Spring 2010

Sensor data-based decision making

Ahmet Soylemezoglu

Department: Engineering Management and Systems Engineering

SENSOR DATA-BASED DECISION MAKING

by

AHMET SOYLEMEZOGLU

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ENGINEERING MANAGEMENT

2010

Approved by

Jagannathan Sarangapani, Co-advisor
Can Saygin, Co-advisor
Cihan H. Dagli
Kenneth M. Ragsdell
Scott Grasman

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following five articles:

Paper 1, A. Soylemezoglu, M. J. Zawodniok, K. Cha, D. Hall, J. Birt, C. Saygin, and J. Sarangapani, *"A Testbed Architecture for Auto-ID Technologies,"* Assembly Automation, vol. 26, no. 2, pp. 127-136, 2006.

Paper 2, M. D. Mills-Harris, A. Soylemezoglu and C. Saygin, *"Adaptive Inventory Management Using RFID Data,"* Int. Journal of Advanced Manufac. Technology, vol. 32, no. 9, pp. 1045-1051, 2007.

Paper 3, A. Soylemezoglu, M. Zawodniok, S. Jagannathan, *"RFID-based Smart Freezer,"* Industrial Electronics, IEEE Transactions on, vol. 56-7, pp. 2347-2356, 2009.

Paper 4, A. Soylemezoglu, S. Jagannathan, C. Saygin, *"Mahalanobis Taguchi System (MTS) as a Prognostics Tool for Rolling Element Bearing Failures,"* Accepted for publication on ASME Journal of Manufacturing Science and Engineering, Manuscript No: 09-1254, 2010.

Paper 5, A. Soylemezoglu, S. Jagannathan, C. Saygin, *"Mahalanobis Taguchi System (MTS) as a Multi-sensor based Prognostics Tool for Centrifugal Pump Failures,"* Intended for submission to IEEE Transactions on Reliability, 2010.

ABSTRACT

Increasing globalization and growing industrial system complexity has amplified the interest in the use of information provided by sensors as a means of improving overall manufacturing system performance and maintainability. However, utilization of sensors can only be effective if the real-time data can be integrated into the necessary business processes, such as production planning, scheduling and execution systems. This integration requires the development of intelligent decision making models that can effectively process the sensor data into information and suggest appropriate actions. To be able to improve the performance of a system, the health of the system also needs to be maintained. In many cases a single sensor type cannot provide sufficient information for complex decision making including diagnostics and prognostics of a system. Therefore, a combination of sensors should be used in an integrated manner in order to achieve desired performance levels. Sensor generated data need to be processed into information through the use of appropriate decision making models in order to improve overall performance.

In this dissertation, which is presented as a collection of five journal papers, several reactive and proactive decision making models that utilize data from single and multi-sensor environments are developed. The first paper presents a testbed architecture for Auto-ID systems. An adaptive inventory management model which utilizes real-time RFID data is developed in the second paper. In the third paper, a complete hardware and inventory management solution, which involves the integration of RFID sensors into an extremely low temperature industrial freezer, is presented. The last two papers in the dissertation deal with diagnostic and prognostic decision making models in order to assure the healthy operation of a manufacturing system and its components. In the fourth paper a Mahalanobis-Taguchi System (MTS) based prognostics tool is developed and it is used to estimate the remaining useful life of rolling element bearings using data acquired from vibration sensors. In the final paper, an MTS based prognostics tool is developed for a centrifugal water pump, which fuses information from multiple types of sensors in order to take diagnostic and prognostics decisions for the pump and its components.

ACKNOWLEDGMENTS

I would like to dedicate this dissertation to my nine month old daughter Faith Senay Soylemezoglu. She has brought such joy and happiness to our lives. I want to thank my dear wife, Mandy Soylemezoglu for all of her support and encouragement throughout my academic endeavors. Without her love and support, writing this dissertation would not have been possible.

I would also like to thank my parents Senay and Taner Soylemezoglu and brother Emre Soylemezoglu and my father-in-law Richard Hodges for always supporting me in whatever path I decided to take and for always being there for me.

I am deeply grateful to my co-advisors Dr. Jag Sarangapani and Dr. Can Saygin, whose encouragement, supervision and support enabled me to develop an understanding of the subject and also develop as a person. My advisors always have gone out of their ways to help me and throughout the dissertation have acted more like older brothers than advisors to me.

I would like to send my heartfelt thanks to the rest of the members of my committee, Dr. Cihan H. Dagli, Dr. Kenneth M. Ragsdell and Dr. Scott Grasman, for all of their support, encouragement, supervision and understanding.

I am also sincerely grateful to the Department of Engineering Management and Systems Engineering for all of their support, generosity and friendship that they shown to me and my family.

Finally, I offer my regards to all of those who supported me in any respect during the completion of this dissertation.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

SECTION

1. INTRODUCTION

Increasing globalization and growing industrial system complexity has amplified the interest in the use of information provided by sensors as a means of improving overall manufacturing system performance and maintainability, and lowering costs. Performance loss in a manufacturing system can either be due to inventory related problems or because of the health condition of the system. Lack of critical inventory items can cripple the entire production/ processes in the manufacturing system; whereas holding too much inventory can generate substantial holding costs for an organization. On the other hand, one cannot maintain performance if the components that comprise the overall manufacturing system are performing below acceptable performance thresholds. The drop in performance of components may be due to sudden failures or slow and gradual wearing of parts. Unfortunately there are no known solutions for the former case, however, with the help of an adequate monitoring, diagnostic and prognostics scheme, the latter case can be prevented.

1.1. INVENTORY MANAGEMENT

In the recent years Radio Frequency Identification (RFID) sensors have gained wide attention due to their capability to provide non-contact object identification and inventory visibility. An RFID system consists of (a) RFID tags, which store the unique item identity and other relevant information about items, (b) readers and antennas, that

interrogate the tags and read the information, and (c) software called middleware, that controls the RFID equipment, manages the data and provides an interface for the distribution and sharing of data with other enterprise applications. Potential benefits of RFID include: (1) instantaneous and automated data entry and monitoring, (2) effective use of labor, (3) real time visibility, and (4) mobile databases (i.e., self-storage of pertinent information of an item on the item itself).

RFID is a member of a broad group of technologies called Auto-ID (automatic identification), which are used to identify objects without human intervention. In general, Auto-ID technologies, advanced sensing capabilities, and recent developments in the area of mobile wireless ad hoc networking provide a potential to establish a data-rich manufacturing environment [1, 2]. Such technological capabilities provide real-time visibility of each single entity in the supply chain; presenting many opportunities for process improvement and re-engineering [3, 4]. On the other hand, the technology also presents various challenges due to the lack of established industrial standards and application roadmaps [5, 6], and the difficulties that arise while dealing with voluminous data in a timely fashion [1].

A typical Auto-ID application requires effective integration of two components. First, business processes (i.e., decision-making models) must be re-engineered to encapsulate Auto-ID data. Business process re-engineering should involve the integration of real-time data into the critical processes such as manufacturing execution, production planning, and scheduling systems. This integration requires the development of intelligent decision making models that can effectively process the sensor data into information and suggest the appropriate actions. For instance, the amount of safety

stocks, which are typically used in traditional inventory management models, can be dramatically reduced if Auto-ID data can be incorporated into the model. The second component is the networking topologies and related scheduling/routing protocols. In order for the decision-making component to be realistic, the network must facilitate effective and efficient routing of Auto-ID data packets. Academic studies usually focus either on manufacturing-specific decision-making (manufacturing engineering and industrial engineering) [7-9]; or on networking (electrical and computer engineering) [10, 11]. Due to the gap between these two sub-components, namely decision-making and networking, the solutions provided to the industry are not directly applicable; further testing on Auto-ID technologies within the proposed solution is usually necessary in order to fine-tune it to the production environment.

In the recent years, Auto-ID technologies, especially RFID has been gaining momentum in inventory management applications [12-14]. The objective in inventory management is to have the necessary inventory items at the right time, at the right amount, and at the right place. Since not all inventory items are of high value, tight tracking of them may not be desired. On the other hand, critical inventory items may require a tighter control. Due to the variety of items in a typical manufacturing environment, inventory items can be categorized in different classes, such as ABC classification, since tracking of all inventory items may not be economically viable. When implemented properly, RFID technology can provide item-level visibility in which electronic tags programmed with unique identification information are attached to "objects" that need to be monitored, tracked, or identified easily when needed; this has been an unreachable region of manufacturing information domain.

Today's inventory management systems operate with Stock Keeping Unit (SKU) level data, which means that hundreds of products can fall under the same SKU that associates brand, size, flavor, etc. and essentially categorizes them under a product type. While SKU-level data provide aggregate inventory levels for inventory items, it falls short when visibility of each item is desired. RFID technology facilitates Electronic Product Codes that provide unique identification for each product inside the enveloping SKU or inventory category [15]. This type of data helps associate production events with each inventory item, which facilitates tighter inventory control that relies on such real-time data.

Organizations that understand these possibilities have become trend-setters by establishing a January 2005 mandate for suppliers, which require RFID implementation on container/pallet size inventory. Such organizations include Wal-Mart and the Department of Defense (DOD). The focus of these mandates are twofold; not only do they force companies to take quick action in fulfilling requirements, but also to look towards the future with overall cost savings in mind. The need to organize and make decisions based on the data provided by the RFID tags is prominent.

To date, research that conjoins RFID technology and item-level inventory management on the shop floor is at a preliminary stage, only inferring benefits upon application. The challenge is to collect RFID data in a timely manner, to process such voluminous data, and to make timely decisions that are tied into manufacturing execution systems. If the challenge is overcome, then the benefits such as waste elimination, inventory reduction, automatic replenishment, stock-out reduction, and overall cost

savings can be easily realized. Therefore, there is a need for RFID data-based effective decision making algorithms that can lead to such benefits.

In addition, past work in the literature [1-14] assume a straightforward deployment of RFID technology will yield 100% satisfactory results in terms of item visibility. However, most of these studies have been performed under "controlled" laboratory settings and upon technology transfer to a real industrial environment, unforeseen problems, such as readability may arise from radio frequency (RF) interference, RF absorbing materials, and environmental conditions [16]. For instance, RFID tags may experience temperatures down to 100 deg F below zero in extremely low temperature environments such as an industrial freezer, or a cold chain warehouse. In such a case, excessive ice buildup on the tags may render the technology unreliable resulting in some of the tags becoming unreadable by RFID readers. Therefore, additional design and development need to be undertaken in areas in such as antenna configuration, antenna power control, etc. in order to provide feasible solutions for the intended application.

1.2. MANUFACTURING SYSTEM HEALTH MANAGEMENT

It is important to note that RFID is only one of many possible sensors that can be "embedded" in business processes in order to improve system performance. Even if the RFID/Auto-ID implementation for inventory management is successful, the performance of the overall system still depends on the physical health condition of its components. Therefore, it is necessary to be able to monitor and diagnose problems that can arise due to the wearing out of components.

As discussed in Section 1.1, RFID sensors can usually provide satisfactory results in order to prevent performance loss due to inventory management problems. However, in the component health management case, a single sensor type usually cannot provide sufficient information complex decision making. Therefore, a combination of sensors should be used in an integrated manner in order to achieve desired performance levels. Some of the various sensors that have been traditionally used to control the physical operations of system components are vibration, acoustic emission, temperature, force, and torque sensors [17, 18]. These sensors can also be utilized to convey information about the health condition of components that they are attached to. If the data acquired from these sensors are properly analyzed, it can be turned into diagnostic and prognostic information about the components. Consequently, this information can be stored on the RFID tags which are attached to the components and provide additional visibility about the manufacturing system which would lead to better decision making.

Most of the work in component health management focuses on monitoring and diagnostics of a single component out of many that comprise a system. For instance rolling element bearings and centrifugal pumps have been extensively studied in order to develop numerical model-based or process history-based monitoring and diagnostic algorithms and techniques since they are often the critical components that directly affect the overall performance of the entire system [19-25]. However, diagnostics by its nature is a reactive approach. In order to prevent catastrophic failures, preventive maintenance strategies are usually employed by industries. Maintenance on machinery is performed on a schedule-basis, which tends to unnecessarily increase downtime, reduce machine life and thus generate redundant costs. In order to rectify this problem, it is necessary to

transition to a more proactive approach, i.e. predictive maintenance. Prognostics are important to predictive maintenance, and it focuses on performance degradation and estimation of remaining useful life (RUL) of components. Therefore, research on prognostics has been gaining momentum in the last decade; however, compared to diagnostics, literature on prognostics is much smaller.

In the literature, there are two approaches for the estimation of RUL. The first approach utilizes physics of failure models that incorporate operating conditions in order to track the damage that components experience. These models require specific mechanistic knowledge and theory relevant to the components under investigation [26, 27]. The second approach to estimation of RUL comprises of pattern recognition models that are derived from historical data collected from the system under investigation. Artificial intelligence techniques such as neural networks, neuro-fuzzy inference systems, self organizing maps, etc. have been applied to the prediction of RUL by many researchers [28, 30].

It should be noted that most of the work done in this area [19-30], focuses on specific failure cases created under controlled laboratory conditions or through computer simulations. In some cases several faults may be present at the same time, one of which would ultimately result in failure. Nevertheless, all of the faults need to be monitored concurrently in order to have an effective prognosis on the RUL. In addition, it is usually much easier to detect damage than to assess its severity and progression. Both physics of failure and damage-progression/performance-degradation models face difficulties when the system under consideration is complex and affected by the operating conditions and the environment. The papers in the literature present several signals; and features that are

derived from these signals, which need to be monitored. If these signals/features show a large variability between experiments, it becomes difficult to select suitable thresholds to be used in monitoring diagnostics and prognostics. In addition, the existing body of work does not present a systematic method which can identify the most pertinent variables and eliminate the redundant ones in order to reduce analysis overhead. Therefore, a new multivariable method is necessary to identify the input variables necessary to be monitored for fault detection, root cause analysis, and fault prognosis.

1.3. ORGANIZATION OF THE DISSERTATION

In this dissertation, which is presented as a collection of five journal papers, several reactive and proactive sensor data-based decision making models that utilize data from single and multi-sensor environments are developed. The first three papers strive to provide solutions for inventory management related problems in order to increase manufacturing system performance. Papers IV and V, delve into condition monitoring, diagnosis and prognosis of components. Table 1.1 and Figure 1.1 show the organization of the dissertation in the form of several papers, and display the overall application area and the level of complexity.

Table 1.1. Organization of the Dissertation

SENSOR DATA-based DECISION MAKING	Paper I: "**A Testbed Architecture for Auto-ID Technologies**", *Assembly Automation*, vol. 26, no. 2, pp. 127-136, 2006.	Paper II: "**Adaptive inventory management using RFID data**", *Int. Journal of Advanced Manufac. Technology*, vol. 32, no. 9, pp. 1045-1051, 2007.	Paper III: "**RFID-based smart freezer**", *Industrial Electronics, IEEE Transactions on*, vol. 56-7, pp. 2347-2356, 2009.	Paper IV: "**Mahalanobis Taguchi System (MTS) as a prognostics tool for rolling element bearing failures**" Accepted for publication on *Journal of Manufacturing Science and Engineering,* Manuscript No: 09-1254	Paper V: "**Mahalanobis Taguchi System (MTS) as a multi-sensor based prognostics tool for centrifugal pump failures**", Intended for submission to *IEEE Systems Journal*
Single / Multi-sensor	N/A	Single Sensor (RFID)	Single Sensor (RFID)	Single sensor (vibration)	Multi-sensor (vibration, temperature, flow, and, pressure)
Sensor Output	Digital	Digital	Digital	Analog	Analog
Reactive / Proactive	N/A	Reactive	Reactive	Proactive	Proactive
Simulation-based / Hardware Implementation	N/A	Simulation-based	Hardware implemented	Hardware implemented	Hardware implemented
Number of Components	N/A	N/A	N/A	1 (rolling element bearings)	3 (impeller, seal, and filter)
Number of Failures / Data Collection Complexity	N/A	N/A	N/A / Moderate	1 / Moderate	3 / High

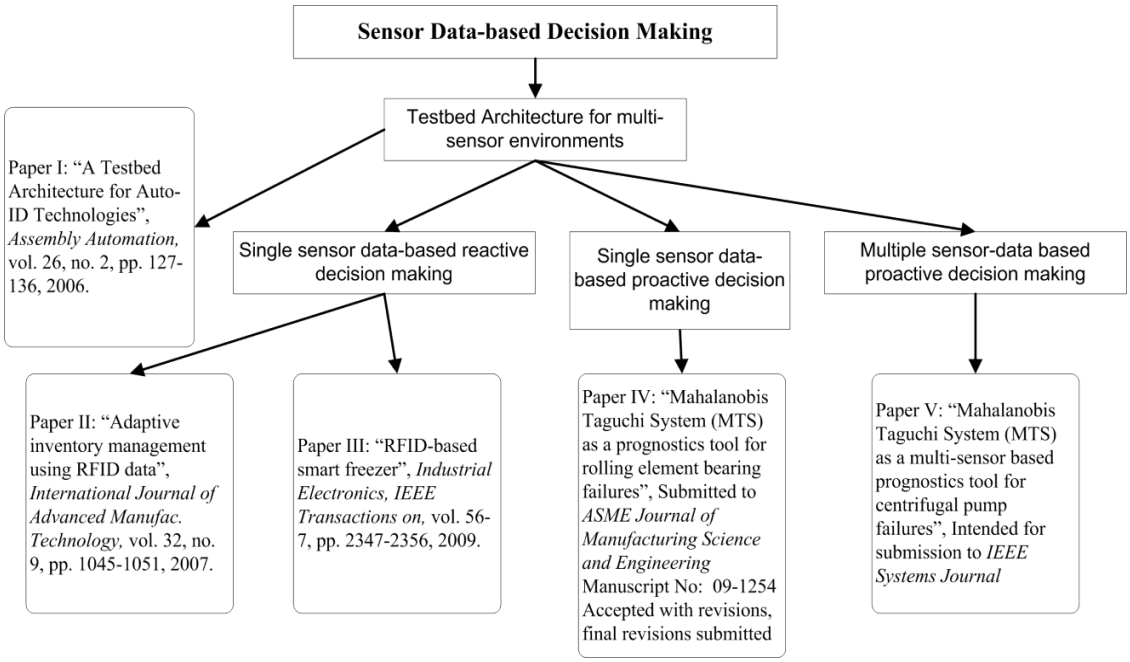

Figure 1.1. Outline of the dissertation

1.4. CONTRIBUTIONS OF THE DISSERTATION

This dissertation introduces several sensor data-based decision making models in single-sensor and multi-sensor environments in order to maintain manufacturing system performance. First, a general testbed architecture which facilitates the hardware/software based testing of the ideas that are developed in the dissertation is presented. Then, two inventory management control models are developed in order to overcome inventory related problems that affect performance. In the final part of the dissertation, a Mahalanobis Taguchi System based component health monitoring, diagnostics and prognostics scheme is developed. The contributions of each paper that constitute the dissertation are listed below.

