transformers was shown to be an important goal. The achievements and shortcomings of existing methods for transformer performance monitoring were presented in Chapter 2. In this chapter, a new system for transformer performance monitoring that is more sensitive to incipient failures is described.

The MIT Transformer Monitoring Project was undertaken with the broad goal of establishing advanced technologies to significantly improve the reliable monitoring of large in-service power transformers, allowing for the detection of incipient failure conditions[20]. A major theme of this project was the premise that trend analysis would facilitate earlier detection of incipient failures. However, it was recognized that the identification of short- and long-term trends in the condition of a transformer first required an understanding of what the normal conditions of a transformer and its signatures are.

In the MIT approach, the present understanding of what is normal is captured in mathematical models—adaptive models, which can be tuned to the observed behavior of the transformer. The use of adaptive mathematical models and trend analysis are the fundamental differences between the MIT approach and previous attempts at on-line transformer performance monitoring.

The Pilot Transformer Test Facility was established as a research environment to enable the development of an integrated system of hardware and software. The result of this development effort is the Pilot Transformer Monitoring System. This system is described in great detail in Chapter 5. In contrast, the chapter presented here gives an overview of the concepts that form the basis of this new approach to transformer monitoring, the structures with which these concepts can be realized, and the processes by which these structures can result in the detection, diagnosis, and prognosis of incipient transformer failures. Section 3.1 explains the philosophy behind the use of adaptive models of normal behavior, the concepts and structures

Table 3.1: Relationships Between Failure Modes and Observable Quantities

Observable Quantities	Failure Modes and Failure Indicators							
	Bent Winding	Core Damage	Cracked Bushing	Electri-fication	Hot Spot	Arcing Short	Gas Bubbles	Contaminated Oil
Moisture-in-Oil			●		●			●
Gas-in-Oil			●	●	●	●	●	●
Partial Discharges			●	●	●	●	●	
Thermal	●	●		●	●			
Vibrations	●	●						
Oil Breakdown							●	●
Electrical					●			

that make an improved monitoring system possible. Section 3.2 describes how the results of these adaptive models can be used to form the basis of a system that, through trend analysis, can prevent catastrophic failures much earlier in their evolution, before serious damage has occurred.

3.1 Philosophy: Concept and Structure

Two issues had to be addressed before an on-line transformer monitoring system could be designed and implemented. These were:

1. Which *quantities* should be measured?

2. How should a *failure* be defined and detected?

The determination of the quantities to be measured started with a detailed literature review and discussions with utility representatives and transformer manufacturers. The results of these inquiries led to the development of a set of observable quantities whose behavior can be affected by one or more typical transformer failure modes. A summary of the relationships between failure modes, or indicators of failures, and observable quantities is given in Table 3.1.

Table 3.1 makes it obvious that there is a great advantage to monitoring multiple signatures of the transformer. Not only does this increase the chance of detecting an incipient failure but, with proper coordination, it may be possible to distinguish between various failure modes. This is what led to the establishment of the goal of developing an integrated monitoring system as differentiated from developing only a set of independent, new and/or improved sensors. Once the development of an integrated transformer monitoring system was defined as a goal, the problem of detecting and diagnosing failures could be addressed.

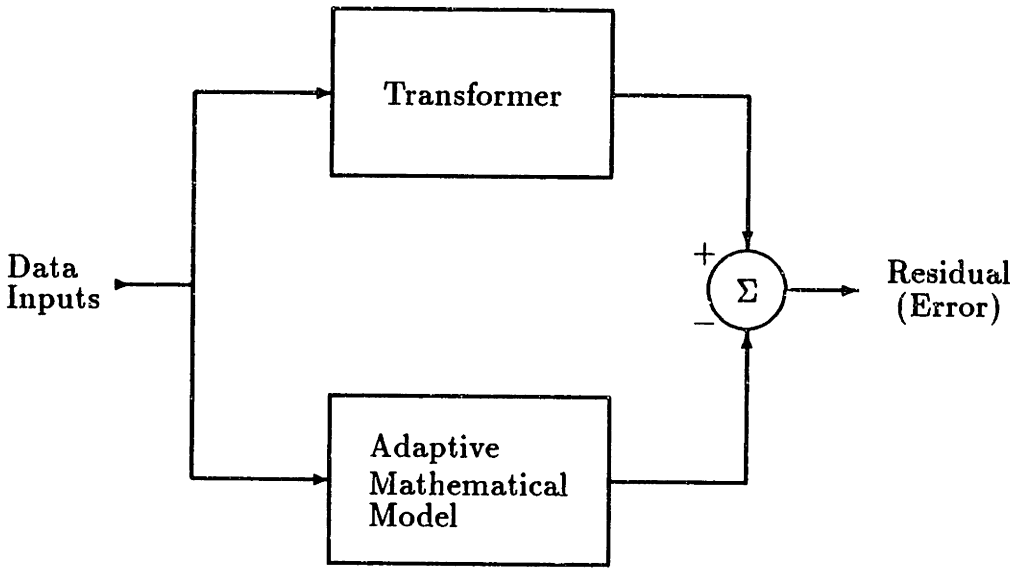


Figure 3.1: Monitoring System Fundamental Operation

3.1.1 Observation vs. Expectation

At the core of MIT's monitoring system is the concept of continuously comparing the actual behavior of the transformer to the expected behavior for the prevailing operating conditions. Figure 3.1 is a graphical representation of this central concept. The expected normal behavior of the transformer is embodied in mathematical models. These models attempt to account for external factors such as load and ambient temperature in predicting the output of a particular sensor. Each mathematical relationship models the behavior of a particular *measurement.* The output of the model is called a *prediction.* In the figure, this is represented by routing the data inputs through the mathematical models. These are the same inputs which determine the actual operation of the transformer; therefore, the data inputs are routed to the transformer, too. The output of the mathematical model is then compared to the actual behavior of the transformer, generating an error term, or a *residual.* If the transformer is behaving normally, then this residual should be small. If the transformer is failing *and the observable quantity is sensitive to the particular failure mode in effect*, then the residual should become significant. Changes in the observable behavior of the transformer can be used to infer its changing internal condition.

Normally, computation of the residual is a simple subtraction

$$residual = measurement - prediction$$

possibly normalized against the size of the measurement. This formulation has the advantage that a measurement that is higher than expected results in a positive residual, while a measurement that is lower than expected results in a negative residual. A residual which indicates close agreement between the measurement

and the prediction is nominally a *zero residual*, though noise and modeling errors make a residual which is identically zero for any great length of time a practical impossibility. Thus, a zero residual is simply any residual which is not considered significant.

3.1.2 Adaptive Models

One problem with the use of mathematical models is that every transformer is different: large power transformers are individually designed to meet a utility's requirements. To handle this difficulty, each model is abstracted to a parameterized model structure which can be tuned to fit the observed behavior of a particular transformer by a suitable selection of parameters. The model structure is chosen to be as widely applicable as possible, and is chosen so that the parameters can be estimated from the observed data. The estimated parameters then characterize a particular subsystem, or *signature* of a given transformer.

The behavior of each transformer is potentially different from other transformers in small, but quantifiable, ways. The parameterized model structures solve this problem by capturing these small differences as differences in the parameters of the model.[1] There is an additional difficulty, however, in that the behavior of an individual transformer may change over time. This occurs as a natural consequence of the normal aging processes that are always at work inside the device (see Chapter 2). As a result, an unchanging model could gradually drift into disagreement with reality even though nothing unexpected had occurred, since aging is not unexpected. *Adaptive models* can handle this problem by periodically updating their parameters to ensure the models continue to fit the observed behavior.

Above, it was noted that parameterized model structures can only be used if the parameters are identifiable from the observed data. Suitable model structures for adaptive modeling must have the additional property that some subset of the parameters must be identifiable from the observed *on-line* data. That is, some parameters must be ascertainable while the transformer is energized and in service. Other parameters can only be estimated or measured while the transformer is out of service, or they are based on the original design specifications of the transformer; such parameters are called *fixed parameters*. In the following discussions, the word "parameters" refers to the non-fixed, or on-line, parameters.

The use of adaptive models immediately begs the question of how one can identify discrepancies between the actual and expected behavior of the transformer when the expectations are constantly being updated to agree with the actual behavior. This question is answered in two ways. First, the parameters are updated infre-

[1] Some differences, such as whether the transformer is shell-type or core-type, or whether or not it is a base-load unit, may require distinct model structures. It may be possible to define "meta-model" structures that can relate the various model structures at a higher conceptual level. Investigations into these possibilities are beyond the scope of this thesis.

quently relative to the time scales of both data acquisition and quickly-developing failure modes. For a failure which manifests itself quickly, the parameters will usually be constant during the failure evolution and the anomaly will be evident in the residuals. (It may also be possible to suppress parameter updates when a possible failure has been detected, so that the residual response will always directly correspond to the detected anomalous behavior. See Chapter 4.) Second, and perhaps more importantly, the history of the changing parameters reflects the changing behavior of the transformer on a longer time-scale.

Above, two requirements for a suitable model structure were given: a wide applicability to various transformer designs, and parameters that could be estimated from the observed behavior. There is a third, as yet unmentioned, requirement—the model structure must not be overspecified, that is, there must be a single unambiguous set of parameters that fits the model structure to the observed data. This requirement insures that parameters are repeatable and stable during normal operation. If the parameters are stable during normal operation of the transformer, then the model can not adapt to changing transformer behavior without a perceptible shift in the parameter values. If the changing behavior does not result in a change in the parameter values (as, for instance, in the case where additional zero-mean noise is introduced to the modeled signal), then the change will be reflected in the behavior of the residual. One way or another, the changing behavior of the transformer will be captured by the system, either in the changing behavior of the parameters, the residual, or both.

3.1.3 The Module

The necessity of being able to *adapt* to a particular transformer is handled by estimating the parameters of the model using actual data from the transformer being monitored. Assuming that a given transformer is normal when new (having passed its initial acceptance tests), the parameters of a model may be estimated *on-line*. The error term, or measurement residual, then reflects the deviation of the transformer from its own normal state in the short term. If the parameters of a model are periodically re-estimated, long-term tracking of the *condition* of that particular signature may be accomplished. These concepts of adaptability and short- and long-term tracking are embodied in the block diagram of a *module* given in Figure 3.2.

A module[19] is implemented primarily in software. A list of definitions pertaining to Figure 3.2 follows.

1. Signals (data) from sensors pass to the *Signal Processor* where any necessary data preparation or reduction steps are performed.

2. Processed data then moves to the *Outlier Detector* where threshold checks for bad data are made; bad data is announced to the human operator and the

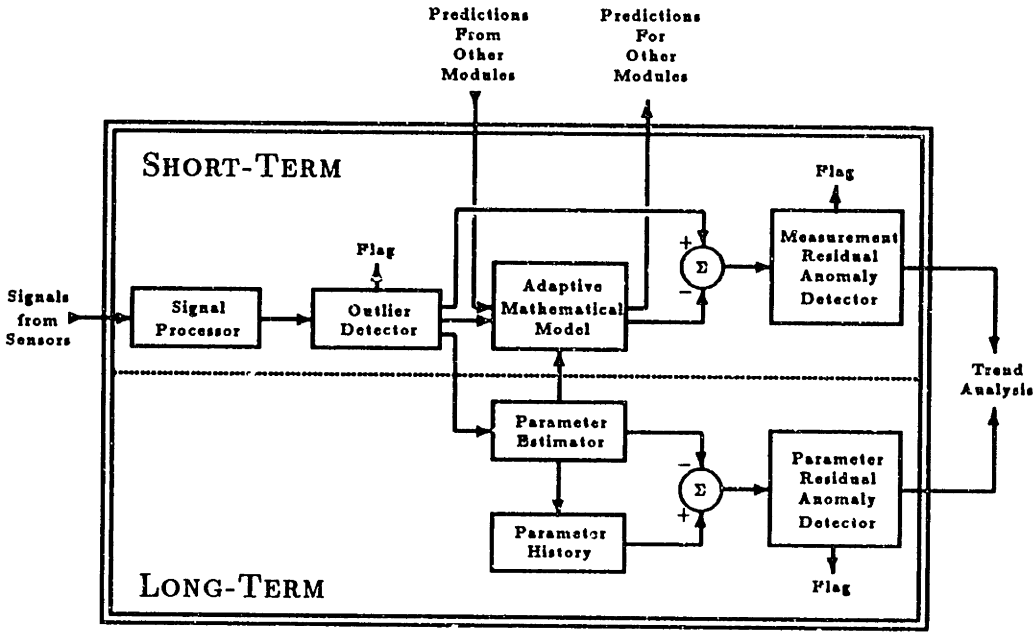


Figure 3.2: Module Block Diagram

detection/diagnosis system with a *Flag.*[2]

3. Validated data is used as the input to a *Model* which predicts the values (of the *Signature* in question) that are expected during normal operation of the device being monitored. Additionally, the model may accept predictions from other modules as inputs and may output predictions for other modules. These additional inputs and outputs are used for compensation purposes, e.g., temperature compensation.

4. Predicted values are compared to measured values in the *Measurement Residual Anomaly Detector.* This block looks for levels, rates-of-change, and patterns which are abnormal. If an abnormality is detected, the human operator and the detection/diagnosis system are alerted with a Flag.

5. Periodically, the parameters of the mathematical equation which makes up the Model are updated, using measured values, through the operation of the *Parameter Estimator* to assure that the Model remains accurate. When the Parameter Estimator operates, it automatically checks the new parameters for validity before installing them. (If the parameters are estimated using

[2]Alternately, both the Signal Processor and the Outlier Detector may be packaged as part of a "smart" sensor. This would result in a greater interchangeability of different types of sensors, improving the modularity of the system. However, sensor self-diagnosis has inherent limitations that guarantee that sensor failures will always be an important consideration during transformer failure diagnosis.

information-poor data, they will not accurately characterize the Signature.) Valid parameters are also passed to the *Parameter History* for use in anomaly detection.

6. The parameters of the Model are then tracked by the *Parameter Residual Anomaly Detector* to discriminate between acceptable changes, such as normal aging, and anomalies caused by incipient failures. As with the Measurement Residual Anomaly detector, this block checks for anomalous levels, rates-of-change, and patterns. When an anomaly is detected, the human operator and the detection/diagnosis system are alerted.

The horizontal dotted line in Figure 3.2 divides the module according to time scales: the top half of the module operates on the minutes-to-hours time scale, and the bottom half operates on the days-to-weeks time scale.

In the intervals between installations of updated parameters (newly estimated parameters that satisfy the parameter validity criteria), the condition of the signature and the accuracy of the model are checked via the measurement residuals. If the measurement residuals are small, the previously estimated parameters still accurately characterize the signature, and the condition of the signature is normal. If the measurement residuals exceed established limits (in level, rate-of change, or pattern), an anomaly is detected even if the measurement residuals return to normal when a new set of valid parameters are installed. In this case, there has been a change in the condition of the signature, but the structure of the model still correctly describes the signature. If the measurement residuals exceed established limits and newly estimated parameters are systematically failing the validity test, the condition of the signature has changed so much that the model structure is itself no longer valid. This is another (probably more serious), form of anomaly.

3.1.4 The Monitoring System

A module exhibits increased sensitivity to incipient failures which affect the condition of a particular signature. This is due to the adaptive model, continuous real-time operation, and the differential comparison technique. Sensitivity to incipient failures can be increased even further by cross-correlating the detection outputs of various modules. To do this, it is necessary to combine these modules in a system which can control and schedule data acquisition, information organization, module operation, detection, diagnosis, prognosis, communications and interfacing with the operator.

The block diagram for such a system is given in Figure 3.3. The system is implemented in a combination of hardware and software, performing the functions listed above while mediating scheduling and data conflicts. The activities of the system blocks include:

- Acquisition of raw data from *Sensors*

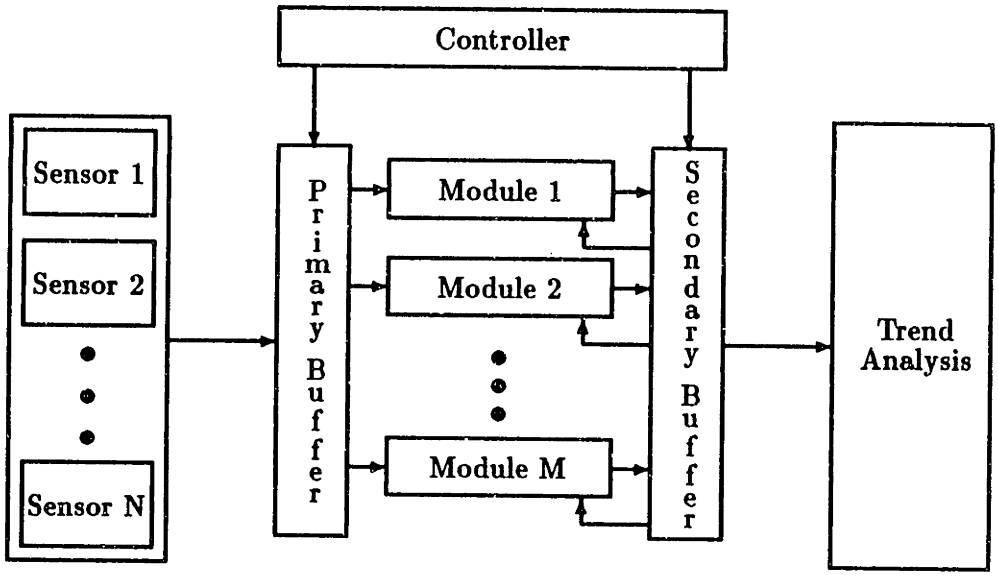


Figure 3.3: System Block Diagram

- Organization of raw data into a time-correlated format in the *Primary Buffer*, thus making the raw data available to the remainder of the system
- Processing raw data in *Modules* to extract information relevant to a determination of whether or not the transformer being monitored is operating normally
- Placement of relevant information from modules into the *Secondary Buffer*, for use by the rest of the system
- Performance of *Trend Analysis* on raw data and relevant information from modules to detect anomalies in the transformer being monitored, diagnose the condition of the transformer, and deliver a prognosis on the future operation of the transformer
- Organization and scheduling all of the above, and providing for an operator interface, through the operation of a *Controller*.

In summary, the MIT-developed monitoring structure is an integrated system with the module as its core. Conceptually, each of the functions of the system operate independently and in parallel, sharing information when required. This functionality permits the overall system to be highly flexible. Since information organization and scheduling of operations are handled by the system, resulting in a well-defined interface between modules and the system, modules may be added or removed easily. Trend analysis integrates the information streams from the individual modules to provide the knowledge upon which diagnostics can be based[19,21,2].

3.2 Trend Analysis: Detection, Diagnosis and Prognosis

Accurate in-service performance monitoring of transformers can realized with the achievement of three goals:

- *Detection* of anomalous (potentially hazardous) changes in the transformer's internal condition
- *Diagnosis* of the present internal condition of the transformer based on detection of anomalies
- *Prognosis* of the future condition of the transformer based on the diagnostic results

In the MIT monitoring system, each of these tasks is supported by the ability of the system to identify trends in the various residual and parameter streams, either individually or in relation to each other. However, while trend analysis is a *tool* employed during detection, diagnosis, and prognosis, "trend analysis" is, for historic reasons, a collective term for these three tasks[22,23].

The remainder of this section will explain how the tasks of detection, diagnosis and prognosis are accomplished in the framework of the MIT Transformer Monitoring System. This will be a general discussion of the issues and difficulties associated with each task. Chapter 5 will detail how this system might be integrated into a traditional program of transformer performance monitoring.

3.2.1 Detection

The first task of trend analysis is the detection of anomalous events that may indicate the presence of an incipient failure. Anomalies are revealed by tracking the behavior of the various residuals and parameters that are generated by the system. The types of anomalies that are watched for are threshold violations, abnormal rates of change, and unusual patterns of behavior.

A threshold violation occurs when a residual or parameter goes beyond the maximum or minimum limits that have been set for that particular data stream. Normally, it is assumed that a residual will hover around zero. However, some model structures lend themselves to the production of residuals with a non-zero mean. It is not required that the normal residual variation be centered at zero; a positive residual value and a negative one do not have to be equally likely. On the surface, it may seem that a residual with a non-zero mean need not be tolerated—the model could include a constant term to offset this mean. But it must be remembered that a suitable model structure can not be over-specified. A constant term may improve the accuracy of the prediction (and decrease the magnitude of the residual) while

rendering the parameters unstable. And stable parameters are a major requirement of a proposed model structure.

Stable parameters allow the monitoring of parameter values to determine when the condition of a signature has changed significantly. As with residuals, thresholds are set for parameter values and violations of these thresholds are considered interesting events. The limits are set so that the normal variation of parameters due to noisy or incomplete data will not generate spurious anomalies. Of course, before the estimated parameters are compared against the limits, statistical tests are run that can verify whether the data used to generate the parameters is an adequate representation of the condition of the signature (see Chapter 4).

Parameters will change over time—that is the reason adaptive modeling is necessary. Because of this, parameter thresholds will have to change over time. Experience suggests that parameters will change very slowly during normal operation, in relation to their natural variation due to noise. This slow change is punctuated by changes in the parameters due to events both internal and external to the transformer, events as varied as through-faults and maintenance shutdowns. Therefore, not much effort as been expended on how to automatically update parameter thresholds based on established and accepted trends in the parameters. However, it is expected that, for some signatures and modules, the magnitude of the parameter trends will necessitate automatic parameter threshold setting. It may also be necessary when there is a constant low-grade failure that is too small to be repaired economically. While not a part of normal operation, and certainly not desired, such a failure mode may be accepted as stable and non-threatening. Once accepted as a necessary evil, the effect of the failure mode on the parameters has to be recognized and discounted.

This leads directly to another class of detection events: rate-of-change anomalies. Once a particular rate, or range of rates, of change in a parameter has been established as normal, a change in this rate may indicate a change in the internal condition of the transformer. Thus, such an event should trigger further diagnosis that can search for more subtle indications of incipient failure.

Rate-of-change detection is not limited to parameters. Rate thresholds are also established for residuals. Residuals often exhibit an oscillation around zero, and so the rate of change also oscillates around zero. Depending on when, during this oscillation, a failure occurs, the failure may be reflected in the rate change before the actual residual threshold is violated. Regardless of any improvement in detection response time, the knowledge of whether a residual gradually drifted out of tolerance or rapidly diverged may be necessary to arrive at a quick initial diagnosis. Rate-of-change monitoring may also reveal when parameter updates are having a great effect on model behavior; this may act as a verification of direct parameter analysis.

All other anticipated detection events are loosely grouped under the classification of pattern detection. It is possible that both parameters and residuals may exhibit identifiable behavior that can not be verified through either threshold or

rate-of-change detection. One possibility is that parameters may show repeatable seasonal variation due to unmodeled factors such as average ambient temperature or humidity. In effect, the system would be applying the concept of models of normal behavior at a higher level, though the model would be qualitative, not quantitative. Another detectable pattern that may prove to be extremely important is the amount of noise present in a measurement or a residual. This could be very helpful in the detection of sensor failures.

The classification of an event as normal or anomalous is dependent on experience with the signature being modeled. As yet, it is impossible to set system thresholds from first principles; the system must be exercised to learn its limits. Automatic learning of these limits, analogous to the automatic learning of model parameters, would be a valuable addition to the system's abilities.

Experience may also lead to the ability to specify dynamic threshold and rate-of-change limits. In other words, the monitoring system may be able to recognize when its sensitivity should be increased, and when it should be decreased. One example: a model may yield very poor predictions immediately following a load change, due to unmodeled transients, yet give accurate predictions at all other times. It would be inappropriate to limit the strength of the model to its weakest link, but it is undesirable to generate spurious anomalies with each load change. The solution may be to have one threshold for load changes and another at all other times. Another example is that many signatures become very stable when the transformer is allowed to enter thermal steady-state. The system should be able to recognize this situation and adjust its sensitivities accordingly.

The major aim of the detection task is to identify anomalous events and bring them to the attention of the diagnostic process. However, every effort should be taken to suppress those events which are due to previously identified conditions or are numeric artifacts. This makes the job easier for the diagnostic process, the subject of the next section.

3.2.2 Diagnosis

Diagnosis is the act of identifying a failure from the observable effects it has on the behavior of the transformer. Diagnosis is necessary because the mere detection of an anomalous event can not form the basis for effective action. There must be an assessment of the type and magnitude of the problem, or even if any problem truly exists. Only with this sort of conclusion can a transformer operator make an informed decision as to whether to shut down the transformer, schedule maintenance, or ignore the situation. Without a diagnostic analysis, the detected events would form a flood of relatively useless information.

The initial impetus for diagnosis comes from the detected anomalies described in the previous section. However, diagnosis is not limited to these occurrences. The diagnostic process should have access to some amount of past data, for data

can take on a new significance when it is reviewed to support or deny a particular hypothesis. For instance, the moment that a residual threshold was violated may be far removed in time from the moment that the residual began the trend. But it may be that the correlation of the start of the trend to other events may yield critical information that would resolve a diagnostic ambiguity.

3.2.2.1 Data Correlation

There are several ways in which diagnostic analysis of module outputs can proceed. The simplest method is the *auto-correlation* of an individual module response. Auto-correlation merely means that the output of a single module is interpreted in isolation from the rest of system. A particular residual or parameter is classified according to its pattern of behavior. The data may evince a step response, exponential rise or decay, etc. There will be some situations in which such a classification will serve to indicate an unambiguous diagnosis.

Multiple auto-correlation is the next level of complexity in analysis. When auto-correlation of an individual module does not reveal an unambiguous diagnosis, several modules can be used to generate competing diagnoses. A voting process is then employed to combine the diagnoses into a single diagnosis or set of possible diagnoses that enjoy a higher degree of confidence than any individual module may be able to generate. This voting process is non-trivial, for a particular module may be unable to confirm or deny an hypothesis simply due to the nature of the relationship between the failure mode and the modeled signature. Thus, negative support and neutral support for a candidate hypothesis have to be distinguished. It would be desirable to allow a full spectrum of support for a hypothesis, from almost certainly true to almost certainly false. (If the hypotheses were certain, no voting would be necessary.)

Multiple auto-correlation allows for the interaction of different modules at a very high level of abstraction. This has a beneficial effect on the modularity and complexity of the monitoring system. Unfortunately, some information may be lost if the changing behavior of the various transformer signatures are only considered in relation to one another at this high level. *Cross-correlation* of the behavior of two or more parameters or residuals from different modules may reveal events that are not readily apparent when analyzing the data streams in isolation. These events (or lack of events) may support or refute competing diagnoses, and thus improve the diagnostic resolution of the monitoring system.

Consider the situation where there are two failure modes and two residuals. Assume both failure modes can be unambiguously *detected* by the first residual but, because both failure modes affect the residual in a similar manner, they can not be unambiguously *diagnosed* using that residual in isolation. If one failure mode demands an immediate (and costly) shutdown of the transformer while the other failure mode can be tolerated, this is an important distinction. Now assume that the second residual displays an observable event at random times during normal

operation, but that a similar event will always be correlated in time with the onset of one failure mode and not the other. No diagnosis (or, rather, only a diagnosis of healthy operation) could be reached analyzing the second residual in isolation. Only by cross-correlating the time-series behavior of the two residuals could both failure modes be detected and *diagnosed.* An example of such a situation will be more fully presented in Chapter 5.

3.2.2.2 Interpretation of Parameters

Models can generally be sorted into two classes: black-box models and physically-based models. Physically-based models are constructed from first principles, possibly using simplifying assumptions. The parameters of such models are well-defined composites of physical parameters of the transformer (including physical parameters of the sensors). Black-box models, on the other hand, are constructed through system identification techniques, with little or no consideration given to the physical principles at work. While the parameters of a black-box model are determined by the physical parameters of the transformer, there is no clear relationship between any physical parameter and a particular parameter of the model.

Even for a physically-based model, however, it may be difficult to interpret the behavior of an individual estimated parameter as a change in a particular physical parameter or set of parameters. This is because a failure mode need not be correlated with a change in any particular physical parameter upon which the model structure is based, but in fact could take the form of the introduction or intensification of an effect that was not previously modeled. Fortunately, it is not necessary to interpret an estimated parameter in isolation.

There are many tools for interpreting parameters. The tools described in the previous section (auto-correlation, multiple auto-correlation, and cross-correlation) do not depend on reasoning from first principles, but can make use of empirical relationships between the behavior of parameters and the presence of incipient failures. Experience is the key. Only with a great deal of experience observing the quantities that the monitoring system generates can come the knowledge to establish these empirical relationships.

This is not to say that diagnosis can not proceed until this store of experience has been fully developed. A great deal of experience already exists and can be taken advantage of. The quantities that are modeled by the monitoring system were chosen because experts recognized that their behavior can reflect the presence of incipient failures. For this same reason, experts have themselves been watching most of these quantities, although not as closely as the monitoring system is able to. Estimated parameters, taken in the context of the model structure from which they were generated, give an excellent handle for reasoning about the overall behavior of the modeled quantity.

3.2.2.3 Human Intervention

Most of the information that is necessary to reach a diagnosis is continuously available to the monitoring system. Most, but not all. At times, it will be necessary to acquire information that is not immediately available. The system must be able to recognize when such external information is warranted, and must be able to use the information effectively.

This adds a significant degree of complexity to the task of diagnosis. It has been assumed that all continuous data is generated at no incremental cost to the system. That is, any piece of data that would improve the accuracy or precision of the eventual diagnosis could, and would, be generated. Any information requiring human intervention, however, can only be requested after assessing the cost of the procedure and the utility of the possible responses. Such assessments are a normal part of many diagnostic programs; medical expert systems often have to weigh the dangers of invasive tests (such as exploratory surgery) versus the value of any possible information[24].

3.2.2.4 Interaction with Parameter Estimation

In addition to the information embodied in residuals and parameters, there is information to be gained from the parameter estimation *process*. It is not always possible to generate adequate parameters from the data that is available. The acceptance or rejection of generated parameters is an algorithmic procedure, tuned to the application at hand. It may be possible to uncover patterns of acceptance that can reveal an underlying problem. Also, a distinction must be made between parameters that are adequate for the generation of predictions and residuals and those that are adequate for the task of diagnosis. The issue of the interaction between diagnosis and parameter estimation (and parameter estimation in general) is discussed more fully in Chapter 4.

3.2.3 Prognosis

The prognosis of large power transformers is not a subject which is very well understood. However, it is widely accepted that a transformer is created with a life expectancy that may be conserved or wasted, depending on how the device is operated[4].

The rate at which this life expectancy is reduced is closely associated with the operational temperature of the transformer, integrated over time (see Chapter 2). With the continuous thermal monitoring available to the MIT monitoring system, it may be possible to track this integral and provide a quantitative indication of the rate at which the transformer is depreciating.

A novel way to perform this forecasting could use the parameters that are already so useful for detection and diagnosis. If slow trends can be identified at the beginning

of the transformer's useful life, these trends can be extrapolated to the end of the transformer's life expectancy. The results of the extrapolation could form a set of thresholds; the rate at which these thresholds are approached could possibly form the basis for a crude prognosis[25, p. 328]. As more experience is gained as to the trajectory of the parameters during the normal lifetime of a transformer, these thresholds could be improved.

The previous discussion concerns the prognosis of an otherwise healthy transformer. More valuable, perhaps, is a prognosis for a transformer which has been diagnosed with a specific failure. For the present, there are no plans to automate the process of prognosis. It may be sufficient to provide a transformer engineer with a diagnosis (or a set of weighted diagnoses), the rationale behind the diagnosis, and access to the raw data upon which the diagnosis was based. Prognosis would proceed with a more structured, usable set of information. And this is a step in the right direction.

3.3 Summary

This chapter has provided a general description of the concepts and issues that make the proposed monitoring system so interesting. The first section presented the philosophy behind the system. Primarily, this is the continuous reconciliation between the observed and expected behavior of the transformer. This continuous comparison resulted in dynamic thresholds that take into consideration such exogenous factors as ambient temperature and load history. Dynamic thresholds improve the sensitivity of detection by, for instance, recognizing the distinction between normal temperature distributions at half load, full load, and overload. A sensor reading that is well within the full range of sensor variation may not be reasonable for the prevailing operating conditions; a transformer performance monitoring system should be able to identify this discrepancy.

Continuous monitoring of multiple inputs enables another valuable tool for performance review: trend analysis. Expert transformer diagnosticians perform some crude trend analysis in interpreting the results of dissolved gas analyses. They often use the results of past analyses to place the current analysis in its proper context—they identify simple trends. Continuous monitoring generates an enormously larger statistical population from which to identify subtle trends that could not possibly be identified via infrequent oil samples. There is also an advantage in that gross trends can be determined with greater accuracy and confidence. These trends, both subtle and gross, have the potential of revealing a great deal about the changing health of the transformer.

Both dynamic thresholds and trend analysis are implemented through the vehicle of adaptive models of normal behavior. Adaptive models are used to highlight long- and short-term changes in transformer response; changes evolving over days or months and changes arising in minutes can be illuminated through the these

models. In the first section, the concept and structure underlying the monitoring system were explained. How the information can, in general, be applied to the three problems of detection of anomalous behavior, diagnosis of incipient failure, and prognosis of transformer condition was discussed in the second section. Presented here were only the general concepts which a successful system should implement. The actual implementation of a monitoring system for large power transformers is presented in more detail in Chapter 5.

Chapter 4

Implementation of the MIT Approach

The basic underpinnings of the MIT approach have been implemented and applied to the Pilot Transformer Test Facility at MIT's High Voltage Research Laboratory[20]. Suitable models have been developed and verified for the test transformer. The Pilot Monitoring System has successfully generated predictions and residuals based on parameters that are automatically estimated from on-line data. The modules currently operational are a vibration module[26,27,28], a thermal module based on the IEEE loading guide[29,30], a thermal module based on the MIT constrained flow model[29], and a combustible gas module. A moisture module exists, but a continuous sensor of moisture content in oil has not been available. However, the module has been verified with off-line measurements.

This chapter gives a detailed presentation of the implementation of the MIT approach to transformer monitoring as it is applied to the Pilot Facility. The first section sketches the system of software that makes use of the data acquisition capabilities of the Pilot Facility to realize the monitoring system. Each of the implemented modules receives special attention; the embedded models that are used to generate predictions and residuals as well as the relations from which the estimated parameters are derived (which are sometimes subtly different from the corresponding models) are presented.

The second section focuses on one portion of the existing monitoring system prototype: parameter estimation. Two versions of the parameter estimation mechanism are presented. In Section 4.2.1, the initial version of parameter estimation, called the *type A implementation*, is discussed in detail, with examples from the modules which use this implementation. Section 4.2.2 discusses a subsequent version, the *type B implementation*, that uses a different approach to gathering data for the estimation process, and discusses why the first version was inadequate for one of the modules. Finally, Section 4.2.3 discusses some suggestions for improvements that may bear investigation.

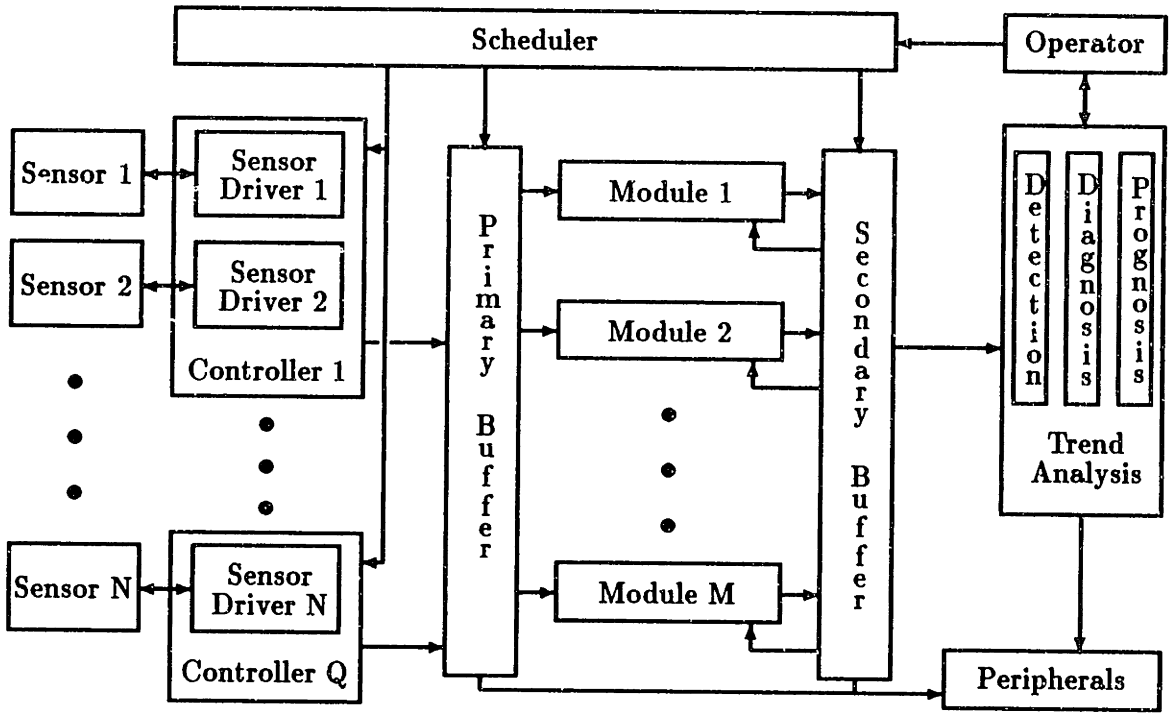


Figure 4.1: Pilot Monitoring System Block Diagram

4.1 Pilot Transformer Monitoring System

This section describes the implementation of a Pilot Transformer Monitoring System using the structure and concepts discussed in the previous chapters. The Pilot Monitoring System developed by MIT is installed in the Pilot Transformer Test Facility in MIT's Building N10. It is a combination of computer hardware and software designed to fulfill the dual functions of data acquisition for model and module development and implementation of an on-line transformer monitoring system. The discussion will first introduce the Pilot Transformer Test Facility, then present a more detailed system block diagram, and finally will proceed into a description of the actual hardware and software.