The contribution of Paper I includes the development of a multi-sensor testbed architecture in order to support the research, development and implementation of Auto-ID technologies in network centric manufacturing environments. The proposed solution utilizes a unique hardware-in-the-loop simulation methodology, which integrates decision making model development with the design of networking topology and data routing/scheduling schemes in order to develop, test and implement viable Auto-ID solution. The paper also presents two case studies that highlight the effective use of RFID technology, its potential benefits, challenges, and deficiencies. The architecture developed in this paper lays the foundation for the successful implementation of hardware solutions and decision making models developed in the subsequent papers.

The contribution of Paper II is the development of a single sensor data-based reactive inventory management decision making model in a purely simulation environment. The paper introduces an adaptive inventory management model, which uses

a forecasting algorithm founded on RFID-based state and event data for tracking and dispatching of time and temperature sensitive materials. The study compares a traditional, static inventory model with the proposed model in terms of cost savings, inventory waste and reduction, and decision making complexity. The performance of the proposed model is validated in a simulation environment, which demonstrates the overall benefits and effectiveness of RFID technologies in providing low cost manufacturing solutions, reduced inventory levels and waste.

Subsequently, single sensor data-based reactive decision making model development has been extended to include hardware implementation in Paper III. Hardware implementation creates additional challenges which are addressed by the development of a novel RFID-based smart freezer using a new inventory management scheme for extremely low temperature environments. The proposed solution utilizes backpressure inventory control, systematic selection of antenna configuration, and antenna power control. The proposed distributed inventory control (DIC) scheme dictates the amount of items transferred through the supply chain. When a high item visibility is ensured, the control scheme maintains the desired level of inventory at each supply chain echelon. The performance of the DIC scheme is guaranteed using a Lyapunov-based analysis. The proposed RFID antenna configuration design methodology coupled with locally asymptotically stable distributed power control (LASDPC) ensures a 99% read rate of items while minimizing the required number of RFID antennas in the confined cold chain environments with non-RF friendly materials. The proposed RFID-based Smart Freezer performance is verified through simulations of supply chain and experiments on an industrial freezer testbed operating at -100°F.

Paper IV introduces additional sensors increasing the overall decision making complexity due to the multi-variate nature of the problem. In addition, the decision making model developed in this paper is proactive, further increasing the complexity. The contribution of this paper is the proposed novel Mahalanobis Taguchi System (MTS) based fault detection, isolation, and prognostics scheme. The proposed data-driven scheme builds upon MTS by utilizing Mahalanobis Distance (MD) based fault clustering for diagnostics and by the use of the progression of MD values over time for prognostics. MD thresholds derived from the clustering analysis are used for fault detection and isolation. When a fault is detected, the prognostics scheme, which monitors the progression of the MD values, is triggered. Then, using a linear approximation, time to failure is estimated. The performance of the scheme has been validated via experiments performed on rolling element bearings inside the spindle headstock of a micro computer numerical control (CNC) machine testbed. The bearings have been instrumented with vibration and temperature sensors and experiments involving healthy and various types of faulty operating conditions have been performed. The analysis is performed in the feature domain by extracting useful features from the raw sensor data. The experiments show that the proposed approach renders satisfactory results for bearing fault detection, isolation and prognostics. Overall, the proposed solution provides a reliable multivariate analysis and real-time decision making tool that; (1) presents a single tool for fault detection, isolation and prognosis, eliminating the need to develop each separately and (2) offers a systematic way to determine the key features, thus reducing analysis overhead. In addition, the MTS-based scheme is process independent and can easily be

implemented on wireless motes and deployed for real-time monitoring, diagnostics and prognostics in a wide variety of industrial environments.

In Paper V, the Mahalanobis Taguchi System is be utilized for fault detection, isolation and prognostics of centrifugal water pump failures. Unlike, the bearings mentioned in paper IV, the centrifugal water pump has multiple components, each with their own failure modes. In addition, experimental data is provided through multiple sensors, which increases the data collection and processing complexity. In addition, the analysis is performed in the data domain, i.e. directly on the raw data collected from the sensors. The contribution of this paper can be summarized as follows: (1) Uses MTS in order to fuse multi-sensor information into a single system performance metric; (2) Builds upon MTS by introducing the use of MD-based fault clusters for fault isolation; (3) Utilizes the progression of MD values over time in order to estimate the remaining useful life of components; and (4) Applies MTS to the domain of centrifugal pump monitoring, diagnostics and prognostics for the first time to the best knowledge of the authors.

REFERENCES

[1] McFarlane, D., Sarma, S., Chirn, J.L., Wong, C.Y., and Ashton, K. "Auto ID systems and intelligent manufacturing control," Engineering Applications of Artificial Intelligence, vol. 16, pp.365-376, 2003.

[2] Lee, Y.M., Cheng, F., and Leung, Y.T. (2004) "Exploring the impact of RFID on supply chain dynamics," Proceedings of the 2004 Winter Simulation Conference, vol. 2, pp. 1145-1152, 2004.

[3] Brewer A., Sloan N., and Landers, T.L. "Intelligent tracking in manufacturing," Journal of Intelligent Manufacturing, vol. 10, No. 3-4, pp. 245 – 250, 1999.

[4] Michael, K. and McCathie, L. "The Pros and Cons of RFID in Supply Chain Management," International Conference on Mobile Business, pp. 623 – 629, 2005.

[5] McFarlane, D. "Auto ID-based control systems: an overview," IEEE International Conference on Systems, Man and Cybernetics, vol. 3, pp. 6-11, 2002.

[6] Penttila, K., Sydanheimo, L., and Kivikoski, M. "Performance development of a high-speed automatic object identification using passive RFID technology," IEEE International Conference on Robotics and Automation, vol. 5, pp. 4864-4868, 2004.

[7] Zaremba, M.B. and Morel, G. "Integration and control of intelligence in distributed manufacturing," Journal of Intelligent Mfg, vol.14, no.1, pp.25-42, 2003.

[8] Naso D. and Turchiano, B. "A coordination strategy for distributed multi-agent manufacturing systems," Int. J. Prod. Res., vol.42, no.12, pp. 2497-2520, 2004.

[9] Gao, Q., Luo, X., and Yang, S. "Stigmergic cooperation mechanism for shop floor control system," Int. J. Adv. Mfg. Tech., vol.25, pp.743-753, 2005.

[10] Agarwal, S., Katz, R.H., Krishnamurthy, S.V., and Dao, S.K. "Distributed power control in ad-hoc wireless networks," 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, vol.2, pp. F59-F66, 2001.

[11] Vaidya, N., Dugar, A., Gupta, S., and Bahl, P. "Distributed Fair Scheduling in a Wireless LAN", IEEE Transactions on Mobile Computing, vol.4, no. 6, pp. 616-629, 2005.

[12] H. K. H. Chow, K. L. Choy, W. B. Lee, and K. C. Lau, "Design of a RFID case-based resource management system for warehouse operations," Expert Sys. with Apps., vol. 30, no. 4, pp. 561-576, 2006.

[13] X. Lian, X. Zhang, Y. Weng, Z. Duan, and A. Z. Duan, "Warehouse logistics control and management system based on RFID," Proc. of the IEEE Int. Conf. on Automation and Logistics, pp. 2907-2912, 2007.

[14] S. Han, H.Lim, J. Lee, "An efficient localization scheme for a differential-driving mobile robot based on RFID system," IEEE Trans. Ind. Elec., vol. 54, no. 6, pp. 3362-3369, 2007.

[15] Gramling, Kathryn, Anthony Bigornia and Tig Gilliam, "IBM Business Consulting Services EPC Forum Survey," MIT Auto-ID Center White Paper, 2003.

[16] R. H. Clarke, D. Twede, J. R. Tazelaar and K. K. Boyer, "Radio frequency identification (RFID) performance: the effect of tag orientation and package contents," Packaging Technology and Science, vol. 19, no. 1, pp. 45-54, 2006.

[17] C. J. Li, and S. M. Wu, "On-Line Detection of Localized Defects in Bearings by Pattern Recognition Analysis," ASME Journal of Engineering for Industry, vol. 111, no. 4, pp. 331-336, 1989.

[18] T. A. Harris, Rolling Bearing Analysis Fourth Edition, New York, NY 10158: Wiley-Interscience, 2001.

[19] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-Network-Based Motor Rolling Bearing Fault Diagnosis," Industrial Electronics, IEEE Transactions on, vol. 47, no. 5, pp. 1060-1069, 2000.

[20] S. Janjarasjitt, H. Ocak, and K. A. Loparo, "Bearing Condition Diagnosis and Prognosis Using Applied Nonlinear Dynamical Analysis of Machine Vibration Signal," Journal of Sound and Vibration, vol. 317, no. 1-2, pp. 112-126, 2008.

[21] P. W. Tse, Y. H. Peng, and R. Yam, "Wavelet Analysis and Envelope Detection for Rolling Element Bearing Fault Diagnosis---Their Effectiveness and Flexibilities," Journal of Vibration and Acoustics, vol. 123, no. 3, pp. 303-310, 2001.

[22] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S. Liang, "Adaptive Prognostics for Rolling Element Bearing Condition," Mechanical Systems and Signal Processing, vol. 13, pp. 103-113, 1999.

[23] E. Lihovd, T. I. Johannessen, C. Steinebach, and M. Rasmussen, "Intelligent Diagnosis and Maintenance Management," Journal of Intelligent Manufacturing, vol. 9, no. 6, pp. 523-537, 1998.

[24] O. Coutier-Delgosha, R. Fortes-Patella, J. L. Reboud, M. Hofmann, and B. Stoffel, "Experimental and Numerical Studies in a Centrifugal Pump with Two-Dimensional Curved Blades in Cavitating Condition," Journal of Fluids Engineering, vol. 125, no. 6, pp. 970-978, 2003.

[25] N. R. Sakthivel, V. Sugumaran, and S. Babudevasenapati, "Vibration Based Fault Diagnosis of Monoblock Centrifugal Pump Using Decision Tree," Expert Systems with Applications, vol. 37, no. 6, pp. 4040-4049, 2010.

[26] A. K. S. Jardine, D. Lin, and D. Banjevic, "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance," Mechanical Systems and Signal Processing, vol. 20, no. 7, pp. 1483-1510, 2006.

[27] Y. Li, T. R. Kurfess, and S. Y. Liang, "Stochastic Prognostics for Rolling Element Bearings," Mechanical Systems and Signal Processing, vol. 14, pp. 747-762, 2000.

[28] P. Wang, and G. Vachtsevanos, "Fault Prognostics Using Dynamic Wavelet Neural Networks," AI in Equipment Service, 2000.

[29] H. Qiu, and J. Lee, "Feature Fusion and Degradation Using Self-Organizing Map," Proceedings Machine Learning and Applications, pp. 107-114, 2004.

[30] R. B. Chinnam, and P. Baruah, "A Neuro-Fuzzy Approach for Estimating Mean Residual Life in Condition-Based Maintenance Systems," International Journal of Materials and Product Technology, vol. 20, no. 1/2/3, pp. 166-179, 2004.

PAPER

1. A TESTBED ARCHITECTURE FOR AUTO-ID TECHNOLOGIES

1.1. INTRODUCTION

The advent of Auto-ID technology has enabled electronic labeling and wireless identification of objects, which facilitates real-time product visibility and accurate tracking at all levels of the product life cycle (McFarlane et al., 2003). From supply chain level business processes to shop floor level manufacturing execution, this technology presents many opportunities for process improvement and re-engineering (Brewer, Sloan, and Landers, 1999; Lee et al., 2004; Michael and McCathie, 2005). On the other hand, it also presents many challenges due to lack of standards and roadmaps to transform Auto-ID technologies into Auto-ID solutions (Engels et al., 2001; McFarlane 2002; Penttila et al., 2004). In spite of its potential advantages, the major challenge is how to manage such voluminous data in a timely fashion. If this can be achieved, then "information" can replace "inventory" on the shop floor.

A typical Auto-ID application requires effective integration of two components. First, business processes (i.e., decision-making models) must be re-engineered to encapsulate Auto-ID data. For instance, the amount of safety stocks, which are typically used in traditional inventory management models, can be dramatically reduced if Auto-ID data can be incorporated into the model. The second component is the networking topologies and related scheduling/routing protocols. In order for the decision-making component to be realistic, the network must facilitate effective and efficient routing of

Auto-ID data packets. Academic studies usually focus either on manufacturing-specific decision-making (manufacturing engineering and industrial engineering) (Zaremba and Morel, 2003; Naso and Turchiano, 2004; Gao et al., 2005) or on networking (electrical and computer engineering) (Jacquet et al., 2001; Agarwal et al., 2001; Vaidya et al., 2005). Due to the gap between these two sub-components, namely decision-making and networking, the solutions provided to the industry are not directly applicable; further testing on Auto-ID technologies within the proposed solution is usually necessary in order to fine-tune it to the production environment.

1.2. UMR'S AUTO-ID TESTBED

An Auto-ID testbed has been established at the University of Missouri-Rolla (UMR) with the objective of providing viable solutions to industry by developing architectures, methodologies, and tools that include both decision-making and networking components, and facilitate effective and efficient collection and use of Auto-ID data for manufacturing applications.

The generic approach developed in the Testbed to provide a viable solution, as depicted in Figure 1.1, includes two major activities so that the research and development work can be carried out in realistic and accurate conditions. In the Decision-Making Module, the current business process is analyzed. After initial data collection about the process, a simulation model is developed in order to carry out "what-if" scenarios. As a result of this approach, various alternative decision-making models that rely on a certain level of Auto-ID data are developed. The information flow required by each model is then communicated to the Networking Module, which mimics, through network

simulation, the possible load on the network. In this module, various data scheduling and routing protocols are developed and tested using simulation. These two modules, Decision-Making and Networking, provide an in-depth analysis of all possible solutions. The most reliable networking scheme and the most promising decision-making model are then combined together for hardware-in-the-loop testing. Finally, a viable solution emerges after the hardware-in-the-loop testing is successfully completed. The solutions developed and validated in UMR's Auto-ID Testbed include a decision-making component and a networking component that supports the industrial application.

In order to avoid long development lead times while implementing the approach depicted in Figure 1.1 for different industries, modular software and hardware components have been developed, which can be integrated easily so that the testbed environment can be re-configured in a timely manner in order to satisfy the need for the particular industrial application.

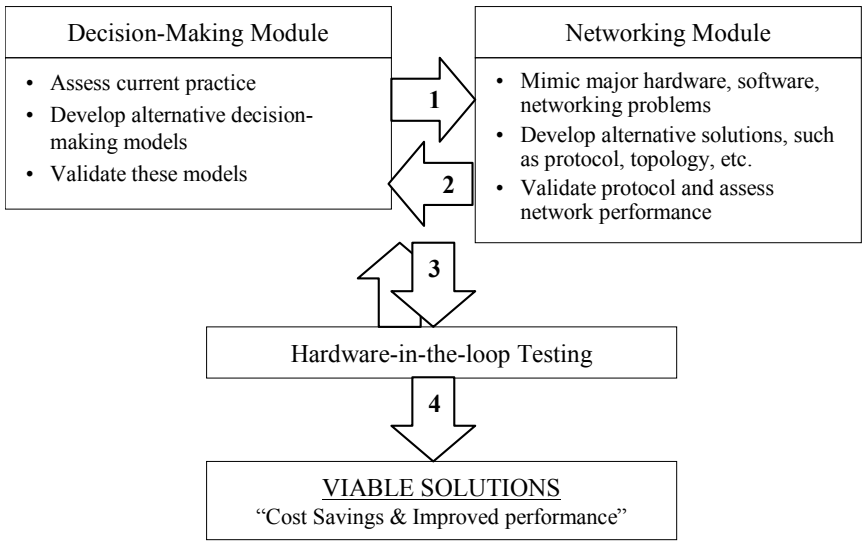

Figure 1.1. UMR's Auto-ID testbed's approach to providing viable solutions

1.3. HARDWARE-IN-THE-LOOP SIMULATION METHODOLOGY

The fundamental motivation behind establishing an Auto-ID testbed at UMR stemmed from the fact that in order to demonstrate the potential benefits of integrating Auto-ID technologies in a manufacturing environment, a testbed needs to be equipped with a variety of industry-grade production equipment and Auto-ID hardware and software tools, which is not a realistic expectation due to two reasons: Such equipment (1) have a very high overall cost and are not usually flexible enough to be utilized in a variety of different configurations for a variety of different industrial applications, therefore they are not cost effective; and (2) do not allow for expedited testing (i.e., testing of different alternative models in a very short time). Therefore, there is a need to develop "virtual" hardware that can mimic the real hardware in a cost-effective way and also allow for developing simulated production scenarios for expedited testing. An additional advantage of such virtual models is that they can be run on distributed computers, similar to a realistic industrial setting, which then allows for networking and communications related testing.

UMR's Auto-ID Testbed is equipped with virtual models that have the capability to communicate with real hardware, which leads to the concept of hardware-in-the-loop simulation models. Such an approach provides a dynamic, controlled testbed environment for developing, testing, and evaluating Auto-ID systems, and it enables development of guidelines for technology transfer.

Virtual models can be used to mimic actual hardware and they can generate a large amount of data, which are essential for developing realistic solutions to industrial problems. For example, data generated by using such a combination of actual hardware

and virtual models can provide a basis for low-cost experimentation on network overloading. In order to make effective decisions with the data generated in a hardware-in-the-loop simulation environment, data must be properly communicated among decision-makers (i.e., controllers), therefore reliable data scheduling and routing schemes are required for such reliable data transfer.

Most real-world systems are too complex to allow realistic models to be evaluated analytically. Traditionally, simulation is used for modeling of such systems where a simple closed form analytical solution through the use of a mathematical model is not possible due to the highly stochastic natures of these systems (Law and Kelton, 1991). In a typical simulation study, a computer is used to evaluate a model numerically, and data are gathered in order to estimate the desired true characteristics of the model. In addition to its typical use, at UMR's Auto-ID Testbed, simulation is used in three other ways:

(1) **Virtual Model (VM) Development**: Simulation is used to develop virtual hardware/equipment models that can communicate with real hardware. When such virtual models are validated, they can be duplicated to form a more complicated production environment, where they can interact with each other, as well as with other hardware, and run in real-time (i.e., unlike typical simulation models that are run in accelerated mode) on distributed computing platforms to represent complex behaviors.

(2) **System Controller (SC) Development**: Since simulation packages have capabilities to facilitate decision-making modeling, simulation models can be used to control real hardware/equipment if they are properly designed to

communicate with the hardware (i.e., I/O capability with the equipment's protocol).

(3) **Networking Simulation (NS):** Auto-ID technology potentially leads to generation of large amounts of data, which can overload the network and cause congestion and information loss. Amount and type of data flow is dictated by the design of decision-making models embedded in a system. Therefore, networking performance, in the presence of such voluminous data transfer, must be evaluated in order to make sure that the decision-making models perform as expected. Network simulators are used to construct virtual networks of varying sizes based on the Auto-ID devices that are being considered for deployment with alternative topologies to find suitable solutions for a particular Auto-ID environment. In addition, network simulators are used in the process of evaluating different communication protocols and in the design of new protocols specifically for Auto-ID networks. One example of such a protocol design is to predict the onset of congestion due to Auto-ID data and to mitigate congestion while ensuring that sufficient data will be transferred for decision making.