4.1.1 Pilot Monitoring System Structure

An implementation of the monitoring system discussed in Chapter 3 involves more detail than presented in the structural diagram of Figure 3.3. This added detail is depicted in the block diagram of Figure 4.1. The blocks in this system diagram are chosen to represent functional pieces of the Pilot Monitoring System; as such, some of the blocks represent hardware, some represent software, and some represent combinations of hardware and software.

Figure 4.1 sees the addition of controllers and sensor drivers to the system block

diagram of Figure 3.3. This is a more accurate depiction of the interface between the sensors and the rest of the monitoring system. The scheduler interacts with the controllers on the time scale of minutes, triggering data acquisition. The controllers, in turn, govern the operation of the sensor drivers on the time scale of microseconds. The controllers are dedicated devices with sufficient resources to guarantee precise and timely data acquisition. The controllers then communicate the results to the primary buffer under the control of the scheduler.

The interaction between the transformer operator and the monitoring system is also presented in Figure 4.1. First, the operator can modify the schedule that is maintained by the scheduler. Judging from the present body of experience, it seems likely that this would be a very infrequent procedure. Other infrequent operator procedures may include sensor adjustments, modification of fixed parameters, etc. (not represented in Figure 4.1). The operator will interact much more frequently with the trend analysis processes (detection, diagnosis, and prognosis). This interaction will take the form of operator review of generated information and operator notification of incipient and catastrophic failures.

The last detail added to Figure 4.1 is the peripherals block. Peripherals are devices that, while not being essential to the operation of the monitoring system, make it more reliable, more convenient, and generally more useful. These include archive storage devices (floppy disks and nine-track tapes), plotting devices, and external communications.

4.1.2 Master Machine Software

The specifications for the monitoring system call for a coordinating element to synchronize the activities of the individual modules. The operation of this coordinating element is required to be independent of the particular actions a module performs and, in fact, independent of the number of modules being coordinated. The specification also calls for the establishment of a mechanism for passing data between various modules, while limiting the constraints on the number and types of modules running. This mechanism will perform the duties of the primary and secondary buffers in the system block diagram.

Together these two requirements necessitate a standardized interface for the modules. It was decided that a module would only be required to perform a given set of actions at a predefined interval. The module would then respond to some trigger from the coordinating element by performing this set of actions, secure in the assumption that the module is synchronized with the system.

For flexibility, each module may also have its own initialization and/or termination code. The initialization code is triggered simply by starting the module. If the initialization fails, the normal trigger is taken as an initialization trigger until it succeeds. There is a separate termination trigger that causes termination code to be executed. The termination code will be executed after the normal set of module

actions until it succeeds, at which time the module exits.

Inter-module communication of data is handled through the file system of the host computer. A limited buffer is provided for efficient retrieval of recent data.

4.1.3 Dispatch Software

The coordinating element consists of a single process that coordinates an arbitrary number of individually compiled programs. The resulting process is alternately referred to as DISPATCH, the scheduler or the synchronization process.

The programs which are coordinated by the synchronization process are referred to as modules. These modules are implemented specifically to fit into this scheme. (The structure of a module is discussed in Section 3.1.3.) Each module is a separately compiled program. Because of this, the set of presently executing modules can be modified with ease, and the addition of new modules has little or no impact on existing modules. The set of modules which is to be run is established through the use of an input file, also referred to as the jobs file. The modules run continuously in the background and are triggered to execute various portions of their code by the synchronization process. DISPATCH can determine the execution status of each module and, if a module is not ready to be triggered at the appropriate time, a count of missed intervals would be incremented. When the module is ready to be triggered, it may perform some processing based on this value. In this way, each module is kept synchronized with the entire system.

4.1.4 Module Software

From a software point of view, a module consists of four parts: an initialization routine, a normal iteration routine, a synchronization error recovery routine and a termination routine. Though a module is a separately-executable program, it must be run by a synchronization program to operate correctly. A set of module utilities have been provided to interface the module with the dispatch process.

MIT chose to develop modules for the following signatures:

- Thermal (IEEE Loading Guide Model)
- Thermal (Constrained Flow Model)
- Winding Vibration (Black-Box Model)
- Dissolved Gas In Oil (Thermal Based Model)
- Dissolved Moisture In Oil (Thermal Based Model)
- Partial Discharges (Electrically Based Model)

Unfortunately, not enough progress was made on the development of an electrically-based sensing-scheme for partial discharge detection to warrant development of a module; therefore, partial discharges will not receive further consideration in this chapter.

4.1.4.1 Module and System Summary

The dispatch process and the module interface have proven to be a flexible mechanism for implementing the various modules. The dispatch process is independent of the functions of the modules under its control. As such, bringing a new or updated module on line is simply a matter of editing an input file to reflect the new set of modules (and their schedules) and re-invoking the dispatch process. Communication between the dispatch process and an individual module follows the same lines regardless of the particular module being driven, modified only by the schedule provided in the input file.

Using the module interface reduces the problem of implementing a new module to implementing just those routines that distinguish one module from another. In effect, one just implements the mathematical model at the heart of the module. All problems of scheduling and communication have been abstracted away.

Each individual module is designed to capture the function of some subsystem of the transformer. THIE3MOD and THMOD handle the thermal system, VIBMOD deals with the windings, and GASMOD and WTHMOD handle the oil and insulation systems. In describing the function of a transformer subsystem, each module embodies a mathematical model of how that system works. The mathematical model may be intended to describe a physical model, such as THMOD's constrained flow model, or may describe an observed functional relationship, such as in the moisture module (WTHMOD). In either case, the mathematical model contains parameters that adapt to observed conditions, to *tune* the module to the actual behavior of the transformer. The design of the module system is intended to simplify the process of inserting a particular model into the system and allow for the maintenance of the adaptive parameters.

4.1.4.2 Thie3mod

One purpose of this module is to detect changes in the thermal system of the transformer, particularly excess heating. A second purpose is to predict unmeasurable temperatures to be used in compensating the models in other modules (e.g., the dissolved gas module). A third, as-yet-unrealized purpose is to enhance loadability by running the model faster than real time to allow the operator to foresee the consequences of operational decisions (e.g., overloading during peak periods).

This module is based on the IEEE/ANSI Loading Guide Models for prediction of top oil temperature and hot spot temperature using ambient temperature and load current as inputs[30]. The standard models have been modified to allow the

top oil model to adapt to the transformer on-line[29]; the hot spot model is not adaptive, relying on parameters measured during initial heat runs.

The model is given in Equation (4.1)[1].

$$\hat{T}_{topoil}[k] = \begin{cases} T_{topoil}[1] & \text{if } k = 1 \\ \theta_1 \times (\hat{T}_{topoil}[k-1] - T_{ambient}[k-1]) + & \\ \quad \theta_2 \times I_{low}[k]^{1.6} + & \\ \quad T_{ambient}[k] & \text{if } k > 1 \end{cases} \tag{4.1}$$

The measured ambient temperature ($T_{ambient}$) and load current (I_{low}) are used to predict the top oil temperature (T_{topoil}) in this dynamic model. In the Pilot Facility, data acquisition and temperature prediction occur at two minute intervals. Embedded in the parameter θ_1 of Equation (4.1) is a time constant that accounts for this sampling interval; DISPATCH ensures that this sampling rate is maintained.

The reader will note that the model is recursive: the predicted temperature, $\hat{T}_{topoil}[k]$, is based on the previous prediction, $\hat{T}_{topoil}[k-1]$. Because of this, the equation contains an initialization clause that (optimistically) sets the first prediction identically equal to the first measured value. The transients introduced by this simplistic initialization are quickly damped out in accordance with the thermal time constant of the transformer.

Every two minutes, as each prediction of the top oil temperature is generated, it is compared to the corresponding measured value. The top oil temperature residual is computed by subtracting off the prediction from the measured value:

$$\tilde{T}_{topoil}[k] = T_{topoil}[k] - \hat{T}_{topoil}[k] \tag{4.2}$$

Note that the residual[2] becomes more positive if the measured value rises with respect to the expected value, and more negative if the measurement falls below expectations. Thus, $\tilde{T}_{topoil}[k]$ will be positive if the top oil is "too hot", and negative if the oil is "too cold".

The residual is compared to level and rate-of-change thresholds. That is, if the residual is too high (or low) or it has changed too fast, an incipient failure has possibly been detected. Once detection has occurred, the responsibility for determining if a failure has, in fact, developed and for categorizing the failure falls to the diagnostic process. It should be stressed that the thresholds mentioned above are *residual* thresholds, not measurement thresholds. Residual thresholds grant the monitoring system a dynamic sensitivity that measurement thresholds can not. However, for completeness, measurement thresholds are supported by the system.

Figure 4.2 shows how the various thresholds may be set. This file is continually interpreted by the monitoring system; the operator of the pilot facility adjusts

[1] In [29], temperatures and temperature rises are indicated by T and θ. In this document, θ is reserved to indicate parameters and all thermal quantities are indicated by T.

[2] The notation $\tilde{x}$ indicates a residual based on the measured and predicted values of x.

```
Entries are of the form

        NAME: RANGE_A ; RANGE_B ; ... (MAX_CHANGE)

A range is of the form

        MINIMUM, MAXIMUM
or
        MAX_ABSOLUTE_VALUE

The maximum change is either an absolute change, represented
by just a number, or a percentage change, represented by a
number followed by '%'

Modification History:

5-19-89 rgtoil:         20.0 -> 20.0; (2.0)
        brgtoil:        20.0 -> 50.0

$

gtoil: 15.0, 135.0
gambient: 10.0, 60.0
ilow: 0.0, 250.0
rgtoil: 20.0; (2.0)
brgtoil: 50.0
```

Figure 4.2: `thie3mod.chk`

the thresholds by modifying the file with a text editor. In the file, everything up to the first '$' is ignored by the system. The operator can make note of any useful information at the top of the file. Shown in Figure 4.2 is a reminder of the format of threshold entries and a modification history of the included thresholds. In a commercial monitoring system, a user-friendly interface would insure that only well-formed entries would be accepted and would automatically generate an audit trail of threshold modifications. The mechanism shown is entirely adequate in a research environment.

Each threshold entry begins with a string that indicates to which quantity the thresholds are to be applied. In Figure 4.2, "gtoil" is one such string. The monitoring system uses this string as an index to search this file. The string "gtoil" denotes the top oil temperature (*generic top oil*), "gambient" is the ambient temperature (*generic ambient*), and "ilow" is the load current (*low-voltage-side current*). Thus, the associated thresholds are measurement thresholds: the top oil temperature can range from 15.0 to 135.0°C without generating any detection events, ambient can range from 10.0 to 60.0°C, and the load can range from 0.0 to 250.0 amps.

The string "rgtoil" indicates the top oil temperature residual. The single value of 20.0 is the threshold for the absolute value of the residual; the residual may vary from −20.0 to 20.0°C. The value in parentheses, 2.0, sets the threshold for the change in the residual from one iteration to the next. If the residual changes by

more than 2°C in either direction in a single step, a flag will be raised. Eventually, a flag will act as a trigger to the diagnostic process, but at present the event is only logged in a file that can be reviewed later by researchers.

The string "brgtoil" refers to what is known as a *baseline residual.* A baseline model is one for which the parameters are not automatically updated as for the adaptive models. This is used as a research tool to highlight the effects of changing parameters, and would not be included in a commercial system.

An operating power transformer is subject to large temperature gradients. Since many processes that occur inside a transformer are dependent on the temperature of their local environments, it is necessary to model unattainable measurements based on available measurements. Use of a suitable thermal model can allow other models to be thermally compensated, increasing their accuracy and, therefore, their usefulness. The ANSI/IEEE loading guide contains a model of the internal winding temperature, a temperature that plays a key role in the processes of the winding. This "hot spot" temperature model is based on the measured top oil temperature and the load current. It is a static model that is used to predict the internal winding temperature for each sampling interval. This predicted temperature is used for thermal compensation of the vibration module (VIBMOD), the combustible gas module (GASMOD), and the moisture module (WTHMOD). The model is given in Equation (4.3):

$$\hat{T}_{wtint}[k] = C_1 \times I_{low}[k]^{1.6} + T_{topoil}[k] \tag{4.3}$$

The parameter C_1 is fixed: it is based on values given in the loading guide.

The top oil temperature predictor parameters are estimated daily using load current, ambient temperature and top oil temperature measurements, from Equation (4.4):

$$\begin{aligned} T_{topoil}[k] - T_{ambient}[k] \;=\; & \theta_1 \times (T_{topoil}[k-1] - T_{ambient}[k-1]) + \\ & \theta_2 \times I_{low}[k]^{1.6} \end{aligned} \tag{4.4}$$

Equation (4.4) shows that the predicted top oil temperature *rise over ambient* is based on the previous prediction of rise and the present load. As discussed in Chapter 4, θ_1 and θ_2 in Equation (4.4) are estimated using a two-day-wide window of data. The (over-specified) set of equations is then solved using the principle of least squares.

In Figure 4.3, there are two sets of thresholds. One set is concerned with the parameter estimation validity measures described in Chapter 4, the other with the parameters themselves. The strings "vk1" and "vk2" label the validity measure thresholds for θ_1 and θ_2, respectively. If either of these thresholds are violated, *both* parameters are rejected to be conservative. If the set of parameters is accepted, each parameter is compared to its thresholds. A flag is raised if any threshold is violated (for the Pilot Facility, an entry is made in the system log).

The validity measure thresholds and the parameter thresholds are determined empirically. Originally, the validity measure thresholds were set at a very high

```
$

vk1: 1.2e-7
vk2: 2.5e-7

k1: 0.9950, 0.9963
k2: 4.3e-5, 5.2e-5
```

Figure 4.3: `thie3prm.chk`

level in order to accept nearly all candidate parameters. The thresholds were then lowered to screen out parameters that yielded inaccurate predictors. These parameters were, in general, those that were estimated from information-poor data. The thresholds that were settled on accept parameters that, with few exceptions, are stable, generate accurate predictions, and are estimated from information-rich data. Once a reasonable number of sets of appropriate parameters had been accepted, the parameter thresholds were set to bracket the ranges of these parameters.

4.1.4.3 Thmod

This module shares all of the goals of THIE3MOD: detection of change in the thermal system of the transformer, the prediction of unmeasurable temperatures to be used to compensate the models in other modules, and to enhance loadability. However, this module uses more accurate models than the IEEE module; physically-based equations have been developed to predict temperatures in and near regions of constrained oil flow, such as cooling ducts in windings, and at locations in the winding bulk[29]. More dynamics are included than in the IEEE models. Three ducts have been instrumented in the test transformer: one specifically constructed for the purposes of experimentation called the artificial duct, and two normal ducts in the high voltage section of the winding, arbitrarily designated the thermocouple-side duct and the accelerometer-side duct. The disadvantage of this module is that it requires oil temperature measurements to be made in regions near the winding, although not actually inside the winding. The models which predict oil temperatures are adaptive, the models which predict winding surface and internal temperatures are partially adaptive.

For THMOD, the measured duct bottom (inlet) oil temperature ($T_{botduct}$) and load current are used to predict duct top (outlet) oil temperature ($T_{topduct}$). Like the embedded model of THIE3MOD, this is a dynamic model which is is triggered every two minutes. The model is given in Equation (4.5).

$$\hat{T}_{topduct}[k] = \begin{cases} T_{topduct}[1] & \text{if } k = 1 \\ \alpha \times T_{botduct}[k] + & \\ \quad \beta \times (\hat{T}_{topduct}[k-1] + T_{botduct}[k-1]) + & \\ \quad \gamma \times I_{low}[k]^2 & \text{if } k > 1 \end{cases} \tag{4.5}$$

```
$

wboil: 15.0, 135.0
wtoil: 15.0, 135.0
wtaccoil: 15.0, 135.0
wbaccoil: 15.0, 135.0
wtthoil: 15.0, 135.0
wbthoil: 15.0, 135.0
ilow: 0.0, 250.0
rwtoil: 6.0; 10.0; (2.0)
brwtoil: -15.0, 15.00
rwtaccoil: 8.0; 12.0; (2.0)
brwtaccoil: -15.0, 15.00
rwtthoil: 8.0; 12.0; (2.0)
brwtthoil: -17.0, 17.00
```

Figure 4.4: `thmod.chk`

The parameters α, β, and γ are adaptive.

Since there are three instrumented ducts, there are three individual sets of parameters used to generate three separate predictions. In each case, the inputs $T_{botduct}$ and $T_{topduct}$ refer to the inlet and outlet temperatures of the specific duct. In almost all respects, the three models are completely distinct and can be thought of as three separate modules. The exception is that all of the parameters are accepted or rejected as a set during parameter estimation.

The duct top oil temperature residual is computed each interval. There are, of course, three residuals—one for each prediction.

$$\tilde{T}_{topduct}[k] = T_{topduct}[k] - \hat{T}_{topduct}[k] \tag{4.6}$$

Figure 4.4 gives the measurement and residual threshold for THMOD. The string "wboil" and "wtoil" label the inlet and outlet oil temperature, respectively, for the artificial duct. The embedded strings "th" and "acc" denote the thermocouple-side duct and the accelerometer-side duct. As with THIE3MOD, a baseline residual is computed for each adaptive model in THMOD.

Also like THIE3MOD, THMOD predicts the internal winding temperature at each interval. The constrained flow model of the "hot spot" temperature is more complicated than the model based on the ANSI/IEEE loading guide, and more accurate. Actually, THMOD's winding interior temperature model is a group of inter-related models, presented in Equations (4.7)–(4.12). The measured duct top and bottom oil temperatures are used to predict the oil temperature at a location within the duct (Equation (4.7)). The height within the duct is chosen to be level with the spot where experience says the winding gets the hottest. This is a static model with a fixed parameter (C_2) indicating the desired height.

$$\hat{T}_{wmtoil}[k] = C_2 \times T_{topduct}[k] + (1 - C_2) \times T_{botduct}[k] \tag{4.7}$$

This predicted duct internal oil temperature ($\hat{T}_{wmtoil}$) and the load current are used to predict the winding surface temperature adjacent to the location in the duct for which the oil temperature was predicted. This is a dynamic model with fixed parameters.

The predicted winding surface temperature ($\hat{T}_{surf}$) and load current are used to predict the internal winding temperature. This dynamic model is partially adaptive; the parameters estimated for the artificial duct's top duct oil temperature predictor (α, β, and γ) are used to improve the internal winding temperature predictor's response.

The internal winding temperature prediction of this module could potentially be used in any of the other modules that need to be compensated for the winding temperature. However, at this time THIE3MOD's prediction is used. This is because the measurements that THMOD uses as inputs require sensors that are generally considered intrusive by utilities. THIE3MOD has the advantage that the sensors it needs are, for many transformers, either already present or readily retrofitted. Experimentation and experience are necessary before it can be seen whether the improved prediction accuracy of the constrained flow model is worth the cost of the more intrusive sensors.

$$\hat{T}_{surf}[k] = C_3 \times \hat{\tilde{T}}_{surf}[k-1] + C_4 \times (I_{low}[k]^2 + C_5 \times \hat{T}_{wmtoil}[k]) \quad (4.8)$$

$$\begin{aligned} \hat{T}'_{topduct}[k] = {} & \alpha \times T_{botduct}[k] + \\ & \beta \times (T_{topduct}[k-1] + T_{botduct}[k-1]) + \\ & \gamma \times C_5 \times (\hat{T}_{surf}[k] - \hat{T}_{wmtoil}[k]) \end{aligned} \quad (4.9)$$

$$\tilde{T}_{surf}[k] = C_6 \times \tilde{T}_{surf}[k-1] + (1 - C_6) \times (T_{topduct}[k] - \hat{T}'_{topduct}[k]) \quad (4.10)$$

$$\hat{\tilde{T}}_{surf}[k] = \hat{T}_{surf}[k] + C_7 \times \tilde{T}_{surf}[k] \quad (4.11)$$

$$\hat{T}'_{wtint}[k] = C_8 \times \hat{T}'_{wtint}[k-1] + C_9 \times \hat{T}_{surf}[k] + C_{10} \times I_{low}[k]^2 \quad (4.12)$$

The parameters for the constrained flow thermal module are estimated daily. The equation from which the parameters are estimated is given in Equation (4.13). Duct top oil temperature predictor parameters are estimated using load current and measured duct top oil temperatures (every 24 hours).

$$\begin{aligned} (T_{topduct}[k] + T_{botduct}[k]) - (T_{topduct}[k-1] + T_{botduct}[k-1]) = {} & \\ \theta_1 \times (T_{topduct}[k] - T_{botduct}[k]) + & \\ \theta_2 \times I_{low}[k]^2 & \end{aligned} \quad (4.13)$$

The reader will note that the parameters that are estimated do not map directly to the parameters that are used in the predictor (Equation (4.5)). The form of the equation given above is a direct interpretation of the physical basis of the model:

```
$

vz1: 8.4e-5
vz2: 9.0e-9
vzth1: 6.0e-5
vzth2: 2.6e-8
vzacc1: 8.7e-5
vzacc2: 3.8e-8

z1: -0.1, -0.07
z2: 7.5e-6, 1.2e-6
zth1: -0.032, -0.023
zth2: 1.24e-5, 1.56e-5
zacc1: -0.032, -0.023
zacc2: 1.1e-5, 1.5e-5
```

Figure 4.5: `thprm.chk`

the change in the average temperature of the duct is linearly related to the duct temperature rise and the winding losses ($I_{low}{}^2$). Isolating the predicted quantity ($T_{topduct}[k]$) on one side of the equation yields:

$$\begin{aligned} \alpha &\triangleq \frac{\theta_1 + 1}{\theta_1 - 1} \\ \beta &\triangleq \frac{1}{1 - \theta_1} \\ \gamma &\triangleq \frac{\theta_2}{1 - \theta_1} \end{aligned}$$

Figure 4.5 displays the validity measure thresholds and parameter thresholds for both parameters of each of the three ducts' model.

4.1.4.4 Gasmod

The purpose of this module is to detect anomalous changes in the dissolved gas content of the oil. The model is partially black-box, partially physically-based, and is intended for use with the Syprotec H-201R Hydran Dissolved Gas Monitor. The Hydran is sensitive to hydrogen, carbon monoxide, acetylene, and ethylene. The module actually runs two models, both predicting the dissolved gas reading of the Hydran. One model uses the measured top oil temperature as its input, the other model uses the predicted internal winding temperature as its input. The models are static and adaptive.

The measured top oil temperature and predicted internal winding temperature are used to make two separate predictions of the Hydran dissolved gas reading. The models are static. In the Pilot Facility the predictions are generated every ten minutes. The embedded model is:

$$\widehat{hydran}[k] = \alpha + \beta \times T_{topoil}[k] + \gamma \times T_{topoil}[k]^2 \qquad (4.14)$$

```
#

gas: 0.0, 500.0; 0.0, 1000.0; 0.0, 2000.0
gtoil: 15.0, 135.0
pwtint: 15.0, 135.0
rgasgtoil: 18.0; 50.0; 100.0; (9.0)
rgaspwtint: 18.0; 50.0; 100.0; (9.0)
brgasgtoil: 100.0
brgaspwtint: 100.0
```

Figure 4.6: **gasmod.chk**

The structure is the same for the predicted internal winding temperature; T_{topoil} is replaced by $\hat{T}_{wtint}$.

A residual is computed for each prediction (i.e., every ten minutes). A separate residual is maintained for both models. Figure 4.6 shows the measurement and residual thresholds that are used for the Pilot Facility. Note that the residual thresholds have three levels of reporting; one threshold is crossed at ± 18 ppm, the next at ± 50 ppm, and the most serious threshold at ± 100 ppm. At present, the Pilot Facility merely notes which threshold has been violated. Eventually, a different action could be specified for each. For instance, at the lowest threshold the violation could merely be logged, while the second threshold could sound an alarm, and the third could trip the transformer. The residual checking also specifies a rate-of-change threshold: ± 9 ppm/interval.

$$\widetilde{hydran}[k] = hydran[k] - \widehat{hydran}[k] \tag{4.15}$$

Model parameters are estimated daily using the measured top oil temperatures and Hydran readings for one model and the predicted internal winding temperature and Hydran readings for the other. The form of the equation used for the estimation is given in Equation (4.16). While the embedded model, Equation (4.14), would seem to indicate that any third-order polynomial would be appropriate, the estimation equation limits the model to a parabola with a minimum at 0°K. The model is limited in this way in order to generate stable parameters; a general third-order polynomial is greatly over-specified.

$$\sqrt{hydran[k]} = \theta_1 + \theta_2 \times T_{topoil}[k] \tag{4.16}$$

$$\begin{aligned} \alpha &\triangleq {\theta_1}^2 \\ \beta &\triangleq 2 \times \theta_1 \times \theta_2 \\ \gamma &\triangleq {\theta_2}^2 \end{aligned}$$

```
#

vxgtoil: 0.02
vygtoil: 2.5e-5
xgtoil: -16, -6
ygtoil: 0.04, 0.07

vxpwtint: 0.02
vypwtint: 2.5e-5
xpwtint: -12, -5
ypwtint: 0.04, 0.06
```

Figure 4.7: **gasprm.chk**

Again, the estimation validity measure thresholds and the parameter thresholds can be set by the operator. Figure 4.7 shows the thresholds currently in force for the Pilot Facility.

4.1.4.5 Vibmod

The purpose of this module is to detect potentially dangerous changes in the physical structure of the winding (e.g., loose wedges) caused by events such as through-faults.

This module uses as its inputs: a core vibration time series signal acquired from an accelerometer mounted on the core, a winding current time series signal taken from a current transformer on the low voltage side which is squared in software, the RMS terminal voltage, and the predicted winding internal temperature. The module performs a Fourier transform on the time series core vibration and load current squared data. The complex Fourier coefficients for the first three harmonics of these signals are input to a black-box model. Based on these inputs the model predicts the Fourier coefficients of the first three harmonics of the winding vibration. The model contains no dynamics but is completely adaptive. The predicted winding vibration Fourier coefficients are compared to measured winding vibration Fourier coefficients (calculated using a time series signal acquired from an accelerometer mounted on the winding) and a measurement residual is computed[28,27,31,26]. Predictions and residuals are generated every ten minutes.

$$\hat{v}_{winding}[k] = (\alpha + \beta \times \hat{T}_{wtint}[k]) \times u[k] + \gamma \times v_{core}[k] \tag{4.17}$$

In Equation (4.17), the winding vibration harmonics ($v_{winding}$), the core vibration harmonics (v_{core}), and the load current harmonics (u) are complex vectors with length 3, and α, β, and γ are complex 3 x 3 matrices. The Pilot Facility monitoring system is only concerned with the first three harmonics (fundamental, second, and third).

The prediction of the winding vibration harmonics, then, can actually be broken down into three models, predicting each of the first three harmonics. When ana-

lyzing the residual response of this module, the prediction is, in fact, dealt with as three separate predictions and three separate residuals. Equation (4.18) indicates how each of the residuals is computed (the index i can be one of *fundamental*, *second*, or *third*). Because the winding vibration can go through huge changes in magnitude in response to changing operating conditions, it was felt that behavior of the residual could be more readily understood if it were normalized against the magnitude of the measured harmonic.

$$\tilde{v}_{winding,i}[k] = \frac{\|v_{winding,i}[k]\| - \|\hat{v}_{winding,i}[k]\|}{\|v_{winding,i}[k]\|} \tag{4.18}$$

Because of difficulties with triggering the data acquisition, the phase information inherent in the vibration measurements was suspect. The effect of this can be seen in the equation above: the winding vibration residuals only track the errors in the amplitude of the prediction. The prediction could be 180° out of phase without affecting the size of the residual. To take this phase information into consideration, an alternate equation for computing the residual could be:

$$\tilde{v}_{winding,i}[k] = \frac{\|v_{winding,i}[k] - \hat{v}_{winding,i}[k]\|}{\|v_{winding,i}[k]\|} \tag{4.19}$$

The equations used to generate parameters for VIBMOD are given below. The vibration module is different from the modules discussed above in that the parameters are not estimated from a window of data spanning two days. Data is collected for an arbitrary amount of time until a set of data that is sufficiently rich in information is achieved. At that time, the parameters are estimated and automatically accepted. This discussion is presented in great detail in Chapter 4.

$$\begin{aligned} v_{winding,fundamental}[k] \;=\; & \theta_1 \times u_{fundamental}[k] + \\ & \theta_2 \times u_{second}[k] + \\ & \theta_3 \times u_{third}[k] + \\ & \theta_4 \times \hat{T}_{wtint}[k] \times u_{fundamental}[k] + \\ & \theta_5 \times \hat{T}_{wtint}[k] \times u_{second}[k] + \\ & \theta_6 \times \hat{T}_{wtint}[k] \times u_{third}[k] + \\ & \theta_7 \times v_{core,fundamental}[k] + \\ & \theta_8 \times v_{core,second}[k] + \\ & \theta_9 \times v_{core,third}[k] \end{aligned} \tag{4.20}$$

$$\begin{aligned} v_{winding,second}[k] \;=\; & \theta_{10} \times u_{fundamental}[k] + \\ & \theta_{11} \times u_{second}[k] + \end{aligned}$$

$$
\begin{aligned}
& \theta_{12} \times u_{third}[k] + \\
& \theta_{13} \times \hat{T}_{wtint}[k] \times u_{fundamental}[k] + \\
& \theta_{14} \times \hat{T}_{wtint}[k] \times u_{second}[k] + \\
& \theta_{15} \times \hat{T}_{wtint}[k] \times u_{third}[k] + \\
& \theta_{16} \times v_{core,fundamental}[k] + \\
& \theta_{17} \times v_{core,second}[k] + \\
& \theta_{18} \times v_{core,third}[k]
\end{aligned} \tag{4.21}
$$

$$
\begin{aligned}
v_{winding,third}[k] \;=\; & \theta_{19} \times u_{fundamental}[k] + \\
& \theta_{20} \times u_{second}[k] + \\
& \theta_{21} \times u_{third}[k] + \\
& \theta_{22} \times \hat{T}_{wtint}[k] \times u_{fundamental}[k] + \\
& \theta_{23} \times \hat{T}_{wtint}[k] \times u_{second}[k] + \\
& \theta_{24} \times \hat{T}_{wtint}[k] \times u_{third}[k] + \\
& \theta_{25} \times v_{core,fundamental}[k] + \\
& \theta_{26} \times v_{core,second}[k] + \\
& \theta_{27} \times v_{core,third}[k]
\end{aligned} \tag{4.22}
$$

$$
\alpha \triangleq \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_{10} & \theta_{11} & \theta_{12} \\ \theta_{19} & \theta_{20} & \theta_{21} \end{bmatrix}
$$

$$
\beta \triangleq \begin{bmatrix} \theta_4 & \theta_5 & \theta_6 \\ \theta_{13} & \theta_{14} & \theta_{15} \\ \theta_{22} & \theta_{23} & \theta_{24} \end{bmatrix}
$$

$$
\gamma \triangleq \begin{bmatrix} \theta_7 & \theta_8 & \theta_9 \\ \theta_{16} & \theta_{17} & \theta_{18} \\ \theta_{25} & \theta_{26} & \theta_{27} \end{bmatrix}
$$

It was found that a single set of parameters did not yield accurate predictions throughout the operating space of the transformer. In particular, it was found that the parameters that worked well for full load and above did not work as well for a lightly loaded transformer. The operating space of transformer was partitioned so that one set of parameters was concerned with loads below 175 amps, and one set of parameters dealt with loads of 175 amps and above. This is not two separate models, though the parameters are estimated and stored separately. Instead it is one model that utilizes two sets of parameters depending on the load on the transformer. Perhaps this is an unnecessary complication and two completely separate models should be maintained. Only further experience can answer this question.

4.1.4.6 Wthmod

The purpose of this module is to detect anomalous changes in the dissolved moisture content of the oil. Such changes (usually an increase) indicate deterioration of the paper insulation due excessive heating and/or acid attack.

This module computes an approximation of oil moisture content based on a temperature reading. Like the combustible gas module, two models are maintained, one based on top oil temperature, and one based on the predicted internal winding temperature[32]. Presently, no residual is calculated on-line, due to the lack of availability of a solid state moisture sensor. Moisture readings are therefore made by hand, as is the measurement residual calculation. The models are static and adaptive, though without on-line moisture measurements parameters can not be estimated automatically.

In the Pilot Facility, the moisture predictions are generated every ten minutes. The model is shown in Equation (4.23). (For the other model, replace T_{topoil} with $\hat{T}_{wtint}$.)

$$\widehat{moisture}[k] = \alpha \times e^{\frac{T_{topoil}[k]}{\beta}} \tag{4.23}$$

The predicted dissolved moisture-in-oil concentrations are compared weekly by hand with actual moisture measurements (measured using the Karl-Fischer method) to compute dissolved moisture measurement residuals.

$$\widetilde{moisture}[k] = moisture[k] - \widehat{moisture}[k] \tag{4.24}$$

Model parameters are estimated infrequently using the following equation:

$$\ln(moisture[k]) = \theta_1 + \theta_2 \times T_{topoil}[k] \tag{4.25}$$

$$\begin{aligned} \alpha &\triangleq e^{\theta_1} \\ \beta &\triangleq \frac{1}{\theta_2} \end{aligned}$$

4.2 Parameter Estimation

There are three distinct activities involved with parameter estimation: data selection, candidate parameter generation, and candidate parameter validation. *Data selection* consists of choosing the set of data from which to generate parameters. *Candidate parameter generation* entails algorithmically processing the chosen data to produce parameters. The final step, *candidate parameter validation*, is needed to verify that the parameters generated for the given model structure do indeed accurately describe the selected data. Two mechanisms for implementing parameter estimation are presented below.

4.2.1 Type A Implementation

The type A implementation has proven to be quite useful for estimating parameters in situations in which the input space of the module is normally adequately sampled in a moderate time interval. When this is true, a set of data rich in information can be provided to the estimation process. Accurate parameters can then be produced without critically stressing the storage and computing resources of the Pilot Transformer Monitoring System.

4.2.1.1 Data Selection

The initial approach to data selection was to take all of the most recent data over a span that could be determined by the system administrator. In other words, if the administrator were to decide to use a single day's worth of data to perform parameter estimation, then the most recent 24 hours of data would be used. All data from this period would be used, and each piece of data in this period would receive equal weight.

At present, the operator-specified time span can be set to multiples of one day. For the Pilot Facility, one day's worth of data is 720 sets of thermal and electrical data (taken every two minutes) and 144 sets of vibration and dissolved gas data (taken every ten minutes). The schedule of parameter estimation currently being followed for the IEEE thermal module (THIE3MOD), the MIT constrained flow thermal module (THMOD), and the combustible gas module (GASMOD) is that the estimation is done daily, using two days' worth of data. This schedule was arrived at as a compromise between accuracy and responsiveness. All three modules are dependent on thermal variation for accurate parameter estimation. The load profile that is considered "normal" for the Pilot Facility generates a reasonably wide range of thermal variation each day (Figure 4.8 shows three days of this discrete daily load cycling).

While one day of load cycling data is sufficient to generate good parameters, it was deemed unwise to sample for parameter estimation at the same rate at which the load was being cycled. Thus, two days was chosen as an appropriate interval: responsive to change, yet not subject to sampling problems. Daily estimation was selected for two reasons. First, the overlapping data sets prevent the oversight of trends that may occur at the edge of an interval. Second, the schedule accommodates itself to a daily routine of human review of parameters, during the initial development stage.