Simulation packages, by default, are designed to collect data on a variety of events and entities defined by the user. Therefore, VM, SC, and NS development strategies, virtual models and controllers are also used as "system monitors" since they have the informational equivalence of real systems, which justifies the approach adopted by the UMR's Auto-ID Testbed. In addition, distributed simulation models provide advantages in modeling the behavior of Auto-ID devices embedded in an application that consists of geographically distributed environments.

The hardware-in-the-loop simulation methodology, developed at UMR's Auto-ID Testbed, is built on three integrated levels of development and testing activities. A typical industrial application (referred to as a "*project"* hereafter) goes through those three levels, as described in Sections 2.3.1 through 2.3.3, before a viable solution is generated. Before starting at Level 1, a conceptual system design is carried out; the application domain analyzed, potential Auto-ID tools determined, material and information flow requirements investigated, and new decision-making models that rely on Auto-ID data are developed. These activities are carried out in close cooperation with the industry client. Once an agreement is reached, the 3-level hardware-in-the-loop simulation methodology starts.

1.3.1. Level 1: Controller Simulation. At Level 1, controllers in the industry client's workshop shop floor are modeled individually and their interaction with possible Auto-ID tools is simulated. Such individual models provide insights in terms of the controllers' capability to handle voluminous Auto-ID based data and to timely respond to changes within its domain.

An example Level 1 application is depicted in Figure 1.2. A virtual replica of a pick-and-place robotic cell, with additional RFID features, is developed in the Arena simulation environment (Rockwell Automation product). The programmable logic controller (PLC), which controls the real robotic cell, is integrated with the virtual cell (i.e., the Arena simulation model) via RSLinx, which is a communication server providing plant-floor device connectivity. The Arena model includes virtual RFID antennas and virtual products with RFID tags. The production objective is to sort parts

according to their types, which is determined in the simulation environment based on virtual antennas and tags.

A typical Level 1 application demonstrates the use of virtual Auto-ID data on the shop floor and the use of simulation as real-time manufacturing execution and control system. Level 1 applications are used for developing proof-of-concept decision-making models in the early stages of an industrial application via system simulation on a single computer. Therefore, it does not allow for networking performance analysis.

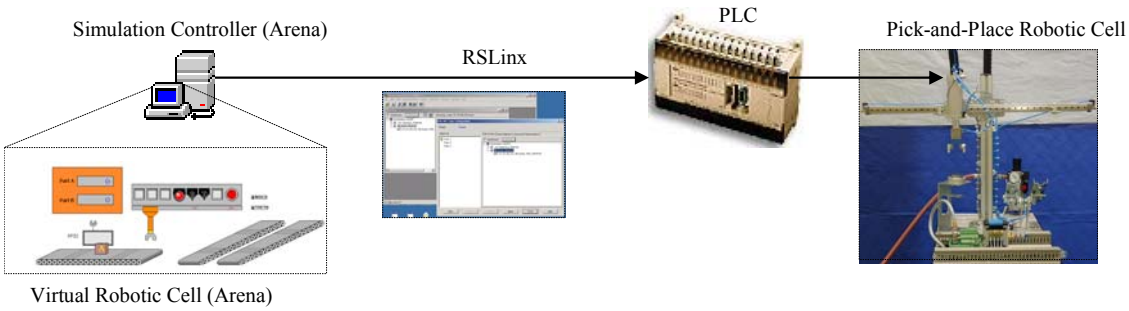

Figure 1.2. Level 1: controller simulation

1.3.2. Level 2: Distributed Controller Simulation. Level 2 is built upon Level 1. Once Level 1 applications are developed on individual computers or mobile devices are verified, they are integrated with each other based on material and information flow determined at the conceptual design stage of a project. Such integration requires communications network design, selecting alternative topologies, and necessary data scheduling and routing protocols. Level 2 allows for experimenting with distributed Level 1 applications that have virtual Auto-ID tools. At this level, alternative decision-making models, along with their communications requirements, can be tested and benchmarked.

Therefore, an initial assessment of the overall system, including a network performance analysis, can be carried out.

Different networking topologies that connect decision-making models, which reside on distributed computers or mobile devices, can be tested. Once the decision-making models are validated in terms of system-level performance measures, a series of different network topologies and communication protocols can be evaluated to identify satisfactory solution for such decision-making models. This is first done by prototyping the decision-making network with existing modular hardware and software. For example, at UMR's Auto-ID Testbed, a set of wireless nodes are built for evaluating different wireless networks and protocols; these plug-and-play nodes include Bluetooth, 802.11, Zigbee, and 900MHz RF radio, each with varying capabilities and limitations in terms of data transfer rate. Such modular devices allow for validation of small scale communication network and provide sufficient information on simulating larger scale networks. If it is not possible to develop a satisfactory networking topology in order to meet the networking performance requirements, referred to as quality of service (QoS), defined in terms of end-to-end delay, information loss, throughput, and delay variation, alternative networking topologies and decision-making models are sought and tested.

Different data scheduling and routing schemes are required in order to meet the user defined QoS, which affects the decision-making models. Therefore, the feedback/revision loop (as shown in Figure 1) between the system simulation and networking simulation continues until satisfactory results, both at system and networking levels, are achieved.

Figure 1.3 depicts an example Level 2 application developed at UMR's Auto-ID Testbed. The application includes a virtual cell and an actual robotic cell. A decision-making model, developed in Arena, is added on these two cells as a higher-level controller with the objective of demonstrating a supervisory control scheme. This application demonstrates the interaction among the three computers/controllers and allows for investigating the networking performance in addition to decision-making model analysis.

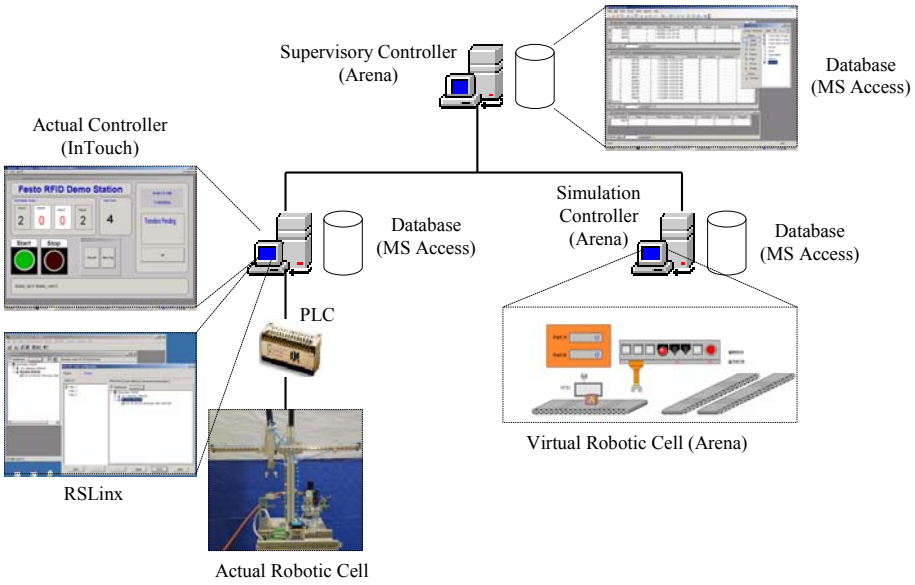

Figure 1.3. Level 2: distributed controller simulation

1.3.3. Level 3: Distributed Control Simulation with Hardware-in-the-loop.

At Level 3, some of the virtual (i.e., simulated) Auto-ID tools developed at Level 2 are replaced with actual Auto-ID hardware/software. Since full scale (i.e., industry level) implementation is not possible at testbed level, a Level 3 application still contains various simulated Auto-ID tools in addition to several actual Auto-ID tools. At Level 3, the

overall objective is to demonstrate how a variety of technology tools, including Auto-ID, can be effectively integrated in a distributed environment built on the modules that have been developed at Level 1 and Level 2.

An example Level 3 application is shown in Figure 1.4. The application includes two physically distributed (i.e., communication is provided via the Internet) production areas; one of them is a virtual (simulated) manufacturing system and the other one is a system with actual RFID antennas and parts with RFID tags. The objective is to monitor the inventory levels at two production areas and to minimize stock-outs by allocating inventory from one area to the other. As shown in the figure, a supervisory controller monitors the status of the two production areas via RFID data and aims to balance the inventory levels. When a critical point in inventory level is reached, the supervisory controller sends a message to the PDA of the inventory dispatcher via wireless communication alerting him of the unbalance between the production areas. In addition to demonstrating how various technology tools can be effectively integrated, this application shows how the inventory visibility provided by real-time RFID data can enhance decision-making on the shop floor.

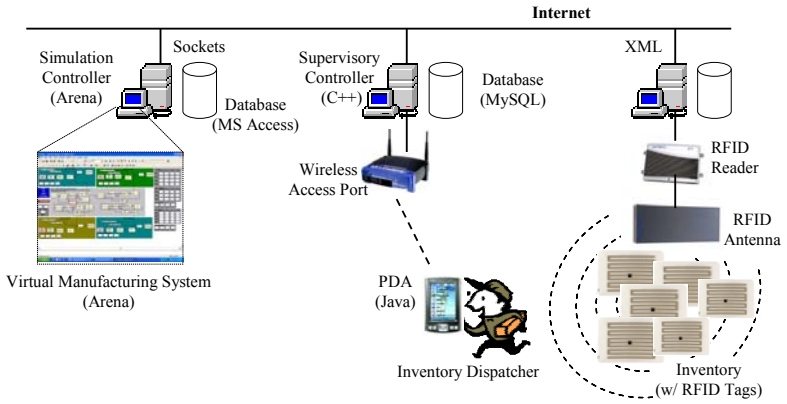

Figure 1.4. Level 3: Distributed controller simulation with Auto-ID hardware-in-the-loop

Most Auto-ID applications in industry are carried out in a turn-key manner, which usually leads to (1) several revisions after implementation, (2) loss in production, (3) purchasing more hardware than necessary, and (4) interconnectedness between legacy systems and the new Auto-ID system due to the unforeseen problems at the design stage. Unlike such an approach, UMR's Testbed provides an evolutionary approach that facilitates incremental testing and evaluation through three integrated levels of the hardware-in-the-loop simulation methodology for validating system and networking-level models so that viable solutions, with shorter implementation lead times, can be developed.

1.4. CASE STUDIES

In this section, two case studies, that have been developed using the hardware-in-the-loop simulation methodology, are presented: (1) RFID data-driven shop floor control, and (2) Adaptive inventory management of time-sensitive materials using RFID data. Through these case studies, the objective is to highlight the integration of decision-making models and the network topology/scheduling/routing models with emphasis on the testbed design.

1.4.1. RFID Data-Driven Shop Floor Control. At UMR's Auto-ID Testbed, a shop floor environment that consists of dock doors, automated guided vehicles, a conveyor system, an automated storage and retrieval system (AS/RS), and an assembly area has been equipped with RFID systems in order to test a variety of decision-making models and networking protocols in the presence of RFID data, as shown in Figure 1.5. A programmable logic controller (PLC) and three computers, which are a cell control

computer, an RFID middleware computer, and a production planning & control (PP&C) computer, are used to run the system.

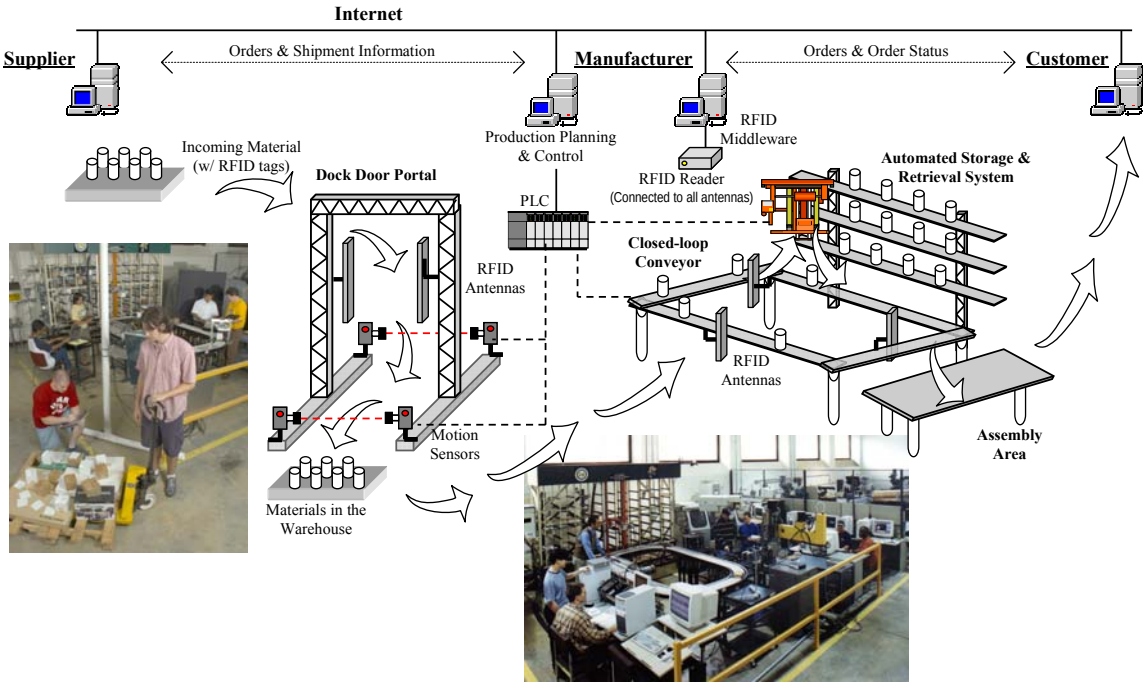

Figure 1.5. RFID data-driven shop floor control

This application is the first step towards a bottom-up RFID implementation for RFID data-based supply chain management. The ultimate objective is to demonstrate how RFID data can be used to streamline supply chain-level business processes integrated with shop floor operations. At its current stage, there are three actors in the environment: Supplier, Manufacturer, and Customer. Each actor is provided with php-based Web pages that facilitate interaction with each other. The supplier receives orders for raw materials automatically from the manufacturer when the baseline inventory level drops below a certain value. The customer site can be used by several customers, similar

to a typical Internet-based shopping site. Each customer can place an order, which includes an assembly of certain number of raw materials stored in the AS/RS.

The focus of this section is on the manufacturer's facility, which includes industry-size production equipment: dock doors, representing its warehouse, a closed loop conveyor that transports raw materials to/from an AS/RS, and an assembly area. The dock door and the conveyor are equipped with RFID antennas, which are connected to an RFID reader. A PLC is used to control the shop floor operations, including the conveyor, AS/RS, and the motion sensors on the dock door to detect inbound and outbound material.

A cell controller computer is used to serve the following purposes: (1) Act as an HMI (human-machine interface) for the operator(s); (2) Execute the control logic that governs the production cell; (3) Store local events in a database; and (4) Handle communication to/from the PLC by means of an OPC server and XML-DA gateway. On another computer an RFID middleware package resides, which serves as a filter of RFID data and facilitates forwarding the data to the correct recipient (system/software).

A production planning and control computer is used to place orders to the suppliers, receive orders from customers, and monitor material flow within the facility. The various computers in this arrangement communicate over the Intranet, which is typically the case in many industrial settings due to the geographical locations of facilities.

In the following sub-sections, problems, challenges, and solution strategies that have been developed to solve those RFID technology-related problems are discussed.

1.4.1.1. Shipping and receiving operations at the dock door. UMR's Auto-ID testbed has two dock doors, which are equipped with RFID antennas, light stacks, and motion sensors. The shipping and receiving operations are carried out by pallet jacks and automated guided vehicles.

Initial implementation was done using pallet jacks, which require manual handling of tagged materials placed on pallets, to investigate read rates. The second implementation was carried out using AGVs that also require wireless communications, which adds more complexity to the overall system. The third application involved operating two dock doors side by side, similar to an actual warehouse setting, in order to investigate the interference and cross-reading of material, which refers to the case where both dock-doors read each other's material due to the range of the RFID antennas, and location and identity information cannot be matched. As a result of the third implementation, an RFID data-based AGV traffic management model was developed.

The basic dock door operation involves the following steps. Each material has an RFID tag and all materials are placed on an industrial pallet. First, the motion sensors detect the direction of the pallet (i.e., inbound or outbound material flow). The information is sent wirelessly to an RFID server (i.e., middleware). It turns on the red light on the light stack, in case of manual material handling, for the dispatcher to stop or it stops the AGV, in case of automated (i.e., unmanned) material handling, at the dock door. Then, the server initiates the RFID tag reading process and passes the readings to the PP&C computer, which compares the order list with the pallet content to determine if right materials at the right amount have been received.

After reading the tags, there can be a mismatch between the order list and the pallet content, which can be due to two possibilities. First, the pallet content may be correct but the readings, due to read-rate problems, can be wrong. Second, the pallet may include incomplete shipment. In either case, the material flow is not disrupted at the dock door but the inventory database is updated to reflect this anomaly. When the pallet reaches its destination on the shop floor, each material on the pallet is manually placed on the conveyor, which is equipped with RFID antennas. On their way to the AS/RS via the conveyor, the tag on each material is read and checked against the order list in the database. Since it is less likely to have a read-rate problem when you have only one tag, the overall decision-making model that monitors material flow between the dock door and the conveyor facilitates "process visibility" in spite of the read-rate problem associated with RFID technology.

The dock doors use Class-0 900 MHz UHF RFID readers and tags, which are known to be prone to interferences (Engels, 2002). High power devices, such as RFID readers, placed in close proximity cause interferences among themselves and for other RF devices (i.e., wireless transmitter for motion sensor). In dense networks, detection range and read rates of RFID readers suffer severely while other lower power RF devices become completely unusable. Similarly, dock doors located next to each other form a dense RFID reader network, and therefore it is not possible to obtain reliable identification data. In addition, when a pallet, loaded with tagged materials, is moved parallel to the dock doors, the antennas on both dock doors read the tags. In other words, it is not possible to detect whether the material is being moved through a dock door or it is being maneuvered in front of the dock doors.

In order to solve both problems, a “read it when you need it” concept has been developed, which is implemented using an on/off power control protocol. The protocol enables turning on and off RFID antennas intelligently to avoid communication signal interference depending on the status of the motion sensors on dock doors. The concept has been implemented as a forklift (or AGV) traffic management system, as shown in Figure 1.6. The antennas are turned on and off whenever it is needed by making sure that only one dock door is active at any time so that cross-reading is avoided. The power control on antennas and management of forklifts are synchronized by the use of light stacks (i.e., similar to traffic lights).

The interference problem can be expanded to a warehouse scale, where large number of RFID devices may operate together with each having unique specifications according to the applications. For example, for a dock door application, on/off time should be frequent and relatively shorter, especially for field motion sensors. On the other hand, for inventory monitoring, waiting time can be long by considering changes are not sensitive to time while reading time should be also be sufficiently long for large item counts. Therefore, a priority based scheduling algorithm is necessary and it was developed to accommodate a “multi-sensor” environment. The algorithm dynamically schedules waiting and read time intervals based on unique requirements of numerous devices and applications.