The reader will note that the choice of schedule is, to some extent, arbitrary. The only requirements are that sufficiently varied data be collected to identify the parameters (i.e., that an adequate portion of the operating space has been visited so that the resulting parameters are representative of the device over the expected operating range), that old data (corresponding to old parameters) not be used so much that changes in parameters are masked, and that parameter estimation be

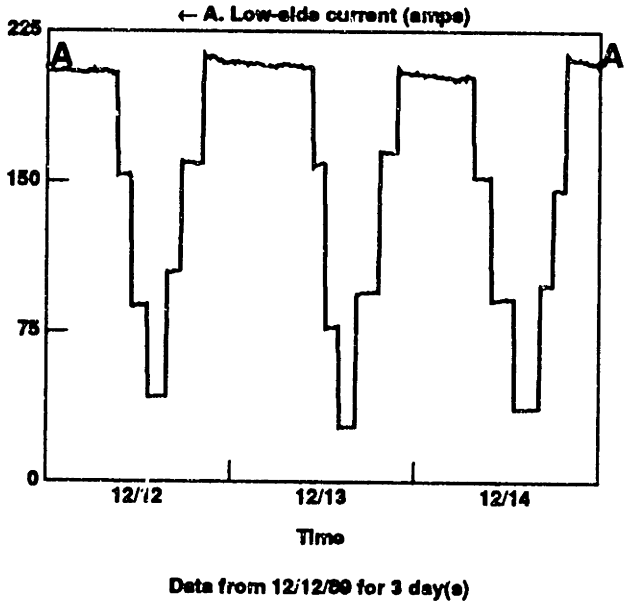


Figure 4.8: Discrete Daily Load Cycle

performed often enough that parameter trends can be detected before failures become serious. The schedule described above meets these requirements for the Pilot Facility. Two days' worth of data usually contains enough information to generate useful parameters. Only the most recent data is used. And, lacking contradictory experience, it is believed that many of the incipient failures that are of concern can be seen to evolve on the time scale of days. Future experience or a different load profile may favor the choice of a new schedule of parameter estimation.

The reader should also note that the choice of estimation schedule is intrinsically related to the model being implemented. Different models (even of the same signature) may require different types or amounts of information in order to identify the parameters. An example may be two models, one with dynamics, one without. It is easy to see that without sufficient dynamic information the transient response can not be estimated, while the same data may be sufficient for the estimation of the steady-state response.

4.2.1.2 Candidate Parameter Generation

To generate parameters, the selected data is passed as input to a multivariate least-squares algorithm[33]. As an example, consider the embedded model of the IEEE

thermal module (THIE3MOD), original presented in Equation (4.5):

$$\hat{T}_{topoil}[k] = \begin{cases} T_{topoil}[1] & \text{if } k = 1 \\ \theta_1 \times (\hat{T}_{topoil}[k-1] - T_{ambient}[k-1]) + \\ \quad \theta_2 \times I_{low}[k]^{1.6} + \\ \quad T_{ambient}[k] & \text{if } k > 1 \end{cases} \tag{4.26}$$

where $\hat{T}_{topoil}$ is the predicted mixed top oil temperature, T_{topoil} is the mixed top oil temperature, $T_{ambient}$ is the ambient temperature, I_{low} is the load current, and θ_1 and θ_2 are estimated parameters. The equation used to estimate θ_1 and θ_2, however, is:

$$\begin{aligned} T_{topoil}[k] - T_{ambient}[k] &= \theta_1 \times (T_{topoil}[k-1] - T_{ambient}[k-1]) + \\ &\quad \theta_2 \times I_{low}[k]^{1.6} \end{aligned} \tag{4.27}$$

The reader will notice that Equation (4.27) consists of a linear combination of terms which are functions of measured values equated to a term which is also a function of measured values. The general form of this equation is:

$$y[k] = \theta_1 \phi_1[k] + \theta_2 \phi_2[k] + \cdots + \theta_j \phi_j[k] \tag{4.28}$$

$y[k]$ is called an *observation*, each $\phi_i[k]$ is called a *regressor*, and each θ_i is an unknown *parameter*. y and each ϕ_i must consist of measurable quantities.

For Equation (4.27):[3]

$$\begin{aligned} y[k] &\triangleq T_{topoil}[k] - T_{ambient}[k] \\ \phi_1[k] &\triangleq T_{topoil}[k-1] - T_{ambient}[k-1] \\ \phi_2[k] &\triangleq I_{low}[k]^{1.6} \end{aligned}$$

Equation (4.28) can be written even more simply as:

$$y[k] = \phi^T[k]\theta \tag{4.29}$$

where:

$$\begin{aligned} \phi^T[k] &\triangleq [\ \phi_1[k]\ \phi_2[k]\ \cdots\ \phi_j[k]\ ] \\ \theta &\triangleq [\ \theta_1\ \theta_2\ \cdots\ \theta_j\ ]^T \end{aligned}$$

The manipulations necessary to transform Equation (4.26) into Equation (4.27) are straightforward: $T_{ambient}[k]$ is subtracted from both sides, and T_{topoil} is used in place of $\hat{T}_{topoil}$. The first operation arises from the fact that the temperature

[3]The notation $\triangleq$ is used to mean "is defined as"

differential between the ambient temperature and the mixed top oil temperature is the actual quantity being modeled. Equation (4.26) is simply a form of the relationship that is suitable for residual generation, because it predicts a measurable quantity (T_{topoil})—parameters can only be predicted for equations based solely on measurable quantities. In addition, T_{topoil} is an appropriate quantity to predict because it is familiar; this temperature is one that has traditionally been monitored to govern the operation of the transformer. The substitution of the measured value T_{topoil} for the corresponding predicted value $\hat{T}_{topoil}$ is the mechanism by which the model is made to conform to the actual behavior of the transformer being monitored.

The parameters are then estimated by finding the least-squares solution (θ_1 and θ_2) to the matrix equation:

$$\begin{bmatrix} T_{topoil}[2] - T_{ambient}[2] \\ T_{topoil}[3] - T_{ambient}[3] \\ \vdots \\ T_{topoil}[k] - T_{ambient}[k] \end{bmatrix} = \begin{bmatrix} T_{topoil}[1] - T_{ambient}[1] & I_{low}[2]^{1.6} \\ T_{topoil}[2] - T_{ambient}[2] & I_{low}[3]^{1.6} \\ \vdots & \vdots \\ T_{topoil}[k-1] - T_{ambient}[k-1] & I_{low}[k]^{1.6} \end{bmatrix} \times \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{4.30}$$

In general form, the equation is:

$$Y[k] = \Phi[k]\theta \tag{4.31}$$

where:

$$\begin{aligned} Y[k] &\triangleq [\ y[m]\ y[m+1]\ \cdots\ y[k]\ ]^T \\ \Phi[k] &\triangleq [\ \phi^T[m]\ \phi^T[m+1]\ \cdots\ \phi^T[k]\ ]^T \end{aligned}$$

and m is the number of states in the underlying model. For example, a model with no dynamics would have $m = 1$, while the thermal model given in Equation (4.27) would have $m = 2$ (the two states k and $k-1$). Y is called the *observation vector*, Φ is called the *regressor matrix*, and θ is called the *parameter vector*.

Candidate parameters are generated by solving Equation (4.31) for θ, given Y and Φ. Since the system of equations represented by Equation (4.31) is over-specified, θ is estimated through the application of the principle of least squares; that is, θ is chosen to minimize the criterion $(Y[k] - \Phi[k]\theta)^T(Y[k] - \Phi[k]\theta)$. Appendix B goes into detail on how this estimation is accomplished. For a particular model, there may be more than one possible equation to use for estimation, and there are certainly other estimation methods (e.g., maximum likelihood[34]). But in general, an appropriate estimation equation is suggested by the model structure itself, and least-squares has proven to be an accurate, computationally tractable

algorithm[33,35,36] in this case. Least-squares is appropriate for zero-mean uncorrelated white noise and is suitable for the model structures that have been chosen; however, it would be inappropriate for some model structures. For a type A implementation, k in Equation (4.30) and the interval of estimation are determined by the system operator. k is the number of samples taken in the specified time span and the most recent time span is used.

4.2.1.3 Candidate Parameter Validation

After candidate parameters have been generated, the procedure must check whether the parameters accurately fit the input data (whether the underlying model structure has changed) and whether the data has enough variation to enable good parameters to be estimated. The latter check must be done to prevent the system from estimating the parameters with too little data. Otherwise, the situation would be analogous to estimating a line from a single point; while a line can be chosen that fits that one point perfectly, the resulting line is completely useless for modeling the output at any other point. Thus, the generated parameters could model the input data perfectly and yet be completely meaningless.

A single measure that would reflect both accuracy of parameters and variation of input was desired. The measure chosen for the type A implementation is as follows:

$$\begin{aligned} \mathrm{E} &= Y - \Phi\hat{\theta} \\ valid &= \frac{\mathrm{E}^T\mathrm{E}}{k+1-m} diag\left(\left(\Phi^T\Phi\right)^{-1}\right) \end{aligned} \tag{4.32}$$

As in Equation (4.31), Φ is the regressor matrix, and Y is the observation vector. $\hat{\theta}$ is the vector of candidate parameters ($\hat{x}$ indicates a prediction or estimation of a quantity x). E is then the vector of differences between the actual and predicted values of the observation vector, using the candidate parameters; it is called the *error vector*. $k+1-m$ is the number of samples used to estimate the parameters (k is the number of samples taken and m is the number of states of the model in question, thus the first $m-1$ samples out of k cannot be used while the model initializes); that is, $k+1-m$ is the length of E, Y or Φ. Thus the term $\mathrm{E}^T\mathrm{E}/(k+1-m)$ reduces to the average squared error, a value which was minimized by the estimation process. This term, then, is a scalar that reflects how accurately the candidate parameters model the data given.

Diag() is merely a function which returns the main diagonal of a matrix as a vector. The matrix passed to this function, $(\Phi^T\Phi)^{-1}$, is called the *covariance* matrix. The elements on the main diagonal of this matrix are the variances of the corresponding element of the parameter vector (i.e., $(\Phi^T\Phi)^{-1}_{ii}$ is the variance of θ_i). These variances have the property that they tend to increase greatly when there is only a small amount of variation in the input, and they stay small when

there is wide variation. So, the error term discussed above measures how well the candidate parameters fit the past data, while the variances measure the likelihood that the parameters will be appropriate for a random set of future data. When multiplied together, these terms form composite values that tend to blow up when the parameters are inaccurate or the input is information-poor. Therefore, a simple threshold for each of these elements may be set to determine acceptable values[33].

In practice, it was found that each element of *valid* could not be interpreted in isolation from the corresponding element of θ. Reviewing the definition of *valid*, it becomes apparent why. If a standard deviation (the square root of the variance) is much larger than the corresponding parameter, regardless of the absolute size of the standard deviation, then this represents a great deal of uncertainty as to the most significant digits of the parameter. For this reason, a slightly different validity criterion was introduced:

$$valid_i' = valid_i/\theta_i \tag{4.33}$$

That is, each element of the validity vector was scaled against the corresponding candidate parameter. This essentially boils down to using the relative variance, rather than the absolute variance. Perhaps this was ill-considered, as the variance is not linearly related to the magnitude of the parameter. Another possible validity measure is:

$$valid_i'' = \sqrt{valid_i}/\theta_i \tag{4.34}$$

making use of the relative standard deviation. However, Equation (4.33) has proven sufficient to the task of screening out inappropriate parameters.

Determination of effective thresholds proceeded by trial and error. Thresholds were initially set at very high levels to accept all parameters. These thresholds were then lowered to screen out inaccurate parameters or parameters based on information-poor data.

4.2.2 Type B Implementation

The type A implementation of parameter estimation was installed in the Pilot Facility and exercised fully. Apart from some small difficulties in determining appropriate parameter validation thresholds, the system required no operator intervention for effective operation. Parameter estimators using this scheme were established for THIE3MOD, THMOD, and GASMOD (called, respectively, THIE3PRM, THPRM, and GASPRM). Unfortunately, this scheme could not work for the vibration module, VIBMOD.

A type A implementation is only applicable to a particular parameter estimation task when the memory capacity of the monitoring system can maintain an information-rich set of data (under a normal load profile) without selective data storage, beyond a first-in-first-out data buffer. The type B implementation was developed to address the problem of insufficient memory capacity, spurred on by the requirements of the vibration module.

4.2.2.1 Need for Extensions

The basic difficulty with estimating vibration parameters for the Pilot Facility was lack of resources— there was not enough memory available to store the amount of data necessary to estimate parameters. This was because the operating point of the transformer, with respect to the inputs to the vibration module, varied so slowly. Data must be taken from a significant percentage of the operating envelope or the generated parameters will unduly reflect the operating point of the transformer at the time of estimation.

Increasing the amount of memory available to store vibration data was not the solution. For one reason, the amount of data in question would make the task of parameter candidate generation computationally unmanageable. But more importantly, there can be no guarantee that, for any given time interval, enough of the operating envelope will be explored during such an interval to estimate parameters that are valid for the entire operating envelope. The solution, it was decided, was to select input to the estimation process more intelligently, picking and retaining data until a more complete picture could be gathered.

4.2.2.2 Data Selection (Bins)

Intelligent data selection was implemented by subdividing the operating envelope into discrete regions. Each region was then assigned a *bin*, a structure for retaining the most recent data from the corresponding region. When the system determines that enough of the operating envelope is represented, the data is used to generate parameters.

Axes In order to subdivide the operating envelope, one must first identify its dimensions, or *axes*. In general, the dimensions of the envelope are the inputs to the module. In practice, it is desirable to choose quantities that are not closely correlated. For example, consider the vibration module. The embedded model of the module is:

$$\hat{v}_w = cu + \alpha T_w u + a v_c \tag{4.35}$$

where u is a complex vector of load current harmonics, T_w is the transformer winding internal temperature (corresponding to the winding hot-spot temperature), v_c is a complex vector of core vibration harmonics, $\hat{v}_w$ is a prediction of the complex vector of winding vibration harmonics, and c, α, and a are matrices of complex parameters.

A brief examination would suggest that the operating envelope of the transformer is best described (for this module) by having a dimension for each of the current harmonics, for each of the core vibration harmonics, and for the hot-spot temperature. With more insight, it becomes clear that the load current harmonics are heavily dependent on the magnitude of the load current, and the core vibration harmonics are mainly determined by the load voltage. Therefore, the operating en-

velope can be described adequately with respect to the vibration module in terms of load current, load voltage and hot-spot temperature.

Using these quantities has many advantages. One advantage is that a smaller number of dimensions is more manageable. Another is that the three quantities chosen— load current, load voltage, and hot-spot temperature— are more nearly orthogonal than the individual harmonics. They are not completely orthogonal since the hot-spot temperature is determined in part by the load current and load voltage, but the hot-spot temperature does have an orthogonal component: ambient temperature. Also, while a high temperature generally corresponds to a high load current, the load current can change much more rapidly than the hot-spot temperature. This means that some portions of the operating envelope will be explored during the transient response to a load change. Perhaps this is not appropriate, since the vibration model is a static one; future experience will reveal if data gathered during large transients should be screened out from the parameter estimation process.

Perhaps the most important advantage to the dimensions chosen is that the quantities are ones with which transformer experts are very familiar. The load ratings and the maximum hot-spot temperature can be read directly off the transformer's nameplate. This means that the normal ranges for these quantities can be immediately determined; the transformer does not have to be run through a lengthy learning period in order to delineate the edges of the operating envelope.

A good choice of axes for describing the operating envelope, then, should be a small number of familiar, orthogonal quantities that correlate well with inputs to the module's embedded model. It is no accident that the chosen quantities fulfill these specifications. Load current, load voltage, and hot-spot temperature have long been sampled or predicted because of the understanding that these values play an important role in determining the operation of the transformer. The modeling effort, in further quantifying their effects, merely reflects this understanding.

Bins Having chosen the appropriate axes, the next step was to sub-divide the operating space with respect to these axes. This was done simply by setting maximum and minimum values for each axis, as well as the number of intervals the resulting range would be divided into. In this manner, the operating space was divided into regions of equal "volume". Due to the one-to-one correspondence between these regions and the structures used to hold the most recent data in each region, both are referred to as *bins*.

Each bin has a number of *slots*, each of which can hold a distinct set of time-correlated data. When data is to be stored into a bin, an empty slot is selected. If no empty slot is available, the slot containing the oldest data in that bin is used instead. Thus, only the most recent data is retained. Multiple slots are used to decrease the influence of noisy data, and provide more data to the candidate generator for greater statistical validity.

4.2.2.3 Candidate Generation Triggering Criterion

Selecting data intelligently raises the question of when there is enough data available to perform parameter estimation. One approach would be to continue triggering parameter estimation using a schedule similar to the one discussed above, in Section 4.2.1.1. A measure of the validity of the resulting parameters would again be used to determine if they should be installed.

Upon consideration, however, it was realized that the degree to which the various bins are filled is a measure of how well the data describes the operation of the transformer throughout the operating space. That is, the more bins or slots that are filled, the greater the information content of the data. It was therefore decided that the number of filled slots would be used directly as the measure of information content on which to trigger parameter estimation.

A consequence of this decision is that it is not necessary to perform the computations described in the section on parameter validation, Section 4.2.1.3. At least, it is not necessary to compute the term that measures the information content of the input. As for the term related to the average size of the squared error, it was decided that, without referring to the magnitude of the signals being modeled, this quantity had little relevance. This is because the size of the vibration signal can change by multiple orders of magnitude depending on what portion of the operating space is being sampled. Thus, the measure of the estimation error can not be directly compared from one estimation to the next, since there is no guarantee that a particular portion of the operating space will be sampled for every estimation. Presently, the residual is normalized against the size of the measured signal for measurement residual analysis. Perhaps a similar measure would be appropriate for determining the suitability of generated parameters.

Thus, the parameters for the vibration module are not currently subject to validation before being installed. The fact that parameters are estimated at all indicates that an information-rich set of data was available as input; essentially, the candidate parameters are pre-approved. Experience has shown that the resulting parameters are reasonably accurate and consistent. One issue to be addressed is that a change in the physical structure of the transformer may invalidate the model structure, rendering it impossible to estimate valid parameters regardless of the information content of the input data. But such a change should be detectable through analysis of the measurement residual, which should be at least as informative as a measure of the error in the estimation.

4.2.2.4 Partitioning the Operating Space

After operating the vibration module for a while, it was determined that a single set of parameters did not seem to adequately describe the vibration behavior over the entire operating space. The accuracy of $\hat{v}_w$ decreased markedly near the rated load of the transformer and above. It was decided to divide the operating space of the

transformer into two subspaces and independently track the parameters associated with each.

In effect, this achieves the replacement of a single vibration model with two models. The two models have the same structure (though this would not necessarily have to be true) but are applied to distinct regions of the operating space. The parameters of the models are tracked separately. In tracking the vibration residual, however, no distinction is made concerning which model gave rise to a particular residual value. For this reason, the two models are embedded in a single module, which chooses between the two as appropriate. If a distinction should prove important to the diagnostic process, then the two models may need to be embedded in separate modules. For example, it may be desirable to set different residual anomaly detection thresholds for the two models.

4.2.2.5 Data Retention

As a final note on this implementation, the entire bin structure is emptied when parameters are estimated successfully. This increases the amount of time required between parameter estimations, but it is a simple way of guaranteeing that any estimated parameters represent current conditions, i.e., that old data will not persist through many estimation iterations.

4.2.3 Suggested Improvements

The two types of implementation of parameter estimation work reasonably well. The type A implementation is appropriate when there are no constraints on the storage capacity and computational speed of the monitoring system in relation to the size and time scale of the estimation problem. That is, it is appropriate when there is a relatively small, well-defined period of time during which sufficient information to perform parameter estimation is usually generated. The type B implementation is needed when this does not hold true, when there is insufficient storage capacity or number-crunching power. In a commercial implementation, the need to keep costs at a minimum (and perhaps the need to support only a single parameter estimation mechanism) may make a bin implementation appropriate for every module.

Both of the implementations discussed above have been installed in the MIT Transformer Monitoring Project Pilot Facility. Several months of experience operating and reviewing the performance of the parameter estimators has revealed some areas where they might be improved with further development. What follows is a discussion of some suggested improvements that may prove worthwhile. These suggestions generally fall into four categories:

data screening — the suppression of data that may adversely affect the estimation process.

expiration of data — the retention and expiration of old data, particularly from infrequently traversed portions of the operating space.

consistent estimation — the suppression of transitions in the estimated parameters that are merely artifacts of the data acquisition process.

interaction with the diagnostic process — the use of diagnostic information for control of the estimation process and the use of information generated by the estimation process for diagnosis.

It should be stressed that the implementations discussed above work adequately for the Pilot Facility (in the absence of failure conditions). The following suggestions are made in anticipation of possible difficulties that may arise in the course of applying the system to full-scale large power transformers and to transformers that are experiencing incipient failures. Only experience will reveal the most promising avenues of investigation.

4.2.3.1 Data Screening

In a nutshell, parameter estimation should not attempt to adapt a model to fit data it was never intended to predict. Examples of such data are measurements from faulty sensors, measurements outside the normal operating envelope of the transformer, or measurements made during large transients for a model that is not structured to handle such transients.

The type A implementation does not perform any data screening. All data collected during the specified time period is used. The type B implementation does screen data that falls outside of a specified operating space. As discussed above, to increase the accuracy of the vibration module, the operating space of the transformer was divided into two sub-regions. Each region has its own bin structure for data selection, and a separate set of parameters is maintained for each. Each sub-region is screening out data from the other sub-region as well as data outside the operating space of the transformer. This results in parameters that are more accurate for both regions.

The vibration module also screens data that is deemed bad. But, since bad data detection is presently a simple comparison of measurements against acceptable ranges of values, this amounts to screening data outside the operating space. No effort is made at this time to screen out data during large thermal transients, despite the fact that the embedded model of the vibration module assumes steady-state thermal conditions.

4.2.3.2 Expiration of Data

Another potentially rich avenue of investigation is to determine the best way of using old data to improve the performance of the parameter estimator. A naive

approach (embodied in the type A implementation) is that no old data should be used, since one goal of estimation is to achieve the most up-to-date value possible. This ignores the fact that the parameters generated should also be as accurate, as statistically valid as possible. The estimated parameters should be "accurate" in the sense that they should remain constant when the physical parameters do so, and change as the physical parameters change (to enable the use of the estimated parameters to track the changing internal condition of the transformer.) There is a trade-off, then, between this definition of accuracy and the definition in which the parameters result in minimal average squared residuals.

This trade-off is thoroughly exercised because of an additional assumption: that there is a single set of parameters for each module that are valid over the entire operating space of the transformer. This assumption is enforced in part during the modeling effort. A model is only accepted if a single set of parameters proves to be adequate. In practice, however, parameters estimated from data taken from one portion of the operating space may not be in strict agreement with those *tuned* to another portion. Taken to an extreme, this may require that the operating space be partitioned into distinct subspaces with individual sets of parameters. VIBMOD is an example of such a situation (see Section 4.2.2.4.)

If the parameters should happen to be tuned first to one operating region and then another, the estimated parameters may appear to undergo a significant transition while, in fact, the physical parameters of the transformer remain constant. A goal of the estimation process, then, is to achieve parameters that are representative of the behavior of the transformer over the expected operating range. This implies that the parameters must be based on data taken throughout the expected range. The type B implementation, of course, was developed as a solution to this problem.

However, the type B implementation makes no guarantees about how long it will take to sufficiently fill the bin structure to enable parameter generation. By the time the bins are full, some of the stored information may be quite old, and perhaps inappropriate. A common method of dealing with old data is to use it, but to *forget* or discount the data over time. An *exponential forgetting factor* can be used to exponentially decrease the weight assigned to a set of data as the set ages. Thus, with a (constant) exponential forgetting factor λ ($0 < \lambda < 1$), the weight given a set of data n intervals old would be λ^n (see Appendix B.) λ can be chosen to tailor the estimation process to the physical time constants of the transformer.

Chosing λ is not trivial; data must not be forgotten faster than the transformer's operating space is traversed, and yet not remembered so long that reasonable rates of change in parameters can not be tracked in real time. It may not even be desirable to forget data from all portions of the operating space at the same rate. Some portions of the space may be visited very infrequently. Data from such regions may have to be retained for a longer time period, i.e., forgotten at a slower rate. This would allow successive iterations of parameter estimation to trigger at an acceptable frequency and yet allow each iteration to see a consistent coverage of the operating

space.

Such an action is only necessary if the information gathered from the remote region increases the amount of diagnostic information embodied in the parameters, or if accurate predictions in that region are critical. A prediction may be critical because of a large *a priori* probability of failure while the transformer is operating in the region in question. This may be the case, for example, when the transformer is very hot or overloaded.

If the remote information is not important to the estimation process, it may be preferable to screen the data from that region even when it is available. This may impair the accuracy of the model in that region. But it is not true that the estimated model has to be *equally* valid over the entire operating range of the transformer. If the diagnostic process can recognize when a large residual is not significant (based on knowledge of the estimation process or experience with the model), this should not have a noticeable effect on the diagnostic accuracy of the monitoring system.

Is it better to have old data from remote regions of the operating space than none? How old is "old"? These are questions that can only be answered through experience, the experience of operating a system such as the one described in Chapter 3 for many hours on many transformers.

4.2.3.3 Consistent Estimation

There is a great deal of information generated as input to the estimation process. It would be advantageous to use as much of this data as possible to decrease the susceptibility of the estimation process to noise in the data. Unfortunately, this data can not be used in its entirety without recognition of how evenly the operation space is represented. A consistent view of the modeled data is important to the diagnostic information content of the resulting parameters. Parameters should not be arbitrarily tuned to a small range of operating conditions.

The type A implementation uses all the data from the specified period with uniform weighting of all data points. Thus, the resulting parameters are heavily influenced by the prevailing conditions during that period. As the prevailing conditions (e.g., ambient temperature) change, artificial transitions are introduced in the parameters. This decreases the monitoring system's sensitivity to changes in the physical parameters that the estimations represent.

The type B implementation deals with this issue by having a limited number of slots available in each bin, and forcing a significant number of bins to be filled before estimation is triggered. Therefore, there is a maximum level of influence any one region can have on the estimation of parameters. But this has the unfortunate effect of discarding information that could be used to improve the monitoring system's performance. An improvement on this implementation may be to use all the data collected, yet ensure that no bin unduly influences the estimation.

The total weight given to a bin is the sum of the weights given to the data points found in that bin. That is, for a bin b_j the weight of that bin at time k, $w_{b_j}[k]$, is

defined as:

$$w_{b_j}[k] \triangleq \sum_{i=1}^{k} w[i] : \phi[i] \in b_j$$

where $w[i]$ is the weight given the data taken at interval i. ("$\phi[i] \in b_j$" indicates that the operating point of the transformer at time i lies the operating subregion defined by b_j.) Assuming the use of a constant exponential forgetting factor λ, this would be:

$$w_{b_j}[k] \triangleq \sum_{i=1}^{k} \lambda^{k-i} : \phi[i] \in b_j$$

If no forgetting factor is used (i.e., uniform unity weighting), the total weight of a bin is simply the number of data points stored in that bin.

In order to ensure that no small set of bins dominate the estimation process, one must prevent any w_{b_j} from becoming too large. The bin implementation described above enforces a strict upper bound on w_{b_j}: the number of slots available in each bin. Since there is no forgetting factor, the weight of a bin is just the number of filled slots in that bin.

It is possible to use all the data collected and yet ensure that any w_{b_j} will saturate at a given limit. This can be done by normalizing the weights of individual data samples (in bins with a total weight exceeding the limit) to result in a total bin weight that is below the limit. This procedure can be performed independent of any other weighting scheme, such as the exponential forgetting described above. Each individual sample will still contribute the same relative weight in characterizing the model in that one bin. But each bin will be limited in its contribution to the complete model.

Each sample weight, $w[i]$, would be rescaled by $w_{max,b_j}/w_{b_j}$. w_{max,b_j} would not have to be constant for all bins b_j, in recognition of the difference in importance of various operation regions. Perhaps w_{max,b_j} need not be a constant for an individual bin as well, but a function of the current (unnormalized) weight of the bin. This function could approach a saturation value exponentially. This would allow more frequently visited bins a marginally greater weight in the estimation process. This, in turn, would allow parameters to be optimized somewhat to changing prevailing conditions while still suppressing day-to-day parameter transitions. The saturation value for an individual bin could be computed based on a long-term frequency count of bin visits.

If there were a problem with storage space for the amount of data generated during the arbitrary length data acquisition interval, the recursive least squares algorithm could be adapted to the purpose. To perform the weighted least squares calculation, two matrices can be used: $\Phi^T[k]W[k]\Phi[k]$ and $\Phi^T[k]W[k]Y[k]$ ($W[k]$ is a diagonal matrix of $w[1]..w[k]$). The following relationships are true:

$$\Phi^T[k]W[k]\Phi[k] = \sum_{i=1}^{k} \phi[i]w[i]\phi^T[i] \tag{4.36}$$

$$\Phi^T[k]W[k]Y[k] = \sum_{i=1}^{k} \phi[i]w[i]y[i] \tag{4.37}$$

These matrices can be defined recursively:

$$\Phi^T W\Phi[k] \triangleq \Phi^T W\Phi[k-1] + \phi[k]w[k]\phi^T[k]$$
$$\Phi^T WY[k] \triangleq \Phi^T WY[k-1] + \phi[k]w[k]y[k]$$

But the following relationships are also true (assuming each valid $\phi[i]$ falls into some bin b_j):

$$\Phi^T[k]W[k]\Phi[k] = \sum_{b_j}\sum_{i=1}^{k} \phi[i]w[i]\phi^T[i] : \phi[i] \in b_j \tag{4.38}$$

$$\Phi^T[k]W[k]Y[k] = \sum_{b_j}\sum_{i=1}^{k} \phi[i]w[i]y[i] : \phi[i] \in b_j \tag{4.39}$$

Thus, a pair of matrices can be defined for each bin as follows:

$$\Phi^T W\Phi_{b_j}[k] \triangleq \begin{cases} \Phi^T W\Phi_{b_j}[k-1] + \phi[k]w[k]\phi^T[k] & \text{if } \phi[k] \in b_j \\ \Phi^T W\Phi_{b_j}[k-1] & \text{otherwise} \end{cases}$$
$$\Phi^T WY_{b_j}[k] \triangleq \begin{cases} \Phi^T WY_{b_j}[k-1] + \phi[k]w[k]y[k] & \text{if } \phi[k] \in b_j \\ \Phi^T WY_{b_j}[k-1] & \text{otherwise} \end{cases}$$

with the intention that:

$$\Phi^T W\Phi[k] \triangleq \sum_{b_j} w_{max,b_j}[k]/w_{b_j}[k] \times \Phi^T W\Phi_{b_j}[k]$$
$$\Phi^T WY[k] \triangleq \sum_{b_j} w_{max,b_j}[k]/w_{b_j}[k] \times \Phi^T WY_{b_j}[k]$$

4.2.3.4 Interaction with the Diagnostic Process

Possibly the most fruitful avenue of research into parameter estimation will be along the lines of developing a method of interaction between the parameter estimator and the diagnostic process. This would be two-way communication; the parameter estimator generates information which is useful for diagnosis, and the status of the diagnostic process may affect the estimation.

Obviously, the estimated parameters themselves are the key pieces of information that the diagnostic process will need from the estimator. Not as obvious, however, is that other information about the estimation process may be important as well. One example is the history of failed estimation attempts; a failure to estimate accurate parameters may distinguish one type of failure from another. The

information content of the input data is another quantity which may be of interest. It is possible that anomalies in the information content, given that the load profile of the transformer is stable, may be used as evidence of sensor malfunction, for example.

In the other direction, the diagnostic process may be able to provide input to the estimation process. For example, if an internal change in the transformer is detected and diagnosed as non-threatening, it may be necessary, as a normal operation, to change the validity thresholds used to accept parameters. This could represent the recognition that a larger excursion from zero in a measurement residual must then be acceptable. This is because the average squared error is a contributing factor in determining the validity thresholds. Another example is that the diagnostic process may deduce that a serious change in parameters has occurred. If this were the case, it would be necessary to forget all old data in order to accurately estimate the new parameters.

4.3 Summary

In this chapter, the Pilot Transformer Monitoring System was described. The first section reviewed the structure that was used to implement the new approach to transformer monitoring and outlined the software that was developed. Each of the modules which are currently operating were fully discussed.

In the second section, two fully implemented parameter estimation schemes were presented. The type A implementation was found to be acceptable for models whose input space was sufficiently explored in a moderate time interval. For such models, the storage and computing resources of the monitoring system were not strenuously taxed.

One module, however, did not fit this template. The type B implementation was developed to allow the estimation of accurate, diagnostically-useful parameters for the vibration module, VIBMOD. The input space of the vibration model was explored very slowly, but the bin structure enabled accurate, consistent parameters to be achieved.

The presentation of the two implementations was followed by a discussion of some areas where further investigation or empirical demonstration may be needed. It should be noted, though, that the suggestions above are based on a relatively short span of experience, compared to the hours of experience that would be necessary to explore and verify them. It can not be overstressed that experience is the key to continued improvement to the monitoring system. Experience will show which suggestions are most worth exploring and, in fact, whether any of them are truly needed.

Chapter 5

Integration of MIT Approach with Traditional Approach

In Chapter 3, the basic concept underlying the MIT approach to transformer monitoring was presented. A structure for an integrated monitoring system applying this concept was then described. Chapter 4 detailed how this monitoring structure is implemented. This chapter will discuss how the new monitoring approach can be used immediately to improve the reliability of large power transformers while the diagnostic capabilities of the system are still being explored.

The first section will review the tasks of detection and diagnosis. This review will provide the framework for the discussions that follow. In the next section, the sensitivity of the MIT transformer monitoring system to incipient failures is used to leverage the known diagnostic power of traditional techniques. Early experience with the test facility yielded some initial insight into how the diagnostic capabilities of the monitoring system might be improved. The third section chronicles these events and their preliminary interpretations.

It became obvious that further work into the diagnosis of incipient transformer failures could not proceed effectively without solid data. The rationale and the broad goals of the ensuing experiments are presented in the final section. This discussion will lay the groundwork for the presentation of the experimental setup and results in Chapter 6.

5.1 Review: Detection and Diagnosis

In the discussion that follows, it is important to maintain the distinction between the detection and diagnosis of incipient failures. Detection of incipient failures entails monitoring the various sensors and recognizing when some sensed quantity is not behaving normally. Diagnosis is a decision process by which the abnormal behavior will, if possible, be explained with some degree of certainty. Some diagnoses will be able to be reached solely on the weight of the detected anomalies; others will require

additional information—either in the form of additional tests, modified sampling schedules, adjusted detection thresholds, or simply more time to see if the condition persists.

A mechanism for the detection of incipient failures has been fully implemented and integrated into the Pilot Transformer Test Facility. The Pilot Monitoring System, including this detection method, is described in great detail in Section 4.1. However, lacking from this description is any mention of an implemented diagnostic system. The purpose of the remainder of this thesis is to determine how such a system might be developed.

Before diagnosis can be considered, a complete understanding of the detection mechanism must be achieved. Anomaly detection is realized through the use of an array of thresholds: measurement thresholds, residual thresholds, and parameter thresholds. Each of these types of thresholds, which monitor absolute levels, are complemented by rate-of-change thresholds that determine if any of these quantities are changing faster than normal.

Measurement thresholds and measurement rate-of-change thresholds are very traditional. Many of these thresholds could be read directly from the transformer's nameplate, or from the ANSI operating guidelines. But static measurement thresholds do not take full advantage of the range of data and experience available to the transformer expert. This is the primary consideration behind the development of the system described in this thesis.

Residual thresholds and parameter thresholds are the simplest method for implementing a detection scheme using adaptive models and knowledge specific to a particular transformer. The transformer signatures that have been modeled were chosen because their behaviors change in the presence of failures (see the matrix of relationships between failure modes and observable quantities, Table 3.1). Therefore, at some point in the evolution of the failure, the residuals and/or the parameters will diverge from past behavior. In a perfect world, healthy transformers would yield residuals consisting of white noise and parameters that were perfectly stable. Threshold anomaly detection would not only be sufficient, it would be optimal. Of course, in the real world residuals are often correlated with module inputs, and parameters are subject to noise and drift.

Each residual and parameter is compared to its corresponding thresholds on a continual basis. Tunable residual and parameter thresholds allow the monitoring system to be further customized to the individual transformer being monitored. For instance, a base-load unit would show a very different set of residual behaviors than a transformer subject to extensive load cycling. Therefore, the residual thresholds for a base-load unit might be much tighter than for other transformers. If a transformer is consistently highly-loaded, then a higher parameter rate of change may be expected and the thresholds would reflect this.

It should be noted that threshold anomaly detection is not the only method for incipient failure detection. One common method involves modeling each possible

type of failure separately and choosing the model that fits the observed behavior most closely as the most likely candidate[25, pp. 263–293]. This approach integrates the detection and diagnostic tasks. Unfortunately, there is not nearly enough experience available to create detailed models of every failure mode. In addition, the diagnostic (and prognostic) information available through parameter estimation are not fully exploited in this approach.

5.2 Integration of Adaptive Models into Present-Day Monitoring

It was shown in Chapter 4 that several diverse modules have been fully implemented and are presently generating predictions, residuals, and parameters for the Pilot Facility. A simple and effective method of anomaly detection has also been implemented: residual and parameter threshold detection. Lacking an effective way of performing a diagnosis using only the information available to the detection system, the modules described in this thesis are relegated to the role of incipient failure detectors—albeit sensitive ones.

It would be hasty to recommend that scheduled oil samples and inspections be replaced by the system described in the previous chapter, but it is reasonable to expect that such a system could eventually handle the task of incipient failure detection. Using the adaptive modules as triggers for the traditional diagnostic techniques presented in Chapter 2 would improve the reliability of large power transformers. Unfortunately, anomalies (as defined by residual and parameter threshold detection) will not correspond directly to dangerous incipient failures; false alarms might be common.

One way of limiting the number of false alarms would be to decrease the sensitivity of the detection system, so that only the most severe anomalies trigger diagnosis. This is unacceptable, however, in that it undermines one of the major strengths of the proposed system: its sensitivity to changes in the condition of the transformer. The alternative is to automate some portion of the diagnostic process to screen out as many of the false alarms as possible before involving the human operators. Of course, one difficulty with this approach is that traditional diagnosis techniques have not made use of many of the quantities that are generated by the modules. Automating the diagnostic process requires that the process itself first be fully understood. The following section makes a rough cut at outlining the process and isolating those factors that must be considered before automation can be achieved.

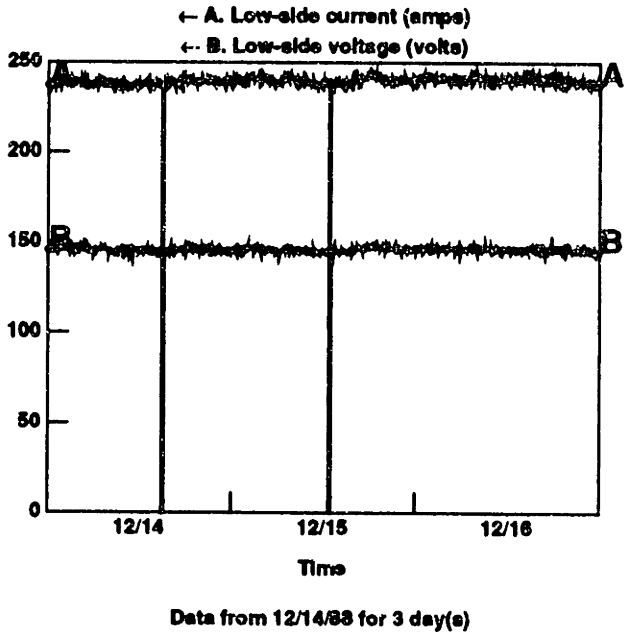


Figure 5.1: Steady-State Operation

5.3 Initial Proposed Advancements to Diagnosis

First, it is necessary to determine how detected anomalies and the raw module outputs (predictions, residuals and parameters) can be used for diagnosis. To understand how the information generated by the MIT monitoring system can best be used to support diagnosis, an example is illuminating. Consider a preliminary experiment that was run early in the history of the Pilot Monitoring System. In this experiment, a heating tape was wrapped around the combustible gas sensor. Power was dissipated in the tape, disrupting the sensor's temperature compensation and introducing a small anomalous heat source into the transformer.

Figure 5.1 shows that the transformer was operating in steady-state at 75% of full load.[1] The heating tape was used to inject approximately 30 W of heat at the tank wall. This is equivalent to about 10% of the losses of the transformer.

Figure 5.2 shows the behavior of the combustible gas residual during this experiment. A large step change in the residual occurred when the heating was started; another step change occurred when the heating was reduced. From later experience, it was learned that, while gas can be generated rapidly enough to produce the positive step change seen here, the gas residual will not maintain that level and then

[1]The arrows (← or →) preceding the curve labels indicate which vertical axis is being used. In Figure 5.1, both curves are plotted against the left-hand axis (although the units differ) so they are marked with a left arrow (←). The units are indicated in parentheses after the curve label.

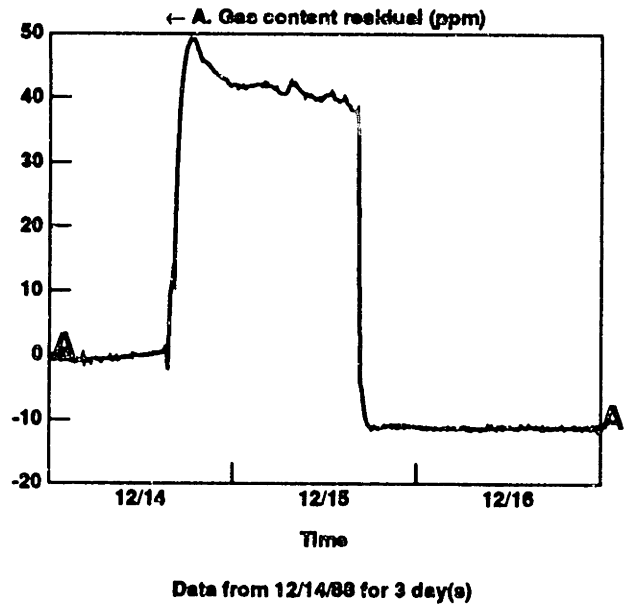


Figure 5.2: Anomalous Combustible Gas Residual

experience a negative step change. Instead, in a gas-space transformer such as was used here, upon reaching a maximum the gas residual will drop exponentially as the oil and gas space try to reach equilibrium. Thus, the residual behavior seen during this experiment is more likely indicative of a sensor failure than a gas-generating transformer failure. The fact that the residual returns to a level ten parts per million *lower* than where it started makes it even more likely that a sensor failure was involved. A single failure that both generates and consumes combustible gas is extremely unlikely.

Figure 5.3 shows the behavior of a thermal residual for the same time period as Figure 5.2. Seen in isolation, this residual would not represent an anomaly, for the entire range of the residual is well within normal bounds. In fact, the residual is generally getting smaller during the period in question (this is normal behavior in response to a change in load that occurred earlier). Looking more closely at the times of the step changes in the gas residual, one may see that the residual rises approximately 0.5°C at the time of the first step change, and falls again at the time of the second.

Again, it must be emphasized that a half a degree rise is not usually significant. The rise only returned the residual to the level it had been at half a day earlier. The rise in temperature is significant only because of the absence of any change in load at that time and because of the coincidence of the gas residual anomaly.

The presence of an unexplained temperature rise in conjunction with the anoma-

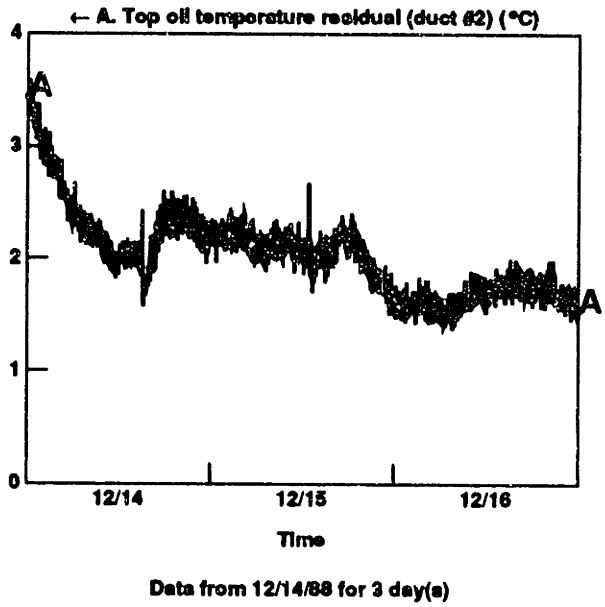


Figure 5.3: Anomalous Thermal Residual

lous gas residual yields a more accurate diagnosis. The gas residual alone indicates only that the problem is likely associated with the sensor rather than the transformer itself. The detection of an unexplained temperature rise, though slight, further indicates that there is most likely a problem with the sensor itself rather than the data acquisition system. It is much easier to hypothesize a coupling between the two events in the physical media of the transformer than in the almost completely distinct data paths of the data acquisition system. There is always a possibility that the temperature rise is completely unrelated. But the close correlation in time of the two events and the principle that the simplest explanation is usually true argue against this possibility.

The missing piece of knowledge necessary to arrive at a relatively precise diagnosis is that the temperature compensation of the combustible gas sensor is imperfect. The conclusion is that a new heat source has been introduced that is interfering with that compensation. Following more operating experience with the thermal module, the size of the thermal residual may be used to reason about the location of the heat source. The embedded model in the thermal module is tuned to detect heat sources in the transformer winding. The small magnitude of the thermal rise may indicate that there is either a small heat source in the winding or a large heat source external to the winding. As a small heat source in the winding is unlikely to disrupt the temperature compensation in the gas sensor, a likely conclusion may be that a large heat source can be found near the gas sensor. Which in this case proves to be true.

As a final note, referring back to Figure 5.1, one notes two "blips" where the current and voltage briefly go to zero. These two events roughly correlate in time with the end points of the gas and thermal residual anomalies. Experience with the operating profile of the Pilot Facility allows one to recognize these blips as maintenance shutdowns. Similar events on other transformers may not lead to the same conclusion if, for example, those transformers are never re-energized that quickly. On a transformer with a different operating profile, then, this same electrical event may be seen as data acquisition error.

However, diagnosis need not rely completely on deductive reasoning in this situation. When it is a question of whether the transformer was de-energized or not, the human operator should have the definitive answer. If the monitoring system were fully integrated into the control room operations, the system would be forewarned about a shutdown. But if diagnosis depends on this fact and it is unknown, the most direct route is to query the operator. In the scenario described above, the correlation between the maintenance shutdowns and the duration of the anomalies would lead one to suspect human intervention. It would not be unreasonable for an automated diagnostic system to seek further information on what activities occurred during the shutdown.

This experiment illustrates some important issues in reasoning about transformer failures.

- The ability to reason about the coincidence of two events in different signatures is vital. It was necessary to locate a correlated event in a second signature to improve the initial diagnosis of sensor failure. The improvement in diagnosis may mean the difference between performing an expensive operation (such as replacing a sensor) or an inexpensive one (such as applying some thermal insulation).

- The system must be able to re-examine recent raw signatures with new sensitivity thresholds to identify events that may have been insignificant at an earlier time. The thermal residual in isolation was uninformative. If the system were set up to flag all such data as interesting, it would be swamped with these "events".

- The system should be able to reason about the relative magnitude of data signals, such as the rise identified in the thermal residual. This is one area in which simulation and actual operational experience are critical. It is possible, though unlikely, that a 0.5°C rise in a thermal residual in so short a time (with no correlated change in load) is always indicative of extraneous heating. Or it is possible that a 50 parts per million change in a gas residual can not manifest itself so quickly, so that it is absolutely certain that the exhibited behavior has to be due to a sensor failure, regardless of the residual's subsequent decay. But these possibilities can not be properly assessed without many hours of operational experience.

- The system should be able to reason about the type of transformer under scrutiny. A piece of knowledge applicable to a transformer with a gas space may not be applicable to one without; it may also be crucial to know whether the transformer is core-type or shell-type.

- The type of sensor being used may be important to the diagnostic process. In the example above, the knowledge of the particular type of sensor's temperature compensation proved to be an important piece of knowledge.

- Module-specific knowledge can play a key role during diagnosis. The thermal residual discussed above is based on a model of the temperature distribution in transformer windings. Knowledge of this model can be used to reason about the sort of events that this module should be more or less sensitive to. This knowledge led to a conclusion about the location of the extraneous heating.

- A transformer-specific database is needed to recognize events that may be unique to the particular transformer. In the example above, it is possible that the duration of the two shutdowns would be unlikely for most transformers. If a utility goes to the expense of de-energizing the transformer, then a more complete maintenance review would be initiated, requiring more time.

- Diagnosis should not be limited to the information which is automatically available to the system. Some diagnoses can be reached quickly, efficiently, or unambiguously if a small amount of external information is supplied. An automated diagnostic system must be able to request this data, and must be able to analyze the relative costs and benefits to determine if the data should be requested.

In Table 5.1, the matrix of relationships between failure modes and observable quantities is updated to reflect the status of the Pilot Monitoring System. The table shows that the implemented modules have the capability to detect many important failure modes and, working in parallel, distinguish one failure mode from another. What the table does not reflect, but is reflected in the discussion above, is the great promise that these modules hold out for diagnosis through working together, and not merely in parallel.

5.4 Need for Experimentation

The utility of the MIT monitoring system is blunted by a lack of experience, by a lack of detailed knowledge as to how the evolution of failures affects the monitored signatures. In creating a polished system for device diagnosis, there is no true substitute for experience. Only thousands of operating hours on many transformers can yield a population of failures large enough to act as a basis for reasoning about

Table 5.1: Relationships Between Failure Modes and Modules

	Failure Modes							
Modules	Bent Winding	Core Damage	Cracked Bushing	Electri-fication	Hot Spot	Arcing Short	Gas Bubbles	Contaminated Oil
Thie3mod	●	●		●	●			
Thmod	●	●		●	●			
Gasmod			●	●	●	●	●	●
Vibmod	●	●						
Wthmod			●		●			●

their diagnoses. But, using a small number of incipient failures, one can identify the paths of reasoning that are used to arrive at a diagnosis from the data that is generated by the monitoring system. That is, one can identify the structure of the *knowledge base* that will be used by the diagnostic system. As incipient failures are investigated, important mechanisms of the diagnostic process are highlighted. If it is found, for instance, that the ability to reason about past behavior of a residual is critical in many different instances, regardless of the particular residual or failure in question, then efforts should be made to support the effective use of that knowledge.

Due to limitations of scope and time on this thesis, there can be little thought of generating a data base sufficient to the task of yielding a complete diagnostic system. The focus of this thesis is instead on the identification of the structure of the knowledge base on which such a system could be built. This will be done by simulating a selected set of failure modes and generalizing from this constrained sample. Thus when the experience of many thousands of hours is digested, the structures and mechanisms necessary to implement that expertise will already be present.

In automating the diagnostic process, the goal is to take full advantage of the automatic anomaly detection. In pursuing this goal, several considerations are important. First, the system should not be inextricably tied to a particular set of modules. It should degrade gracefully when modules are removed, and be easily extended to include new or improved modules. Second, the information necessary to adapt the system to a different transformer design should be automatically generated where possible (in the case of adaptive parameters), and explicitly stated and localized where it is not. The third consideration is that the system should be able to take advantage of redundant information, such as the underlying module structure common to all the modules. Fourth, the system must be able to handle ill-specified knowledge and incomplete data in reaching a diagnosis. Finally, the system must be able to incorporate the knowledge of transformer experts and conventional diagnosis methods.

To address these considerations, the choice was made to use an *expert system.*

This technology offers a way to provide modularity, extensibility, and maintainability while capturing the expertise of the transformer engineer[37,38].

Unfortunately, there is not yet enough experience with the types of quantities that are generated by the MIT monitoring system to develop a knowledge base that would support the diagnosis of real-world failures. However, the analysis of the preliminary experiment described above shows that significant insight into the diagnostic problem can be gained through case studies of individual failures. To this end, incipient failures were simulated in the Pilot Facility to reveal the knowledge that was used to perform diagnosis, and how that knowledge was used. From this constrained set, the framework for an automated diagnosis system can be projected. The experimental setup and results are presented in Chapter 6. Careful analysis of these results have revealed effective mechanisms for transformer diagnosis.

5.5 Summary

This chapter has explored how to make the best use of the existing MIT Transformer Monitoring System. In the first section, the completed status of incipient failure detection was contrasted to the undeveloped state of automated diagnosis. A brief discussion on the manner in which failure detection using adaptive models of normal behavior could be integrated with traditional diagnosis followed. Early experience with the diagnostic capabilities of the system was presented. This early experience shaped the goals of this thesis towards gathering experimental data that could be used as a template for advanced automated diagnosis.

Chapter 6

Experimental Results

A reliable diagnostic system must be based on a solid body of experience. The body of knowledge concerning large power transformers is rich and useful, yet the type of information needed for an adaptive-model-based approach is lacking. Little is known of the short-term behavior of transformers in the face of incipient failures, and there is almost no experience available concerning the particular parameters that the MIT system estimates and monitors.

There is a statistical component involved in diagnosis; diagnostic methods employ the concepts "usually", "often", and "probably". There is a basic uncertainty involved that can only be addressed through statistics, based on experience. Despite this fact, case studies can illuminate the problems of diagnosis and reveal how diagnosis should proceed, even if *a priori* likelihoods can not be evaluated.

This chapter presents the experimental results of this research. First, there is a brief discussion of the philosophy behind the chosen approach to broadening the understanding of the diagnostic capabilities of the monitoring system. Next, the strategy behind the selection of the specific experiments conducted is revealed. The physical setup is described in the following section. Finally, the results are presented, including some analysis of what these results signify.

6.1 Philosophy

There can be no substitute for experience to prove out the concepts and structures that have been described in the preceding chapters. Unfortunately, a sufficient body of experience does not yet exist to fully evaluate the diagnostic capabilities of the proposed system. These techniques, however, hold such promise for the advancement of incipient failure diagnosis of large power transformers that research into model-based diagnosis is desirable. To this end, the experiments described below were chosen to explore the diagnostic information revealed by the monitoring system.

It was not intended that this thesis would yield a definitive diagnostic system.

The experiments were intended as a case study of how a system of adaptive models of normal behavior could be applied to the task of transformer diagnosis. The preliminary experiment described in Chapter 5 showed how this approach to diagnostic research can yield deep insights. Additional case studies will highlight the shared structure of the diagnostic process, as well as reveal the range of analytic tools that are required for diagnosis. Identifying the shared structure is the first step towards implementing a diagnostic system that is flexible and maintainable.

By actually performing the diagnosis, and carefully considering each step of the process, the underlying strategy of transformer diagnosis can be revealed. Understanding this strategy is critical to the generation of a system whose results will have meaning for a transformer engineer or operator. But the adaptive model-based trend analysis tools described in the previous chapters have not been fully integrated into the state of the art. Chapter 5 described how the residuals and parameters would be used for detection and hinted how they might be used for diagnosis and prognosis. But only by applying these methods to actual data can the worth of a set of residuals and parameters be evaluated.

Efforts have been made during this research to clearly define the relationships between the numerical data collected and the physical parameters of the Pilot Facility. This has been done to ensure that the analysis and reasoning brought to bear here can be applied in an analogous manner to a full-size power transformer.

6.2 Strategy of Selection

Having determined the importance of gaining experience with the transformer monitoring system's response to failures, the problem was reduced to choosing an appropriate set of failure modes to simulate. There were a number of criteria to consider in making this selection. These criteria included constraints on time and money, the relative importance with which the utilities viewed each failure mode, the extent to which the pilot monitoring system was suited to detecting a particular failure mode, the risk of physical damage to the Pilot Facility, repeatability, ease of simulation, and non-disruptive (on-line) control of the simulation.

The time constraint was governed by the author's goals and the sponsors needs. It was determined that each of the tasks of hardware implementation, experimentation, and data analysis should be accomplished in a matter of months. The incremental cost of this research program (beyond the operating budget of the Pilot Facility) was dominated by the cost of maintaining a graduate student.

Communication with the sponsoring utilities focused the search onto gassing failures and through faults. Gassing failures include arcing, partial discharges, and general or local overheating. Arcing was chosen because a straight-forward implementation mechanism was available (a spark gap with an external power supply), the arcing could be controlled externally without altering the operating profile of the transformer, and the arcing could be closely reproduced (after making allowances

for a rise in the total gas content of the transformer). However, to minimize the potential risk to the transformer, the experiment was limited to arcing in the oil, rather than near the cellulose.

One result of through faults that the Pilot Facility is particularly suited to detect is winding deformation due to the huge electromechanical forces generated by the events. A method of applying a force to the winding that would change the winding in an impermanent way was needed. This would change the stress patterns in the winding in much the same way that a loosening clamp would.

The next section describes the apparatus that were designed to simulate the chosen failure modes. The issues of physical risk to the test transformer, repeatability, and on-line control are closely related to the choices made during the design phase. These issues are touched on in the discussion of the setup.

6.3 Setup

First, a description of a proposed experiment to investigate the monitoring system's response to changing mechanical stresses on the winding is presented. Due to time and cost constraints, this experiment had to be abandoned. The proposal is included in the hopes that it may be carried out in the future, and so that the reader can have another example of an appropriate experiment for the investigation of the diagnostic capabilities of the MIT's Transformer Monitoring System.

The implemented experiment on the diagnosis of gas generation in transformers is presented next. Results are given in the subsequent section. Included are numerical analyses of the data to verify the physical processes at work.

6.3.1 Winding Deformation

An additional consideration in designing an apparatus that would change the transformer's vibration response was that no new mechanical coupling between the winding and the core or the winding and the tank could be introduced. This would change the vibration response of the winding in a way that would be difficult to relate to an actual failure mode. This eliminated any device that was braced against the core or tank in order to apply a force to the winding. It also seemed to eliminate a direct mechanical connection between the apparatus and its external controls. However, it was decided that a hydraulic control system would not transmit a significant amount of vibration to the tank, and thus was acceptable.

Unable to *push* on the winding, it was decided that the solution was to *squeeze* the winding. The forces exerted by the winding clamps have already been shown to affect the vibration response in definite, detectable ways[28,27]. Variable control of the squeezing would allow emulation of slow changes due to the normal loosening of winding clamps, as well as sudden changes due to through faults.

One suggested method for squeezing the winding was to gird the winding with a cinch strap. The cinch might be tightened by pulling on the two ends of the strap with a motor or a hydraulic piston. However, making this control small enough to fit in the transformer tank and attaching the ends of the strap securely would pose some difficulties. In addition, the forces applied to the transformer would be symmetric about the circumference of the winding. This may lead to an unnatural, and potentially misleading, simulation of the failure modes seen under normal circumstances. A proposed solution is to use the cinch strap as a brace for a device that would apply a localized pressure to the winding.

As described above, a strap would be used to encircle the winding. But instead of adjusting the tension on the strap by lengthening or shortening the strap, a hydraulic piston would be inserted between the strap and the winding. This proposed device is shown in Figure 6.1. This method has the advantages that no new mechanical couplings are introduced between the winding and the rest of the transformer, and that the resulting forces are asymmetric. Hydraulic jacks with very short maximum extensions but with significant lift are commercially available. Transformer oil could be used as hydraulic fluid in order to eliminate the danger of contamination of the transformer from the hydraulic system.

In order for the winding deformation test apparatus to be truly successful, it is desirable that the deformation of the winding be repeatable. Any stretching of the strap must be elastic, as must any deflection of the winding. The harsh conditions inside the transformer (with the presence of high temperatures, oil, and strong electric fields) require that the materials be known to be stable in such an environment. Organic materials are relatively stable under these conditions. A cotton luggage strap was tested on a spare pole-type distribution transformer. The results were inconclusive. It was detected that the tension on the strap decreased over time, suggesting plastic stretching of the strap. Permanent deformation of the winding would also lead to reduced strap tension when the hydraulic piston was fully retracted, but this was ruled out by separate measurements of the deflection of the winding surface.

Plastic stretching was suggested, but that did not take into account that the cotton strap had become saturated with oil from contact with the winding. The clamping structure had been deemed adequate for the dry strap, but there may have been some gradual slippage after the strap was lubricated by the oil. Whether the reduced tension was due to plastic stretching or clamp slippage, the effect was avoided if the extension of the piston (and thus the tension on the strap) was limited.

Unfortunately, it was decided that the simulation apparatus described above could not be installed in an acceptable amount of time. Installing the cinch strap would necessitate de-tanking the transformer. Once the winding had been exposed to air, the lengthy process of drying out the winding would be required. In addition, the long down time for the Pilot Facility would have disrupted other parallel research projects. And finally, if the transformer were to be de-tanked, access to the

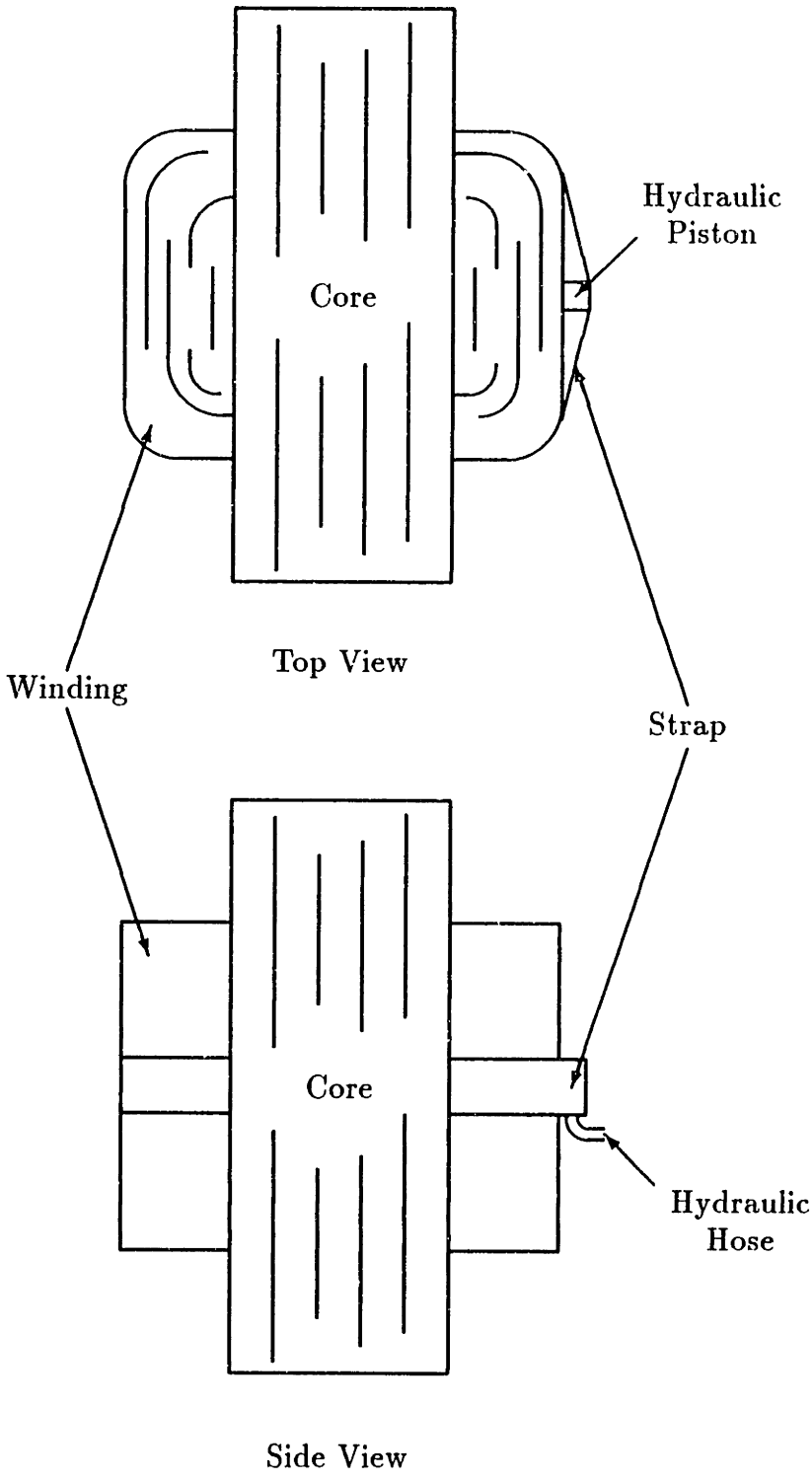


Figure 6.1: Proposed Winding Deformation Test Apparatus

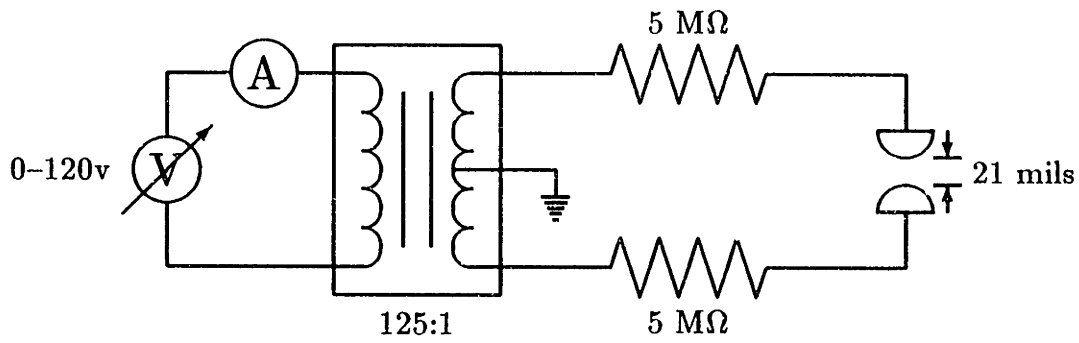


Figure 6.2: Arcing Simulation Circuit

transformer would have to have been coordinated with other researchers to make their own modifications.

This approach to vibration module verification has been incorporated into a continuing research project at MIT. This research is entitled "Transformer Monitoring Using Vibration Analysis".

6.3.2 Arcing

An arcing failure was simulated in the Pilot Facility by inserting an automotive spark plug into the test transformer. To avoid permanent damage to the transformer, the spark plug was located away from the winding and the high voltage leads. A schematic of the circuit used is given in Figure 6.2. A 120 VAC variable supply was connected to a neon-sign transformer, rated at 15 kV/30 mA. An inductive ammeter was used to measure the input current to the neon-sign transformer, which is center grounded. High-voltage ceramic resistors were used to limit the current to the transformer's rated amperage. 5 MΩ of resistance (two (2) 10 MΩ ceramic resistors in parallel) was attached to each bushing of the neon-sign transformer. The balanced resistance on a center-grounded transformer ensured that, when the device was actually arcing, the voltage of the electrodes was near the system's ground. This was done to minimize the chance of an arc from the simulation apparatus to some other portion of the test transformer. The spark gap distance was set to 21 mils. This gap distance was found, through trial and error, to be small enough to yield a spark reliably at the attainable voltages, yet large enough to avoid fouling by the carbonization of the oil.

The spark plug was positioned in the transformer by screwing it into a piece of Lexan, which was suspended from a fiberglass scaffold at the top of the transformer tank. Lexan was chosen because it was non-conducting, easily machinable, and

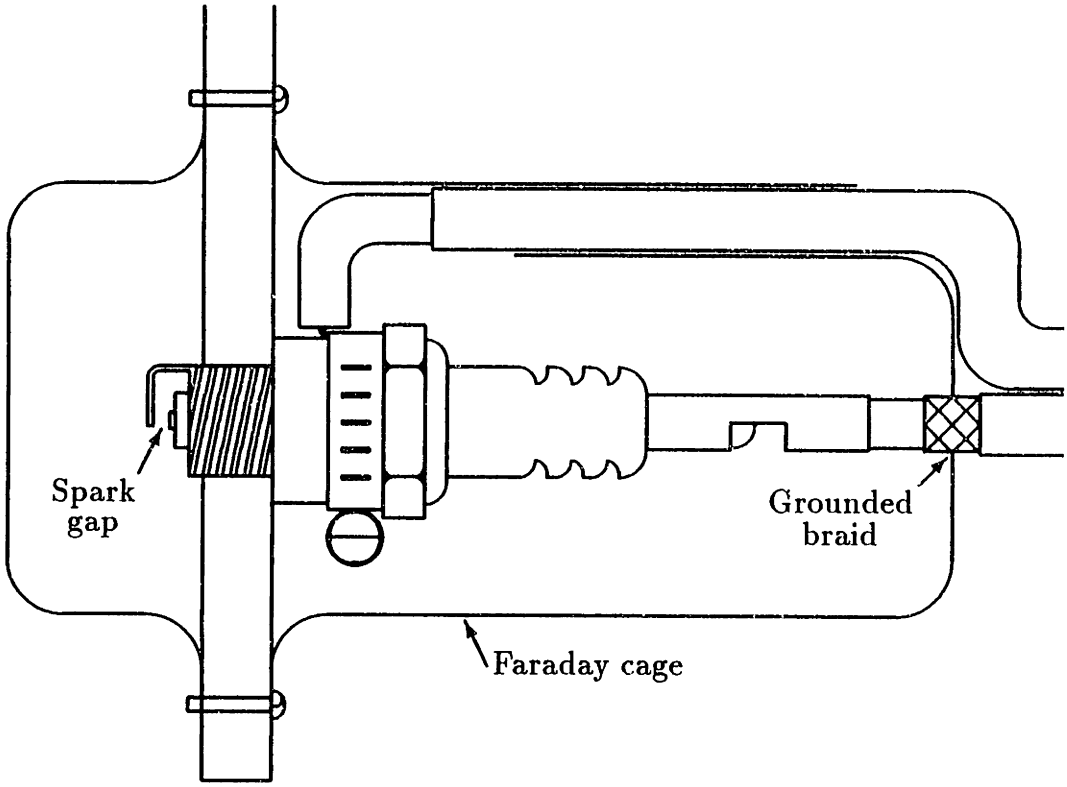


Figure 6.3: Shielded Spark Plug for Arcing Simulation

would not become pliant in the temperature range to which the transformer was exposed. The spark plug was shielded and the shielding was grounded in order to protect the other instrumentation.

The shielded spark plug is shown in Figure 6.3. The shielding, or Faraday cage, was built from copper mesh. The cage was constructed from two pieces, one piece forming a cap over the spark plug's electrodes, the other shielding the spark plug's body. The two pieces were fixed in place and connected electrically with screws that passed through the Lexan. Copper sheet was added where the two mesh forms met to provide extra shielding. The copper sheet was held in place, and all sharp edges covered, by a framework of copper and brass tubing. The tubing was soldered together with silver solder to withstand the temperatures of the transformer.

Power was supplied to the spark plug through two RG-59/U coaxial cables. The copper braid of each cable was grounded externally. One of the braids was then connected to the Faraday cage to ground the unit. The other braid was insulated from the cage to prevent circulating currents in the shielding, though the braid extended well into the shielded region. Thus, the entire sparking apparatus was shielded from well outside the transformer tank all the way to the spark plug. This decreased the chance that electromagnetic interference would damage or otherwise disturb the other monitoring instrumentation.

While the transformer tank was open for the installation of the arcing simulation apparatus, the integrity of the tank and lid were improved. A few small leaks were identified and sealed. When the installation of the device described above was completed, it was found that a significant amount of gas was escaping down the length of the coaxial cable between the braid and the outer insulation. The cables were subsequently sealed where the braid was exposed for the grounding connection. After these leaks were dealt with, the tank was able to maintain a positive pressure. This indicated that any detected changes in gas concentrations were due to activity inside the transformer, and not atmospheric contamination.

6.4 Results

In this section, some of the results collected from the MIT Pilot Transformer Test Facility are presented. In Section 6.4.1, the initial condition of the transformer following the installation of the simulation hardware is documented. Following this presentation, three arcing experiments are reported. The first experiment was designed to investigate the response of the combustible gas module residual during a gas-producing event under steady-state operation of the transformer. The next experiment explored GASMOD's residual behavior during normal load cycling, and explored the subsequent model adaptation. The final experiment presented here focused on the behavior of the combustible gas module's parameters in the presence of low rates of gas generation.

For this discussion, the reader should be familiar with some details of the dissolved combustible gas module (GASMOD) presented in Chapter 5. The equation used to estimate the parameters of this module was given in Equation (4.16). Squaring both sides of the equation, the result is:

$$hydran[k] = (\theta_1 + \theta_2 \times T_{topoil}[k])^2 \tag{6.1}$$

GASMOD also supports a model that uses the predicted winding interior temperature ($\hat{T}_{wtint}$) instead of the mixed top oil temperature (T_{topoil}). To simplify the discussion, only the model presented in Equation (6.1) will be considered. Little effort has been expended on evaluating the relative advantages and disadvantages of the two models.

For the Pilot Facility, parameters are automatically generated daily using two days' worth of data. The parameters are generated shortly after midnight; parameters accepted on a Friday, for example, represent a measure of the condition of the transformer for Wednesday and Thursday of that week.

6.4.1 De-gas and Equilibration

After the spark plug was installed in the test transformer, the transformer oil was subjected to a high vacuum to lower the dissolved gas concentrations as much as

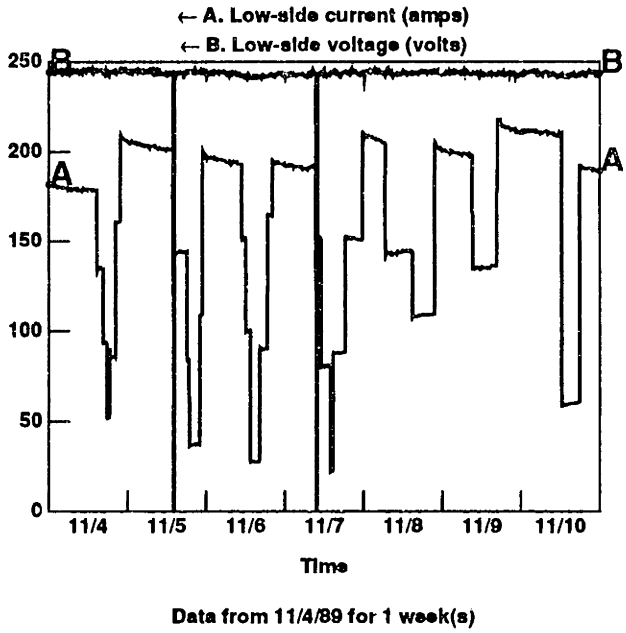


Figure 6.4: Load and Excitation after De-gas

possible. The transformer was set at a moderate load to establish a convective flow in the oil. This process also verified the gas integrity of the transformer tank and lid. This is important to discount the possibility that changing gas concentration levels in later experiments were due to contamination with the atmosphere.