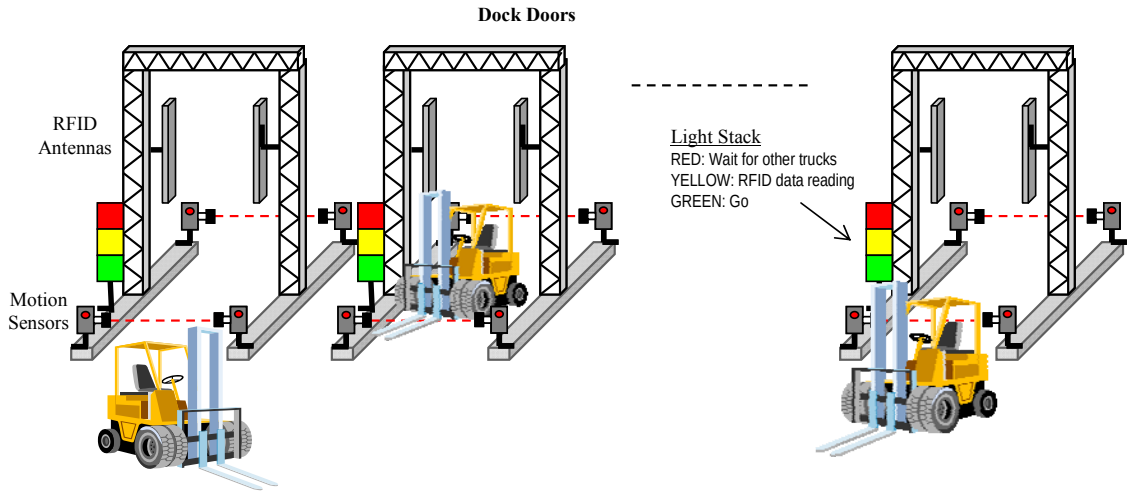

Figure 1.6. An RFID data-based forklift traffic management system

1.4.1.2. Integration of RFID middleware with PLC. At UMR's Auto-ID testbed, the overall objective is to provide seamless data flow among different controllers in order to facilitate visibility on the shop floor. The efforts to integrate the RFID middleware with the PLC at the Testbed showed that an intermediary level of control between the RFID reader/middleware and the equipment control system (PLC) was essential in order to execute RFID data-based events on the shop floor. In this application, the intermediary control facilitates communication between the PLC and the RFID middleware and processes the RFID data received from the middleware, sending commands or data to the PLC, as required by the manufacturing events. For the PLC to be able to react quickly to RFID read events, the data must be transferred throughout the network in a timely manner.

In this application, the conveyor, AS/RS, and dock door sensors are controlled via a GE Fanuc PLC. An OPC server (Kepware) is employed at the cell level to allow communications to/from the PLC and the cell control computer. The cell controller

provides an HMI for the operator, local control logic for the cell, and database storage of local events. A separate server hosts the RFID middleware. Since the middleware is designed to communicate using the XML-DA protocol, an OPC-DA to XML-DA gateway (Kassle) was employed on the cell controller in order to translate between OPC protocols (herein referred to as OPC server).

When a tagged material is placed on the conveyor, it is detected by a sensor on the conveyor line just before passing under an RFID antenna. This sensor is connected as an input to the PLC and its state is externally visible through the OPC server as an OPC tag. An RFID read cycle is utilized due to the OPC tag trigger capability built into the RFID middleware. The RFID middleware continually checks the state of the antenna sensor OPC tag and when a material is sensed, an RFID read cycle commences. This filtered RFID data is then sent from the middleware server through the intranet to the cell control computer. This controlled read cycle helps to reduce interference between multiple RFID antennas in close proximity and also allows the system to detect items on the conveyor with missing or unreadable RFID tags, which strengthens visibility on the shop floor.

Upon receiving the RFID data from the middleware server, the cell control computer then processes the data into information useful at the cell level. The cell control computer is in a unique position, which allows it to communicate with both lower level (PLC) and higher level systems (PP&C). This capability allows the cell control computer to associate the RFID tag read with a material type and transform this information into actions needed to be taken by the PLC. The necessary data is then forwarded to the PLC through the OPC server.

This arrangement of decoupling the equipment control functions (PLC) from the higher level processes lends itself to an evolutionary approach to the implementation of RFID data onto the shop floor. Existing systems and controls can be refitted to make use of RFID data without a major interruption in production.

1.4.2. Adaptive Inventory Management. Inventory management of time-sensitive materials is very critical due to their shelf lives. One primary concern is to insure that the required materials are available at all times to the operators. The lack of time-sensitive materials results in loss of production and in turn loss of profits. On the other hand, those materials that are not used in production within their lives expire and become another cost factor. In addition, disposal of time-sensitive materials, once they reach their shelf life, to prevent their usage on a product is also major concern.

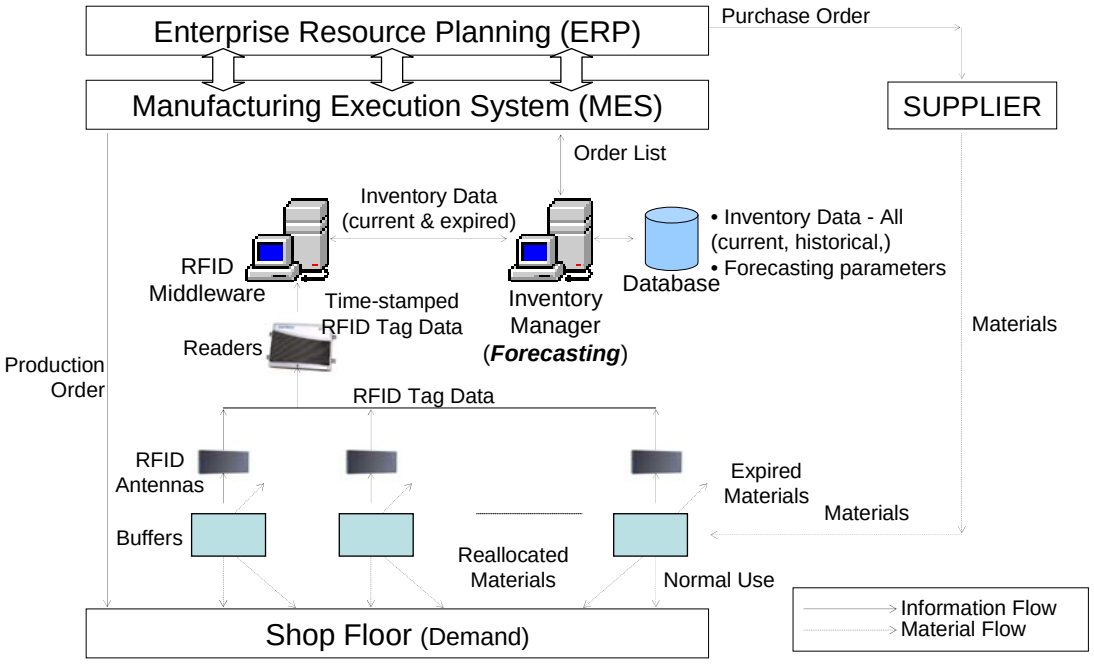

Figure 1.7. Adaptive inventory management based on RFID data

The implementation, as show in Figure 1.7, included two major tasks:

(1) A simulation-based business case that demonstrated the overall cost savings when real-time RFID data were used for inventory management was developed. To achieve this goal, a trend-adjusted exponential smoothing algorithm (i.e., forecasting model) was developed. The algorithm looks at the difference between the current inventory level based on real-time RFID data and the forecast (i.e., predicted demand) in order to determine the amount of material that needs to be ordered. The simulation model included 20 buffers and approximately 6,000 virtual RFID-tagged time-sensitive materials. A gap analysis, which compared the *as-is* practice with the RFID data-based inventory management on the basis of amount of expired, ordered, and re-allocated (among buffers) materials, was carried out. The gap analysis showed that approximately $500K per year would be saved by implementing the proposed model.

(2) A prototype network was built with one real and several virtual RFID readers, multiple databases, RIFD middleware/decision-maker, and a PDA unit. Several scenarios were created to test the basic connectivity and reliability. A larger scale network simulation was then constructed to model the real manufacturing environment. The simulation model involved one hundred RFID readers connected in a star topology, and network traffic was simulated based on real data captured at an RFID reader. The model was then expanded to emulate multiple routes (i.e., duplicate routes), network bandwidth test (i.e., 10/100/1000Mbps), and wireless network test (i.e., 802.11a/b/g). Average delay, packet drop, and

average throughput were used to evaluate different topologies. An example of the network simulation results is shown in Table 1.1.

Table 1.1. Network Simulation Results

Performance Measure	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Link bandwidth	10Mbps	100Mbps	1000Mbps	802.11g (54Mbps)
Packet drop	2923	0	0	345
Average delay (Seconds)	0.237	0.0157	0.0132	0.0179
Link utilization	100%	44.6%	4.6%	99%

1.5. DISCUSSION

UMR's Auto-ID Testbed is currently being used to address various research challenges. It has been experienced that the read rates on RFID antennas vary between 70-80% depending on product types, number of products, and orientation of the products with respect to each other on a pallet. If the read rate problem cannot be solved, then the ultimate goal of RFID, which is "100% visibility on the shop floor", cannot be realized. While the RFID vendors are working on improving the technology, the industry users should understand the fact that for a successful RFID implementation, their processes must be 100% visible. RFID technology cannot magically bring order to a chaotic shop floor environment; the processes must be redesigned to make them lean and visible.

To avoid the above-mentioned potential deficiency of RFID technology, the applications at UMR's Auto-ID Testbed have been designed to operate on various check points so as to improve the visibility of the overall process. For instance, when raw materials are received at the dock door, the antennas read the RFID tags on the materials

while they are on a pallet. If the pallet content does not match the order list, the pallet is not rejected immediately since the problem can be caused by the read rate of the antenna, not necessarily due to a mismatch between the order list and the pallet content. The raw materials are then transported to the production area and each material is loaded on a pallet on the conveyor. Each pallet goes through two RFID antennas to verify its identity. After each material is identified, it is stored in the AS/RS. If a material cannot be identified after it is circulated on the conveyor twice, then it is pushed out of the conveyor for manual identification. If there is an unknown material, then it is sent back to the supplier. Similarly, if there are materials missing on the pallet, then the supplier is notified.

Another potential problem with RFID antennas is that when they are placed in closed proximity, a tag can be read by multiple antennas. In this case, the actual location of the tag (i.e., the product) cannot be determined. For example, when the dock door is placed very close to the conveyor, the antennas on the dock door and the conveyor can read "all" the tags in the environment; a tag on a pallet at the dock door simultaneously appears on the conveyor. To avoid this problem, the environment must be equipped with additional sensors. The testbed demonstrates a lean approach: "read it when you need it". Unless there is an event that requires automatic identification, then the antennas should not be powered. Event detection can be done by integrating sensors into the environment and interpreting the sensor data together with RFID data. Therefore, it is important to scrutinize RFID implementations as potential "multi-sensor" applications, where such sensors need to be carefully determined in order to improve the visibility of the business process.

1.6. CONCLUSIONS

UMR's Auto-ID Testbed provides a flexible test environment that can be used for evaluating Auto-ID technologies and solutions for industry. The testing environment, as well as the testing conditions, can be varied easily in order to match the needs of the industry. The testbed can be used as a low-cost technology development and assessment platform (proof-of-concept) due to the availability of reconfigurable models, which facilitate rapid model development, hardware/software benchmarking, and implementation.

UMR's Auto-ID Testbed provides viable solutions to industry due to its integrated approach; decision-making models and hardware/software/networking issues are considered simultaneously. The testbed allows the researchers and industry practitioners to investigate the operational limits and possible failure behaviors of a variety of Auto-ID technologies. Especially, the effective use of simulated scenarios in the presence of hardware leads to realistic solutions that are greatly valued by industry.

The distributed hardware-in-the-loop simulation environment at the UMR's Auto-ID Testbed is intended to investigate, with increasing level of system complexity, the potential benefits of Auto-ID technologies using simulation.

1.7. REFERENCES

Agarwal, S., Katz, R.H., Krishnamurthy, S.V., and Dao, S.K. (2001) "Distributed power control in ad-hoc wireless networks," 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Vol.2, pp.F59-F66.

Brewer A., Sloan N., and Landers, T.L. (1999) "Intelligent tracking in manufacturing," Journal of Intelligent Manufacturing, Vol. 10, No. 3-4, pp. 245 – 250.

Engels, D.W. (2002), "The reader collision problem," MIT-AUTOID-WH-002, MIT Auto-ID Center.

Engels, D.W., Foley, J., Waldrop, J., Sarma, S.E., and Brock, D. (2001), "The networked physical world: an automated identification architecture," Proceedings of the Second IEEE Workshop on Internet Applications (WIAPP 2001), pp. 76-77.

Gao, Q., Luo, X., and Yang, S. (2005) "Stigmergic cooperation mechanism for shop floor control system," Int. J. Adv. Mfg. Tech., Vol.25, pp.743-753.

Jacquet, P., Muhlethaler, P., Clausen, T., Laouiti, A., Qayyum, A., and Viennot, L. (2001) "Optimized link state routing protocol for ad hoc networks," Proceedings of the IEEE International Multi Topic Conference, pp.62 – 68.

Law, A.M. and Kelton, W.D. (1991) "Simulation Modeling and Analysis," McGraw Hill, New York.

Lee, Y.M., Cheng, F., and Leung, Y.T. (2004) "Exploring the impact of RFID on supply chain dynamics," Proceedings of the 2004 Winter Simulation Conference, Vol. 2, pp. 1145-1152.

McFarlane, D. (2002), "Auto ID-based control systems: an overview," IEEE International Conference on Systems, Man and Cybernetics, Vol. 3, pp. 6-11.

McFarlane, D., Sarma, S., Chirn, J.L., Wong, C.Y., and Ashton, K. (2003) "Auto ID systems and intelligent manufacturing control," Engineering Applications of Artificial Intelligence, Vol. 16, pp.365-376.

Michael, K. and McCathie, L. (2005) "The Pros and Cons of RFID in Supply Chain Management," International Conference on Mobile Business, pp. 623 – 629.

Naso D. and Turchiano, B. (2004) "A coordination strategy for distributed multi-agent manufacturing systems," Int. J. Prod. Res., Vol.42, No.12, pp.2497-2520.

Penttila, K., Sydanheimo, L., and Kivikoski, M. (2004) "Performance development of a high-speed automatic object identification using passive RFID technology," IEEE International Conference on Robotics and Automation, (ICRA '04), Vol. 5, pp. 4864 – 4868.

Vaidya, N., Dugar, A., Gupta, S., and Bahl, P. (2005) "Distributed Fair Scheduling in a Wireless LAN," IEEE Transactions on Mobile Computing, Vol.4, No. 6, pp. 616 – 629.

Zaremba, M.B. and Morel, G. (2003) "Integration and control of intelligence in distributed manufacturing," Journal of Intelligent Mfg, Vol.14, No.1, pp.25-42.

2. ADAPTIVE INVENTORY MANAGEMENT USING RFID DATA

2.1. INTRODUCTION

The advent of automated identification (Auto-ID) technology has enabled electronic labeling and wireless identification of objects, which facilitates real-time product visibility and accurate tracking at all levels of the product life cycle [1]. From supply chain level business processes to shop floor level manufacturing execution, this technology presents many opportunities for process improvement and re-engineering [2-4]. On the other hand, it also presents many challenges due to lack of standards and roadmaps to transform Auto-ID technologies into Auto-ID solutions [5-7]. In spite of its potential advantages, the major challenge is how to manage such voluminous data in a timely fashion. If this can be achieved, then "information" can replace "inventory" on the shop floor.

The objective in inventory management is to have the necessary inventory items at the right time, at the right amount, and at the right place. Since not all inventory items are of high value, tight tracking of them may not be desired. On the other hand, critical inventory items may require a tighter control. Due to the variety of items in a typical manufacturing environment, inventory items can be categorized in different classes, such as ABC classification, since tracking of all inventory items may not be economically viable. When implemented properly, Radio Frequency Identification (RFID) technology can provide item-level visibility in which electronic tags programmed with unique identification information are attached to "objects" that need to be monitored, tracked, or

identified easily when needed; this has been an unreachable region of manufacturing information domain.

Today's inventory management systems operate with Stock Keeping Unit (SKU) level data, which means that hundreds of products can fall under the same SKU that associates brand, size, flavor, etc. and essentially categorizes them under a product type. While SKU-level data provide aggregate inventory levels for inventory items, it falls short when visibility of each item is desired. RFID technology facilitates Electronic Product Codes that provide unique identification for each product inside the enveloping SKU or inventory category [8]. This type of data helps associate production events with each inventory item, which facilitates tighter inventory control that relies on such real-time data.

Organizations that understand these possibilities have become trend-setters by establishing a January 2005 mandate for suppliers, which require RFID implementation on container/pallet size inventory. Such organizations include Wal-Mart and the Department of Defense (DOD). The focus of these mandates are twofold; not only do they force companies to take quick action in fulfilling requirements, but also to look towards the future with overall cost savings in mind. The need to organize and make decisions based on the data provided by the RFID tags is prominent.

To date, research that conjoins RFID technology and item-level inventory management on the shop floor is at a preliminary stage, only inferring benefits upon application. The challenge is to collect RFID data in a timely manner, to process such voluminous data, and to make timely decisions that are tied into manufacturing execution systems. If the challenge is overcome, then the benefits such as waste elimination,

inventory reduction, automatic replenishment, stock-out reduction, and overall cost savings can be easily realized. Therefore, there is a need for RFID data-based effective decision making algorithms that can lead to such benefits. In this study, a forecast-integrated inventory management model that relies on real-time RFID data is presented. The goal of this research is threefold: (1) to model and analyze the decisions made to manage inventory levels of time sensitive materials in a shop floor manufacturing environment; (2) to investigate new decision making algorithms that substantiate the use of RFID for real-time visibility; and (3) to explore the impact of fundamental control parameters on performance measures.

This paper is organized as follows. A brief literature survey is presented in Section 2.2. Section 2.3 discussed the problem definition, model development, and performance measures. Experimentation is presented in Section 2.4, which includes detailed comparison of inventory models based on statistical analysis. Finally, conclusions are given in Section 2.5.

2.2. LITERATURE REVIEW

The momentum behind RFID is growing. With increased awareness and improved technology, RFID systems are being implemented throughout a surprisingly wide variety of industry sectors. This trend is particularly prevalent within supply chain management where limitations exist due to the current traditional methods used. These limitations include the inaccuracies of product identification throughout the product life cycle, the daunting and difficult task of actually tracking the product, and uniquely identifying each part rather than product type [1]. This leads to the realization of specific

benefits of using RFID within the manufacturing supply chain, such as product tracking accuracy, reduced inventory levels and automated inventory management. McFarlane [9] goes further to explain a potential long term benefit with respect to holonic behavior such as customized products and self-organizing production, distribution and inventory systems. He also explains in more detail the importance of product identity to enhance the visibility or observability of a product through its life cycle. Chappell et al. [10] give an in-depth look at the business potential of Auto-ID on manufacturing, as well as other associated benefits. These include increasing capacity utilization and yield, reducing cycle time, increasing labor productivity, improving product quality, preventative maintenance, reducing product obsolescence costs, tracking and managing spare parts inventory, facilitating statistical process control, enabling lot/batch track and trace, ensuring worker safety, reducing returns and warranty claims, and reducing scrap.