The test transformer was brought back on line in early November, 1989. Load cycling was begun immediately, with the intention of generating parameters (Figure 6.4). During the de-gas operation, the Hydran was disabled to avoid exposing the device to the high vacuum used. Flow to the Hydran was cut off, which resulted in a reading of 0 ppm. (The sensor consumes the gases to which it is sensitive. When fresh oil is prevented from reaching it, the combustible gases are depleted and the device "starves".) Flow was restored to the Hydran on November 4, 1989. It was immediately apparent that the concentrations of combustible gases in the transformer oil were successfully reduced. This can be seen in Figure 6.5.

However, Figure 6.5 also shows that the combustible gas content of the transformer rose significantly in the following days, reaching levels about 60% of those before the de-gas. Some combustible gas may have been trapped in the winding in much the same way that moisture is held. But the majority of the gas finding its way into the oil of the transformer tank probably came from the expansion tank connected to the gas space. The oil and the gas space in the expansion tank had not been exposed to the atmosphere during the installation procedure, and thus did not pose the same danger of moisture and oxygen contamination that the oil

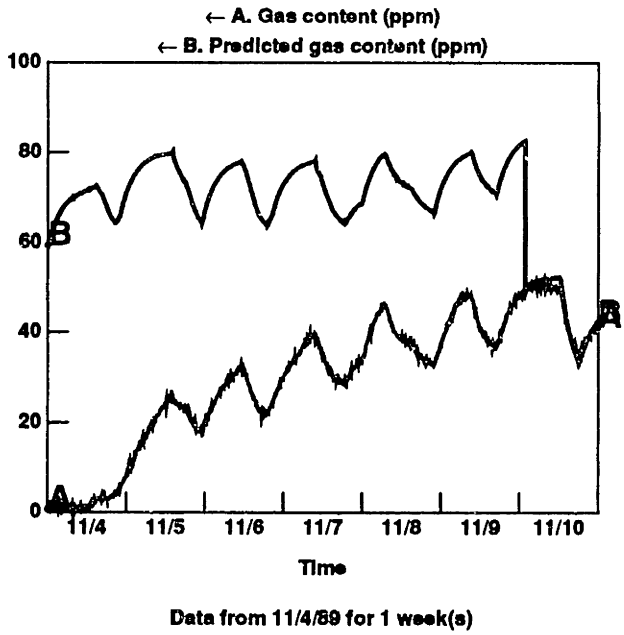


Figure 6.5: Hydran Response after De-gas (Predicted and Actual)

in the transformer tank did. So, the expansion tank was not subject to the de-gas operation.

Also shown in Figure 6.5 is the predicted combustible gas content. The predictions were initially much higher than measured values, indicating that the dissolved gas content of the oil had been reduced. The parameters had not had a chance to adapt to the new condition of the transformer. The parameters were not updated during the de-gas operation because the Hydran yielded a constant zero reading while the oil flow to the sensor was shut off. During the de-gas operation, the temperatures inside the transformer were relatively constant. This situation yielded data that was very poor in information content, resulting in the rejection of the candidate parameters.

Once the flow was restored to the Hydran and load cycling resumed, however, it still took several days for the parameters to be updated. The parameter update can be seen quite clearly from the data of November 10, since the condition of the transformer had changed markedly since the last parameter update. Both the actual and predicted response of the Hydran show that there was significant daily variation due to temperature. Why, then, was parameter acceptance delayed until November 10?

The upward trend in the Hydran readings was not due to general heating of the transformer. The thermal data shows this, but it can also be seen in the predicted response of the gas sensor, in Figure 6.5. It was because of this uncorrelated

trend in the gas levels that the model could not properly adapt to the condition of the transformer. The module assumes that there is a time-invariant relationship between gas content and temperature during the period of parameter estimation.

In Chapter 4, the criteria used to accept or reject candidate parameters was shown to be based on a measure of the information content of the data used for estimation, and a measure of how well the resulting model fit this data. The data collected between November 4 and November 9 was sufficiently rich, under normal circumstances, to generate acceptable parameters. But because of the uncorrelated trend in the dissolved gas level, the model could not adequately fit the data. This points up one shortcoming of the mechanism used to accept parameters: the lack of distinction between rejected parameters based on information-poor data and those that failed to fit the data. For, while the estimated parameters could not fit the data as closely as is normal, they may have yielded valuable diagnostic information about the changing condition of the transformer.

This is an example of a conflict between the two uses to which parameters are put. Parameters are used to generate predictions for residual analysis, and are used directly in parameter trend analysis. Accepting parameters that are based on data that violates the model structure of the module may greatly complicate the residual analysis, yet those same candidate parameters may reveal a dangerous trend. The conflict only arises if one assumes that the parameters used for residual analysis and those used for parameter analysis must be identical. This assumption should receive further scrutiny.

Figure 6.6 shows the behavior of GASMOD's residual during the same period. Note that the residual is displayed at the same magnification that was used to display the measured and predicted behavior of the sensor. Normally, the residual is examined at much higher magnification to identify small features of the signal. In the situation here, however, the larger trend dominates the behavior of the residual; smaller features are relatively insignificant in the face of this trend.

The residual is negative because the actual measured gas content is lower than the predicted amount. Subtracting the prediction from the measured value almost completely accounts for the daily variation of the reading. The residual clearly and cleanly shows the trend in the gas content of the oil. When the module finally succeeds in adapting the model to the observed data, the residual returns to approximately zero.

The residual can be closely modeled as an exponential with a time constant of 40 hours. The first hypothesis was that this time constant was the result of the diffusion rate of hydrogen from the expansion tank into the gas blanket of the transformer, and that this was obscuring the time constant of the diffusion of hydrogen from the gas blanket into the oil. To determine the validity of this hypothesis, the time constant of the diffusion between the two tanks was computed using Fick's equation

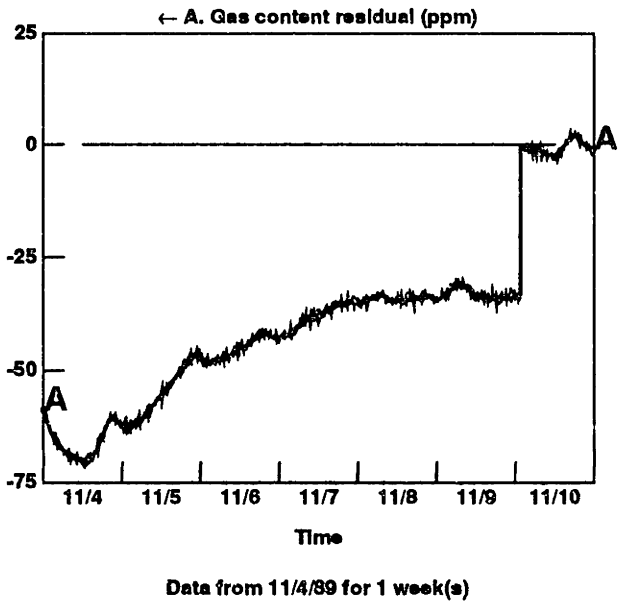


Figure 6.6: GASMOD Residual Response After De-gas

of diffusion,[39, p. 6]

$$\frac{w}{A} = -D\frac{\partial c}{\partial x}, \tag{6.2}$$

where $\frac{w}{A}$ is the molar mass flux in the hose connecting the two tanks (mole/sec m^2), c is the concentration of hydrogen in the hose as a function of position and time (mole/m^3), and D is the coefficient of diffusion (m^2/sec).

As a simplification, the concentration gradient in the hose is assumed to be linear. Thus,

$$\frac{\partial c}{\partial x} = \frac{c_2 - c_1}{L}, \tag{6.3}$$

where c_2 is the concentration of hydrogen in the expansion tank and c_1 is the concentration of hydrogen in the gas blanket of the transformer, as a function of time, and L is the length of the hose. Next, note that the change in concentration is equal to the change in mass divided by the volume

$$\frac{dc_1}{dt} = -\frac{1}{V_1}w \tag{6.4}$$

or

$$\frac{dc_1}{dt} = \frac{AD}{V_1 L}(c_2 - c_1) \tag{6.5}$$

where A is the cross-sectional area of the hose, and V_1 is the volume of the gas blanket.

Since the transformer is a closed system, mass is conserved. Thus,

$$c_1 V_1 + c_2 V_2 + \frac{c_1 + c_2}{2} AL = N \tag{6.6}$$

where N is a constant. The third term is negligible since AL, the volume of the hose, is small, so the equation can be reduced to

$$c_1 V_1 + c_2 V_2 = N \tag{6.7}$$

Solving for c_2,

$$c_2 = \frac{N - c_1 V_1}{V_2} \tag{6.8}$$

Substituting Equation (6.8) into Equation (6.5) yields

$$\frac{dc_1}{dt} = -\frac{AD(V_1 + V_2)}{V_1 V_2 L} c_1 + \frac{ADN}{V_1 V_2 L} \tag{6.9}$$

Therefore, c_1 is an exponential with a time constant τ such that

$$\tau = \frac{V_1 V_2 L}{AD(V_1 + V_2)} \tag{6.10}$$

Evaluating τ for the Pilot Facility yields τ equal to 1.9×10^8 sec, or approximately 6 years. Obviously, the observed transient in the dissolved gas content of the oil was not due to diffusion of hydrogen through the hose.

Having eliminated diffusion from the expansion tank, the question became one of discovering from whence the gas entering the oil came, and what physical properties determined the time constant. At this point there was not enough information to decide, but the transient was ascribed to the equilibration of the convectively mixed oil and the gas blanket, as a working hypothesis. This still leaves unanswered the question of where the combustible gas in the gas blanket came from, since the blanket was supposedly pure nitrogen. A tentative explanation is that when the valve between the transformer and the expansion tank was opened, the expansion tank was at a slightly higher pressure. This would have forced hydrogen-laden gas down the hose to equalize the pressure, providing a source for combustible gas in the oil.

The residual succeeded in tracking the changing state of the transformer over the course of a few days. Looking at the parameters spanning this time period reveals the effect of the de-gas, and shows the parameters start to return to their original values, even though the parameters were not successfully estimated for most of this period. Figure 6.7 shows the behavior of one of the gas module's parameters (parameter θ_1 of Equation (6.1)) in context with estimates of the parameter from before the de-gas.

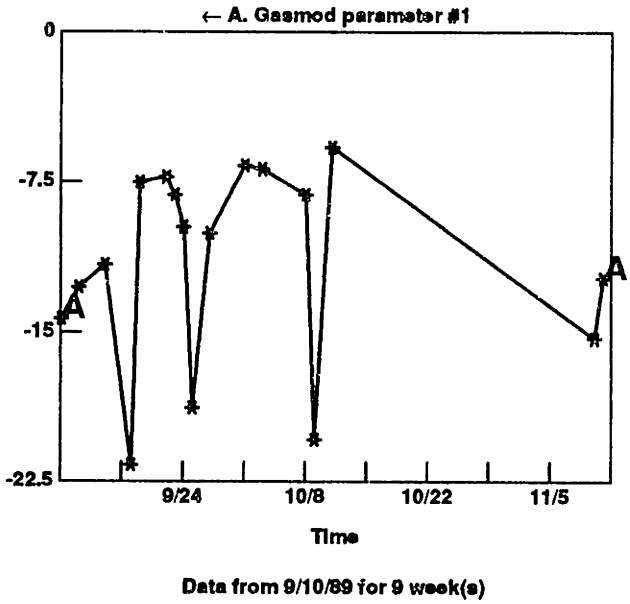


Figure 6.7: GASMOD Parameter Response After De-gas

The three "spikes" in the parameter on September 18, September 25, and October 9 in Figure 6.7 represent information-poor parameters that were accepted erroneously. Each of these dates is a Monday, meaning that the parameters were generated using data from Saturday and Sunday. The load cycling of the transformer is not automated, thus little or no information is generated over the weekend when no one is present to adjust the load. This points out another shortcoming of the parameter acceptance scheme: the multiplicative nature of the acceptance criteria means that, with a degenerate set of data, the data fit can be so close as to disable the measure of information content. That is, if the data can be fit almost perfectly, the curve fit error will be approximately zero. The product of the two terms may then approach zero, even though the measure of information content may be relatively high (signifying information-poor data). Following this period, the acceptance criteria were tightened, so that such parameters would have a much smaller chance of being accepted.

The plot in Figure 6.7 shows that the parameters were sensitive to the de-gas operation and the subsequent equilibration. The following experiments were then conducted to verify the monitoring system's sensitivity to high- and low-energy arcing.

6.4.2 Residual Response Experiment

The arcing equipment described above granted very little control over the intensity of the arc produced. The spark gap distance and the values of the current-limiting resistors could not be changed. The voltage supplied to the neon-sign transformer was variable, but it was found that, because of the high dielectric strength of the transformer oil, arcing could only be reliably initiated near the neon-sign transformer's output voltage rating.

However, at this point, it was not the *ratios* of combustible gases produced that were of concern, but total quantity and rate of production. These variables could be controlled through the duration and scheduling of the arcing events. Thus, a high-intensity arcing event was simulated through a single long (tens to hundreds of seconds) spark; an incipient arcing failure was simulated with brief three-second arcs distributed over several days.

The first experiment described here consisted of a short twelve-second arc, followed a half an hour later by a sixty-second arc. Two hours later, a seventy-two-second arc was produced. These arcs were generated while the transformer was in thermal steady-state. The load and excitation for the week beginning at this time are shown in Figure 6.8. The load was held essentially constant for several days following the arcs. Though the transformer was shut down several times for maintenance and to draw oil samples for gas chromatography for comparison, the load was restored to its previous level each time. This introduced some thermal transients into the transformer, but the temperature was re-established at the same steady-state value.

The results of these arcing events are shown in Figure 6.9. The initial twelve-second arc was generated at 11:30 on November 14. A close examination of the residual data shows a rise in the residual of approximately 1.5 ppm, correlated in time with the arc. This correlation in time includes a twenty-minute time lag due to mixing and mass transport effects. While encouraging, such a small rise would be lost in the noise.

The sixty-second arc was generated at noon. Ten minutes later, the Hydran's reading was virtually unchanged. At fifteen minutes after the event, the measured dissolved gas content had risen by approximately 5 ppm. At twenty minutes, the reading had stabilized with a rise of approximately 10 ppm. Under normal utility procedures, an absolute rise of 10 ppm would be ignored. Such a change is well within the normal daily variation of the Hydran reading. Some progressive utilities have alarms that react to abnormal rates of change, but this event would likely go unnoticed.

Simply by sampling the Hydran output every two minutes and presenting the results graphically, one can detect and analyze an event that would be missed completely using traditional monitoring techniques. That is not to say that this event represents a threat to the transformer. As an isolated event, it would simply be

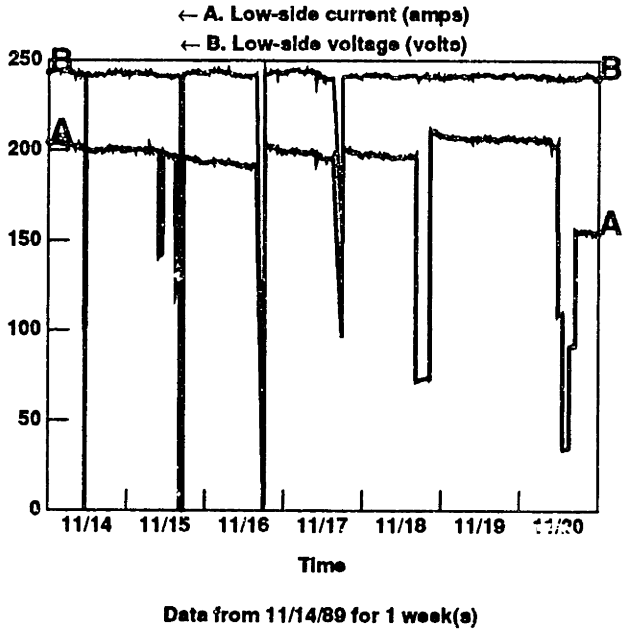


Figure 6.8: Load and Excitation During Steady-State

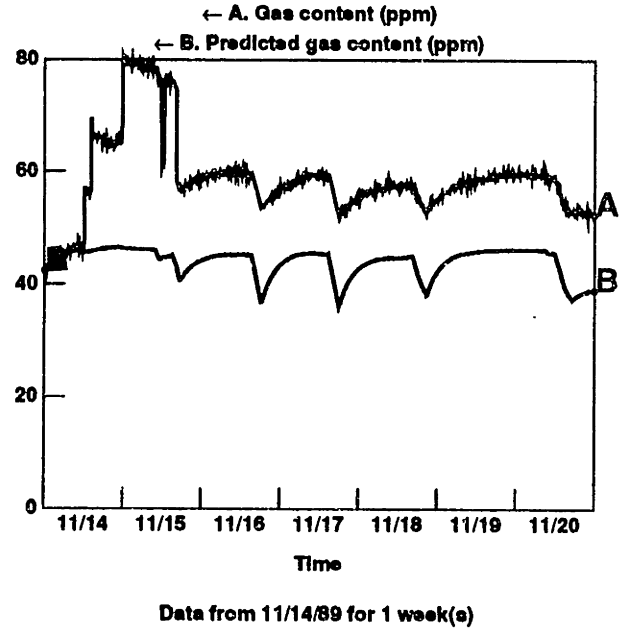


Figure 6.9: Hydran Response to Arc During Steady-State

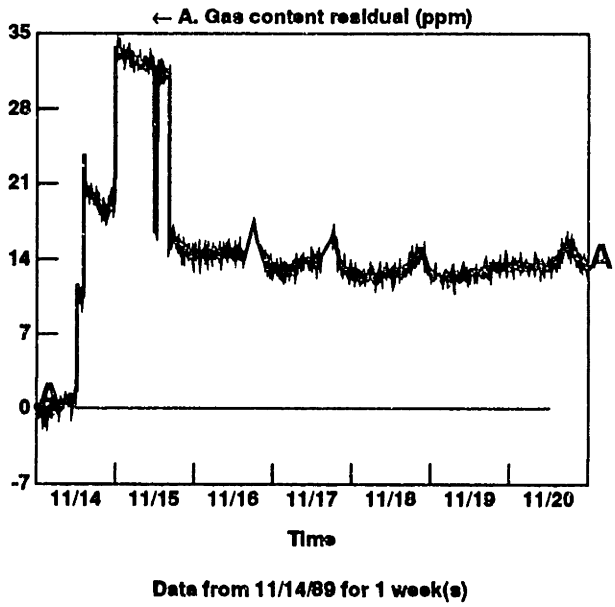


Figure 6.10: GASMOD Residual Response to Arc During Steady-State

noted and the operation of the transformer would most likely continue unchanged. Since such data from large power transformers is scarce, it is quite possible that such events happen many times each year. But they are only detected as an aggregate rise in the total combustible gas content of the oil. However unlikely it is that this one event represents a dangerous situation, it is true that for a brief time conditions existed in the transformer to support the generation of an arc. Detecting such an event could focus attention on an incipient failure before it can progress very far.

In this situation, the raw Hydran data was sufficient to detect an anomaly, as the rapid 10-ppm rise ran contrary to experience. Turning attention to the associated residual confirms that the anomalous behavior of the Hydran was not caused by thermal effects. Figure 6.10 displays the residual in question. The rise of 1.5 ppm that was overlooked before now gains significance. Instead of having a single isolated gassing event, there is a possibility that there were two events—the second being more intense or of longer duration.

After reaching a maximum value twenty minutes after the spark plug was operated, the Hydran reading began to slowly decay. This is expected behavior for a transformer with a gas blanket. Initially, the gas produced by the arc dissolves in the volume of oil in that area. If this slug of gas-rich oil circulates to the gas sensor before mixing thoroughly, then an artificially high reading may result that does not represent the gas content of the entire transformer. Quickly this gas-rich oil will mix with the rest of the oil. The gas in the oil, however, can not equilibrate

with the gas blanket as quickly. The decay of the Hydran reading may reflect the movement of combustible gas into the gas space.

At 2 P.M., the arcing device was run for seventy-two seconds. The twenty-minute time lag was confirmed. The "overshoot" due to incomplete mixing can be seen more clearly in this event. The maximum rise was approximately 13 ppm. Again, the absolute gas quantities that are being produced would not attract the attention of a transformer operator. No alarms would be triggered and the events would not even be noted. But, with the MIT monitoring scheme (and, to this point in the discussion, only measurement and residual analysis have been considered), three gassing events were detected, each more severe than the last.

It is still unclear whether such a scenario represents a significant threat to the health of a power transformer. More experience is necessary to evaluate the danger—experience with full-scale power transformers, experience with transformers that fail catastrophically. But it is certainly clear that this scenario would warrant close attention by the transformer operator, with load reduction as a possible result.

The data appears to indicate that a great deal of gassing occurred just before midnight the same day, with other events on November 15. However, as was pointed out in the discussion of the preliminary anomalous heating experiment in Chapter 5, a negative step change in the Hydran reading is indicative of a sensor failure. The fact that the magnitudes of the negative step changes are correlated with those of the positive steps strengthens this interpretation. This problem was eventually traced to a grounding problem in the monitoring instrumentation.

This sensor failure illustrates the strength of a continuous, on-line monitoring system. By considering these events in context, and applying knowledge about the equipment being monitored and the equipment doing the monitoring, one can differentiate between two events that would be considered identical using static thresholds. The fact that one event represents a repairable sensor malfunction and the other a potentially dangerous incipient failure makes the example all the more telling.

6.4.3 Model Adaptation Experiment

The next arcing experiment was also conducted during thermal steady state. However, in this experiment, after the gassing was detected and the steady-state behavior observed in the previous experiment confirmed, load cycling was resumed. The load and excitation of the test transformer for this experiment is presented in Figure 6.11.

The duration of the arc on November 27 was 360 seconds. This arc caused an initial rise in the gas content reading of approximately 50 ppm, as seen in Figure 6.12. This is consistent with the amount of gas generation seen in the first experiment. In the previous experiment, the first two arcs totaled seventy-two seconds in length,

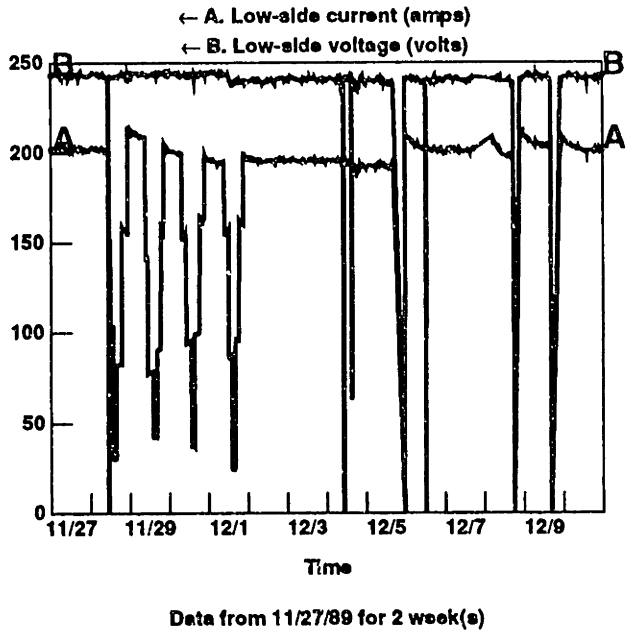


Figure 6.11: Load and Excitation Model Adaptation

which was the duration of the third arc. Each seventy-two seconds of arcing caused a rise of about 10 ppm. At six minutes in length, the arc on November 27 was five times as long, and generated roughly five times as much gas.

The effect of load cycling can be seen in the Hydran response. Even though the general trend of the gas content is downwards after the arcing, there are periods of time where the gas content is actually rising. The residual confirms that these rises are due entirely to thermal cycling of the transformer and do not mask additional gas generation. Figure 6.13 shows the residual.

After two days of load cycling, the module was able to adapt to the new condition of the gas content signature despite the fact that there was an uncorrelated trend in the input data. However, Figure 6.13 indicates that the parameters accepted on December 1 were already out of date when they were installed. Thus, the prediction was too high until parameters based on more stable conditions could be estimated on December 3.

Examining only the Hydran data of Figure 6.12, it is difficult to determine when the gas content of the oil achieved equilibrium. By December 3 and 4, it seemed as though the gas content has stabilized. The measurements and predictions converge due to the parameter update of December 3, and the residual seems stable at zero. Looking at a longer time span, however, it can be seen that the residual continues to diverge, albeit at a slower rate than that immediately following the gas generation. By December 10, the residual had fallen to a value less than -8 ppm with no

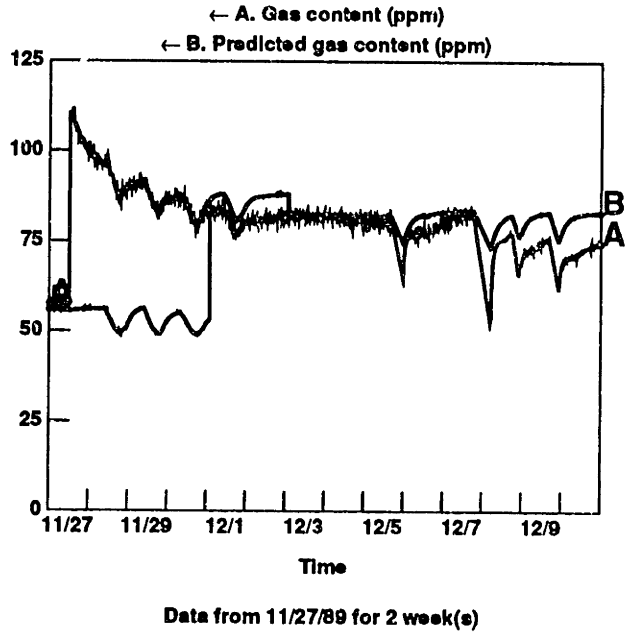


Figure 6.12: Hydran Response to Arc (Steady-State and Load Cycling)

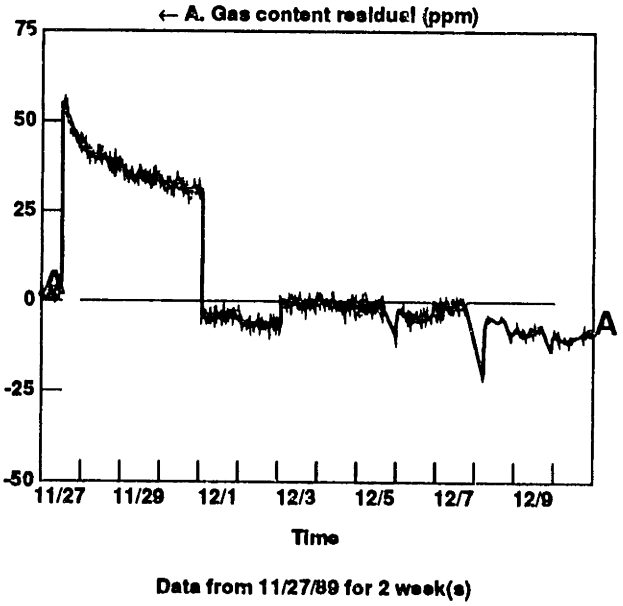


Figure 6.13: GASMOD Residual Response to Arc (Steady-State and Load Cycling)

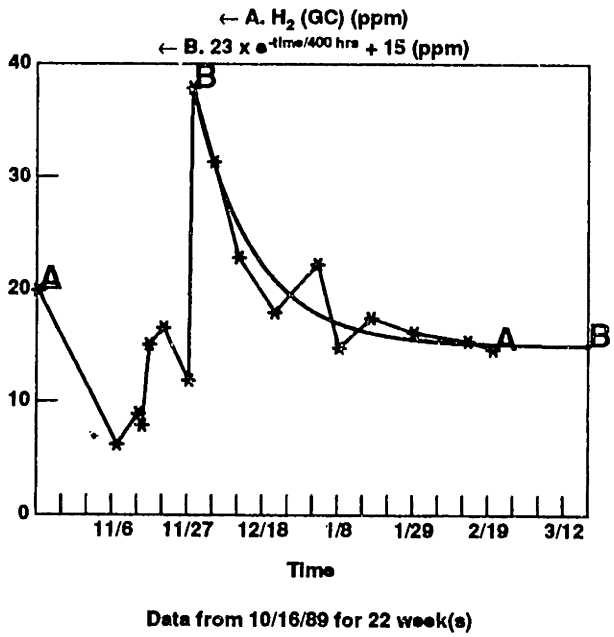


Figure 6.14: Gas Chromatograph H_2 Response to Arc

intervening parameter update.

The gas chromatograph provides a check on the results generated by the Hydran. Figure 6.14 shows twenty-two weeks of GC hydrogen readings, taken weekly. Fitted to this data is shown an exponential with a time constant of 400 hours. This agrees with neither the observed time constant of 40 hours for the equilibration following the re-energization of the transformer, nor with the calculated time constant of 6 years for diffusion through the hose between the gas blanket and the expansion tank. Interestingly, if one examines the Hydran data for this same period, one can note the 400-hour time constant but there is a faster transient that is left unaccounted for. Subtracting off the slow trend, an faster exponential remains—one with a time constant of approximately 40 hours, in agreement with the previously observed behavior.

The Hydran is sensitive to more than one gas. Thus, the presence of two time constants suggested that two or more gases were equilibrating at different rates simultaneously. This explanation was rejected for two reasons: the GC showed that only hydrogen was produced in sufficient quantities to effect the magnitude of changes seen, and hydrogen—having the lowest molecular weight—should have the fastest time constant for diffusion.

Casting about for an explanation of the 400-hour time constant, it was determined that it could be explained very well by the daily cycle of thermal expansion of the gas blanket into the expansion tank. As the transformer gets warm each day, the hydrogen-rich gas in the gas blanket expands into the expansion tank.

The expansion tank maintains a relatively stable temperature. The hydrogen-rich gas mixes with the gas in the expansion tank, raising the concentration of hydrogen there. When the transformer cools, the pressure in the transformer falls and the gas from the expansion tank is drawn into the transformer. This lowers the concentration of hydrogen in the gas space.

To simulate this situation, one first makes use of the ideal gas law:

$$PV = nRT \tag{6.11}$$

Each volume obeys this law. Thus,

$$P_1V_1 = n_1RT_1 \tag{6.12}$$

$$P_2V_2 = n_2RT_2 \tag{6.13}$$

The subscript 1 refers to the gas blanket of the transformer, 2 refers to the gas space of the expansion tank. But, since the two are connected, $P_1 = P_2$, so

$$\frac{n_1T_1}{V_1} = \frac{n_2T_2}{V_2} \tag{6.14}$$

Mass is conserved, so

$$n_1 + n_2 = N \tag{6.15}$$

ignoring the contents of the hose.

Using Equation (6.14) and Equation (6.15) yields

$$n_1 = \frac{NT_2V_1}{T_1V_2 + T_2V_1} \tag{6.16}$$

which is a function of T_1, the temperature of the gas blanket, since T_2 is essentially constant.

When the gas in the gas blanket expands, a some number of moles of gas are forced into the expansion tank. At the minimum temperature of the transformer, $T_{1,min}$, there is some number of moles of gas in the gas blanket, denoted by $n_1(T_{1,min})$. At the maximum temperature, there is some smaller number of moles present, $n_1(T_{1,max})$. At the maximum temperature, the ratio of gases in the gas blanket remains unchanged from the when the transformer was last cool. As seen above, there is essentially no diffusion along the hose, and the assumption is made that the temperature of the transformer is monotonically increasing for part of the day and monotonically decreasing for the rest of the day. Thus, while the transformer is heating, no gas will enter the gas blanket from the expansion tank.

So, the ratio of gases in the gas blanket remains unchanged, but the absolute concentration of hydrogen has dropped by the ratio of number of total moles remaining to the initial number of moles. This is modeled with the following equation:

$$c_{1,max}[k] = \frac{n_1(T_{1,max})}{n_1(T_{1,min})}c_{1,min}[k-1] \tag{6.17}$$

where $c_{1,max}[k]$ is the concentration of hydrogen in the gas blanket at its maximum temperature. Conservation of mass requires that

$$c_{2,max}[k] = \frac{N_{H_2} - V_1 c_{1,max}[k]}{V_2} \tag{6.18}$$

where N_{H_2} is the number of moles of hydrogen in the system.

A similar argument applies to the cooling phase of the daily cycle, yielding the equations

$$c_{1,min}[k] = \frac{N_{H_2} - V_2 c_{2,min}[k]}{V_1} \tag{6.19}$$

$$c_{2,min}[k] = \frac{N - n_1(T_{1,min})}{N - n_1(T_{1,max})} c_{2,max}[k-1] \tag{6.20}$$

Combining these equations and simplifying results in

$$c_1[k] = \frac{T_{min}}{T_{max}} c_1[k-1] + \frac{(T_{max} - T_{min})T_2}{T_{max}(V_2 T_{min} + V_1 T_2)} N_{H_2} \tag{6.21}$$

with a sampling interval of a day. The time constant, τ, of the process can be calculated from this equation with the following relation:

$$\tau = -\frac{24 \text{ hours}}{\ln\left(\frac{T_{min}}{T_{max}}\right)} \tag{6.22}$$

Choosing T_{min} equal to 330°K and T_{min} equal to 350°K yields a time constant of 408 hours. This agrees very well with the observed behavior. This calculated time constant is relatively insensitive to the absolute magnitudes of T_{min} and T_{max} in the normal temperature range achieved by the transformer, but it is sensitive to the difference $T_{min} - T_{max}$. A change in this difference of 5°K can lead to a change in the time constant of nearly 20%.

Having explained the 400-hour time constant, the faster 40-hour time constant ren ains to verified as the time constant of diffusion between the oil and the gas space. The convective flow of the oil provides a turbulent mixing of the oil; the gas space is also well mixed. Under these conditions, diffusion occurs in a small boundary layer in the oil and a much smaller boundary layer in the gas. The diffusion of gas through this boundary layer can be modeled as an infinite series of exponentials if the thickness of the layer is known. The precise conditions internal to the transformer need to be known in order to determine the thickness of the oil boundary layer. At this time, the details of the oil flow are not known, rendering a direct calculation impossible. But the diffusion, though consisting of an infinite number of exponential time constants, is dominated by a single time constant, τ, that satisfies the relation

$$\tau = \frac{L^2}{D} \tag{6.23}$$

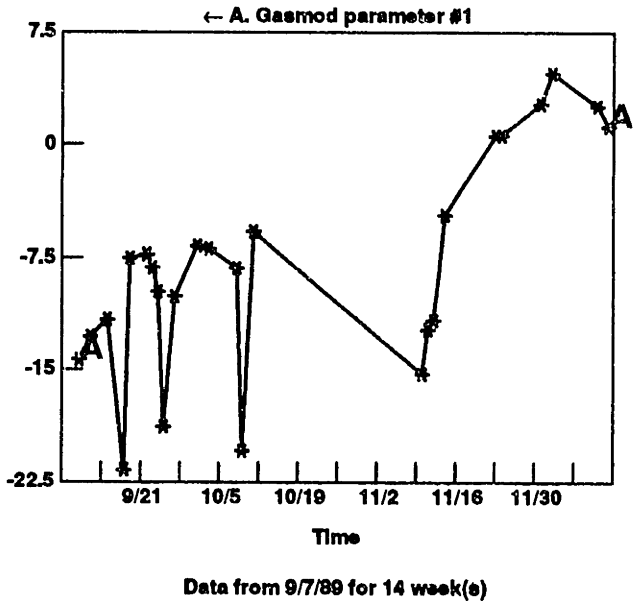


Figure 6.15: GASMOD Parameter Response to Arc

where L is the thickness of the layer, and D is the coefficient of diffusion of hydrogen in transformer oil. Given an observed τ and the value of D, then, one can estimate L and decide if the proposed process makes physical sense. Using an extremely rough estimate of D as 5×10^{-11} m^2/sec, the boundary layer thickness is estimated as 8.5 mm, a reasonable value.

Modeling the transport of hydrogen from the oil to the gas space with a time constant of 40 hours, and from the gas space to the expansion tank with a time constant of 400 hours, the Hydran response following the arcing can be accounted for quite closely. The agreement is improved if it is assumed that some amount of gas went directly into the gas space in the form of bubbles. Preliminary experiments showed that a significant amount of bubbles were created by the arcing apparatus.

The gas module parameters follow the decay, shown in Figure 6.15. Only parameter θ_1 of Equation (6.1) is shown here. Parameter θ_2 of the same equation is closely correlated with this parameter, though the two are inversely related. Surprisingly, the estimate of θ_1 generated on December 3 is larger than that of December 1, despite the fact that the dissolved gas content of the oil was higher on the first of the month. This can be explained by the fact that the trend in the gas content which was uncorrelated with temperature was more significant on December 1. The final experiment described here demonstrated how to improve the correlation between the behavior of the parameters and the actual gas content of the oil.

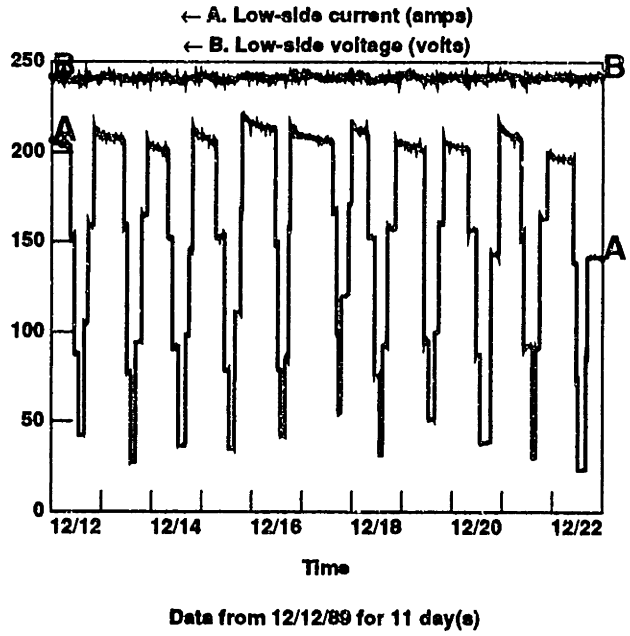


Figure 6.16: Load and Excitation During Low-Intensity Arcing

6.4.4 Parameter Trend Analysis Experiment

This last experiment was designed to investigate the behavior of the gas module parameters in the face of low-intensity arcing. Brief three-second arcs were generated over a period of five days from December 14 to December 18. Ten of these arcs were produced the first day, five arcs the second day, three arcs the next day, four arcs on the fourth day, and six three-second arcs on the last day.

The load and excitation for this period are shown in Figure 6.16. A reasonably consistent load cycle was maintained throughout this experiment. Rough guidelines as to the timing and magnitude of the load changes insured that the loading approximated a large daily load cycle. Random variation was introduced through arbitrary selection of loads from pre-determined load ranges and from flexibility in the schedule of load changes. For this experiment, this load profile was uninterrupted on the weekend.

The response of the Hydran is shown in Figure 6.17. The range of the sensor's output has been magnified to more clearly display the slight overall rise in measured combustible gas content of the oil. However, using only this data, it would be impossible to prove that the actual gas content of the oil has changed. This small rise could, in fact, be due to an increase in the temperature of the transformer—perhaps due to abnormal loading or a rise in ambient temperature. Even on those days when the average reading of the sensor is significantly raised, most of the day's readings are lower than the maximum reached on "normal" days.

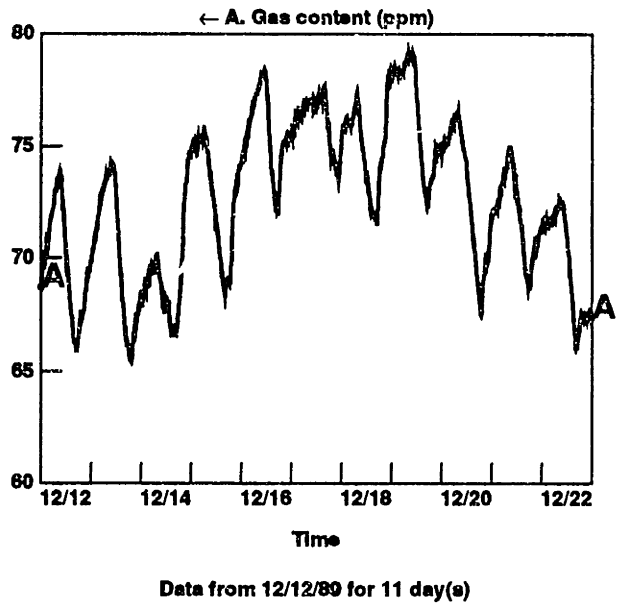


Figure 6.17: Hydran Response During Low-Intensity Arcing

Looking at the thermal data for this same period, in Figure 6.18, the possibility that the rise in the Hydran response is due solely to temperature is not ruled out. The top oil temperature on December 16–17 reaches a higher maximum and minimum than usual. (This was due to slightly abnormal loading.) Without a mathematical model, it is impossible for a transformer operator to recognize this gas content as anomalous with any degree of certainty.

Even armed with an adaptive model of the gas sensor's response, the arcing does not become immediately obvious. In this experiment, the gassing rate was so small that the uncorrelated trend due to arcing did not necessarily disrupt the estimation and acceptance of new parameters. Thus, new parameters were installed the morning of December 17 that caused the residual to once again return to near zero. Figure 6.19 is a plot of the residual.

Figure 6.19 shows that if the rate of gas generation is too slow then the residual will not necessarily reveal the changing condition of the transformer—residual analysis is targeted at faster processes. This is the very reason that parameter trend analysis is so important for transformer diagnosis. The changing values of the parameters capture the changing behavior of the transformer.

In order to evaluate how the parameters could best be evaluated, the parameters were estimated on a continuous basis from this experiment's data. This was done using a moving two-day-wide window of data that was updated for each new piece of data. The results for one of these parameters is given in Figure 6.20. The rise

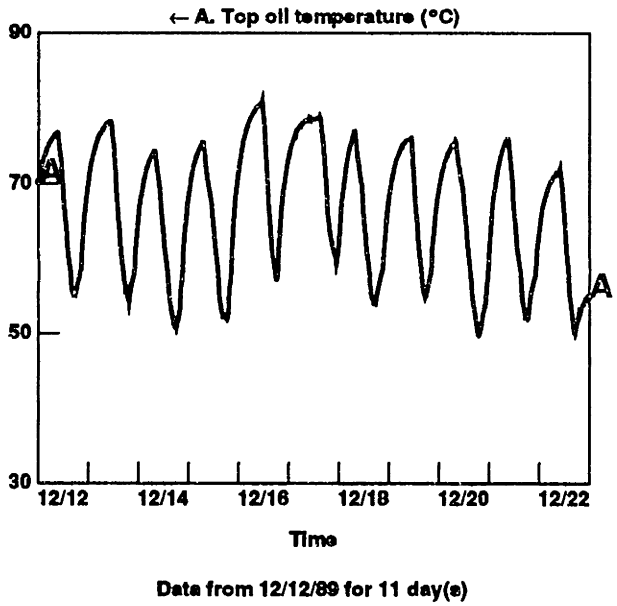


Figure 6.18: Thermal Cycling During Low-Intensity Arcing

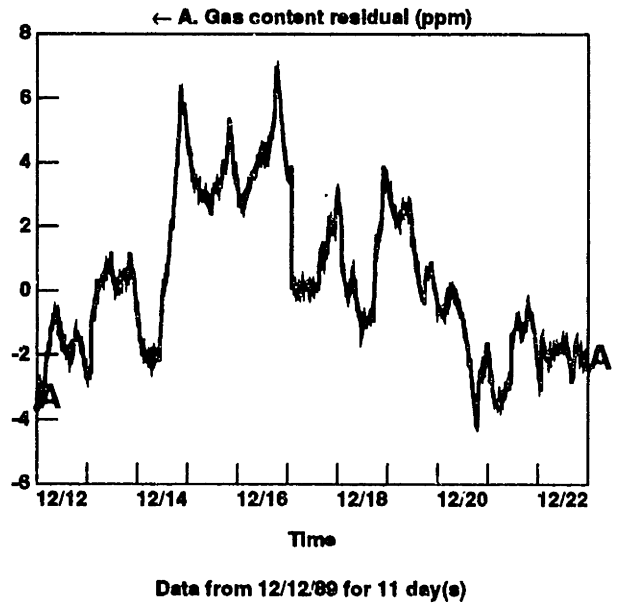


Figure 6.19: Gas Residual During Low-Intensity Arcing

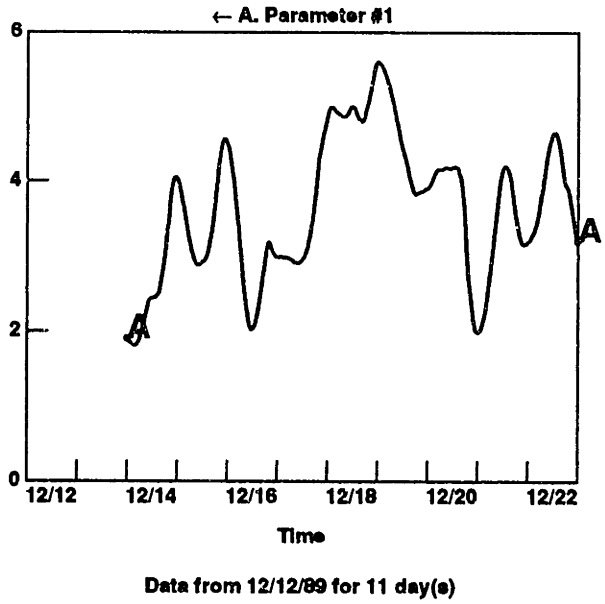


Figure 6.20: Adaptive Gas Content Model Continuous Parameter Response

in parameter value is on the same order of magnitude as the daily variation of the parameter. It is still difficult to see from this data how the gas content of the oil changed, and thus it is difficult to interpret what has occurred.

The next step was to investigate the behavior of the model in the model input/output space. Because the dissolved combustible gas model has a single input and a single output, the model space can be represented in a straightforward fashion in two-dimensions. Each model determined by the continuously estimated parameters of Figure 6.20 was mapped into the model space formed by the top oil temperature and the Hydran output. Displayed in quick succession, this images formed a motion picture that clearly demonstrated the changing behavior of the estimated model.

Figure 6.21 summarizes the changing behavior exhibited by this sequence. Each line spanning the displayed input (temperature) range represents a model estimated from a different two-day window, though most likely overlapping with the input window of another displayed model. The models are displayed as straight lines rather than the more accurate parabolic sections. For the sensor currently installed in the Pilot Facility and for the temperature range shown, a straight line is adequate for purposes of discussion.

The small black arrows denote how one estimated model transformed into the next. The large white arrow indicates the overall behavior of the model with passing time. Generally the model was rising, estimating higher Hydran readings for

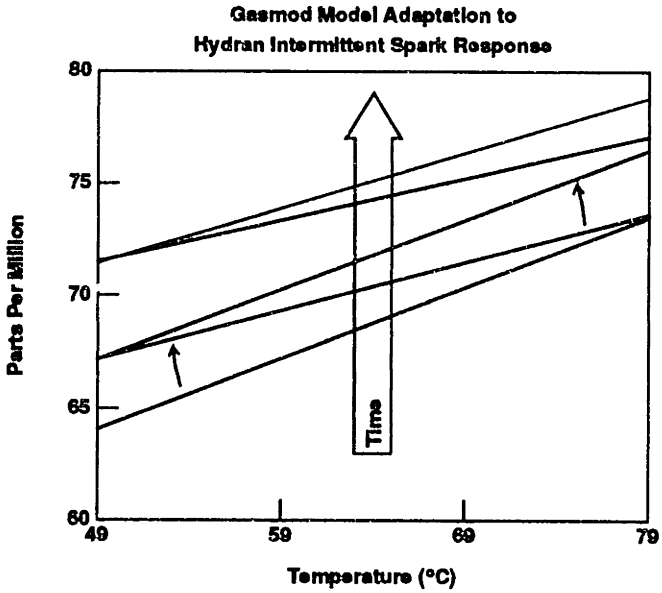


Figure 6.21: Gas Content Model Response in Model Space

the same temperature at a later time. But the model was being updated based on data taken at different times, seemingly changing the temperature dependence of the sensor. That is, when the most recent data was taken at a low temperature, the discrepancy between the readings at lower and higher temperatures was minimized. This is because, though a given gas content would yield a lower reading at a lower temperature, the actual gas content had risen while the temperature was falling. On the other hand, when the most recent data was taken at a higher temperature, the temperature dependence of the sensor seemed more severe, since both the temperature and the actual gas content had risen. While the temperature dependence of the sensor appeared to oscillate, the estimated parameters oscillated as well.

At the extremes of the temperature range, the estimated Hydran reading oscillated widely. But at the middle of the range of inputs, the model response was much more stable. Figure 6.22 shows the output of the estimated models simulated at 65°C. The behavior of the gas content of the oil can be seen very clearly. While the gassing continued, the gas content increased. When the arcing ceased, the equilibration of gas with the gas space began to dominate the behavior of the gas content. Using this thermally-normalized response, the estimated parameters can be more easily related to the actual gas content of the oil.

The technique of thermal normalization was tested on the parameters that had been accepted by the Pilot Transformer Monitoring System. This produced the data shown in Figure 6.23. For comparison, a composite of gas chromatography

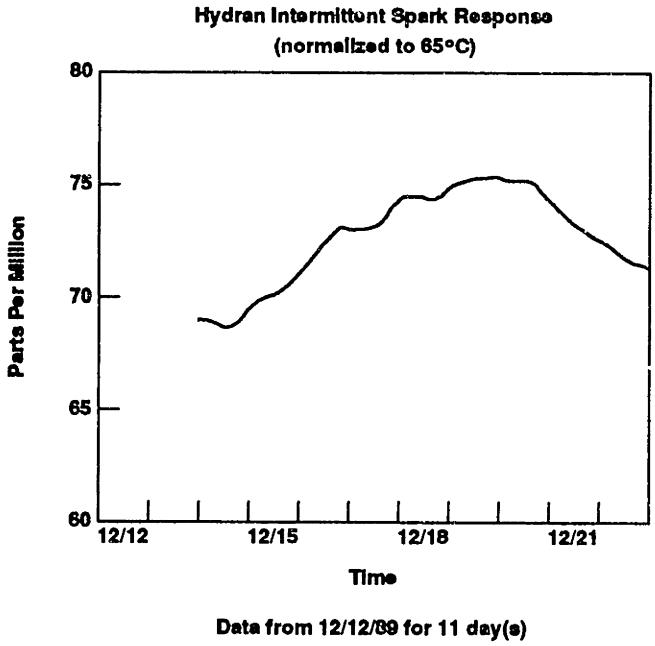


Figure 6.22: Thermally-Normalized Response to Low-Intensity Arc

results is also plotted. The composite value equals

$$H_2 + 0.15\ CO + 0.08\ C_2H_2 + 0.01\ C_2H_4$$

. It should be noted that each data point from the gas chromatograph represents several hours of technician time. In actual practice, such data is expensive and, thus, very infrequent. The normalized parameters, on the other hand, are estimated automatically from data which is continuously monitored; there is virtually no marginal cost associated with each data point.

Comparing the two sets of data, it can be seen that they agree reasonably well. The offset between the two sets had been previously recognized and accepted. Both data streams reveal the effect of the de-gas and the first two arcs. However, only the estimated parameters reveal the presence of a slow gassing event.

Figure 6.23 shows that the normalized parameters are much more robust in the face of inappropriately accepted parameters. The three sets of parameters accepted on Mondays before the criteria was tightened cause much smaller extraneous anomalies. This can be explained by considering that if the transformer is thermally stable at, for example, 70°C, then the parameters that result will not match the behavior of the sensor over the full temperature range, but will yield a reasonable estimate of the response at 65°C. This points up the danger of considering only the normalized parameters—doing so assumes that the temperature dependence of the sensor will not change.

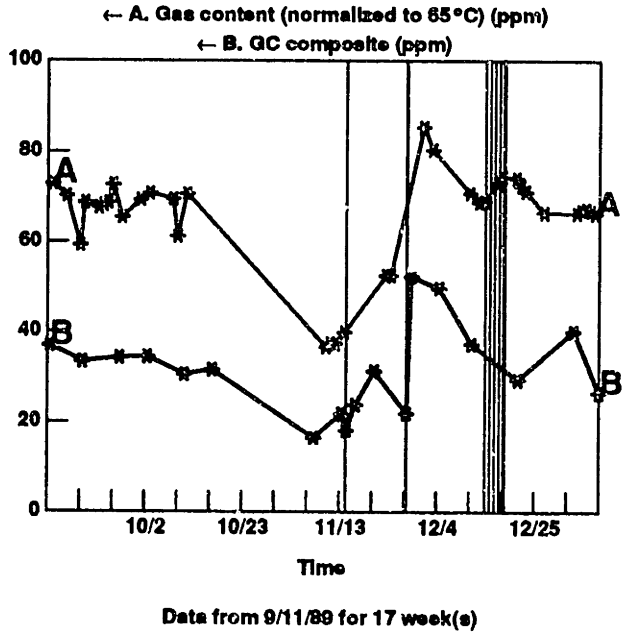


Figure 6.23: Trends in Thermally-Normalized Model and GC Readings

6.5 Summary

This chapter presented the results of several experiments designed to explore the capabilities of the proposed monitoring system. Included was a discussion of why this approach was necessary, as well as how the experiments were chosen. The experimental setup was then described. Finally, the results themselves were presented, along with numerical analyses to justify their physical interpretation.

Chapter 7

Proposed MIT Transformer Monitoring System: Advanced Diagnosis

In this chapter, a structure for an automated transformer diagnostic system will be developed from the results of the previous chapters. Section 7.1 itemizes the categories of knowledge which have been identified as important from both traditional monitoring techniques and the application of adaptive models of normal behavior. In Section 7.2, a detailed analysis of three of the experiments described previously will be presented. This analysis stresses how the diagnosis attained from an individual anomalous event would evolve under the influence of continuously arriving data. The next section, Section 7.3, discusses the common structure of much of the knowledge that is used in diagnosis and how this knowledge might be represented.

7.1 Categories of Knowledge

The transformer expert applies several different types of knowledge to the problem of incipient failure diagnosis. This knowledge can be divided into three broad classes:

- General transformer knowledge
- Site-specific knowledge about the particular transformer being monitored
- "Common sense" knowledge.

Common sense knowledge can not be used effectively by an automated system. For instance, a human expert may know, from knowledge of the location of the transformer, about the environmental conditions to which the transformer has likely been exposed during its lifetime. Soot, acid rain, local wildlife and even vandalism might be inferred from the site of the transformer. Including all of these factors in an automated diagnostic system would not be cost-effective.

Each of the first two classes, general and site-specific knowledge, can be divided into several categories. In this section, these categories will be itemized, with examples from the case studies.

Additional categories of knowledge are introduced that derive from the adaptive models of normal behavior that have been promoted during this research. These new types of knowledge can also be classified in terms of general and site-specific information.

7.1.1 General Knowledge about Transformers

Some categories of knowledge can be applied to transformer diagnosis independently of any detailed knowledge of the design of the particular transformer under consideration. Such categories include:

- Physical properties of materials
- Failure statistics
- Traditional monitoring
- Sensor behavior
- Costs

7.1.1.1 Physical Properties of Materials

Two physical properties which have been used in analyzing the experimental results of the last chapter were the coefficient of diffusion of hydrogen in nitrogen, and Oommen's distribution coefficient for hydrogen in equilibrium with transformer oil. These are basic properties of the materials in question.

7.1.1.2 Failure Statistics

Failure statistics are necessary if the transformer expert is to be able to evaluate the *a priori* likelihoods of various failures. Relative probabilities allow the expert to assess the likelihood of a given diagnosis. Some of the classic diagnostic methods have these failure statistics built in. The Rogers Ratio Method, for example, divides a four-dimensional ratio space into discrete subspaces, some of which correspond to particular diagnoses. But the process of determining how to partition the space is based on the statistical study of a large number of dissolved gas analyses.

7.1.1.3 Traditional Monitoring

The monitoring techniques discussed in Chapter 2 are not expected to be replaced by the new methods described in this thesis. Rather, the different methods can be used together to form a monitoring protocol that will greatly improve incipient failure detection and diagnosis in large power transformers, hence improving their reliability, while still remaining cost-effective.

7.1.1.4 Sensor Behavior/Limitations

Though seldom included in discussions of transformer diagnosis, the limitations of a particular type of sensor are an important consideration. For example, dissolved gas analysis is a key component of many utilities' monitoring efforts, but an abnormal gas chromatograph result is always suspect. In practice, an abnormal analysis must be confirmed with a second analysis before any action is undertaken[40,41]. This is not necessarily a vote of no confidence for the sensor itself, but for the entire sampling and analysis process. In fact, most false abnormal results can be traced to the sampling process.

7.1.1.5 Costs

The costs associated with any possible action affect the decisions made by the expert. However, like failure statistics, costs are seldom considered directly. Instead, they have been compiled into the normal response to a given situation. Only during a major decision, such as whether or not to shut down a transformer, would costs receive explicit attention. Specific costs might be site-specific—it would cost more to send a technician to a remote location to examine a transformer—but in general the cost structure will not change drastically from transformer to transformer.

7.1.2 Site-Specific Knowledge about a Particular Transformer

In addition to general knowledge about transformers, an expert requires specific information about the transformer being monitored. Some site-specific categories are:

- Design parameters
- Operational profile
- Maintenance history

7.1.2.1 Design Parameters

Design parameters include a wide range of characteristics that are chosen to meet the specifications of the utility. These characteristics include such factors as the use

of convective or forced flow for cooling, the presence or absence of a gas space, or full load rating. These characteristics are not always chosen explicitly, but may arise as a consequence of a number of related design decisions. For instance, the thermal time constant of the transformer is not the result of a single deliberate choice, but is derived from the design considerations that ensure a maximum temperature rise with the given cooling mechanism.

7.1.2.2 Operational Profile

The operational profile of a transformer is an important piece of information for diagnosis. It has been shown here that a detailed schedule of loading levels can be used to improve the engineer's understanding of the internal conditions of the transformer. Such detail is not always necessary. Correlations have been discovered between gassing events and the use of the tertiary winding, without reference to precise loading information. Knowledge that a transformer is being used with an arc furnace can modify which failure modes receive the most attention.

7.1.2.3 Maintenance History

The maintenance history of a transformer can help identify the types of failures to which the given transformer is most prone. The maintenance history of transformers with similar designs by the same manufacturer may be useful in the same way. In addition, this record may include known problems that either were not repaired or may not have been repaired properly.

7.1.3 Knowledge Derived from Adaptive Models of Normal Response

Applying adaptive models of normal transformer behavior to the task of transformer performance monitoring introduces new categories of knowledge that the traditional transformer expert does not use. These categories can also be grouped as general (applicable to many transformers) or site-specific (applicable only to the monitored transformer). Two important categories of general knowledge associated with adaptive models are:

- Parameter interpretation
- Residual interpretation

An important category of site-specific knowledge is:

- Estimation requirements

7.1.3.1 Residual Interpretation

Residual interpretation is a critical component for diagnosis. This is intimately related to the model structure associated with the residual. Consider the two thermal models presently implemented: the IEEE loading guide model and the MIT constrained flow model. Both predict essentially the same quantity—the temperature of the oil at the top of the transformer. Both recognize that this temperature is governed, at least in part, by the amount of heat being generated by the transformer, a quantity that may change during an incipient failure. However, the IEEE model is based on global measurements of the transformer, such as the temperature of the mixed top oil, while the MIT model is based on local measurements, such as the temperature of the oil leaving a particular duct.

Because of these differences, identical residual anomalies arising from these two modules would result in very different interpretations. The IEEE model is based on a number of different physical processes. Heat transfer between the winding and the oil, heat transfer between the oil and the tank, and heat transfer between the tank and the environment all play a role. An anomaly could represent a disturbance in any one of these processes. The MIT model focuses on the heat transfer between the oil and the winding, in a specific physical location. Thus, the set of failures to which one module is sensitive is not identical to that of the other. In addition, for the failures to which both are sensitive, identical magnitudes of anomaly may represent very different magnitudes of fault severity.

7.1.3.2 Parameter Interpretation

Identifying parameter trends is not sufficient for performing diagnosis. It is necessary to be able to interpret the parameters and their trends. For some models, estimated parameters can be equated to familiar physical quantities. For others, interpreting parameter trends may be dependent on developing empirical relationships between different types or magnitude of trends and particular failure modes.

Physical Parameters The thermal module based on the IEEE loading guide models, THIE3MOD, is one module whose parameters are amenable to physical interpretation. Equation (4.4) presented the equation for estimating these parameters as:

$$T_{topoil}[k] - T_{ambient}[k] = \theta_1 \times (T_{topoil}[k-1] - T_{ambient}[k-1]) + \theta_2 \times I_{low}[k]^{1.6} \quad (7.1)$$

A half of a year's worth of THIE3MOD parameters are presented in Figure 7.1. Note that curve A, labeled "IEEE thermal parameter #1" (θ_1 of Equation (7.1)), is plotted against the left-hand axis, as indicated by the left arrow ($\leftarrow$), while curve B (θ_2) is plotted against the right-hand axis ($\rightarrow$). Also, curve A is dimensionless,

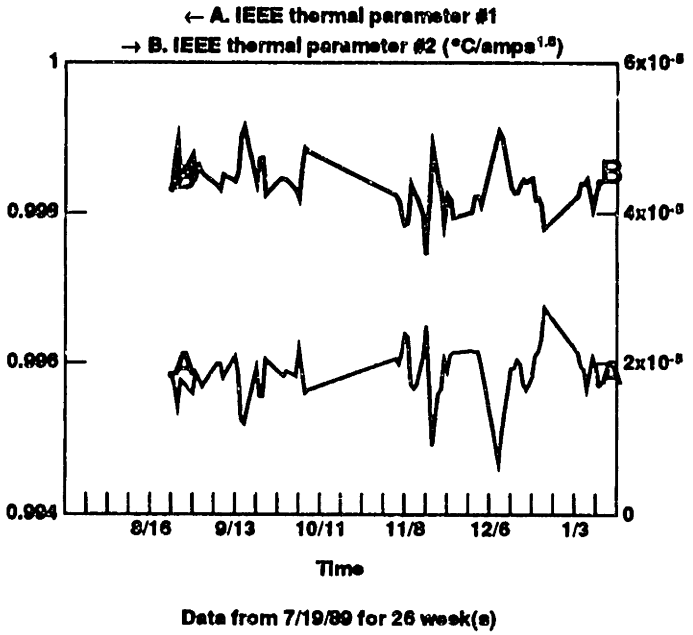


Figure 7.1: THIE3MOD Parameters

so no units are indicated, while curve B has dimensions of °C/amps$^{1.6}$, indicated in parentheses following the curve label.

It is difficult to interpret the parameters as presented in Figure 7.1. Both curves exhibit some noise—in fact, this noise appears to be highly correlated. As displayed, the signal-to-noise ratio of the two curve appears to be similar. Displaying curve A on a scale that includes the origin, however, shows the noise in the estimation of θ_1 to be almost non-existent; θ_1 is almost identical to unity. Which view is correct? (It will be shown below that it is the difference between θ_1 and unity that is the critical quantity.) Learning to interpret these parameters from experience would take a significant amount of time, and it is unclear how difficult it would be to apply the resulting interpretation to different transformer designs. Knowing the relationship of these quantities to physical parameters speeds the learning process and makes the end result more robust.

For purposes of parameter estimation, Equation (7.1) is used. However, the actual temperature rise equation is:[29]

$$T_{topoil}[k] - T_{ambient}[k] = \left(\frac{\mathrm{t}_o}{\tau + \mathrm{t}_o}\right) \times (T_{topoil}[k-1] - T_{ambient}[k-1]) + \left(\frac{\tau\, T_{fl}}{(\tau + \mathrm{t}_o)\, I_{rated}^{1.6}}\right) \times I_{low}[k]^{1.6} \quad (7.2)$$

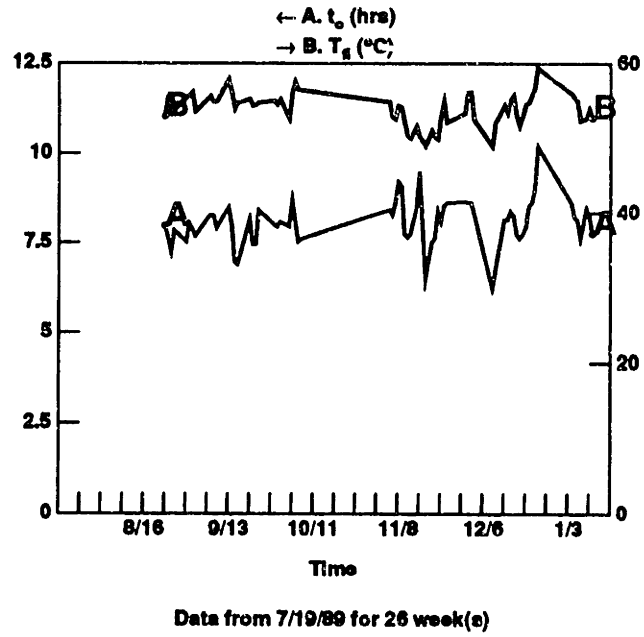


Figure 7.2: THIE3MOD Parameters in "Physical" Form

where t_o is the oil time constant[1], τ is the sampling interval, T_{fl} is the full-load top oil temperature rise above ambient temperature, and I_{rated} is the transformer's rated load. The parameters τ and I_{rated} are fixed parameters; for the Pilot Facility, τ is two minutes and I_{rated} is 208 amps. Solving for t_o and T_{fl} yields:

$$t_o = \frac{\theta_1 \tau}{1 - \theta_1} \tag{7.3}$$

$$T_{fl} = \frac{\theta_2 I_{rated}^{1.6}}{1 - \theta_1} \tag{7.4}$$

The estimated parameters of Figure 7.1 were manipulated according to these relations and the estimates of these physical parameters are shown in Figure 7.2. Note that the full-load top oil temperature rise over ambient temperature, T_{fl}, is plotted against the right-hand dependent axis.

The data shown in Figure 7.2 indicates that the estimates of θ_1 and θ_2 that are being generated by the Pilot Facility's monitoring system correspond to relatively stable estimates of t_o and T_{fl}. Experience is still necessary to tell if the small changes in these estimates represent significant trends, or if they can be linked to changes in the load profile. (Beware of parameters accepted on a Monday!) But

[1]This notation conflicts with that of [29], where t_o is written T_o and T_{fl} is written θ_{fl}. In this thesis, θ is reserved for parameters and T is reserved for temperatures and temperature rises.

this experience may already exist, and it is reasonable to suppose that it would apply across a broad range of transformer designs.

Composite Physical Parameters Even when a model is physically-based, it is not always possible to isolate a particular physical parameter from the models parameters. Equation (4.13), the equation used to estimate THMOD's parameters, is:

$$\begin{aligned}(T_{topduct}[k] + T_{botduct}[k]) - (T_{topduct}[k-1] + T_{botduct}[k-1]) = \\ \theta_1 \times (T_{topduct}[k] - T_{botduct}[k]) + \\ \theta_2 \times I_{low}[k]^2\end{aligned} \tag{7.5}$$

Archer's explanation of the physical basis for this model[29] reveals that

$$\theta_1 = \frac{-2Q_d\tau}{V_d} \tag{7.6}$$

$$\theta_2 = \frac{2g\tau}{\rho c_p V_d} \tag{7.7}$$

where Q_d is the volume rate of flow through the duct, τ is the sampling interval, V_d is the duct volume, g is a proportionality constant that reflects the fraction of the total winding losses which is associated with the duct, ρ is the density of the oil, and c_p is the specific heat of oil.

Focusing on θ_2, it can be seen that it would be difficult to isolate any particular physical parameter. The parameter τ is known, and c_p probably would not change significantly, even in response to fouling of the transformer oil by the byproducts of oil and cellulose degradation. The oil density, ρ, is temperature-dependent, but the oil temperatures around the duct are known, and perhaps ρ could be approximated. This leaves g and V_d as two unknown quantities. V_d might be measured when the transformer is commissioned, but it could change significantly over time—particularly in response to a through-fault.

So, assuming that c_p will remain constant and ρ can be approximated, one is left with the relation

$$\theta_2 \propto \frac{g}{V_d} \tag{7.8}$$

It does not seem possible at this time to isolate g or V_d. However, that is not required to support physical interpretation of the parameter behavior. For instance, an increase in θ_2 could represent either an increase in g or a decrease in V_d. Evidence from other sources may be used to further refine this analysis. If V_d changed enough to be detected by a change in θ_2 of THMOD, this event should also be detected by the vibration module.

Black-Box Parameters Black-box parameters are not related to any set of physical parameters in a known way. Interpretation of these parameters will rely heavily on future experience. However, this is not to say that these parameters will not be immediately useful for diagnosis. The situation described above is a good example. VIBMOD's model of the vibration response of the winding is a black-box model, but the model's parameters are stable. If the parameters do not change much then this is strong evidence that there has been no coil movement. In the future, it may be possible to interpret what a change in parameters actually means. But for now, whether or not the parameters change is important diagnostic information.

7.1.3.3 Estimation Requirements

Recognizing the estimation requirements of a module is necessary for determining when a failure to estimate acceptable parameters is indicative of an incipient transformer failure and when it is not. A transformer will not always exercise its full range of behavior during each estimation interval. This is why the bin-storage implementation of parameter estimation of Chapter 4 was developed. Most of the existing models are driven by temperature, which is in turn driven by load. Thus, when the transformer goes through its normal daily load cycle, it is expected that the transformer will exhibit its normal range of behavior, leading to good parameters.

Failure to estimate acceptable parameters, then, may represent a significant piece of diagnostic information. It could even be used for detection, acting as a trigger for diagnosis. A failure to estimate parameters can usually be traced to a underlying cause: an increase in noise in one of the measured signals or a trend in the targeted output that is uncorrelated to the model's inputs, for instance. These underlying anomalies might be identified more quickly through residual analysis, but if not, reacting to unexpected failures in estimation may improve the response time of the monitoring system to incipient failures.

7.1.4 Summary of Categories of Knowledge

By limiting the domain of the diagnostic system to the types of knowledge discussed above, the complication of having to model common sense is eliminated. In the resultant limited domain, the categories of knowledge can be grouped as general or site-specific. In applying this monitoring system to a particular transformer, the general knowledge that has been compiled over time can be applied unchanged. The site-specific knowledge, on the other hand, has to be either supplied during the installation procedure or learned by the system. Fortunately, most of this information would be available at the time of installation.

7.2 Diagnosis Process

In this section, the results of the experiments described previously will be reviewed, with a focus on how a diagnosis would be developed. Three experiments will be considered: the thermally-induced sensor malfunction (the preliminary experiment), the residual response experiment (the 12-, 60-, and 72-second arcs), and the parameter trend analysis experiment (the intermittent 3-second arcs).

7.2.1 Thermally-Induced Sensor Malfunction

In this experiment, a heating tape was placed around the combustible gas sensor to disrupt its thermal compensation. The reader will remember, from Chapter 5, that the resulting anomaly in the gas residual was correlated in time with a small rise in the constrained flow thermal residual, which was uncorrelated in time with any change in load. In a system of automated diagnosis based on the MIT approach to transformer monitoring, the sequence of events and decisions might have occurred as follows:

0:00 **A combustible gas anomaly is detected.** Both the level and rate-of-change thresholds for the gas residual are violated. Figure 5.2, showing the residual, is reprinted in Figure 7.3. At this point, the whole range of gas-generating failure modes become viable diagnoses. Sensor malfunction is included in the list of possible diagnoses.

When an anomaly occurs in one measurement, residual, or parameter stream, the diagnostic system should check for correlated events in other streams. The presence or absence of correlated events can be used to refine the diagnosis. Because of the time lags introduced by the physical constants of the transformer, these correlated events may include events that occurred before the anomaly was detected.

Load information is one stream that is always of interest. Large increases in load can drastically change the dielectric stress throughout the transformer, possibly acting as a trigger for any one of a number of failure modes. Large changes in load also set up thermal transients in the transformer; this might increase mechanical stress as components expand or contract, or might disturb oil circulation patterns enough to change the immediate source of oil reaching the various sensors.

Looking back at recent load data reveals that the load had risen to its present level from no load only 25 minutes previously. Looking back a little further reveals a correlated drop in load just a few minutes before that. While there is perhaps a small chance (when examining the load data in isolation) that these correlated changes in load could be due to an intermittent sensor malfunction, any such chance becomes vanishingly small when the excitation data

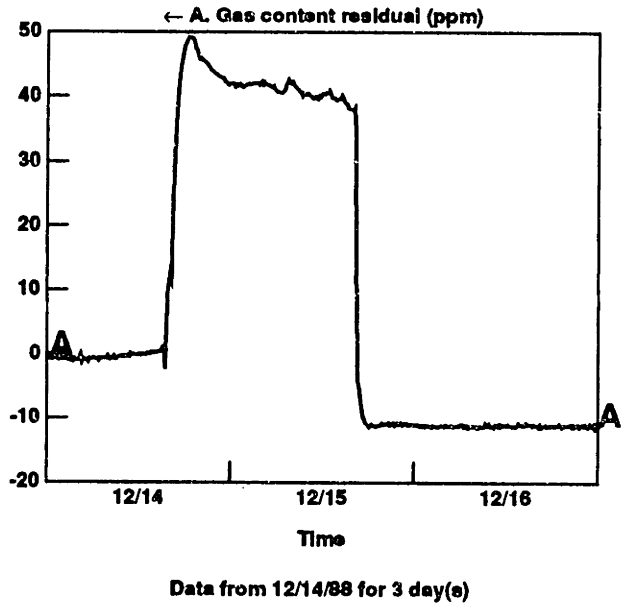


Figure 7.3: Thermally-Induced Gas Residual Anomaly

is considered. A brief drop to no excitation is exactly correlated in time with the brief drop to no load. This is the signature of a maintenance shutdown for the Pilot Facility. Whenever a researcher enters the transformer bay, whether to take an oil sample or adjust a sensor, the high voltage is turned off. It is reasonable to expect that a commercial transformer monitoring system would receive independent confirmation that a scheduled shutdown had occurred.