Another perspective on RFID and supply chain is established by Brewer, Sloan, and Landers [2], where more of the same benefits are explained. In addition to the advantage of RFID, Brewer et al. also discuss the potential coupling of RFID with GPS to provide more information about products in transit. Moran, Ayub, and McFarlane [11] take a closer look into process improvement within a retail company via the "Use Case Approach," described in more detail under Moran, McFarlane, and Milne [12]. The emphasis of their research is directed towards improvement of processes directly related to product handling, storage and exchange; including product reorder, product receipt and put-away, product recount, replenishment, and reverse logistics. The potential benefits of implementing RFID technology through those specific areas comprise of the increase in productivity, reduction in cost, improvement in service, and strategic value.

Additionally, Moran, McFarlane, and Milne [12] explain that increased production occurs due to increased inventory visibility that helps to utilize warehouse facilities and assets, as well as, improved stock outs and more accurate reordering. Cost reduction is associated with the detection of shrinkage and reduction in waste due to improved product rotation. Garcia et al. [13] also present the benefits of RFID technology in assorted material handling systems within the supply chain. Their analysis brings about techniques and solutions, as well as examples of implementing these systems.

As it can be seen throughout the research the management of inventory should be robust yet sensitive enough to change with respect to varying demand patterns. This type of agility is important to all types of organizations and systems. There are many different types of methods that can be used to predict that type of behavior but it is important to tailor the type of forecasting technique or algorithm to the current or desired system.

2.2.1. Problem Definition. Inventory management of time-sensitive materials is very critical due to their shelf lives. One primary concern is to insure that the required materials are available at all times to the operators. The lack of time-sensitive materials results in loss of production and in turn loss of profits. On the other hand, those materials that are not used in production within their lives expire and become another cost factor. In addition, disposal of time-sensitive materials, once they reach their shelf life, to prevent their usage on a product is also major concern.

The inventory models that are developed and compared in this paper has been developed for a manufacturing company based on its shop floor that consists of 18 operator stations with each having a storage area to store time-sensitive materials. There are 23 different types of time-sensitive materials, each containing an identification

number, container size, lot number, manufacture date, and expiration date. The baseline (BL) inventory capacity for each storage area is established by consulting the operators in that specific area and determined by an estimation based upon the quantity of time-sensitive materials used in production within a week. Each storage area contains different quantities and types of time-sensitive materials, specifically tailored to that particular job, and has a capacity of 300 units of time-sensitive materials.

Expeditors, the personnel responsible for inventory management on the shop floor, manually check each storage area, dispose of expired inventory items, and generate a purchase order based on the difference between the baseline and the current inventory. The lead time between placing the orders and receiving the time-sensitive materials is in the range of 3-5 days. When the shipment arrives the expeditor distributes the materials to one storage area at a time. The operators can use multiple materials at a time in order to perform their individual assigned tasks. During an unexpected increase in production demand, a storage area can run out of necessary materials and the operator may obtain the required material(s) from a neighboring storage area, which is called as reallocation.

The primary objective of the manufacturing company was to meet the production demand at the highest possible rate (i.e., maximize service level) while avoiding excessive waste due to expired materials.

2.2.2. Model Development. After several site visits to the manufacturing company and conducting interviews with operators, expeditors, and managers, three inventory management scenarios were developed. The scenarios, including the current practice at the manufacturing company, are as follows:

$SC_{Manual,BL}$: This scenario includes the existing manufacturing environment and directly mimics the current inventory management practice, which is heavily manual. The current practice relies on baseline (BL) inventory levels that have been fixed prior to production for each storage area. On a weekly basis, the inventory level at each storage area is checked and an order is placed to bring the inventory level back to its baseline level. $SC_{Manual,BL}$ is used as a reference for comparing the other three scenarios.

$SC_{RFID,BL}$: In this scenario, it is assumed that the inventory items are tagged (i.e., RFID tags); therefore there is theoretically 100% visibility in terms of inventory levels. Fixed BL values, similar to $SC_{Manual,BL}$, are used for ordering purposes. The objective is to determine how much value RFID data can add if it is implemented only for monitoring purposes, not for dynamically adjusting the inventory levels.

$SC_{RFID,BL/2}$: Similar to the first two scenarios, this scenario is also built on fixed baseline inventory levels. The only difference is that the baseline inventory levels are cut in half in order to investigate the effect of reduced inventory level. For instance, this scenario should allow for tighter inventory control due to real-time product visibility since all inventory items are tagged but would require frequent reallocations among storage areas since inventory levels are reduced.

$SC_{RFID,\alpha\beta}$: In this scenario, an adaptive inventory scheme, which is based on a forecasting model, has been integrated into a decision-making framework to replace the manual inventory operations and to avoid the inflexibility of fixed baselines. Since RFID, in theory, provides 100% inventory visibility, the only two stochastic parameters within the problem definition are purchasing lead time and production demand. The objective is to utilize the RFID data in the best possible way in order to handle uncertainties in the

purchasing lead time and adapt to fluctuations in the production demand. Therefore, a trend-adjusted exponential smoothing algorithm has been adopted. It uses two parameters, α and β, as coefficients for the average production demand and its trend, respectively. The adaptive inventory scheme looks at the difference between the current inventory level of a particular inventory item at a storage area and the associated forecast (i.e., predicted demand) in order to determine the amount of material that needs to be ordered. The equations for the forecasting algorithm are given below:

$$A_{t,i,j} = \alpha D_{t,i,j} + (1-\alpha)(A_{t-1,i,j} + T_{t-1,i,j}) \quad (1)$$

$$T_{t,i,j} = \beta(A_{t,i,j} - A_{t-1,i,j}) + (1-\beta)T_{t-1,i,j} \quad (2)$$

$$F_{t+1,i,j} = A_{t,i,j} + T_{t,i,j} \quad (3)$$

Where:

$D_{t,i,j}$ = demand in period t for material i at storage area j.

$A_{t,i,j}$ = exponentially smoothed average of the series in period t for material i at storage area j.

$T_{t,i,j}$ = exponentially smoothed average of the trend in period t for material i at storage area j.

α = smoothing parameter for the average ($0 < \alpha < 1$)

β = smoothing parameter for the trend ($0 < \beta < 1$)

$F_{t+1,i,j}$ = predicted demand for material i at storage area j in period t+1

2.2.3. Performance Measures. There are six performance measures derived to evaluate the overall performance of the inventory models. These are defined as expired, normal use, ordered, stock-out, reallocated, and service level. Expired refers to the total

number of time-sensitive materials that expire, normal use is the total number of time-sensitive materials the operators request and are able to use in production, ordered is the total number of time-sensitive materials ordered, stock-out is the total number of time-sensitive materials requested by an operator but were not in stock at the time of the demand, and reallocated is the total number of time-sensitive materials an operator obtains from a neighboring storage area (i.e., not his storage area). Service level is the ratio of normal use to demand, which is a percentage. For instance, a 90% service level means that 90% of the production demand has been met. Therefore, a higher percentage is desired.

2.3. EXPERIMENTATION

Each of the four scenarios ($SC_{Manual,BL}$, $SC_{RFID,BL}$, $SC_{RFID,BL/2}$, and $SC_{RFID,\alpha\beta}$) is implemented as a simulation model on a simulation platform, which includes Rockwell ARENA™ simulation package, Microsoft Visual Basic™, and Microsoft Access™. Production system-related events and entities are modeled in ARENA™. A database structure is developed using Microsoft Access™, and the decision-making modules, as well as the data collection procedures, are implemented in Visual Basic™. The ARENA and Visual Basic modules are integrated to run seamlessly.

In the $SC_{RFID,\alpha\beta}$ model, the levels of the smoothing parameters (α and β) are set to 0.2, 0.5, or 0.8, and 9 sub-models are developed, which includes all the possible level-combinations (3x3) of these parameters in order to explore the impact of α and β settings on the performance.

In each scenario, the production environment includes 18 storage areas and 23 different types of time-sensitive inventory items. Each scenario is run for 10 replications for a total simulation period of 7 months. The amount of time-sensitive inventory items in the system varies between 4,500 and 5,500 at any given time during simulation. The production demand on each inventory item is generated based on a Poisson distribution estimated from the historical data provided by the manufacturing company.

The statistical analysis of the performance obtained in each scenario is carried out using analysis of variance (ANOVA) and paired-t test based on 95% confidence interval. ANOVA is performed on the sub-models of $SC_{RFID,\alpha\beta}$ in order to evaluate the impact of the value of smoothing parameters. First, the best performing sub-model $SC_{RFID,\alpha\beta\_best}$ is selected as a result of ANOVA. Then, $SC_{Manual,BL}$, $SC_{RFID,BL}$, $SC_{RFID,BL/2}$, and $SC_{RFID,\alpha\beta\_best}$ are compared using paired t-test.

2.3.1. $SC_{RFID,\alpha\beta}$ Sub-models. Based on their average number of inventory items among 10 replications, the simulation results of the nine sub-models of $SC_{RFID,\alpha\beta}$ are ranked under six performance measures, as shown in Table 2.1. In order to facilitate better understanding of the relative performance of each sub-model, the results under each performance measure, except for reallocated, are assigned a relative weight from 1 to 9 that shows the rank of the sub-model under that performance measure. The higher the relative weight, the worse the performance of the sub-model (i.e., 1 is the best and 9 is the worst).

RFID enables reallocations among storage areas by providing visibility in terms of inventory levels. As it is shown in Table 2.1, the sub-model with $\alpha = 0.2$ and $\beta = 0.2$ yields the lowest number of reallocations while $\alpha = 0.8$ and $\beta = 0.5$ yields the highest

number. In general, reallocations are done in order to use the existing inventory efficiently. On the other hand, it should be noted that reallocations may be costly if the operators themselves are walking over to a neighboring storage area to pick up an inventory item; it may lead to excessive non-productive operator time. Therefore, the numbers shown in the reallocated column must be taken into consideration with an emphasis on reallocation cost, which has been considered to be proportional to the number of reallocations in this study. Therefore, a lower number of reallocations is desired.

Table 2.1. Forecasting Control Parameters vs. Performance Measures

Control Parameters		**Performance Measures (measured as number of inventory items)**					
α	**β**	**Expired**	**Normal Use**	**Ordered**	**Stock-out**	**Service Level**	**Reallocated**
0.2	0.2	8*	1	8	1	1 (96%)	7,185
0.2	0.5	2	4	1	4	3	14,508
0.2	0.8	5	5	2	5	5	16,700
0.5	0.2	3	3	4	3	4	15,044
0.5	0.5	7	9	3	9	9 (91.5%)	18,175
0.5	0.8	4	7	5	7	7	19,257
0.8	0.2	1	8	6	8	8	18,077
0.8	0.5	6	6	7	6	6	19,903
0.8	0.8	9	2	9	2	2	18,930
Difference**		150	2,900	2,100	2,850	4.5%	13,000

Each level-combination of the control parameters, α and β (i.e., sub-models), yields a different performance for different performance measures. For instance, the sub-model, where α = 0.8 and β = 0.2, yields the best result for *expired*. On the other hand,

the same sub-model ranks number 8 for *normal use*, *stock-out*, and *service-level*. Similar comparisons can be made for the other sub-models. The relationship among the performance measures needs to be taken into consideration when analyzing the results. For instance, the more that is ordered the more inventory items are likely to expire and the less operators will run out of stock because more inventory items are potentially available for production.

In order to select the best α-β combination among nine combinations, it is important to be able to make a thorough analysis of the results across all six performance measures. In other words, it may be misleading to evaluate each performance measure in isolation from others. Therefore, the difference row in Table 2.1 should be taken into consideration. For instance, the column for the number of *expired* inventory items shows only a difference of 150 units of inventory items between the best ($\alpha = 0.8$ and $\beta = 0.2$) and the worst ($\alpha = 0.8$ and $\beta = 0.8$) sub-models, which is much less than the difference given under the *normal use*, *ordered*, and *stock-out* columns, which is in the order of a couple of thousands.

Overall, the proposed forecast-integrated model yields approximately the same amount of expired time-sensitive materials regardless of the α-β levels; as shown with a variation of only 150 units of inventory items. There is, however, a substantial difference among the sub-models with regards to *normal use*, *ordered*, *stock-out*, and *service level*: 2900, 2100, 2850, and 4.5%, respectively. The difference for stock-out is nearly the total amount of inventory items that were unavailable during the entire simulation for $SC_{RFID,\alpha=0.2,\beta=0.2}$, therefore; normal use and service level will have a large difference among level-combination since they are inversely proportional to stock-out. Although

service level is the primary concern, the other performance measures need to be taken into consideration to ensure the system performs as desired at the possible lowest cost. A high amount of inventory items expiring in conjunction to the lowest service level would not be cost effective. Therefore, a balance of the two would lead to a more robust inventory model.

When deciding the best α-β combination, it is important to understand the direct influence they have on the overall forecast. When α is small, the ordering forecast puts less emphasis on D_t and more emphasis on A_{t-1} and T_{t-1}; therefore A_t becomes a larger value resulting in more time-sensitive materials ordered due to a larger forecast value. The reverse is true with α; a larger number would put more emphasis on D_t and less on A_{t-1} and T_{t-1}; therefore making the forecast and ordered time-sensitive materials more responsive to recent demands instead of a previous average and more likely to order less.

When β is small, it places more emphasis on T_{t-1}. Conversely, a larger β would result in the difference between A_t and A_{t-1} to be accentuated. Depending on the value of α and ultimately A_t, this can cause the T_t value be large or small. For example, if α was large, the forecasting calculation would be emphasized by A_{t-1} and T_{t-1} and the final value of A_t would closely resemble them. From there, T_t is calculated with an emphasis on the difference between A_t and A_{t-1} or T_{t-1}. If β is small then the emphasis stays on A_t and A_{t-1}. There is a high probability that the value for A_t would be large if α was small and in conjunction with a small β the forecast would result in a large number.

Overall, since $\alpha = 0.2$ and $\beta = 0.2$ yield the highest service level with the lowest number of reallocations, $SC_{RFID,\alpha=02,\beta=0.2}$ is selected to be the best combination of all nine sub-models (i.e., $SC_{RFID,\alpha\beta\_best} = SC_{RFID,\alpha=02,\beta=0.2}$).

2.3.2. Comparison of $SC_{Manual,BL}$, $SC_{RFID,BL}$, $SC_{RFID,BL/2}$, and $SC_{RFID,\alpha=0.2,\beta=0.2}$

The results of the paired t-test analysis conducted on $SC_{Manual,BL}$, $SC_{RFID,BL}$, $SC_{RFID,BL/2}$, and $SC_{RFID,\ \alpha=0.2,\beta=0.2}$ are shown in Tables 2.2 through 2.4. A complete comparison of the means for each performance measure is given in Table 2.5. The results show that RFID implementation combined with the right decision-making model improves the performance.

Table 2.2. Paired t-test for $SC_{RFID,\alpha=0.2,\beta=0.2}$ versus $SC_{Manual,BL}$

Performance Measure	**$SC_{RFID,\alpha=0.2,\beta=0.2}$**	**$SC_{Manual,BL}$**
Service Level	✓	
Expired	✓	
Normal Use	✓	
Ordered	✓	
Stock-out	✓	
Reallocated	Inconclusive	

Table 2.3. Paired t-test for $SC_{RFID,\alpha=0.2,\beta=0.2}$ versus $SC_{RFID,BL}$

Performance Measure	**$SC_{RFID,\alpha=0.2,\beta=0.2}$**	**$SC_{RFID,BL}$**
Service Level		✓
Expired	✓	
Normal Use		✓
Ordered	✓	
Stock-out		✓
Reallocated		✓

Table 2.4. Paired t-test for $SC_{RFID,\alpha=0.2,\beta=0.2}$ versus $SC_{RFID,BL/2}$

Performance Measure	**$SC_{RFID,\alpha=0.2,\beta=0.2}$**	**$SC_{RFID,BL/2}$**
Service Level	✓	
Expired		✓
Normal Use	✓	
Ordered		✓
Stock-out	✓	
Reallocated		✓

The primary performance measure is service level, which needs to be maximized, while the overall cost is to be reduced. As shown in Table 2.5, $SC_{RFID,BL}$ outperforms $SC_{RFID,\alpha=0.2,\beta=0.2}$ on the basis of service life (97% and 95% service levels, respectively). On the other hand, $SC_{RFID,BL}$ orders approximately 8,000 more inventory items than $SC_{RFID,\alpha=0.2,\beta=0.2}$. In addition, $SC_{RFID,BL}$ leads to expiration of approximately 3,700 more inventory items and approximately 4,500 more reallocations than $SC_{RFID,\alpha=0.2,\beta=0.2}$. At this point, a detailed cost analysis with the manufacturing company led to a conclusion that $SC_{RFID,\alpha=0.2,\beta=0.2}$ would be more cost effective than $SC_{RFID,BL}$. This conclusion could have been reversed if the conditions in the manufacturing company were different.

Table 2.5. Comparison of the Means

Scenario	**Service Level**	**Expired ***	**Normal Use***	**Stock-out***	**Reallocated***	**Ordered***
$SC_{Manual,BL}$	91%	9,310	59,897	6,179	7,046	74,363
$SC_{RFID,BL}$	97%	5,685	64,036	1,984	11,762	74,529
$SC_{RFID,BL/2}$	84%	1,631	55,524	10,629	28,288	59,302
$SC_{RFID,\alpha=0.2,\beta=0.2}$	95%	1,957	62,999	3,265	7,185	66,437

2.4. CONCLUSIONS

In this paper, inventory management of time-sensitive materials using RFID data is studied. In addition to two static inventory models that rely on fixed baseline inventory levels, a dynamic, forecast-integrated inventory model is proposed. The impact of integrating RFID technologies with inventory control on the shop floor is discussed. The results show that the forecast-integrated model can adapt to the system dynamics more effectively than the current practice.

Since the performance of the forecast-integrated inventory model depends on the levels of the smoothing parameters α and β, a full factorial experiment is conducted to determine the best levels for the given production scenarios. Analysis of simulation results using ANOVA showed that $\alpha = 0.2$ and $\beta = 0.2$ outperformed all other combinations on the service level, which has been the fundamental performance measure due to the high value of the product. Even though the service level was the primary concern, each performance measure was taken into consideration to ensure the system would perform as desired. For example, the highest service level at a certain (α, β) setting may not be desirable if it yields a high amount of expired time-sensitive materials. Therefore, the "best" (α, β) setting can be selected by seeking a balance among all performance measures, such as in $SC_{RFID,\alpha=0.2,\beta=0.2}$. In this study, after the best levels of α and β are established, the performance of $SC_{RFID,\alpha=0.2,\beta=0.2}$ is compared to $SC_{Manual,BL}$. The results of the paired t-test analysis validated the advantage of using the proposed adaptive inventory management model by proving the ability to lower manufacturing costs, reduce inventory levels, and prevent excessive waste in typical manufacturing environments where RFID technologies can be utilized.

The results of this paper are relevant to all organizations concerned with improving their shop floor operations by reducing their inventory levels through the adoption of RFID. Through simulation, this paper demonstrates that there are opportunities for RFID technology to provide significant benefits on the shop floor, well beyond the automation-oriented advantages, such as labor savings. The scenarios that have been developed and tested in this paper were customized towards the needs of the manufacturing company; thus they might not be completely realistic to all organizations. However, the results do show the potential of RFID in a manufacturing environment, which can further be expanded to supply chain level.