The fact that a shutdown occurred is relatively neutral evidence, neither strongly supporting nor discrediting any particular diagnosis. On the one hand, there may be some slight support for sensor failure, as some maintenance operations can disturb the leads between the data acquisition system and the sensors themselves. Usually associated with this type of disturbance is an increase in the noise content of the sampled signal. A measure of noise content is not immediately available, but will require further measurements. On the other hand, the fact that the load was only returned to the load level that existed before the shutdown does not mean that the load increase could not have been a contributing factor to a gassing failure. The thermal, mechanical, and dielectric transients induced by the shutdown may have pushed a degraded insulation system over the edge. These two considerations pretty much cancel each other out, with sensor malfunction perhaps having a slight advantage.

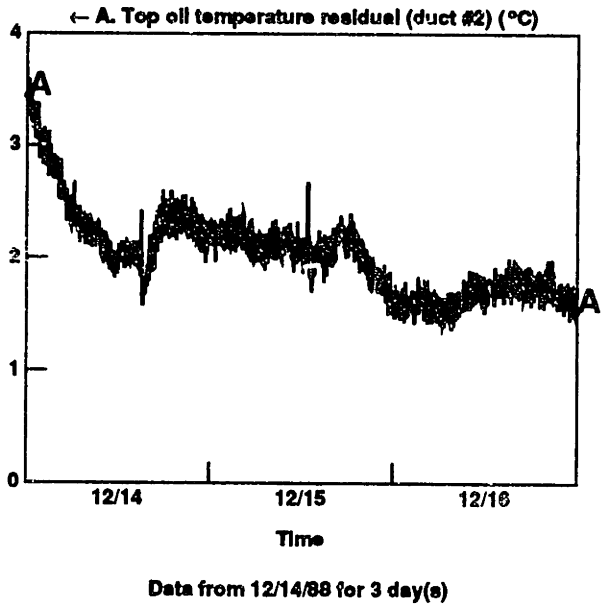


Figure 7.4: Thermal Residual for Thermally-Induced Gas Anomaly

It might be helpful to differentiate between different maintenance operations. Some operations, such as oil sampling, are commonplace; mistakes that affect the performance of the transformer or any of its sensors are unlikely. Other operations, such as adjusting a sensor, are less common; mistakes may be more likely—and these mistakes would probably affect the operation of the given sensor. Knowing what operation was performed may help in determining if a sensor malfunction is a good diagnosis. Querying the transformer operator when this information could be helpful is one option; specifying the maintenance operation as part of the shutdown procedure is another. The latter solution would be preferred in order to limit the amount of human interaction needed during an anomaly.

There are no anomalous residuals besides the gas residual at this time. The thermal residuals showed small "blips" correlated with the the maintenance shutdown. However, this is not unusual, given the large change in load associated with a shutdown. One of the constrained-flow-model top oil temperature residuals is seen in Figure 7.4 (originally shown in Figure 5.3).

The fact that the thermal residuals and thermal parameters were normal before this gas content anomaly occurred would tend to rule out general overheating. Given more experience, it is possible that localized overheating could also be eliminated from consideration; it is not yet possible to say whether localized overheating (due to circulating currents, for example) could cause

such significant gassing without first affecting the thermal residuals.

A continuous moisture residual is not being generated at this time, but its response to this experiment might have been very informative. Of course, the extraneous external heating that actually caused the anomaly would not affect the moisture content of the oil, but this lack of response would have indicated that the failure did not include degradation of the cellulose. In a true gas-generating failure that involves cellulose, gases and moisture would be evolved at related rates. Immediately following the failure, the status of the oil should be an accurate indicator of the amounts of moisture and combustible gases produced.

The gas residual is the only anomalous data stream. Since the other residuals do not exhibit anomalous behavior at this time, it is possible to eliminate some classes of sensor malfunction. This is because there are shared elements in the data acquisition system that would affect more than one measurement were one of them to fail. For example, if an analog-to-digital converter were to fail, any measurement making use of that converter would be affected, possibly in a highly correlated—and thus very revealing—way.

The initial diagnosis is arcing, partial discharge, or sensor malfunction. There is little initial evidence to distinguish between these possibilities. In fact, localized overheating and even general overheating are not eliminated as candidates, merely evaluated as less likely.

At this point, there does not seem to be any immediate danger. If the gas content residual continues to rise at the same rate, there is some time before a dangerous amount of gas will be accumulated. The recommended course of action is to wait and see how the situation develops. Depending on the location of the transformer and the availability of technicians, among other factors, a dissolved gas analysis might be ordered. Failure statistics may enter into the decision; how often does such a situation as this evolve into a catastrophic failure?

0:00–3:00 **The gas residual continues to grow. The thermal residual grows as well.** During the first three hours following the detected anomaly, the gas residual continues to grow at the same rate. From limited experience, this seems to rule out one type of sensor malfunction associated with a faulty connection between the sensor and the data acquisition system. This failure does manifest as a drift in the measured output, but it is unlikely that this rate of drift would stay constant over such an extended period. This type of sensor failure is usually accompanied by an increase in the amount of noise in the measurement.

The thermal residual also increases. This increase was not considered significant for the first half hour or so, as the residual may still be responding to

disturbances in the convective flow due to the maintenance shutdown. However, it soon became apparent that this was an actual rise in the residual, especially in contrast to the decay of the residual that preceded the observed gas anomaly. (This decay was in response to a load change on the previous day; it is an accepted modeling error associated with this particular model structure.) It should be stressed that this thermal residual is not violating any level or rate-of-change detection thresholds. Its behavior is only identified as anomalous because of its correlation in time with an anomaly that *was* detected—the gas anomaly.

Looking back on the beginnings of the thermal anomaly and correlating with the gas anomaly, it can be seen that the two began at essentially the same time. It is possible that this thermal anomaiy and the gas anomaly are unrelated, but logic suggests that this is extremely unlikely. Chances are that the two anomalies have the same root cause and, considering the close coincidence of the onset of each, they most likely have tne same immediate cause.

Assuming that the gas and thermal anomalies have the same immediate cause and assuming that gas anomaly is not due to a sensor malfunction, it may be possible to triangulate the position of the failure. An observed rate of gas generation implies an approximate rate of energy dissipation, and thus fault temperature. Knowing the fault temperature, one might get a rough estimate of how far from the sensor the fault must be to produce the observed anomaly. The Pilot Facility has three instrumented ducts; a monitoring system for a full-sized power transformer might have many more. As more sensors are added, the additional information may narrow the possible location of a detected failure to a very small area. This might require a relatively detailed knowledge of the temperature distribution in the transformer.

If, on the other hand, a thermally-induced sensor malfunction *is* assumed, then triangulation may still be possible. Given the response of the Hydran sensor, a hypothesized failure location implies a particular fault temperature and thus a temperature distribution. This distribution must agree with the observed thermal residuals. If there is no location that is consistent with the observed behavior, then a diagnosis of two unrelated failure modes, or of some type of sensor malfunction that is not thermally-induced, may gain credence.

3:00 **The gas and thermal residuals peak.** It becomes possible (after a few sensor readings) to evaluate the noise component of the gas residual which, it turns out, has not changed appreciably. This, along with the fact that there is a definite correlated event in the thermal residual, make a diagnosis of a loose-connection sensor malfunction extremely improbable.

Reviewing the load data reveals that the load has remained constant throughout the observed anomaly. The fact that the gas residual reached a maximum

independent of any change in load might be used to argue against a diagnosis of overheating. True, some causes of overheating could be come and go independent of load; solar magnetic disturbances and the loss and later repair of forced cooling pumps are two examples. However, overheating is often caused by the degradation of the solid insulation and movement of the coils, two situations that can not spontaneously repair themselves. At the risk of being repetitious, only experience can reveal the relative probabilities of the two sets of events.

3:20 **The gas residual begins to decay abnormally.** Examining Figure 7.3, it can be seen that the gas residual decays from a value near 50 ppm to less than 40 ppm. The residual does not exhibit the "overshoot" seen in the other 50 ppm rise (after the 360-sec arc). In addition, the transition between a rising residual and a decaying one is not as abrupt as seen in the other experiments. Neither of this points is very telling, as the observed residual response when the gas generation stops is most likely dependent on the nature and location of the fault. What is telling, however, is the time constant of the decay, which does not match those observed in the other experiments.

23:30 **The gas residual drops sharply.** There is no known transformer failure mode that could account for such a rapid drop in the measured dissolved gas content of the oil, apart from a sensor malfunction. Even exposed to a hard vacuum, the oil could not release its dissolved gas so quickly.

It is a virtual certainty that the drop in the gas residual is due to a sensor malfunction. There are two possibilities: the previously observed anomaly was a manifestation of the same sensor malfunction, or the two events were unrelated. Since there is an explanation that can account for all the observed effects, this becomes the primary diagnosis. This diagnosis is, of course, extraneous heating near the Hydran.

This final diagnosis had an extremely small *a priori* probability. How often will a Hydran be exposed to such a large external heat source? It is not unreasonable to believe that this diagnosis might not even be included in the set of possible diagnoses for an automated diagnostic system. Because of the sharp drop in the residual, sensor malfunction would probably still be the prime candidate. But, the presence of a confirmed, correlated thermal anomaly would likely lower the certainty with which the diagnosis is put forward. So much so, in fact, that a diagnosis of a gassing failure with an unrelated sensor malfunction may become the leading candidate.

25:00–28:00 **Gas residual remains stable and the thermal residual drops.** The amount that the thermal residual has dropped during this period is approximately equal to the amount that it had risen previously. The rate at which it is falling is also comparable to the the rate of rise.

The stability of the gas residual following its rapid fall supports the hypothesis that no gassing has actually occurred. Also, the start of the thermal anomaly was correlated with the start of the gas anomaly, and the end of the thermal anomaly is correlated with an almost certain dissolved gas sensor malfunction. This tends to confirm that a single sensor malfunction was responsible for the entire sequence of events. If the possibility of a thermally-induced sensor malfunction were considered, it would be accepted as the most likely diagnosis.

7.2.2 Residual Response Experiment

Three arcs were generated in this experiment. The first arc was 12 sec in duration, the second was 60 sec, and the third was 72 sec. Gas generation only became apparent after the second arc, with a rise in the gas residual of about 10 ppm. The effects of the third arc were also readily apparent, with a rise of another 10 ppm. Later, there was a sensor malfunction which exhibited behavior that was superficially similar to the effects of an arc. During this sequence of events, diagnosis might have proceeded past the following milestones:

0:00 **A combustible gas residual anomaly is detected.** As in the last example, an anomaly is detected in the gas residual using level and rate-of-change thresholds. Figure 6.10, showing this residual, is displayed again in Figure 7.5. (From knowledge unavailable to the diagnostic process, it is known that the anomaly was detected approximately 20 min after the second arc was generated.) Arcing, partial discharges, localized overheating, and sensor malfunction again topped the list of candidate diagnoses.

A maintenance shutdown occurred shortly before the anomaly. This might be interpreted as slight support for sensor malfunction for the reasons discussed in the previous example. Load information from this period is displayed in Figure 7.6, as originally displayed in Figure 6.8.

No other residual is exhibiting anomalous behavior. None of the thermal or vibration residuals is behaving in an obviously anomalous way. Again, some types of sensor malfunction can be eliminated from consideration, as explained in the previous example.

The recent history of the gas residual reveals a possible precursor to the observed anomaly. In Chapter 6, it was mentioned that the first arc in this series was followed by a very small rise in the gas residual. Even knowing when in the residual stream to look for this event, it is unclear whether this slight rise is significant. Without outside knowledge as to when the first arc occurred, this tiny anomaly might not even be detected.

Nonetheless, this example illustrates that there may be events that, in the normal course of detection, are overlooked but may loom very large during diagnosis. It may be possible to compare characteristics of the original observed

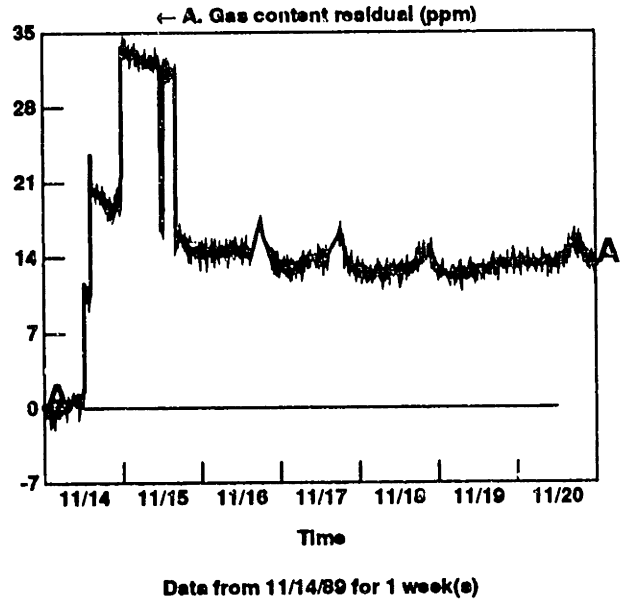


Figure 7.5: Low-Intensity Arcing GASMOD Residual

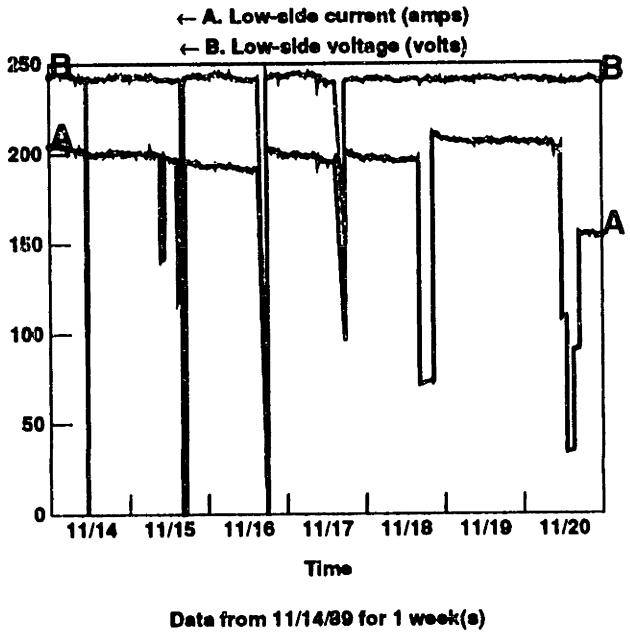


Figure 7.6: Low-Intensity Arcing GASMOD Residual

anomaly and its precursor to reveal the evolution of the incipient failure. In this example, the difference in magnitudes of the step rises may be significant. Or perhaps the interval between the two events could be important. Of course, little meaning can be read into this particular example because both the magnitude and interval were chosen by humans, unconstrained to simulating a physically realizable failure evolution.

2:00 **Another gassing event is detected.** The rate-of-change threshold for the gas residual is again violated a couple of hours after the first observed anomaly. The result is another 10 ppm rise beyond the first rise. The measurement level threshold is, of course, violated as well, but it has been so since the initial detected anomaly.

Comparisons between the two 10-ppm anomalies reveal that both occurred abruptly, making the transition from one level to the other in two sampling intervals (20 minutes). The second arc had a greater "overshoot"; the difference between the peak value and the next sample is significant, but a slower rate of decay is then established. It is surmised that this overshoot is due to incomplete mixing of the gas in the oil. If this is the case, the overshoot of the previous anomaly may not have been as evident because the sensor output may have been sampled farther from the peak value. A faster sampling rate during a failure might help locate the fault; the longer the oil circulation path between the fault and the sensor, the less severe the overshoot. Additional combustible gas sensors and a detailed knowledge of the circulation paths in the transformer might be used to triangulate the location of the failure. At this point, this is merely supposition and would need to be confirmed through experimentation. The marginal value of this information would also have to be considered.

The other residuals remained normal. By this time in the previous example, the existence of a thermal anomaly had been revealed. While this is strong evidence against a thermally-induced sensor malfunction, it is does not unequivocally eliminate that diagnosis. The smaller size of the observed gas residual anomalies (compared to the previous example) means that a less significant heat source would be needed to disrupt the sensor. The smaller heat source might not be enough to generate a thermal anomaly. The reader is reminded that the model which generates the thermal residual is crafted to reveal the thermal characteristics of the winding; its ability to highlight thermal anomalies away from the winding is limited.

2:30 **The combustible gas residual begins to decay normally.** As discussed in Chapter 6, the idea of a "normal" decay is based on the fact that the Pilot Facility's test transformer has a gas space and that the gas space is connected to an expansion tank. In many modern transformers, the oil is not

in contact with a gas space and thus no equilibration will take place.[2] In those transformers, no decay of the residual would be expected.

11:30 **Still another gas residual anomaly is detected.** (This anomaly is not the result of an arc simulation. Thus, this is a true test of the diagnostic capabilities of the transformer monitoring system.) The rate-of-change threshold has again revealed the event. This event has the same basic characteristics as the previous two anomalies: it is nearly a step change, and the magnitude is approximately the same. There is no overshoot, but the initial event showed that this is not necessary. In contrast to the previous events, the onset of this anomaly is not as abrupt; this would most likely be considered insignificant in light of the basic similarities among the three major anomalies.

The diagnosis at this point would have been three gas-generating events of roughly equal severity. With more experience, the gas-generating events could perhaps be identified as arcs.

11:50 **The gas residual again begins to decay.** As this is expected, it did not change the previous diagnosis.

22:20 **There is a sharp drop in the gas residual.** This throws doubt on all previous anomalous gas data. There is almost certainly a sensor malfunction involved. On the other hand, the expected decay in the gas residual *was* observed following the anomalies. So, either the transformer *and* the sensor are malfunctioning—an unlikely event[3]—or the sensor is malfunctioning in a manner that simulates a gas generating event in a transformer with a gas space—also an unlikely event. Based on limited experience, the former seems less unlikely.

Thus, a diagnosis of two unrelated failures becames the prime candidate by default. Assuming this to be true, when did the sensor begin to malfunction? Experience with the Pilot Facility's data acquisition system suggests that a sensor can acquire an incorrect offset. The relative behavior of the sensor remains essentially correct but the absolute readings are off. The negative step change in the gas content measurement was either this offset being established or corrected. The magnitude of this change exactly corresponds to the positive step change of the latest anomaly. This, in conjunction with the slightly different character of that anomaly, makes it most likely that there were only two gas-generating events—though there is a significant chance that the drop in the gas measurement is the first manifestation of the sensor malfunction.

[2]There may be some adsorption of the gas by the solid insulation, though this does not seem to have been a major effect with the Pilot Facility.

[3]Or not so unlikely. A lightning strike could damage both the transformer and the sensor, but the damage would likely manifest more closely in time. The possible behavior of a sensor subject to such abuse is not known at this time.

23:30 **The gas residual returns to its previous level.** This does not greatly change the previous analysis, but merely confirms that the sensor is switching between two states. From past experience, it can be assumed that one of those states represents correct operation. The exact correspondence of the magnitude of this gas residual transition with the two previous transitions make them all suspect. The diagnosis: two gas-generating failures (arcs) and a sensor malfunction.

1:02:40 **The gas residual again drops rapidly.** The magnitude of this drop again corresponds to the latest transitions. This makes the previous analysis all the more certain.

7.2.3 Parameter Trend Analysis Experiment

This experiment was designed to simulate a small incipient failure, generating combustible gases at a very slow rate. While this experiment exercised the gas module, GASMOD, it is illustrative of how adaptive parameters from any of the proposed model structures might be useful. An automated diagnosis process might conform to the following scenario:

0:00 **A gas parameter anomaly is detected.** The thermally-normalized parameters that were developed in Chapter 6 are shown in Figure 7.7 (originally displayed in Figure 6.23). (The vertical lines indicate days on which arcing was simulated.)

On the morning of December 17, parameters were estimated that did not fit with the developing trend. At this time, the transformer was still reacting to the major disturbance injected into the gas/oil system on November 27—the 360-sec arc. The normalized parameter was falling and seemed to be approaching equilibrium. The normalized parameter of December 17 represents a significant deviation from this trend. Perhaps equally significant is the fact that parameters had not been successfully estimated for several days previous to this, despite sufficient thermal cycling.

Recent gas residual data indicates that gas generation had likely started three days previously. Figure 7.8 shows the residual from the period in question (from Figure 6.19). The residual rose fairly sharply on December 14, though not so sharply or so long as to be detected as an abnormality. This rise appeared to cease late that day. The residual did not appreciably rise nor decay for the next two days, but merely went through its normal daily cycle (due to normal small modeling errors).

The rise on December 14, uncorrelated as it was to temperature, explains why parameters were not successfully estimated on December 15 and 16. But at this point in the sequence, it was unclear whether this rise represented an

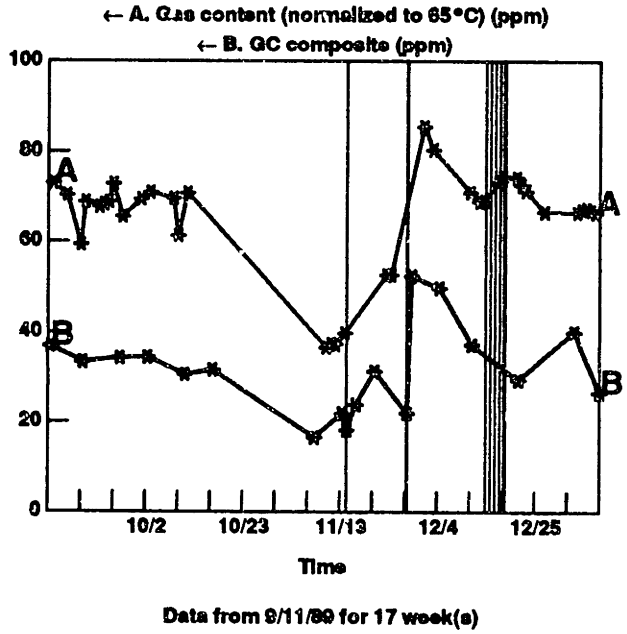


Figure 7.7: Thermally-Normalized Gas Parameters

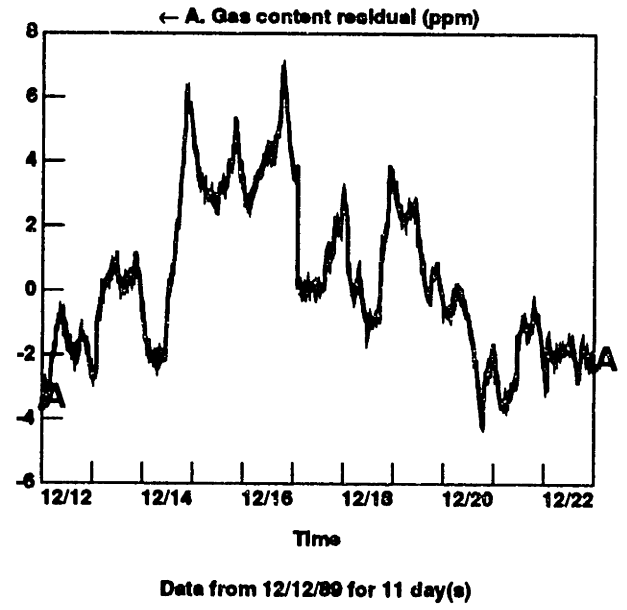


Figure 7.8: Gas Residual During Low-Intensity Arcing

actual gas-generating event or it was another sensor malfunction. On the one hand, the rise was not as quickly achieved as previous sensor malfunctions but, on the other hand, the expected decay did not manifest.

The diagnosis based on data gathered up to this point almost equally supports a sensor malfunction or a slight slow gas generation event. Even if the previous sensor malfunction (identified in Section 7.2.2) were included as a piece of data, the results would still be inconclusive. True, the sensor now has a history of malfunction, but some basic characteristics of the alleged malfunction are different—namely, the rate of drift and the eventual magnitude of the offset.

In either case, the situation is not serious. The total amount of gas generated is minimal, if in fact gas has been generated. And if the sensor is malfunctioning, it is not far off and is still tracking the behavior of the gas content of the oil in the normal manner.

1:00:00 **New parameters confirm the anomaly.** The GASMOD parameters of December 18 confirm the presence of the anomaly detected the previous day. The thermally-normalized parameter reveals that these new parameters represent a slightly *higher* gas content than before. This implies that gassing continued beyond that noted from December 14. Apparently, gas generation continued, but at a slower rate.

These inferences were by no means certain; the increase in the thermally-normalized parameter is extremely small. It is possible that the increase can be ascribed to noise in the estimation process, or to incompletely modeled effects such as changes in ambient temperature. Further experience should reveal at what level such increases become significant.

4:00:00–18:00:00 **Additional parameters describe a response to gas generation.** As can be seen in Figure 7.7, the parameters from December 21 on trace out the expected response of a gas-space transformer to a gas-generating fault. This confirms that diagnosis.

But the fact that the gassing stopped and that the equilibration process then could be discerned was not required for diagnosis. If the gassing had not stopped, the parameters would have continued to rise to reflect the increasing gas content. This is not dependent on the rate of gas generation. The rate of parameter change would then reflect the rate of gas generation and thus the severity of the fault. This experiment exercised this process in the span of a few days, but incipient failures taking weeks or months to develop could be handled equally well.

In this experiment, it was possible to identify changing rates of gas generation from one day to the next. If the gas generation is slower, it may be necessary to watch trends for many days before concluding that the rate of generation has

changed. But by monitoring these changing rates it may be possible to gauge the changing severity of the fault, or even to identify operating conditions which aggravate the fault.

7.3 Knowledge Representation

The examples of the previous section illustrate the thought processes that take place in the mind of a transformer expert. The next goal is to represent the knowledge embedded in the decisions described above in a way that an automated system can use. Over the last decade, a number of computer programs, called expert systems, have been developed that are designed to solve knowledge-based problems—problems that are knowledge-intensive, rather than data-intensive.

Many of these expert systems accomplish their goals by maintaining a distinction between the knowledge necessary to solve the problem and the way in which that knowledge is used. Such programs can be thought of as consisting of two pieces: the knowledge base and the inference engine. The knowledge base might record, for example, that A and B together imply C; the inference engine would then be capable of deducing that C is true, given that A and B are both true. Commercial expert system shells exist wherein working inference engines are made available with an empty knowledge base. The user then employs his knowledge of the problem domain to construct a suitable knowledge base.

There are many different expert system shells with many different features. But there are two major types of expert systems: rule-based and frame-based. For this discussion, a rule-based system will be assumed. In a rule-based system, the knowledge base is encoded as a series of *if-then*, or *production, rules*; the example above would be encoded as

IF A AND
B
THEN C

Expert systems almost always have some way of handling uncertain information. In the previous example, A is not always known for certain and A and B being true might not always mean that C is true. Some systems handle this by allowing assertions to be accompanied with a certainty factor. The situation becomes more complicated when two or more certainty factors must be combined, particularly when the assertions conflict. Qualitatively, if D implies F is "likely", and E implies F is "likely", then if D and E are both true, F should be more than just "likely"—perhaps "very likely". On the other hand, if G implies that F is "unlikely" and G is also true, just how certain is F? As one may imagine, tailoring a system such as this to mimic the behavior of an expert is more an art than a science.

In the following discussion, the issue of uncertain knowledge will be glossed over to some extent. The focus of the discussion will be on how to represent the type

of knowledge employed in Section 7.2. The nontrivial problem of how to effectively combine the separate bits of expertise will be left to others. Certainty factors are not the only mechanism for dealing with uncertainty; one method that is gathering a following is the use of *fuzzy sets* to model uncertainty. Even certainty factors come in different flavors; some are combined as if they were probabilities, using Bayes' Theorem, others measure positive and negative support separately. Below, qualitative terms such as "likely" and "almost certain" will be employed.

If a clause in the "if" portion of a rule is followed by a certainty factor, then the clause must be asserted with that level of certainty before the rule will be triggered. If a clause in the "then" portion of a rule has a certainty factor, then the clause will be asserted with the given certainty factor. In this discussion, the effect of asserting a clause to be true more than once will be left undefined, except to note that the resulting certainty will be at least as great as the more certain assertion. If the certainty of a clause is not specified, it will be assumed to be definite.

The time lags between related events are accommodated by the structure

X WITHIN Y MINUTES

meaning that the assertion X was true within the last Y minutes. If this assertion is true, then

X WITHIN Z MINUTES

is true as well, for any value of Z greater than Y.

In this section, the residual response experiment of Section 7.2.2 will be examined further. The focus here is on how the knowledge used in this experiment would be represented. The diagnosis is examined at two points during the evolution of the failure: immediately following the detection of the gas anomaly, and immediately following the last rapid drop of the gas residual.

7.3.1 Initial Diagnosis

When the detection process discovers an anomaly, it must trigger the diagnostic process to determine if the anomaly represents a serious threat to the transformer. When the expert system is initiated, it will first apply its primitives to the various data streams to classify their behaviors. The result is a list of assertions. For the residual response experiment, a partial listing of these assertions might be:

ANOMALY
GAS_RESIDUAL is HIGH
MAGNITUDE of GAS_RESIDUAL_STEP1 is 10.1
GAS_RESIDUAL_STEP2 WITHIN 30 MINUTES
MAGNITUDE of GAS_RESIDUAL_STEP2 is 1.5
MAINTENANCE WITHIN 1 HOUR
LOAD_STEP1 WITHIN 1 HOUR
LOAD_STEP2 WITHIN 1.5 HOURS
THERMAL_RESIDUAL is ZERO
VIBRATION_RESIDUAL is ZERO
⋮

Each of these assertions is the result of a straightforward analysis of a single data stream, with the exception of *MAINTENANCE*. A *MAINTENANCE* assertion would probably be supported by either an on-line maintenance log or an algorithmic primitive that explicit looks for instances of correlated no load and no excitation.

The assertions listed above can be interpreted as follows:

- An anomaly was detected.
- The gas residual has exceeded its maximum threshold.
- The gas residual just experienced a step change of 10.1 ppm.
- The gas residual experienced a step change 30 minutes ago.
- The gas residual step change 30 minutes ago was 1.5 ppm.
- A maintenance shutdown occurred an hour ago.
- A load change occurred an hour ago.
- A load change occurred an hour and a half ago.
- The thermal residual is not anomalous.
- The vibration residual is not anomalous.

⋮

The assertions *GAS_RESIDUAL_STEP1* and *GAS_RESIDUAL_STEP2* simply differentiate between two events, the step change that forced the residual over the detection threshold and the earlier, relatively insignificant step change. *LOAD_STEP1* and *LOAD_STEP2* refer to the two load changes that, when taken together and correlated with the excitation data, were interpreted as a maintenance shutdown.

Generating the assertions listed above does not consist of merely applying the level and rate-of-change thresholds used by the detection process. As seen in the

thermally-induced sensor malfunction experiment, anomalous behavior that would be ignored by the detection mechanism can be useful to the diagnosis process. The primitives that generate these basic assertions have to be sophisticated enough to recognize *patterns* of behavior and not just thresholds. This lowest-level primitives might only identify simple features such as local maxima and minima and inflection points; there might be another layer of interpretation not represented here that transforms these features into the assertions listed above. For instance, the "step changes" referred to previously might actually consist of a short series of data points that have been grouped together as a single event. This discussion should not imply that each event must be re-identified each time diagnosis is triggered; this would represent inefficient access of old data. Instead, some bookkeeping mechanism would be necessary to prevent the same analyses from being conducted repeatedly during the evolution of a failure. Once an event has been identified, it could be stored in some ready-to-use format that could be referenced as basic facts. Assume, for now, that the individual data streams can be classified so as to generate the list of assertions shown.

The rules that make up the knowledge base of the diagnostic expert system can now be used to combine these assertions to support other assertions. An expert system may employ different strategies to select the appropriate rules to trigger to progress toward the eventual conclusions. (*Backward* and *forward chaining* are two examples.) The mechanism of rule selection will be ignored in this discussion, except to note that if two rules that are presented here appear to conflict, then the more "specific" rule would be used. For example, consider a rule that encodes the knowledge that maintenance increases the chance that the data acquisition system has been damaged:

IF *ANOMALY* AND
MAINTENANCE WITHIN 2 DAYS
THEN *DIAGNOSIS is* SENSOR_MALFUNCTION (possible)

The sooner the anomaly occurs following a maintenance procedure, the more likely it is that the maintenance procedure caused a sensor malfunction. One might encode this with two more rules:

IF *ANOMALY* AND
MAINTENANCE WITHIN 2 HOURS
THEN *DIAGNOSIS is* SENSOR_MALFUNCTION (likely)

and

IF *ANOMALY* AND
MAINTENANCE WITHIN 2 MINUTES
THEN *DIAGNOSIS is* SENSOR_MALFUNCTION (almost certain)

If the anomaly occurs 30 minutes after a maintenance shutdown then the most specific rule that applies out of these three is the second: sensor malfunction is

"likely". The third rule does not apply because the anomaly did not occur within two minutes; the first rule does not apply because it is less specific then the second.

One other rule that might be triggered is:

IF	*GAS_RESIDUAL is* HIGH	AND
	LOAD_CHANGE WITHIN 1 HOUR	AND
	LOAD_CHANGE is LARGE	
THEN	*DIAGNOSIS is* ARC (likely)	AND
	DIAGNOSIS is PARTIAL_DISCHARGE (likely)	AND
	DIAGNOSIS is OVERHEATING (possible)	

At this stage of the knowledge engineering process, no other rules seem to apply to the situation that arises immediately after the gas anomaly is first detected. However, it is interesting to note that one rule that is *not* triggered is:

IF	*GAS_RESIDUAL is* HIGH	AND
	THERMAL_RESIDUAL is HIGH	
THEN	*DIAGNOSIS is* OVERHEATING (likely)	

In other words, since the thermal residual is not behaving abnormally, overheating can not be considered likely, though it is possible (from the previous rule).

Thus the initial diagnosis is that arcing, partial discharge, and sensor malfunction are all likely, and overheating is possible. If the certainties associated with the first three diagnoses could be further differentiated, it would be necessary to adopt a system with greater flexibility for expressing certainties.

7.3.2 Final Diagnosis

A partial list of the assertions produced by applying algorithmic analyses to the individual data streams would include the following:

ANOMALY
GAS_RESIDUAL is HIGH
MAGNITUDE of GAS_RESIDUAL_STEP1 is -16.3
GAS_RESIDUAL_STEP2 WITHIN 3.2 HOURS
MAGNITUDE of GAS_RESIDUAL_STEP2 is 15.2
GAS_RESIDUAL_STEP3 WITHIN 4.3 HOURS
MAGNITUDE of GAS_RESIDUAL_STEP3 is -15.4
THERMAL_RESIDUAL is ZERO
VIBRATION_RESIDUAL is ZERO
⋮

which would be interpreted as:

- An anomaly was detected.

- The gas residual has exceeded its maximum threshold.
- The gas residual just experienced a step change of -16.3 ppm.
- The gas residual experienced a step change 3.2 hours ago.
- The gas residual step change 3.2 hours ago was 15.2 ppm.
- The gas residual experienced a step change 4.3 hours ago.
- The gas residual step change 4.3 hours ago was -15.4 ppm.
- The thermal residual is not anomalous.
- The vibration residual is not anomalous.

⋮

A critical piece of knowledge is encoded as:

IF *MAGNITUDE of GAS_RESIDUAL_STEP is* NEGATIVE
THEN *DIAGNOSIS is* GAS_SENSOR_MALFUNCTION (almost certain)

That is, if there is a negative step change in the gas residual, then there is almost certainly something wrong with the gas sensor.

The next bit of expertise that might be applied is the fact that if two step changes in a single data stream and close in time are equal in absolute magnitude but opposite in sign, and one step change is diagnosed as being due to sensor malfunction, then the other step change is likely due to a sensor malfunction. In rule form:

IF *DATA1_STEP1* AND
DIAGNOSIS of DATA1_STEP1 is SENSOR_MALFUNCTION AND
DATA1_STEP2 WITHIN 2 DAYS AND
MAGNITUDE of DATA1_STEP1 is near
$-$(*MAGNITUDE of DATA1_STEP2*)
THEN *DIAGNOSIS of DATA1_STEP2 is*
SENSOR_MALFUNCTION (very likely)

The reader should note that this rule does not directly reference the gas residual. Instead, the data stream is referred to as *DATA1*. The inference engine of the expert system would be responsible for resolving *DATA1* as *GAS_RESIDUAL* in this example. Likewise, *STEP1* and *STEP2* are place savers that must be resolved to reference different events. So, if this rule were applied to the events *GAS_RESIDUAL_STEP2* (diagnosed as being caused by a sensor malfunction) and *GAS_RESIDUAL_STEP3*, then *DATA1_STEP1* would map to *GAS_RESIDUAL_STEP2* and *DATA1_STEP2* would map to *GAS_RESIDUAL_STEP3*. In this way, a single general rule can replace many specific rules that reference particular data streams.