2.5. REFERENCES

[1] McFarlane, Duncan, Sanjay Sarma, Jin Lung Chirn, C.Y. Wong, and Kevin Ashton, "Auto ID systems and Intelligent Manufacturing control," Engineering Applications of Artificial Intelligence. Vol.16, 2003, pp.365-376.

[2] Brewer, Alexander, Nancy Sloan and Thomas L. Landers, "Intelligent tracking in Manufacturing," Journal of Intelligent Manufacturing. Vol.10, 1999, pp.245-250.

[3] Lee, Young M., Feng Cheng, and Ying Tat Leung, "Exploring the impact of RFID on supply chain dynamics," Proceedings of the 2004 Winter Simulation Conference, Vol.2, 2004, pp.1145-1152.

[4] Michael, Katina and Luke McCathie, "The Pros and Cons of RFID in Supply Chain Management," International Conference on Mobile Business (IEEE Computer Society), 2005, pp.623 – 629.

[5] Engels, Daniel W., Joseph Foley, James Waldrop, Sanjay E. Sarma, and David Brock, "The networked physical world: an automated identification architecture," Proceedings of the Second IEEE Workshop on Internet Applications, WIAPP 2001, pp.76-77.

[6] McFarlane, Duncan, "Auto ID-based control systems: an overview," IEEE International Conference on Systems, Man and Cybernetics, Vol.3, 2002, pp.6-11.

[7] Penttila, Katariina, Lauri Sydanheimo, and Markku Kivikoski, "Performance development of a high-speed automatic object identification using passive RFID technology," IEEE International Conference on Robotics and Automation, Vol. 5, 2004, pp. 4864 – 4868, New Orleans, LA.

[8] Gramling, Kathryn, Anthony Bigornia and Tig Gilliam, "IBM Business Consulting Services EPC Forum Survey," MIT Auto-ID Center White Paper, 2003.

[9] McFarlane, Duncan, "The Impact of Product Identity on Industrial Control Part1: "See More, Do More…," MIT Auto-ID Center White Paper, 2003.
[10]Chappell, Gavin, Lyle Ginsbur, Paul Schmidt, Jeff Smith, and Joseph Tobolski, "Accenture Auto-ID on the Line: The Value of Auto-ID Technology in Manufacturing," MIT Auto-ID Center White Paper, 2003.
[11]Moran, Humberto J., Sana Ayub, Duncan McFarlane, "Auto-ID Use Case: Improving Inventory Visibility in Retail Company – Impact on Existing Procedures and Information Systems," MIT Auto-ID Center White Paper, 2004.
[12]Moran, Humberto J., Duncan McFarlane, and Timothy P. Milne, "Use Case Approach for determining the impact of Auto-ID implementations on Business Information Systems," MIT Auto-ID Center White Paper, 2003.
[13]Garcia, Andres, Duncan McFarlane, Martyn Fletcher and Alan Thorne, "Auto-ID in Materials handling," MIT Auto-ID Center White Paper, 2003.

3. RFID-BASED SMART FREEZER

3.1. INTRODUCTION

Inventory control of time and temperature sensitive materials (TTSM) is a vital problem that needs attention for various industries, for example composite manufacturers and food producers. Such items have to be stored in a temperature-controlled environment and used within a limited time, otherwise, they become waste. Effective management of TTSM will reduce inventory levels and prevent usage of expired materials thus reducing costs.

The goal of inventory management is to make the required quantities of items at the right time and location. Today, in inventory management, barcodes are widely used; however, this technology needs that a line-of-sight is maintained for the items besides the items are scanned one at a time. In addition, barcode-based systems have problems under low temperature environments. Additionally, tight control of certain item types that are critical for production may be required. Consequently, tracking of all inventory items through traditional methods may not be economically feasible.

These limitations can be mitigated with RFID technology, where electronic tags programmed with unique identification code are attached to items. However, in order to provide item-level visibility through monitoring, tracking and identification of items, a proper RFID implementation is necessary.

The application of RFID to inventory management was discussed in many papers. In [1], inventory management of TTSM using RFID data is presented under ambient temperature environments. In [2], an RFID based resource management system, which

integrates RFID, case based reasoning and pure integral linear programming is presented. Lian et al. [3] presents the hardware and software issues associated with warehouse logistics control and management system based on RFID in order to improve efficiency. Han et. al [4] utilize RFID for accurate localization of mobile robots in automated warehouses for inventory transportation.

The availability of item-level visibility also provides opportunities to control another type of inventory, namely work in process inventory [5, 6, 7]. In addition, RFID coupled with other sensing schemes helps close the loop on product life-cycle management (PLM) by providing visibility of product related information such as usage data, disposal conditions, etc., from cradle to grave. Jun, et al. [8], discuss issues regarding PLM, whereas Yang et al. [9] develop a component-based software framework for PLM. Within the scope of inventory management, RFID enables smart shelves [9, 10] and smart freezers.

However, the work in [1-11] assumes that a straightforward deployment of RFID technology providing almost 100% read rates under "controlled" laboratory settings. Upon technology transfer, unforeseen problems, such as readability may arise from radio frequency (RF) interference, RF absorbing materials, and environmental conditions [12]. For instance, in the proposed smart freezer, the low 100 deg F below zero temperatures cause excessive ice buildup rendering RFID technology unreliable. In addition, RF absorbing materials render low read rates and longer reading times.

Therefore in this paper, a novel scheme of obtaining high read rates within a reasonable reading time for inventory management at extremely low temperature environments using passive RFID tags is introduced. This overall scheme consists of: a) a

backpressure-based inventory management scheme for reducing waste and to attain cost savings; b) a selection scheme for placement and number of antennas inside the freezer; and c) a locally asymptotically stable distributed power control (LASDPC) scheme or simply DPC scheme for antenna transmission in order to obtain a high read rate required for backpressure-based inventory management. Lyapunov approach [13, 14] is utilized to show the stability of the inventory control and the DPC schemes. The overall method has been experimentally tested and verified on hardware. Both simulation and experimental results are included to justify theoretical conclusions.

3.2. METHODOLOGY

In this paper, a novel distributed inventory control (DIC) scheme for balancing the inventory levels at each layer at a desired target baseline value using the backpressure mechanism along the supply chain is proposed. Fig. 3.1 illustrates the supply chain structure with buffers, for example freezers on the shop floor, with local inventory control logic at each layer. This scheme considers only a single type of material flow. However, additional types of materials can be accommodated by using the same control logic for each material. The details of the DIC are presented in Subsection 1 under the assumption that there is full visibility

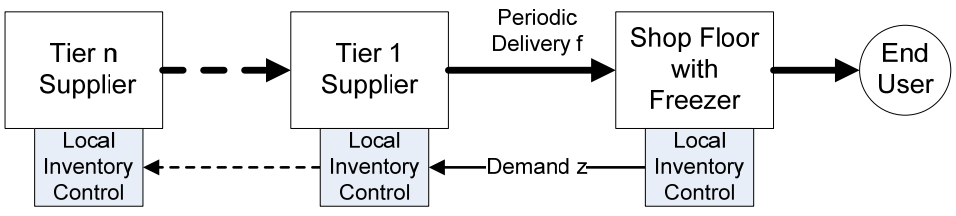

Figure 3.1. Supply chain with back pressure mechanism

In the subsequent section, antenna placement and power control discussion is presented in order to attain the 99-100% read rate requirement. The process of selecting the number of antennas and their location is presented in Section 3.2.3. Additionally, a novel mechanism that controls the power of the antennas to reduce the RF interference among them is proposed in Section 3.2.4. The proposed antenna placement mechanism together with the DPC will be referred hereafter as the Missouri S&T scheme, which is shown to reduce the number of antennas inside the freezer, while guaranteeing 99% read rates by eliminating all nulls.

The proposed DIC scheme expects a high item visibility. The proposed design methodology attains 99-100% read rates by considering several factors:

- Number of antennas and their placements inside the freezer was altered to attain the RF coverage;
- The power on each antenna is varied using the proposed DPC, which reduces RF interferences and collisions in the backscatter direction.

The design methodology is given in Sections 3.2.2, 3.2.3, and 3.2.3.

3.2.1. Inventory Management Across the Supply Chain.

3.2.1.1. Prediction. This paper proposes a new model to predict the instantaneous demand given by

$$f(k) = [\alpha(k-1)\beta(k-1)]\begin{bmatrix} f(k-1) \\ D(k-1) \end{bmatrix} = \theta^T(k-1)\phi(k-1) \tag{1}$$

where $f(k)$ is the demand at time k, $\alpha(k-1)$ and $\beta(k-1)$ are online tunable parameters, and $D(k-1)$ is the actual measured demand at the time instant *(k-1)*. The error in prediction is then expressed as

$$e(k-1) = f(k-1) - D(k-2) \tag{2}$$

Selection of appropriate values for α and β is essential in order to improve the accuracy of the prediction model. Traditionally, these values are selected through offline analysis of historical data. In contrast, the proposed scheme performs online tuning of these parameters to guarantee stability as presented in Theorem 1.

Theorem 1: Consider the instantaneous demand expressed as (1) with the error dynamics given in (2). Let the parameters update be provided by

$$\theta(k) = \theta(k-1) + \gamma\phi(k-1)e(k-1) \tag{3}$$

where $0 < \gamma < 1$ is the gain factor, then the instantaneous demand approaches the measured value asymptotically.■

Subsequently, the instantaneous value is utilized by the inventory controller, to create a backpressure signal, *z,* as shown in Fig. 4.1. The signal is propagated along the supply chain to adjust the delivery amount periodically. Next, the analysis of the backpressure mechanism is discussed.

3.2.1.2. Backpressure mechanism in the supply chain. In this section, inventory level dynamics in a supply chain are modeled using difference equations. In order to maintain target inventory levels, the DIC scheme is developed employing control theory. The performance of DIC is mathematically studied and the error bound on inventory level is obtained through the Lyapunov-based analysis.

At a particular echelon *j* in a supply chain, the inventory level will decrease when materials are used or shipped out to next echelon in the supply chain. When the change depends on the usage, it has to be estimated using (1). Otherwise, the change is dictated by the backpressure signal, $z_{i,j-1}$, which indicates the delivery request from the subsequent echelon, *j*-1. Then, the demand is fulfilled by delivery, *f*, shipped out from the current echelon of the supply chain. On the other hand, the inventory will increase when the

material is delivered from the supplier up the supply chain based on demand z_{ij}. Thus in a supply chain environment, the inventory levels change due to the fluctuation in demand.

Overall, the change of the inventory level inside a freezer in the supply chain at echelon *j* can be expressed as

$$q_{ij}(k+1) = Sat_Q\left[q_{ij}(k) + z_{ij}(k) - f_{ij}(k) + e_{dij}(k)\right] \tag{4}$$

where $q_{ij}(k)$ is inventory level at freezer *i* at time *k*, $z_{ij}(k)$ is the required delivery for freezer *i* for period *k*, $f_{ij}(k)$ is the usage rate of the freezer *I*, and Sat_Q is saturation function modeling the limited capacity of freezers. The fluctuation in the demand, $e_{dij}(k)$, is due to many factors, for example just in time manufacturing, machine break down, etc.

Remark 1: In a supply chain a negative delivery size, $z_{ij}(k)$ corresponds to a reduction in supply from the baseline level. Reduction is necessary to minimize wastage. ■

At any time instant, materials inside the freezer will be shipped back to the supplier. Hence, the calculated delivery request is modified to exclude this case and actual delivery request $z_{ij}^{'}(k)$ can be expressed as

$$z_{ij}^{'}(k) = \begin{cases} z_{ij}(k) & \text{if } z_i(k) > 0 \\ 0 & \text{if } z_i(k) \le 0 \end{cases} \tag{5}$$

Remark 2: In the case when $z_{ij}(k)$ is positive, the baseline will be completed with the instant demand. In contrast, for the case of negative $z_{ij}(k)$, there will be no shipment; hence the inventory level error cannot be directly controlled.■

Define the desired level of inventory as q_{tij}. Then the error in inventory level is equal to $e_{ij}(k) = q_{tij} - q_{ij}(k)$. This value at the next time instant can be obtained as

$$e_{ij}(k+1) = q_{tij} - q_{ij}(k+1) = e_{ij}(k) - z_{ij}(k) + f_{ij}(k) - e_{dij}(k) \tag{6}$$

The requested delivery level $z_{ij}(k)$ should minimize the error in inventory level for the next time instant. In other words

$$q_{tij} - \left[q_{ij}(k) + z_{ij}(k) - f_{ij}(k) + e_{dij}(k) \right] = 0 \tag{7}$$

Unfortunately, the exact demand $f_{ij}(k)$ is unknown for the time period *k* (future) due to uncertainties such as usage rates etc. By using an estimated demand value, backpressure control signal can be obtained as

$$z_{ij}(k) = \left[\hat{f}_{ij}(k) + k_v e_{ij}(k) \right] \tag{8}$$

where $\hat{f}_{ij}(k)$ is the predicted demand for the freezer *i* during the time interval *k* in echelon *j*. Then, the expected error in inventory level at time *k+1* is equal to

$$e_{ij}(k+1) = q_{tij} - q_{ij}(k+1) = e_{ij}(k) - k_v e_{ij}(k) + \tilde{f}_{ij}(k) - e_{dij}(k) \tag{9}$$

where $\tilde{f}_{ij}(\cdot) = f_{ij}(\cdot) - \hat{f}_{ij}(\cdot)$ is an error in demand prediction. Next, the bound on error in inventory level is shown.

Theorem 2 (General case): Consider the desired inventory level, q_{tij}, to be finite, and the demand fluctuation bound, e_{dM}, to be equal to zero. Let the delivery quantity for (4) be given by (8) with the delivery quantity being estimated properly such that the approximation error $\tilde{f}_{ij}(\cdot)$ is bounded above by f_M. Then, the inventory level backpressure system is uniformly ultimately bounded provided $0 < k_v < 1$.

Proof: Consider Lyapunov function candidate $J = \left[e_{ij}(k) \right]^2$. Then, the first difference is

$$\Delta J=\left[k_v e_{ij}(k)+\tilde{f}_{ij}(k)+e_{dij}(k)\right]^2-\left[e_{ij}(k)\right]^2 \tag{10}$$

The stability condition $\Delta J \le 0$ is satisfied if and only if

$$\|e\|>(f_M+d_M)/(1-k_{v\max}) \tag{11}$$

When this condition is satisfied, the first difference of Lyapunov function candidate is negative for all k. Hence, the closed-loop system is uniformly ultimately bounded.■

Remark 3: The bound on error in inventory level depends on $k_{v\max}$ as shown in (11). The bound decreases as $k_{v\max}$ approaches zero thus yielding a smaller error.■

Remark 4: Theorem 1 demonstrates that the actual inventory level converges close to the target value provided the errors in demand estimation and variation being bounded.■

3.2.1.3. Multi-freezer environment. In a typical industrial environment, there may be more than one freezer as shown in Fig. 4.2. Then, the overall inventory of the shop floor is expressed as a set of equations given by (4), where i=1,…, n. For simplicity, the following vector form represents the shop floor inventory levels at echelon j

$$Q_j(k+1)=Q_j(k)+Z_j(k)+F_j(k)+D_j(k) \tag{12}$$

where

$$Q_j(k)=\begin{bmatrix} q_{1j}(k) \\ q_{2j}(k) \\ \vdots \\ q_{nj}(k) \end{bmatrix}, Z_j(k)=\begin{bmatrix} z_{1j}(k) \\ z_{2j}(k) \\ \vdots \\ z_{nj}(k) \end{bmatrix}, F_j(k)=\begin{bmatrix} f_{1j}(k) \\ f_{2j}(k) \\ \vdots \\ f_{nj}(k) \end{bmatrix}, D_j(k)=\begin{bmatrix} e_{d1j}(k) \\ e_{d2j}(k) \\ \vdots \\ e_{dnj}(k) \end{bmatrix}, \tag{13}$$

In such a case, the error in inventory at the shop floor, *e(k)*, is equal to sum of error magnitudes at each freezer

$$e_j(k)=\sum_i \left|e_{ij}(k)\right| \tag{14}$$

3.2.1.4. Inventory balancing. In the multi-freezer scenario, as shown in Fig.3.2**,** items can be quickly moved between the freezers to balance inventory, minimize wastage thus saving cost. Due to the co-location of freezers, the load balancing mechanism can yield almost immediate (relative to delivery period) correction of inventory levels among freezers thus reducing wastage.

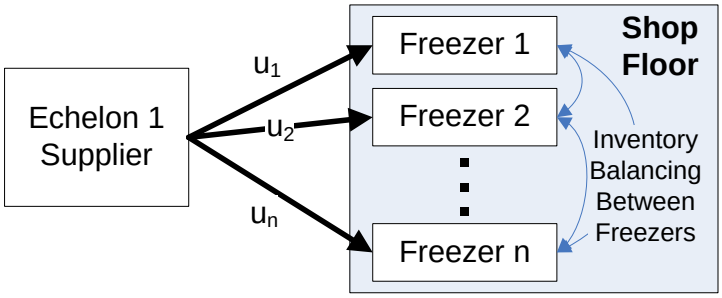

Figure 3.2. Inventory balancing between multiple freezers

In the case of load balancing, the individual inventory levels, the demand and the delivery/usage levels for all freezers can be combined as

$$q_{Tj}(k)=\sum_i q_{ij}(k),\ z_{Tj}(k)=\sum_i z_{ij}(k),\ f_{Tj}(k)=\sum_i f_{ij}(k),\ e_{dTj}(k)=\sum_i e_{dij}(k) \tag{15}$$

where $q_{Tj}(k)$ is total inventory level on the shop floor at time k for echelon j, $z_{Tj}(k)$ is the corresponding total demand, $f_{Tj}(k)$ is the total usage or delivery shipped out of the current supply chain level, and $e_{dTj}(k)$ is the total variation of the demand. Therefore (12) can be written as

$$q_{Tj}(k+1)=q_{Tj}(k)+z_{Tj}(k)-f_{Tj}(k)+e_{dTj}(k) \tag{16}$$

The load balancing is performed internally in order to minimize inventory errors at each freezer. Consequently, the total error in the inventory level at the echelon j is given by $e_{Tj}(k) = q_{Tij} - q_{Tj}(k)$. In the case of individual inventory controllers, the demand z is fulfilled only when it is a positive value. To prevent excess inventory being shipped back, shifting the load between the freezers can offset the negative demand z as explained next.

First, consider the total demand at echelon 'j'. In the case of independent inventory control at each freezer, the total demand $z_{T\_NLB,j}(k)$ would include only the positive demand values of the individual freezers at that echelon. In contrast, in the case of inventory balancing, the total demand will include the negative demand as well, which in turn will reduce total demand $z_{T\_LB,j}(k)$ as

$$\begin{aligned} z_{T\_NLB,j}(k) &= \sum_{i,\, z_i(k)>0} z_{ij}(k) + \sum_{i,\, z_i(k)\le 0} 0 = \sum_{i,\, z_i(k)>0} z_{ij}(k) \\ z_{T\_LB,j}(k) &= \sum_{i,\, z_i(k)>0} z_{ij}(k) + \sum_{i,\, z_i(k)\le 0} z_{ij}(k) \end{aligned} \tag{17}$$

When there is no negative demand, that is $\sum_{i,\, z_i(k)\le 0} z_{ij}(k) = 0, \text{ for all } j = 1,...,m$, the total demand is equal for both scenarios. However, if there is at least one freezer with negative demand, the corresponding excess inventory will be used to fulfill the positive demand of the other freezers at that echelon thus reducing the total order or target demand. Thus, the total order in the case of load balancing scenario will be smaller or at least equal to the total demand observed in the case of no load balancing. This total demand satisfies

$$0 \le \sum_{i,\, z_i(k) \le 0} z_{ij}(k) \quad \Rightarrow \quad z_{T\_LB,j}(k) \le z_{T\_NLB,j}(k) \tag{18}$$

By using (17) and (18), it can be shown that the load balancing can reduce the total error in the inventory levels corresponding to the negative demand, $z_{Tj}(k)$, by shifting inventory among the freezers. In contrast, without load balancing, wastage due to excess inventory at one freezer cannot be redirected to the freezers needing inventory. Additionally, the errors due to demand variation at the freezers may cancel each other out. For instance, in the worst case scenario, the individual errors will be equal for both scenarios with and without load balancing.

In other words, the total error in inventory levels for load balancing scenario is smaller or equal to the scenario when the total error in inventory levels without load balancing as

$$\sum_i \left| e_{ij}(k) \right| \ge \left| \sum_i e_{ij}(k) \right| \tag{19}$$

3.2.2. Space for Antennas in a Freezer. Initial experiments have demonstrated that the number of antennas and their orientation affect the read rates besides reducing capacity due to antenna placement. However, such problems can be easily overcome by reducing the number of items in the baseline when the 99% visibility is ensured. At present, the freezers are filled up regardless of real usage thus leading to high baselines and more wastage. Full visibility through improved read-rates allows more accurate usage levels and less waste. However, if for any reason, such high baselines are needed, the Missouri S&T Smart Freezer scheme can still provide 99% read rates for a fully loaded freezer.

3.2.3. Antenna Configuration. The selection process for finding the necessary number of antennas and their localization inside the freezer in order to achieve full visibility is illustrated in Fig 3.3. First, a realistic RF coverage pattern has to be obtained for the utilized antennas in the target environment.

Next, the individual antenna coverage pattern is used to determine the number and positions of antennas that provide the RF coverage of the entire freezer cavity. The selection of location and the number of antennas required for attaining such high read rates with low reading times is performed iteratively by adding and positioning one antenna at a time. When the final antenna configuration is acquired, hardware tests can be used to verify that each antenna reads all the tags within its designated area. Finally, the proposed power control scheme is applied online in order to minimize interference among the antennas while ensuring the desired coverage inside the freezer. The proposed scheme to obtain antenna configuration is described next.

3.2.3.1. Acquisition of antenna pattern. First an ideal radiation pattern is calculated and modeled by assuming no obstructions and interference. Next, the propagation model is used to calculate the coverage in a non-ideal environment. The radiation pattern can be obtained from antenna vendor, through measurements in an anechoic chamber, or by using existing models from the literature [15].

For example, a dipole antenna radiation pattern can be calculated 15] using $E_{dipole}(\theta,r) = -jI_0 e^{j(\omega t - kr)} \sin\theta / (2\pi\varepsilon_0 cr)$ where E_{dipole} is far electric field (or radiated electromagnetic field) of a half-wave dipole at point A, θ is an Euler angle between direction of dipole and the direction to point A, r distance between the dipole and the point A, I_0 is the maximum current passing through the dipole, ε_0 is the permittivity of

vacuum , c speed of light in vacuum, λ is wavelength, $\omega = 2\pi F$ is the pulsation at frequency F, and $k = 2\pi/\lambda$ is wave-number. The radiation patterns are utilized to determine the antenna gains which are subsequently used to study the attenuation caused by the freezer walls, the material of the stored items, and the freezer cavity. This information will be further used to study the overall coverage, nulls and thus to determine the number of antennas.

The maximum coverage around an antenna is calculated using signal strength equation as

$$P_{bs} = K_1 \cdot P_i / r_{i-t}^{4q} = g_{ii} \cdot P_i \tag{20}$$

where q is environment dependent variable considering path loss (depends on stored material properties), $r_{i\text{-}t}$ is the distance between reader's antenna and the grid point, and K_1 is the constant that includes reader and tag antenna gains, modulation indexing and wavelength, derived in [16]. The gain from the radiation pattern is used to calculate K_1 for each direction. Assuming maximum tranmsssion power, $P_{\max}$, and reader's sensitivity level, γ, the maximum read range can be calculated from (20) as

$$r_{\max}(\phi) = \sqrt[4q]{G_A(\phi) K_1 P_{\max} / \gamma} \tag{21}$$

where ϕ is the angle around antenna, $G_A(\phi)$ is the reader's antenna gain at an angle ϕ.

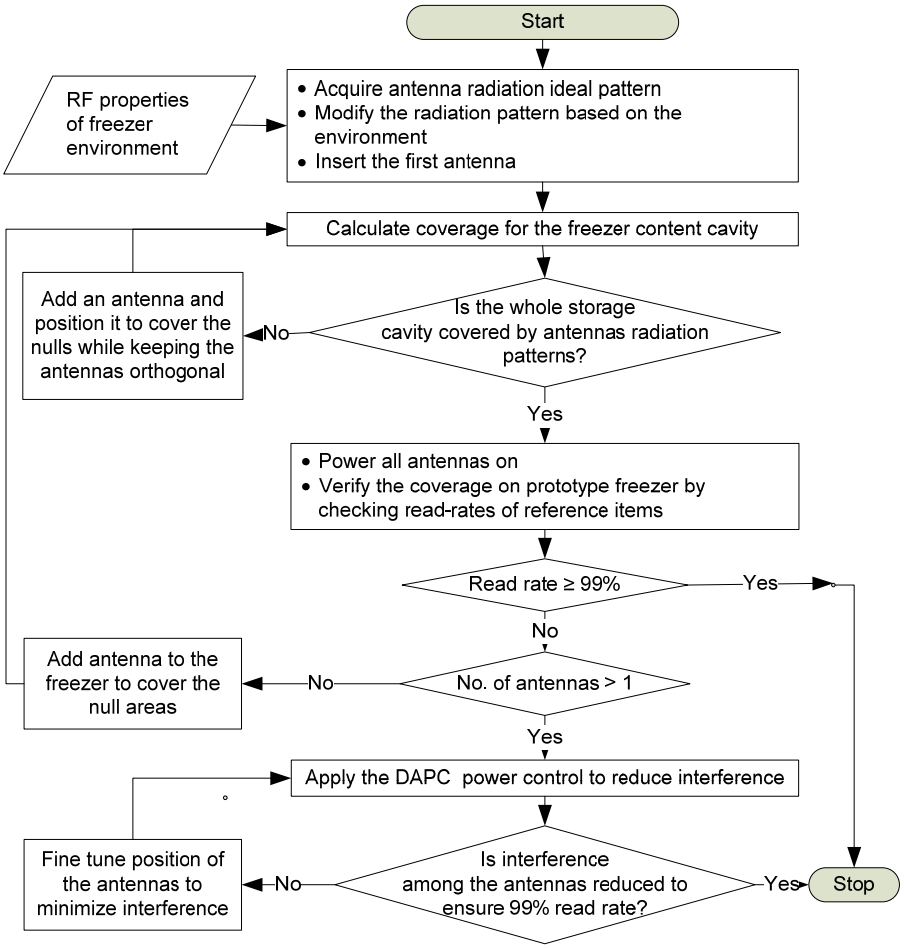

Figure 3.3. Antenna selection configuration for Missouri S&T smart freezer

Remark 5: In the initial step, a maximum power for each antenna is considered. However, operating the antennas simultaneously at their maximum power will cause collisions among the antennas thus rendering a signifcant reduction in coverage and read rates while increasing the read times. As a consequence, DPC is needed to minimize collisions and to ensure the coverage for the given antenna configuration.■

The parameter q strongly depends on the environment and has to be determined for a particular type of stored material, operating temperature, etc. For example, when the RF-absorbing material such as water is used, the attenuation of the RF signal over

distance increases thus reducing coverage and creating nulls. Therefore, more antennas are needed.

3.2.3.2. Selection of antenna configuration to cover the freezer. The procedure begins with a single antenna. The coverage for the entire freezer is calculated by dividing the freezer volume into a grid of reference points spaced at predefined intervals. For each grid point, the signal strength is calculated from each RFID antenna using the propagation model (20) which includes the absorption properties of the materials.

When the coverage for entire freezer cannot be achieved with maximum antenna power, then additional antennas are added one at a time in orthogonal configuration to cover the nulls. Each antenna is placed to provide best coverage based on their radiation pattern. When a new antenna is added, only the location of this antenna is varied, since the location of the existing antennas had been already established.

Remark 6: The orientation of a tag with reference to antenna alters radar cross-section of the tag's antenna. By placing the RFID reader antennas orthogonally the combined radar cross-section will improve since the maximum value among the considered antennas will apply as

$$C_T(\bullet) = \left\{ \max_i (C_i(\phi)) \right\}_{\phi=0..2\pi} \tag{22}$$

where $C_i(\phi)$ is the tag radar cross-section for an angle ϕ in reference to i^{th} RFID reader antenna, and $C_T(\bullet)$ is combined radar cross-section for all RFID reader antennas.■

3.2.3.3. Testbed validation of the antenna configuration. The overall capability of the freezer setup is validated on the hardware. The reference tags are placed at worst possible locations (e.g. in the middle of the item stack, at the corner of the designated cavity). The antennas are powered one at a time and the reference tags are read.

3.2.3.4. Fine tuning of antennas location to reduce interference. Because of the limited space inside a freezer, the coverage obtained from Steps 1 and 2 will be reduced due to the interference/collisions from other antennas operating simultaneously. Then the proposed locally asymptotically stable distributed power control (LASDPC) scheme or simply DPC is used to minimize the interference while attaining the desired coverage. The DPC selects suitable power while ensuring there is no overlap in RF coverage among the antennas at all times. This essentially eliminates the collisions among the antennas. The proposed scheme ensures asymptotic stability as presented in Section 3.2.4 in contrast with [14] where a bounded stability is shown. This asymptotic stability is needed for attaining 99% read rates and low read times. If the LASDPC alone cannot mitigate the interference, then the antenna position is fine-tuned. Once the consistent 99% read-rate is achieved, the current configuration of antennas becomes the solution.

3.2.4. Proposed Distributed Power Control Scheme. In RFID systems, a tag will be detected provided the ratio of the backscatter signal received by the reader is above the target signal-to-noise ratio (SNR) [14]. The SNR is defined as $P_{bs}(k)/I_i(k) > \gamma$, where P_{bs} is the backscatter power from a tag at the time instant k, $I_i(k)$ is the interference at the tag backscatter frequency, and γ is the minimal SNR required to correctly decode the backscatter signal. The SNR state equation for an antenna can now be defined as

$$y_i(k+1)=\theta_i^T(k)\psi_i(k)+P_i(k+1)\cdot\beta_i/I_i(k) \tag{23}$$

where $y_i(k)$ is the SNR value, $\theta_i^T(l)=\begin{bmatrix}\alpha_i(l) & r_i(l)\end{bmatrix}$ is a vector of unknown parameters, and $\psi_i(l)=\begin{bmatrix}y_i(l) & \omega_i(l)\end{bmatrix}^T$ is the regression vector, $P_i(k+1)$ is power at time $k+1$, and β_i is signal loss.

Define the SNR error as $e_i(k)=\gamma-y_i(k)$. By selecting the antenna transmission power as

$$P_i(k+1)=\frac{I_i(k)}{\beta_i^{-1}}\left[-\hat{\theta}_i(k)\psi_i(k)+\gamma_i+k_{vi}e_i(k)+\frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c_i}\right] \tag{24}$$

where $c_i>0$ and λ are design constants, one can ensure locally asymptotically stable system. The proposed update law for the channel parameter estimate, $\hat{\theta}(k)$, is given by

$$\hat{\theta}(k+1)=\hat{\theta}(k)+\alpha\ \psi_i e_i^T(k+1) \tag{25}$$

After combining and defining $\Psi_1(k)=\tilde{\theta}_i^T\psi_i$ and $\varepsilon(k)=\varepsilon_i(k)$, the closed loop error system in terms of SNR becomes

$$e_i(k+1)=k_{vi}e_i(k)+\Psi_1(k)+\frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c_i}+\varepsilon(k) \tag{26}$$

Assume a conic region that satisfies $\varepsilon^T\varepsilon\le 2\varepsilon_1^T\varepsilon_1\le\varepsilon_M=\rho(\|e_i(k)\|)e_i^T(k)e_i(k)$, where $\rho(\|e_i(k)\|)=\lambda^2$, the following theorem guarantees local asymptotic stability of the SNR error dynamics even with channel estimation errors.

Theorem 3 (General Case): Consider the SNR state equation (21) with the parameter update law given by (23) in the presence of bounded channel uncertainties. Let

the power update be provided by (22). Then the SNR error $e_i(k)$ and the channel parameter estimation errors, $\tilde{\theta}(k)$, respectively are locally asymptotically stable.

Proof: See Appendix.

3.3. SIMULATION AND EXPERIMENTAL RESULTS

3.3.1. Simulation Study. In order to verify the effectiveness of the proposed inventory control scheme, several simulation scenarios were completed in Matlab. The simulation scenarios depict an industrial shop floor with 8 freezers. Actual demand for each freezer was modeled using an assortment of functions, for example a step function and a sinusoidal function. Demand variation $e_{dij}(k)$ was implemented using a uniform random variable with range (-10, 10). Total error in the inventory level and total delivery request were selected as the performance metrics. The overall goal is to reduce the total error since high error leads to wastage of materials due to overstocking, or disruption of production due to insufficient inventory. Additionally, the total delivery request is observed in order to identify when excess amount of items is ordered.

The proposed scheme is compared with exponential weighted moving average (EWMA) method [1]. One level of the EWMA smoothing parameter α=(0.9) and two levels of the smoothing parameter β=(0.1, 0.9) were tested. Total of 3 cases were simulated, with and without the distributed inventory control scheme and inventory balancing. Each scenario was repeated 500 times and the results were averaged in order to account for the effects of the stochasticity in the system.

Table 3.1 summarizes the results of the simulation study. The proposed scheme performs on par with the best-case scenario using EWMA. However, the EWMA

parameters need to be manually selected based on experiments [1] since they have a direct effect on the accuracy of the demand prediction. In contrast, the proposed demand prediction model tunes these parameters online in order to reduce prediction error.

When the DIC backpressure signal is applied, the total error in inventory level is reduced by up to 15% for EWMA and 34% for the proposed scheme over the No–DIC scenario. By employing the backpressure DIC signal, the performance for each demand prediction scheme is further improved since the ordering policy for the supply chain compensates for demand fluctuations. It is important to note that, the main improvement in the performance between EWMA and proposed scheme is the direct result of the online tuning that eliminates the need for offline manual training, and can adopt parameters to changing demand patterns without user intervention. Further performance improvement is observed when all the schemes are implemented with DIC coupled with inventory load balancing. For instance, Table 3.1 shows that the DIC with inventory balancing reduces errors in inventory levels by 62-70% when compared to DIC without inventory load balancing.

Table 3.1. Simulation Results

	Proposed Scheme		**EWMA** $\alpha = 0.9, \beta = 0.1$		**EWMA** $\alpha = 0.9, \beta = 0.9$	
	Total Error in Inv. Level	Delivery Request	Total Error in Inv. Level	Delivery Request	Total Error in Inv. Level	Delivery Request
No - DIC	15029	25107	11476	24696	12282	24711
DIC w/o Inventory Balancing	9833	24674	9860	24723	10347	24722
DIC w/ Inventory Balancing	2989	24718	2982	24702	3880	25867

3.3.2. Hardware Experiments. In the hardware portion of the experimentation, impact of RFID reader and antenna types, number and placement of antennas, reading times and operational temperature on the read rate were investigated. Each experimental scenario was repeated at least five times to ensure consistency in the results. The containers were kept in the freezer for a prolonged time to study impact of exposure to low temperatures and ice buildup.

Placement of antennas inside the freezer that can deliver 99% read rates were identified using the proposed methodology in Section 3.2.3. First, following Step 1, the theoretical antenna coverage has been altered to accommodate the presence of ferrous materials inside the freezer. Next, antennas were successively added and the coverage was tested as per Step 2. The intermediate results are shown below. The final configuration, as shown in Fig. 3.4, was achieved by fine-tuning the locations of 6 antennas connected to an Alien ALR9800 reader as per Step 4.

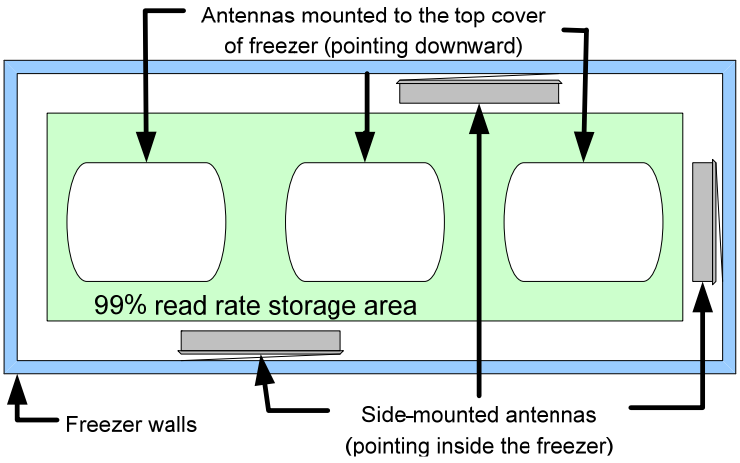

Figure 3.4. RFID antenna locations for smart freezer solution

The reader was operated with a maximum transmission power of 30dBm. Tracked items are tagged with Gen 2 tags wrapped around the containers. The proposed LASDPC scheme was implemented at the reader. The antenna configuration design and confined freezer area resulted in the power control scheme operating in a simplified on/off mode, where the power control combats the multipath fading alone.

The experimental setup used in this work comprised of:

- AR400 Matrics Class 1 Gen 2 RFID reader with dual directional antennas
- ALR 9800 Alien Class 1 Gen2 RFID reader with circularly polarized antennas

The freezer was loaded with 276 containers including 120 containers with non-RF-friendly (iron oxide) materials. Non-RF friendly materials absorb the RF signal and negatively affect the overall read rates.

AD-220 Gen 2 tags from Avery Dennison were attached to the containers.

Series of experiments have been conducted in order to evaluate the impact of each step of the proposed antenna configuration design scheme. First, a baseline scenario with

simple antenna placement is presented. Next, the steps to determine the desired antenna number and placement are applied taking into account signal attenuation due to ferrous materials, frost, and tag orientation. Finally, the proposed power control is applied and tested in varying operating conditions. The worst case of -100F operating temperature has been thoroughly tested over a period of 3 months.