Referring back to Figure 7.5, the reader should note that if this new rule were applied indiscriminately it would result in all the large gas residual step changes being diagnosed as sensor failures. Entirely by coincidence, the step changes resulting from the real gas generating events have approximately the same magnitude as those arise from the sensor malfunction. The "real" step changes were 10.1 and 13.1 ppm (2.1 ppm of which was "overshoot"); the others were 15.2, 15.4, 15.2, and 16.3 ppm. There is a significant difference between these two sets of magnitudes, but this difference is not so large that the first two events should be diagnosed as true gassing failures based solely on their magnitudes.

There must be some restriction on the application of this last rule. A negative step change should only correlate to one positive step change. When the data acquisition system fails in the manner seen here it tends to bounce between two states; the offset may drift until a stable erroneous offset is established, but after this the transition between a zero offset and a error offset is very abrupt. Thus, erroneous step changes in a data signal alternate in sign. The specificity requirement would cause the first negative step change to be correlated with the immediately previous positive step change because their absolute magnitudes are "nearest" to each other.

Of course, while the rule above might be used to characterize a particular type of sensor malfunction, the human expert probably applies a more general rule as well: if one event has been diagnosed as a sensor malfunction, then sensor malfunction gains some credence as an explanation for all other anomalous events.

IF	*ANOMALY1*	AND
	ANOMALY2	AND
	DIAGNOSIS of ANOMALY1 is SENSOR_MALFUNCTION	
THEN	*DIAGNOSIS of ANOMALY2 is* SENSOR_MALFUNCTION (possible)	

In this situation, *ANOMALY2* applies to each of the other observed anomalies. The last rule should only be applied once for each negative step change; this rule should be applied as many times as needed. The knowledge representation needs some way of denoting this—perhaps with a keyword such as EVERY, as in

IF	*ANOMALY1*	AND
	EVERY *ANOMALY2*	AND
	⋮	

Finally, there should be some rule that encodes the fact that if the decay following a gas residual positive step change behaves normally, then the step change is probably due to an actual gas-generating event.

IF *RATE_OF_CHANGE of GAS_RESIDUAL is* NORMAL AND
GAS_RESIDUAL_STEP WITHIN 30 MINUTES
THEN *DIAGNOSIS is* ARC (likely) AND
DIAGNOSIS is PARTIAL_DISCHARGE (likely) AND
DIAGNOSIS is OVERHEATING (likely)

7.3.3 Prognosis and Recommended Treatment

The discussion thus far has focused on diagnosis. This is completely appropriate, as an automated monitoring system can not generate truly useful prognoses and recommended courses of treatment until diagnosis can be performed accurately and reliably. At this time, rules for governing the recommendation of treatments must necessarily be crude. An example:

IF *DIAGNOSIS is* ARC (almost_certain) AND
GAS_CONTENT is CRITICAL
THEN *NOTIFY_OPERATOR_TRIP_RECOMMENDATION*

If the dissolved gas content measurement has reached critical values, and it is almost certainly due to an arc, the operator should be notified that the transformer should be taken off line.

If the gas content has not yet reached the critical range, the recommendation to the operator might take the form of a warning that a serious condition exists and a request to run further tests:

IF *DIAGNOSIS is* ARC (likely) AND
GAS_CONTENT is HIGH
THEN *NOTIFY_OPERATOR_WARNING* AND
REQUEST_DISSOLVED_GAS_ANALYSIS

The operator is made aware of the situation, and can then make his own evaluation of the status of the transformer and his own estimate of the remaining life of the transformer. In addition, a dissolved gas analysis is requested. If the operator is uncertain as to whether to continue running the transformer, the results of the oil analysis could decide the matter. This rule makes no explicit mention of the cost of conducting a dissolved gas analysis; the cost has been "compiled" into the rule. If this rule were implemented as presented, it would be because experience has shown that when a severe arc is likely a dissolved gas analysis is worthwhile. If this rule were to be applied to many different transformers with widely varying cost structures, or perhaps to a single transformer with a variable cost structure, this information would have to be made explicit.

7.4 Summary

This chapter represents the first step towards an implementation of an automated transformer diagnosis system based on adaptive models of normal behavior. First, those categories of knowledge that have proven useful for the diagnostic task were identified. They were divided into two broad classes: general transformer knowledge and site-specific knowledge. These categories were drawn from both the traditional methods of monitoring transformers as well as the new techniques derived from the MIT approach to transformer anomaly detection. To explore how these bits and pieces of transformer knowledge could be applied to the problem of diagnosis, a number of case studies were analyzed in great detail. Three of the experiments were dissected: the thermally induced gas sensor malfunction, the residual response experiment (with actual sensor malfunction), and the parameter trend analysis experiment. Finally, the issue of how to represent this knowledge was addressed. The diagnosis process associated with the residual response experiment was broken into its component decisions at two different points of the evolution of the failure.

Chapter 8

Conclusions, Recommendations, and Other Applications

8.1 Conclusions

The experiments and analyses described in this thesis confirm three crucial points:

- Continuous monitoring of transformer performance in conjunction with models of normal behavior can form the basis of a responsive and sensitive detection mechanism for incipient transformer failures.
- These tools also provide sufficient information to distinguish between competing diagnoses—most importantly, between sensor malfunctions and "real" transformer failures.
- The categories of knowledge necessary to perform this diagnosis and the processes by which this knowledge is applied are amenable to automation.

That the MIT transformer monitoring system is quick to respond to developing failures is enabled by its ability to continuously monitor transformer performance. But continuous monitoring does not in itself guarantee responsiveness. Instead, responsiveness is keyed to the sensitivity of the detection mechanism described herein which is enabled by the use of models of normal behavior.

The Hydran 201R dissolved combustible gas sensor is one example. The Hydran continuously monitors the dissolved gas content of the transformer oil. But, traditionally, the sensor is tied to a static threshold detector and the continuous data is discarded. This system is responsive to large faults but, due to its lack of sensitivity to small anomalies, it is not responsive to incipient failures. The experiments presented in the previous chapters show that integrating this sensor into the MIT transformer monitoring system greatly increases the sensitivity to anomalies in the sensor's output. Failures can be detected in the early stages of development, buying valuable time in which to plan for replacement or repair.

With increased sensitivity comes an increased chance for extraneous detected anomalies—anomalies that do not map to actual failures. For this reason, accurate diagnosis is essential if the system is to be effective. The most critical distinction that must be made is between mere sensor malfunctions and situations that pose an actual threat to the continued health of the transformer. Human intervention can be justified in helping to gauge the severity of an actual failure. But if the system sounds too many false alarms, then the detection sensitivity would have to be decreased, crippling the effectiveness of the system.

As was shown in this thesis, the data generated through the use of adaptive models of normal behavior is sufficiently informative to diagnose sensor malfunctions and incipient transformer failures. A great deal more experience is necessary to determine how well this system would distinguish between specific transformer failure modes. But it must be stressed that it is as least as difficult to diagnose a sensor malfunction as any other failure mode, and this—perhaps most crucial—diagnosis *can* be supported by residual and parameter analysis.

The manner in which diagnostic knowledge is analyzed suggests that it is possible to automate the diagnosis process. Some of the correlation tests that were assumed to be primitives of the expert system laid out in Chapter 7—accessing past anomalies, or identifying lags between anomalies manifesting in two different signatures, for instance—raise some nontrivial implementation issues, but these difficulties are not insurmountable. With these primitives in hand, a rule-based expert system would be able to diagnosis incipient transformer failures.

8.2 Recommendations for Future Work

There are four major directions in which this research could continue:

- Development of the body of experience from which a diagnostic knowledge base can be drawn
- Implementation of the diagnostic process which has been outlined in this thesis
- Model development and improvement of the techniques used to render them adaptive
- Application of other forms of artificial intelligence to the device monitoring problem.

8.2.1 Experience

Both the breadth and depth of the knowledge base sketched out in this thesis should be addressed in future research. The gassing failures that were focused on here are only a portion of the entire range of possible failure modes. Broadening the

knowledge base to include failures such as through-faults, that influence a different set of modules than were affected in these experiments, perhaps in unforeseen ways, may reveal general diagnostic rules compared to which the rules discussed here are merely special cases. At the very least, broadening the knowledge base will allow the implementation of a diagnostic system that can handle a wider range of failures.

It is also needed to deepen the knowledge base in the limited domain of gassing failures that have been explored in the performed experiments. While the arcing simulation experiments introduced some variation in the observed failure behavior, there were some fundamental similarities that could not be escaped: the gas was produced by an arc, in a single location, involving a single material (the oil), and not interacting with any other failure mode. For instance, the behavior of the Hydran 201R in response to the arcing experiments was, each time, dominated by the mass transport characteristics of hydrogen in the oil or the gas space. Perhaps if the gassing fault produced sufficient carbon monoxide to influence the behavior of the sensor, it would be possible to conclude that paper was involved.

The most effective way to flesh out the knowledge base might be to establish a clearinghouse for data from the electric utilities. A system for the detection of incipient transformer failures based on the system described herein has been developed by J.W. Harley, Inc. As these products proliferate, there is the potential for a great deal of data to become available for analysis. This data will represent several different transformer designs and (eventually) several different failure modes. There is also the advantage that the data which is desired greatly, that which concerns the most common failure modes, will tend to become available first. Settling issues such as confidentiality and incentives now may make this a possibility later.

8.2.2 Implementation

"The best laid plans..." The process of implementing a diagnostic expert system based on adaptive models will certainly reveal traps and pitfalls that have not been foreseen. The actual implementation of a system that can duplicate the reasoning set forth in Section 7.2 and Section 7.3 would likely reveal most of the difficulties inherent in a system designed to handle the entire range of transformer failures.

8.2.3 Model Development and Model Adaptation

A great deal of knowledge and effort by members of the MIT Transformer Monitoring Project is embodied in each of the models that were presented here. There are other models, other modules, which had to be put aside to focus on the selected set. These others models describe electrification, partial discharges, and non-destructive breakdown, among other phenomena. Each of these models could be developed into a module that would increase the robustness of the monitoring system and increase the chance that an incipient failure would be detected before serious damage could

occur.

Model development does not have to limited to new models. Existing models might be further refined. This refinement could involve, for instance, adding dynamics to static models. But an exciting alternative is the development of *adaptive model structures*, to complement the use of adaptive models. As an example, consider a signature that is in steady-state for long periods of times divided by periods of dynamic behavior. A steady-state model would not be sufficient for monitoring the dynamic behavior of the transformer, while it would be impossible to get good estimates of the parameters of a dynamic model during periods of steady-state operation. The solution is to have an adaptive model structure; the model structure contains both steady-state and dynamic terms, but the parameters of the dynamic portion of the model are assumed to be fixed during periods of stable operation.

Other suggestions for the improvement of the parameter estimation mechanism are given in Section 4.2.3. During the experiments conducted in the course of this thesis, an additional improvement was discovered. This improvement concerns the acceptance of candidate parameters. In accepting a set of parameters, a single statistical test is applied which uses a composite measure of the average squared error associated with the candidate parameters and a measure of the information content of the data used in performing the estimation. Unfortunately, this blurs the distinction between parameters that are rejected because of low information content and those that are rejected due to some anomaly invalidating the model structure. Parameters estimated from little or no information have no place in the system presented in this document, but, as has been shown, parameters derived from anomalous data may be informative during diagnosis. Thus, it would be desirable to allow the diagnostic process to access this latter type of rejected parameter.

Finally, the application of neural nets to transformer diagnosis is an interesting topic for future investigation. Given sufficient learning examples, neural nets can be trained to associate input patterns with target patterns. If the input patterns consisted of residual and parameter information, and the output pattern represented a diagnosis selection, the neural net can learn to diagnosis transformers using residual and parameter behavior. With a sufficient database of failure information, it would seem that diagnostic system development costs could be cut to zero. However, this is not likely. Large power transformers are designed to a utility's specific needs; many design decisions are made on per-unit basis or as part of a very limited run of transformers. In addition, transformer technology continues to evolve—not rapidly, by some industries' standards, but rapidly enough when compared to the lifetime of a transformer. Thus, there is probably not enough information on the failure behavior of a particular transformer to reliably apply the naive approach presented above.

8.2.4 Further Artificial Intelligence Techniques

A more judicious application of neural nets to the problem of diagnosis might be to use them to recognize specific patterns in particular residual or parameter streams. Thus, neural nets could be an implementation technique for realizing the correlation primitives that would support the diagnostic process.

Neural nets might also play a role in detection, performing some of the same functions that are fulfilled by the linear models upon which the monitoring system is based. Linear models have the advantage that stable parameters can be estimated—parameters which, in some instances, correlate to physical parameters of the transformer. Knowledge of the behavior of these estimated parameters has been shown to reflect the transformer's internal condition. But the linear models are usually approximations of the behavior of the associated signal; the associated residuals are not pure noise, but exhibit correlated behavior of their own. Neural nets are not limited to simulating linear models but can, with a reasonable amount of computation, fit higher-order systems. This may result in more sensitive residuals. Parameters that can be interpreted physically are a valuable resource for diagnosis, but perhaps those signatures which are monitored using black-box linear models would be better served through the use of neural nets. In these cases it is difficult to determine when a parameter change is significant, so perhaps the focus should be on improved residual analysis.

8.3 Other Applications

While this thesis has dealt exclusively with the monitoring of large power transformers, many of the issues and techniques considered herein can be applied to the broader domain of general device monitoring. Power transformers were an ideal application for the proposed approach to device monitoring for many reasons:

- the economics of power generation and distribution warranted the research and development costs
- traditional transformer monitoring techniques do not take advantage of the power and low cost of modern computers—real, immediate benefits can be achieved with the introduction of an on-line incipient failure detection system
- transformers were sufficiently well-understood to support the development of adaptive mathematical models of their behavior
- transformer behavior could be reproduced in a laboratory environment, to some degree of accuracy
- transformer expertise was available

This last point is perhaps the most critical, for each of the reasons which precede it could be used to support other device monitoring tasks equally well. The monitoring techniques described here could be applied to other power apparatus such as rotating machines and circuit breakers, as well as gas turbines, chemical plants, and nuclear power plants.

Appendix A

Pilot Transformer Test Facility

The center of the Pilot Facility is a 50 kVA, 240/8000 volt, single-phase, oil-filled, pole-type transformer. This transformer is known as the *test transformer*. The tank and transformer have been modified with the installation of numerous sensors; the tank does, however, retain its original gas space (sealed to the atmosphere and filled with dry nitrogen). The transformer has also been provided with a forced-oil circulation system to allow external control of heating and cooling. Excitation voltage and load current can be set independently. The test transformer is connected in parallel with a second, identical pole-type transformer. Variable loading to 150% of rated current at full voltage is achieved by using a third, smaller transformer to inductively drive circulating current through the two pole-type transformers. By controlling the phase of the circulating current, the test transformer may be made to look as if it is supplying real and reactive power to a load.

The 50 kVA size units were chosen to be large enough to have space for the needed sensors and to generate substantial core and winding losses during load cycles; yet small enough to allow easily-made changes to the monitoring structure, as well as to fit inside the laboratory building.

A.1 Hardware Overview

The original goal was to implement a monitoring system on a personal computer. It became clear, however, as a data acquisition system was designed, that some sort of multi-tasking, multi-processing computer environment was necessary. The tasks to be executed, from data acquisition on microsecond time scales to data analysis on a daily time scale required more computational power and flexibility than one personal computer was capable of delivering. Consequently, a basic hardware structure of two IBM AT compatible personal computers was settled on.

One of the AT clones was designated the *Master Machine*. It runs at 8 mHz under the IBM Xenix (Version 2.0) operating system. (Xenix is a version of UNIX.) This machine provides a multi-user, multi-tasking environment for the coordination

and control of a data acquisition subsystem as well as processing the resulting data. The Master Machine has a number of peripherals attached to it including a printer, a modem, a color monitor, dual 20-megabyte fixed disk drives, a 9-track open reel tape drive, dual floppy drives, an additional user terminal (with provisions for other serial devices), as well as a data acquisition subsystem.

The data acquisition subsystem is another IBM AT compatible machine, running at 6 mHz under MS-DOS 3.10 and coupled to a Keithley Data Acquisition and Control - Series 500 Measurement and Control System. The AT compatible, called the *Acquisition Machine* has a 20-megabyte fixed disk drive, dual floppy drives, an EGA video card, and monochrome video display. The system board has its memory split into two 512k blocks. The first block is used as DOS base memory. The second block is addressed above the system ROMs as extended memory and is used for a RAM disk. Other than a drive controller and a video display adapter, the only additional board in the expansion bus is the interface to the Keithley System 500 modular data acquisition system. This combination is responsible for obtaining temperatures from twenty-three (23) thermocouples, vibration signals from two (2) accelerometers, high- and low-side current and voltage wave forms and root-mean-square (RMS) values, and the dissolved combustible gas concentration from a Syprotec H-201R Hydran monitor. This subsystem is controlled by the Master Machine using an RS232 serial line. Data is transmitted in batch every few minutes from the Acquisition Machine to the Master Machine over a second RS232 line.

While the data acquisition system is flexible enough to accommodate a variety of sampling rates, the work described in this thesis makes use of temperatures and RMS values of the high- and low-side currents and voltages sampled every two minutes; all other data is sampled at ten-minute intervals.

All of the analog data being acquired from the Pilot Facility's test transformer is handled by the above-mentioned Keithley Series 500 System operating in conjunction with the Acquisition Machine. The Keithley System consists of a self-contained chassis and motherboard with slots to accommodate ten (10) plug-in circuit boards. The slots accept a variety of boards designed to perform various data input and output, or control functions. The data acquisition chassis interfaces with the Acquisition Machine through a cable (or an MIT-developed optic link) which connects to the interface card plugged into one of the Acquisition Machine's expansion slots.

This particular data acquisition system was chosen because of its extreme versatility, large number of available channels, and superior temperature measurement circuitry.

The combination of the Master Machine, Acquisition Machine, and Keithley System forms a loosely-coupled multi-tasking, multi-processor computer system.

A.2 Acquisition Machine Software

Operation of the Keithley System 500 is governed by software running on the Acquisition Machine. This software is a combination of commercial and custom-written code. Fundamental operation of the System 500 is performed by a software package supplied by Keithley. This package is called SOFT500, and it operates as a superset of commands in the interpretive BASIC language environment. The data acquisition routines, or drivers, are therefore, custom-written BASIC programs with embedded SOFT500 commands.

Data acquired by the System 500/Acquisition Machine combination is preprocessed in the Acquisition Machine to cut down on the data transfer requirements of the overall monitoring system. Preprocessing involves computation of RMS values, averaging, scaling, and other data reduction operations. Preprocessing is done with compiled routines written in C to increase computation speed and aid portability. After preprocessing, the reduced data is transferred to the Master Machine for further processing and analysis.

A.3 Master Machine Operating System

The operating system chosen for the Master Machine is UNIX. UNIX is a well-established multi-tasking operating system developed by AT&T Bell Laboratories. The current version is UNIX System V. It is available on many different computers and provides good support for the C programming language. The wide availability of UNIX System V and C means that software written in C or embedded with UNIX system commands is not restricted to one computer. If written properly, the software is quite portable. Furthermore, UNIX contains many system commands useful to the Pilot Monitoring System, and is based on a file system structure which easily lends itself to the buffering and shared information demanded by the monitoring system.

The version of UNIX chosen for the Pilot Monitoring System is IBM Xenix (Version 2.0). IBM Xenix was picked because, among several UNIX operating systems available for AT's and compatibles at the time of selection (1987), it was the only system with proven reliability.

Appendix B

Parameter Estimation Theory

B.1 Problem Formulation

This appendix discusses the background of parameter estimation in greater detail than is developed in the body of the thesis. Particularly, models of the form:

$$y[k] = \phi_1[k]\theta_1 + \phi_2[k]\theta_2 + \cdots + \phi_n[k]\theta_n \tag{B.1}$$

are discussed, where y, ϕ_1, ϕ_2, ..., and ϕ_n are discrete-time functions of observable quantities and θ_1, θ_2, ..., and θ_n are unknown parameters. $y[k]$ is called an *observation* and the various ϕ_i's are *regressors*. This can be written in a simpler notation as:

$$y[k] = \phi^T[k]\theta \tag{B.2}$$

where:

$$\begin{aligned} \phi^T[k] &\triangleq [\ \phi_1[k]\ \phi_2[k]\ \cdots\ \phi_n[k]\ ] \\ \theta &\triangleq [\ \theta_1\ \theta_2\ \cdots\ \theta_n\ ]^T \end{aligned}$$

Pairs $\{\ (y[i],\ \phi[i]) :\ i = 1,\ 2,\ \ldots,\ k\ \}$ are collected. By defining:

$$\begin{aligned} Y[k] &\triangleq [\ y[1]\ y[2]\ \cdots\ y[k]\ ]^T \\ \Phi[k] &\triangleq [\ \phi^T[1]\ \phi^T[2]\ \cdots\ \phi^T[k]\ ]^T \end{aligned}$$

one can write:

$$Y[k] = \Phi[k]\theta \tag{B.3}$$

Estimating the parameters of the model, then, is reduced to solving Equation (B.3) for θ, given $Y[k]$ and $\Phi[k]$.

B.2 Least Squares

Since the system of equations in Equation (B.3) is over-specified, θ can be approximated by applying the principle of least squares. The least-squares estimate of θ is achieved by finding some set of parameters $\hat{\theta}[k]$ that minimize the quantity

$$\left(Y[k] - \Phi[k]\hat{\theta}[k]\right)^T \left(Y[k] - \Phi[k]\hat{\theta}[k]\right)$$

In [33], Åström and Wittenmark prove that parameters $\hat{\theta}[k]$ minimize the least-squares criterion when:

$$\hat{\theta}[k] = (\Phi^T[k]\Phi[k])^{-1}\Phi^T[k]Y[k] = P[k]\Phi^T[k]Y[k] \tag{B.4}$$

provided that the matrix $\Phi^T[k]\Phi[k]$ is nonsingular. The condition that this matrix be invertible is called an *excitation condition.* The inverse matrix, $P[k]$, is called the *covariance matrix.* Written in terms of the individual sample pair, Equation (B.4) is:

$$\hat{\theta}[k] = (\sum_{i=1}^{k} \phi[k]\phi^T[k])^{-1} \sum_{i=1}^{k} \phi[k]y[k] \tag{B.5}$$

Equation (B.5) can be written recursively as:

$$\left.\begin{array}{rcl} \hat{\theta}[k] & = & \hat{\theta}[k-1] + K[k]\epsilon[k] \\ \epsilon[k] & = & y[k] - \phi^T[k]\hat{\theta}[k-1] \\ K[k] & = & P[k]\phi[k] = P[k-1]\phi[k](I + \phi^T[k]P[k-1]\phi[k])^{-1} \\ P[k] & = & (I - K[k]\phi^T[k])P[k-1] \end{array}\right\} \tag{B.6}$$

This formulation, called Recursive Least Squares (RLS), is developed in detail in [33]. The reader will note that the new parameters $\hat{\theta}[k]$ are generated from the old parameters ($\hat{\theta}[k-1]$) by applying a correction factor that is proportional to the size of the error in the model at time k, $\epsilon[k]$. The computation of $K[k]$ requires an inversion of $(I + \phi^T[k]P[k-1]\phi[k])$, but this matrix has rank equal to the number of outputs of the system. In other words, for a system with but a single output the matrix to be inverted reduces to a scalar.

B.3 Time-varying Parameters

The foregoing discussion assumes that the physical parameters, θ, are constant. If, in fact, the parameters change over time, the algorithm presented above will be inadequate. As the algorithm converges on an estimate and the covariance matrix $P[k]$ becomes small, it becomes increasingly unresponsive to new (and possibly conflicting) information. If the parameters change abruptly, yet infrequently, a simple solution is to periodically reset the covariance matrix to αI, where α is some

large number[33]. The algorithm will then converge to the current set of parameters, which may have changed since the previous estimation.

If the parameters vary slowly over time, a common solution is to apply an *exponential forgetting factor*, λ, such that current data is given unit weight while older data is weighted by λ^n (where n is the age of the data in discrete time). This leads to the algorithm:[33,36]

$$\left.\begin{array}{rcl} \hat{\theta}[k] & = & \hat{\theta}[k-1] + K[k]\epsilon[k] \\ \epsilon[k] & = & y[k] - \phi^T[k]\hat{\theta}[k-1] \\ K[k] & = & P[k]\phi[k] = P[k-1]\phi[k](\lambda I + \phi^T[k]P[k-1]\phi[k])^{-1} \\ P[k] & = & (I - K[k]\phi^T[k])P[k-1]/\lambda \end{array}\right\} \quad \text{(B.7)}$$

Fortescue, et al. [35] note that the RLS algorithm with constant forgetting factor runs into problems during periods when there is very little new dynamic information. During these periods, the covariance matrix may become very large, causing the system to become very sensitive to small perturbations or numerical errors. The authors introduce a variable forgetting factor, $\lambda[k]$, such that:

$$\left.\begin{array}{rcl} \lambda[k] & = & 1 - 1/N[k] \\ N[k] & = & \Sigma_0/[1 - \phi[k-1]^T P[k-1]\phi[k-1]]\epsilon^2[k] \end{array}\right\} \quad \text{(B.8)}$$

This modification is intended to ensure that a constant amount of information is used to estimate the parameters. As the information content of the data goes down, $N[k]$ grows larger and λ approaches 1. Therefore, information is forgotten more slowly, preserving the amount of information embodied in the estimate.

Σ_0 is a constant that controls the speed of adaptation of the algorithm. A suggested guideline for chosing this constant is:

$$\Sigma_0 \triangleq \sigma_0^2 N_0$$

where σ_0^2 is the expected measurement noise variance (based on knowledge of the physical process) and N_0 is the nominal asymptotic memory length (which would be the number of data points used to estimate parameters for a stationary process).

B.4 Weighted Least Squares

The RLS algorithm with forgetting factors discussed in Section B.3 is one example of a weighted least squares algorithm. In that algorithm, the exponential weighting is used to weight the more recent data more heavily, in order to force the estimated parameters to reflect the current state of the modeled system. Non-uniform weighting can, in general, be used to indicate which data should be modeled more accurately. Data is often chosen to receive heavier weighting because it is felt to represent the "actual" state of the modeled process. This may be because the data

is more recent, as in the example above, or because of some external knowledge of the amount of error in the data acquisition procedure at different sampling times.

Weighting can also be used to *tune* the model's performance for specific input conditions. This may be done in recognition of the differing frequencies at which various portions of the model's input space are visited in practice. Or it may be because certain sets of input conditions correspond to critical output ranges of the physical system.

For weighted least squares, the quantity to be minimized is:

$$\sum_{i=1}^{k} w_i \epsilon^2[i]$$

As presented in [34], the solution to the weighted least squares equation is:

$$\hat{\theta}[k] = (\Phi^T[k]W[k]\Phi[k])^{-1}\Phi^T[k]W[k]Y[k] \tag{B.9}$$

where $W[k]$ is a diagonal *weighting matrix* such that:

$$W[k] = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_k \end{bmatrix} \tag{B.10}$$

In this notation, RLS with (constant) exponential forgetting factor would have a recursively defined weighting matrix of:

$$W[k] = \begin{bmatrix} \lambda W[k-1] & \\ & 1 \end{bmatrix} \tag{B.11}$$

Note that the weighting of a particular set of data is not constant over time, but an exponentially decreasing function $w_i[k]$. Therefore, $w_i[k]$ can be defined recursively as:

$$w_i[k] = \lambda w_i[k-1]$$

Therefore the weight of a particular set of data is λ^n, where n is the age of that set.

B.5 Robustness

In [33], Åström and Wittenmark point out that the least squares method is fundamentally susceptible to a single large error in the acquired data. This is because the Gaussian assumptions upon which the method is based assign very low probabilities to large errors. The authors suggest replacing the estimator of Equation (B.7) with a relation such as:

$$\hat{\theta}[k] = \hat{\theta}[k-1] + K[k]f(\epsilon[k]) \tag{B.12}$$

where $f(\epsilon)$ is a function that varies linearly with small ϵ, but varies more slowly with large ϵ. An example of such an equation is:

$$f(\epsilon) = \frac{\epsilon}{1 + a|\epsilon|} \tag{B.13}$$

With the use of such an estimator, large errors in the input to the estimation routine will not have as drastic a consequence to the estimates. The estimation algorithm is then called *robust* in the presence of outliers.

Bibliography

[1] J. S. Forrest. "Opening Address". *Conference on Diagnostic Testing of High Voltage Power Apparatus in Service, Part II*, pages 4–11, Institution of Electrical Engineers, London, England, March 6–8 1973.

[2] Thomas H. Crowley, Wayne H. Hagman, Richard D. Tabors, and Chathan M. Cooke. "Expert System for On-Line Monitoring of Large Power Transformers". *EPRI Conference on Expert Systems Applications for the Electric Power Industry*, Orlando, FL, June 5–8 1989.

[3] D. J. Allan. "Editorial". *Electra*, No. 114:3–6, October 1987.

[4] S. D. Meyers, J. J. Kelly, and R. H. Parrish. *Fifty Years: A Guide to Transformer Maintenance.* S. D. Meyers, Inc., Akron, OH, 1981.

[5] A. C. Franklin and D. P. Franklin. *The J & P Transformer Book.* Butterworth and Co., Boston, MA, 11th edition, 1983.

[6] W. J. McNutt, T. O. Rouse, and G. H. Kaufman, Sr. "Mathematical Modelling of Bubble Evolution in Transformers". *IEEE Transactions on Power Apparatus and Systems*, PAS-104(2):477–487, February 1985.

[7] Chathan M. Cooke et al. *Final Report, First Year Studies on Trend Analysis—Performance Monitoring of Transformers.* LEES Technical Report TR-85-007, Laboratory for Electromagnetic and Electronic Systems, Massachusetts Institute of Technology, Cambridge, MA, November 1985.

[8] H. House et al., editors. *Conference on Diagnostic Testing of High Voltage Power Apparatus in Service, Part I*, Institution of Electrical Engineers, London, England, March 6–8 1973.

[9] M. Kurtz, G. C. Stone, P. Daechael, and B. K. Gupta. "Fault Anticipator for Substation Equipment". *IEEE Transactions on Power Delivery*, PWRD-2(3):722–724, July 1987.

[10] R. E. James, F. E. Trick, B. T. Phung, and P. A. White. "Interpretation of Partial Discharge Quantities as Measured at the Terminals of HV Power

Transformers". *IEEE Transactions on Electrical Insulation*, EI-21(4):629–638, August 1986.

[11] F. C. Pratt. "Diagnostic Methods for Transformers in Service". Paper 12-06, *1986 CIGRE International Conference on Large High Voltage Electric Systems*, August 27–September 4 1986.

[12] T. D. Poyser. "An On-Line Microprocessor Based Transformer Analysis System to Improve the Availability and Utilization of Power Transformers". *IEEE Transactions on Power Apparatus and Systems*, PAS-102(4):957–962, April 1983.

[13] T. D. Poyser, D. A. Yannucci, J. B. Templeton, and B. N. Lenderking. "On-Line Monitoring of Power Transformers". *IEEE Transactions on Power Apparatus and Systems*, PAS-104(1):207–211, January 1985.

[14] R. R. Rogers. *Concepts Used in the Development of the IEEE and IEC Codes for the Interpretation of Incipient Faults in Power Transformers by Dissolved Gas in Oil Analysis.* Technical Report, Central Electricity Generating Board, Transmission Development and Construction Division, Guilford, England.

[15] Joseph B. DiGiorgio, Fredi Jakob, T. J. Haupert, and Mason Sperry. "Cut Failures by Analyzing Transformer Oil". *Electrical World*, 52–54, October 1 1979.

[16] *Transformer Fault Detection and Diagnosis Using RuleMaster by Radian.* Technical Report, Radian Corporation.

[17] Jill R. Howes. "TOGA—Transformer Oil Gas Analyst: the Evolution of an Expert System". *EPRI Conference on Expert Systems Applications for the Electric Power Industry*, Orlando, FL, June 5–8 1989.

[18] H. E. Dijk. *Exformer, an Expert System for Transformer Faults Diagnosis.* Technical Report, N. V. KEMA, The Netherlands.

[19] Fred C. Schweppe. *In Service Transformer Monitoring.* Internal Report, M.I.T. Laboratory for Electromagnetic and Electronic Systems, December 1986.

[20] James R. Melcher and Chathan M. Cooke, Co-Principal Investigators. *Trend Analysis: Performance Monitoring of Transformers.* Cambridge, MA, October 1988. Final Report, Laboratory for Electromagnetic and Electronic Systems and Energy Laboratory Electric Utility Program, M.I.T.

[21] Wayne H. Hagman, Thomas H. Crowley, Richard D. Tabors, and Fred C. Schweppe. "An Adaptive Transformer Monitoring System". *International*

Symposium for Demonstrations of Expert System Applications to the Power Industry, Montreal, Canada, May 7–12 1989. Sponsored by Hydro-Québec.

[22] James R. Melcher and Chathan M. Cooke, Co-Principal Investigators. *Trend Analysis—Performance Monitoring of Transformers.* Research Proposal, Energy Laboratory and the Laboratory for Electromagnetic and Electronic Systems, M.I.T., Cambridge, MA, August 1983.

[23] Fred C. Schweppe and Richard D. Tabors. *Trend Analysis for the Transformer Project.* Internal Report, M.I.T. Laboratory for Electromagnetic and Electronic Systems, October 1984.

[24] G. Anthony Gorry, Jerome P. Kassirer, Alvin Essig, and William B. Schwartz. "Decision Analysis as the Basis for Computer-Aided Management of Acute Renal Failure". *The American Journal of Medicine*, 55:473–484, 1973.

[25] L. F. Pau. *Failure Diagnosis and Performance Monitoring. Control and Systems Theory*, Volume 11, Marcel Dekker, Inc., New York, NY, 1981.

[26] Daljit Singh. *The Impact of the Political and Economic Environment on Policy Choices—The Case of the Electric Power System in the United States and India.* Master's thesis, M.I.T. Technology and Policy Program, February 1989.

[27] Daniel J. McCarthy. *An Adaptive Model for Vibration Monitoring of Power Transformers.* Master's thesis, M.I.T. Department of Electrical Engineering and Computer Science, May 1987.

[28] Juan C. Lavalle. *Failure Detection in Transformers Using Vibrational Analysis.* Master's thesis, M.I.T. Department of Electrical Engineering and Computer Science, September 1986.

[29] Dale S. Archer. *An Adaptive Thermal Module for Transformer Monitoring.* Master's thesis, M.I.T. Department of Electrical Engineering and Computer Science, August 1988.

[30] *Guide for Loading Oil Immersed Distribution and Power Transformers Rated in Excess of 100 MVA.* IEEE PES Standards, July 1982.

[31] James L. Kirtley Jr., Juan C. Lavalle, and Daniel J. McCarthy. "Acoustic Monitoring of Transformer Structures". Paper 820-03, *CIGRE Symposium on Monitoring for HV Systems*, May 1987.

[32] M. C. Zaretsky. *Parameter Estimation Using Microdielectrometry With Application to Transformer Monitoring.* ScD thesis, M.I.T. Department of Electrical Engineering and Computer Science, November 1987.

[33] Karl Johan Åström and Björn Wittenmark. *Adaptive Control.* Addison-Wesley Publishing Company, Reading, MA, 1989.

[34] Gene F. Franklin and J. David Powell. *Digital Control of Dynamic Systems.* Addison-Wesley Publishing Co., Reading, MA, 1980.

[35] T. R. Fortescue, L. S. Kershenbaum, and B. E. Ydstie. "Implementation of Self-Tuning Regulators with Variable Forgetting Factors". *Automatica*, 17(6):831–835, 1981.

[36] Peter Andersson. "Adaptive Forgetting in Recursive Identification through Multiple Models". *International Journal of Control*, 42(5):1175–1193, 1985.

[37] Bruce G. Buchanan and Edward H. Shortliffe. *Rule-Based Expert Systems.* Addison-Wesley Publishing Company, Reading, MA, 1985.

[38] Donald A. Waterman. *A Guide to Expert Systems.* Addison-Wesley Publishing Company, Reading, MA, 1985.

[39] Warren M. Rohsenow and Harry Choi. *Heat, Mass, and Momentum Transfer. Prentice-Hall Series in Engineering of the Physical Sciences*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1961.

[40] Private communication with Bruce Davison, Boston Edison Company, Boston MA.

[41] Private communication with Anne Hueston, Boston Edison Company, Boston MA.