3.3.2.1. Results for antenna propagation pattern calculations. The antenna specification provides basic data about its radiation pattern including beam width and gains. For example, ALR-9610 circular antenna provides 65º of 3dB-beamwidth. Consequently, the coverage pattern assumes a cone shape that creates nulls for tags located close to antenna but off the beam center. In the proposed design methodology additional antennas cover the nulls. Second major parameter used in the creation of coverage pattern is the effective read range. The parameters of the model (20) have to be set appropriately for the given environment. Inside a fully loaded freezer the content can absorb RF signal thus increasing path loss parameter from q=1 (free space) to q=3 for an iron oxide used in the experiments. Consequently, the read range reduces to 3ft. Furthermore, the orientation of the tag with respect to the antenna's beam direction can reduce tags gain by up to 3dB. Based on the read range model and experimental verification, the 3dB loss corresponds to a distance of about 4 inches. Hence, the final coverage pattern assumes the shape of a cone with a height of 26in.

3.3.2.2. Experiments with various antenna configurations. Table 3.2 was obtained during the initial testing of RFID-enabled freezer. The study shows inadequate read rates for simple antenna configuration for both Alien and MATRICS antennas. In general, the read rates increase with the number of antennas since the additional antennas

can be positioned to cover the nulls. However, read times appear to be quite high which can be mitigated using DPC as presented in Table 3.3.

Moreover, the read rates improved for the 'multi-static' Alien antennas since they can switch between transmission and reception whereas a Matrics antenna includes two static elements. As a result, each Alien antenna act as two sets of fixed function antennas although physically there is one. This information should not be confused in the Table 3.2 between the manufacturers. Next, the proposed design methodology from Sections 3.2.2, 3.2.3 and 3.2.4 was applied and experimentally tested.

By using the flow chart in Fig. 3.3, the adequate number of antennas and their location was calculated to be equal to six and their respective locations are shown in Figure 3.4.

Table 3.2. Performance Comparison of Matrix and Alien Equipment

Reader	**Antenna**	**Duration**	**Read rate**		
			1-Antenna	**2- antennas**	**4-Antennas**
Matrics AR-400	AN-100 Dual Directional	180 sec	**34-50 %**	**68-75 %**	**82-86%**
Alien ALR-9800	ALR-9610 Circular	180 sec	**59-63%**	**75-77 %**	**84-88 %**

*Each Matrics antenna include a separate transmitting and receiving element

Remark 7: The Matrics reader requires that each antenna set contain two elements (transmitter and receiver). Consequently, the complete freezer setup for Matrics reader would require 12 antenna elements, which reduces the storage volume significantly. Hence, Alien readers were used only.■

Remark 8: The introduction of antennas into the freezer reduces its volumetric capacity. The actual reduction depends upon the freezer and antenna dimensions. In the proposed freezer (So-Low C85-9 interior: 47”x16”x20”), the reduction of the freezer capacity is equal to 28.5% due to 2in. spacing at each freezer wall mounted with antennas.■

3.3.3. Results for Varying Temperature. The selected antenna configuration was validated at several operating freezer temperatures with proposed antenna placement and power control. The temperature varied from 20° F to −100° F, which is a typical range of temperatures used in the composite industry. Table 3.3 presents the summary of results. Missouri S&T scheme renders 99% read rates regardless of freezer temperature since the methodology ensures coverage in the presence of severe fading and interference while ensuring a low read time of 40 sec.

Table 3.3. Read Rates

Reader and Antennas	# of antennas	Duration	Temp. (° F)	No Power Control Random Placement	Power Control Random Placement	Power Control & Antenna Placement
Alien Alr-9800 ALR-9610 Circular	6	40 sec.	20	81%	94%	99%
			-20	80%	93%	99%
			-40	78%	93%	99%
			-60	75%	89%	99%
			-100	73%	88%	99%

3.4. CONCLUSIONS

This paper introduces a novel DIC scheme and a methodology to identify the location and number of antennas required for attaining full visibility. The proposed DIC scheme with inventory balancing reduces errors in inventory level when contrasted against a simple demand prediction scheme. The performance of the DIC scheme was demonstrated through Lyapunov approach. In order to achieve the 99% read rates, Missouri S&T smart freezer solution was utilized. The experimental results demonstrate 99% read rates within 40 seconds at −100°F temperature with six antennas. Overall, the DIC scheme in concert with the smart freezer design delivers a reliable and efficient supply chain solution.

3.5. REFERENCES

[1] M. D. Mills-Harris, A. Soylemezoglu, and C. Saygin, "Adaptive inventory management using RFID data," International Journal of Advanced Manufac. Technology, vol. 32, no. 9, pp. 1045-1051, 2007.
[2] H. K. H. Chow, K. L. Choy, W. B. Lee, and K. C. Lau, "Design of a RFID case-based resource management system for warehouse operations," Expert Sys. with Apps., vol. 30, no. 4, pp. 561-576, 2006.
[3] X. Lian, X. Zhang, Y. Weng, Z. Duan, and A. Z. Duan, "Warehouse logistics control and management system based on RFID," Proc. of the IEEE Int. Conf. on Automation and Logistics, pp. 2907-2912, 2007.
[4] S. Han, H.Lim, J. Lee, "An efficient localization scheme for a differential-driving mobile robot based on RFID system," IEEE Trans. Ind. Elec., vol. 54, no. 6, pp. 3362-3369, 2007.
[5] E. Budak, B. Catay, I. Tekin, H. Yenigun, M. Abbak, S. Drannikov, and O. Simsek, "Design of an RFID-based Manufacturing Monitoring and Analysis System," RFID Eurasia, pp. 1-6, 2007.
[6] G. Q. Huang, Y. F. Zhang, and P. Y. Jiang, "RFID-based wireless manufacturing for walking-worker assembly islands with fixed-position layouts," Robotics and Computer-Integrated Manufacturing, vol. 23, no. 4, pp. 469-477, 2007.
[7] P. Vrba, F. Macurek, and V. Marík, "Using radio frequency identification in agent-based control systems for industrial applications," Eng. Apps. of Artificial Intelligence, vol. 21, no. 3, pp. 331-342, 2008.

[8] H-B. Jun, D. Kiritsis, and P. Xirouchakis, “Research issues on closed-loop PLM,” Computers in Industry, vol. 58, no. 8-9, pp. 855-868, 2007.
[9] X. Yang, P. R. Moore, C-B. Wong, and J-S. Pu, “A component-based software framework for product lifecycle information management for consumer products,” Consumer Electronics, vol. 53, no. 3, pp. 1195-1203, 2007.
[10]M. J. Imburgia, "The role of RFID within EDI: Building a competitive advantage in the supply chain," Proc. IEEE Int. Conf. on Service Operations, Logistics and Informatics, pp. 1047-1052, 2006.
[11]C. Bornhovd, T. Lin, S. Haller, and J. Schaper, “Integrating automatic data acquisition with business processes experiences with SAP's auto-ID infrastructure,” Proc. Int. Conf. on Very Large Data Bases, vol. 30, pp. 1182-1188, 2004.
[12]R. H. Clarke, D. Twede, J. R. Tazelaar and K. K. Boyer, “Radio frequency identification (RFID) performance: the effect of tag orientation and package contents,” Packaging Technology and Science, vol. 19, no. 1, pp. 45-54, 2006.
[13]G. G. Yen and L. Ho, "Online multiple-model-based fault diagnosis and accommodation," IEEE Trans. Ind. Elec., vol. 50, no. 2, pp. 296-312, 2003.
[14]K. Cha, A. Ramachandran, and S. Jagannathan, “Adaptive and probabilistic power control algorithms in dense RFID networks,” Int. Journal of Distributed Sensor Networks, vol. 4, no. 4, pp. 347-368, 2008.
[15]M. Fogiel, “The Electronic Communications Problem Solver,” New York, NY, 10018: Research and Education Association, 1988.

4. MAHALANOBIS TAGUCHI SYSTEM (MTS) AS A PROGNOSTICS TOOL FOR ROLLING ELEMENT BEARING FAILURES

4.1. INTRODUCTION

According to the Motor Reliability Working Group and the Electric Power Research Institute [1], the most common failure mode of electric motors is bearing failure, which accounts for over 40% of all machine failures. Therefore, monitoring, diagnosis, and prognosis (MDP) of bearing failures is vital for reliable operation of industrial equipment. Rolling element bearings consist of four separate parts, as shown in Figure 4.1: (1) outer race, (2) inner race, (3) cage, and (4) rolling elements

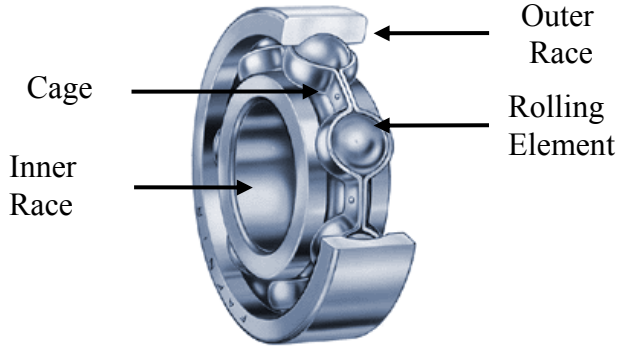

Figure 4.1. Components of a rolling element bearing
(image courtesy of Timken Co.)

Of the various sensory data commonly used for machine monitoring and evaluation purposes (e.g. vibration, acoustic emission, temperature, force, and torque, etc.), vibration signals are directly related to a machine's structural dynamics and

working conditions. Faults that typically occur on rolling element bearings are due to localized defects on the inner race, outer race, rolling element and the cage of the bearing [2, 3]. As a result, impact vibrations are generated as the rolling elements pass over the surface of the defect. These vibrations are governed by the geometry of the bearing, the operating speed, the actual location of the defect and the natural dampening of the machinery; the total effects of which form what is called the bearing characteristic frequency (BCF) [4].

An in-depth discussion of complete rolling element bearing vibration features with mathematical formulations is provided by Harris [3], Hochmann and Bechhoefer [5], and Li et al. [6]. Consequently, vibration signals are commonly used as an effective indicator of potential machine failure. In this study, features derived from bearing vibration signals are also utilized.

Rolling element bearings have been extensively studied in the literature for the last three decades in order to develop algorithms and techniques for bearing fault MDP. Janjarasjitt et al. [7] present a modified computational algorithm for the correlation integral, which is a popular nonlinear algorithm for time series analysis. The partial correlation integral is applied to bearing vibration data to examine the nonlinear dynamics of bearing system throughout the life of a bearing. The test system in the paper includes two bearings operating under an axial load mounted with accelerometers; and vibration data are recorded. Once the partial correlation integral algorithm is applied to these sets of data, the results show that the vibration characteristics during the normal operating period of a bearing are dominated by nonlinear deterministic signals.

As the bearing progresses towards failure, the nonlinear features start to be dominated by stochastic signal features. The authors conclude that the partial correlation integral and the dimensional exponent provide useful diagnostic and prognostic information for monitoring and detecting inner and outer race failures of bearings in rotating machinery through complexity analysis of the bearing vibration data. Wu et al. [8] utilize frequency spectrum analysis in order to detect localized defects of rolling element bearings. They improve upon the high frequency resonance technique, also called envelope analysis, which is limited by the requirement of bearing operation under constant speed. However, rotating machinery may operate in varying speeds due to varying loads, run-ups and run-downs. The authors state that if variations in speed are present, additional hardware such as tachometers and key phasors would be required to use the envelope analysis.

In order to overcome this limitation they propose a mathematical model that employs instantaneous frequency estimations. They test their scheme on two bearings that are mounted on the shaft of an electrical motor. The bearings are instrumented with accelerometers and one of the bearings is seeded with an inner race fault, while the other one has an outer race fault. The results show that the proposed scheme successfully detects bearing faults without the need for the additional hardware.

Most of the time, impact vibrations generated by bearing faults exhibit low energy and can easily be overwhelmed by noise with higher energy levels other sources of vibration such as the other structural components of the machinery under examination. Due to the wide variety and complexity of machinery in industrial manufacturing environments, assessment of the status of a machine directly from raw vibration signals

or signals analyzed in the time or frequency domains [9, 10] can become complicated. In some cases it may even be entirely unfeasible [11]. This is due to limitations of the time domain and frequency domain procedures. Hence, various time-frequency analysis methods including wavelet transforms have been developed in order to overcome these limitations. Tse et al. [12], demonstrate the advantages of wavelets by performing a comparison between a wavelet based approach and FFT with envelope detection (ED) approach based on experimental data. In their testbed several ball bearings are instrumented with accelerometers for recording vibration data and a tachometer for measuring rotational speeds.

The test bearings were seeded with inner and outer race faults, in addition to a fault induced on the rollers themselves. The results showed that both FFT with ED and wavelet analysis were successful in detecting the outer race defect condition, where as the inner race and the rolling element defects were successfully identified by the wavelet analysis only.

Lou and Loparo [13], propose a new technique that uses the wavelet transform coupled with an adaptive neural fuzzy inference system (ANFIS). ANFIS is a model that maps inputs through input membership functions (MFs) and associated parameters, and then through output MFs to outputs. Detailed information on ANFIS can be seen in [14]. Their experimental system consists of bearings mounted on the shaft of a three-phase induction motor. From the test system, they collect vibration signals for various healthy bearings and bearings with pre-seeded fault conditions. They use the wavelet transform on the normalized data in order the define feature vectors that represent the condition of the bearings. Next, they train an ANFIS as a pattern classifier and test out an assortment

of bearing fault conditions. In their conclusion they state that ANFIS coupled with wavelet analysis could be used to identify the time of occurrence and the degree of severity of a fault, which would enable prognostics of bearings.

All of the papers discussed so far deal with monitoring and diagnosis of bearing failures. Diagnostics, by its nature is a reactive approach. In order to prevent catastrophic failures, preventive maintenance strategies are usually employed by industries. Maintenance on machinery is performed on a schedule-basis, which tends to unnecessarily increase downtime, reduce machine life and thus generate redundant costs. In order to rectify this problem, it is necessary to transition to a more proactive approach, i.e. predictive maintenance. Prognostics are important to predictive maintenance, and it focuses on performance degradation and estimation of remaining useful life (RUL) of components. Therefore, research on prognostics has been gaining momentum in the last decade; however, compared to diagnostics, literature on prognostics is much smaller.

In the literature, there are two approaches for the estimation of remaining useful life. The first approach utilizes physics of failure models that incorporate operating conditions in order to track the damage that components experience. These models require specific mechanistic knowledge and theory relevant to the components under investigation [15]. In their papers, Li et al. employ two defect propagation models developed through mechanistic modeling in order to estimate RUL for rolling element bearings [16, 17]. Li et al. [18], introduce a single surface defect propagation model along with an adaptive prognostic methodology for the prediction of bearing defect growth. In computer simulation and real-life testing, they show that their adaptive methodology is able to effectively predict the defect propagation of bearings without the need for a priori

knowledge of the prognostic model parameters. Luo et al. [19], present an integrated prognostics process based on an interacting multiple model approach. The prognostics process is derived from model-based simulations under nominal and degraded conditions. time-averaged mode probabilities.

The second approach to estimation of RUL comprises of pattern recognition models that are derived from historical data collected from the system under investigation. Artificial intelligence techniques such as neural networks, neuro-fuzzy inference systems, self organizing maps, etc. have been applied to the prediction of RUL by many researchers. Wang and Vachtsevanos [20], discuss the components of a prognostic architecture, such as network structure, learning algorithms, performance assessment, etc. in their study. They develop a prognostic scheme that draws on dynamic neural networks. They test a defective bearing with an inner race crack and validate the prognostics capability of their scheme. Qiu and Lee [21], apply the wavelet transform and self organizing maps in order to facilitate bearing prognostics. They collect data from a testbed that consists of four bearings mounted on an AC motor shaft operation at 2000 rpm. A radial load of 6000lbs is applied to the shaft through a spring mechanism where 3 sets of test are run until failure. Vibration data are collected and analyzed periodically. The results demonstrate that by monitoring the trajectory of bearing conditions within the SOM, it is possible to predict the remaining life of bearings. Finally, Chinnam and Baruah [22], present a neuro-fuzzy approach in a situation where failure data and failure definition models are unavailable but domain expert knowledge with strong experiential knowledge is available.

It should be noted that most of the work done in this area focuses on specific failure cases created under controlled laboratory conditions or through computer simulations. In some cases several bearing faults may be present at the same time, one of which would ultimately result in failure. Nevertheless, all of the faults need to be monitored concurrently in order to have an effective prognosis on the RUL. In addition, it is usually much easier to detect damage than to assess its severity and progression. Both physics of failure and damage-progression/performance-degradation models face difficulties when the system under consideration is complex and affected by the operating conditions and the environment. The papers in the literature present several signals; and features that are derived from these signals, which need to be monitored. If these signals/features show a large variability between experiments, it becomes difficult to select suitable thresholds to be used in rolling element bearing MDP. In addition, the existing body of work does not present a systematic method which can identify the most pertinent variables and eliminate the redundant ones in order to reduce analysis overhead. Therefore, a new multivariable method is necessary to identify the input variables necessary to be monitored for fault detection, root cause analysis, and fault prognosis.

In this paper, rolling element bearings are used as a case study in order to develop a Mahalanobis-Taguchi System (MTS) based fault detection, isolation, and prognostic scheme for real-time decision making. MTS is a multivariate pattern recognition tool, which provides the background to combine all pertinent information about a system into a single metric using Mahalanobis Distance, and it also presents a methodical way of determining the key features required for analysis drawing on the Taguchi methods.

MTS has been widely used in various diagnostic applications that deal with data classification. In MTS, Mahalanobis Distance (MD) is used to identify the degree of abnormality of data, whereas the Taguchi methods, which utilize orthogonal arrays (OA) and signal-to-noise ratio (S/N), are used to evaluate the accuracy of predictions and to optimize the system [23]. MD, introduced by P.C. Mahalanobis in 1936 [24], is a multivariate generalized measure used to determine the distance of a data point to the mean of a group. MD is measured in terms of the standard deviations from the mean of the samples and provides a statistical measure of how well the unknown data set matches with the ideal one. The advantage of the MD is that it is sensitive to the intervariable changes in the reference data. Therefore, it has traditionally been used to classify observations into different groups and diagnosis [25].

The benefits of MTS as a pattern recognition and data classification tool can be summarized as follows:

- A robust methodology that is insensitive to variations in multidimensional systems.
- Can handle many different types of data sets and effectively consolidates the data into a useful metric.
- Implementation of MTS requires limited knowledge of statistics.
- Relies typically on simple arithmetic, contextual knowledge, and intuition.
- Its success has been demonstrated in various practical applications.

In their paper, Chinnam et al. [26], use ten features derived from two degradation signals (thrust force and torque) in a drilling operation in order to determine the condition of the drilling tool. The authors utilize MD in order to generate a measurement scale and

then they use the Taguchi methods to reduce the number of features down to six. They use the 99th percentile of the MDs from the normal group as a threshold, which shows a superior performance when compared with the any of the individual features.

Wang et al. [27] use the iris data, which is often used in statistical textbooks, and credit card data from a Taiwanese bank in order to display the effectiveness of MTS in data classification. They compare the performance of MTS with discriminant and stepwise discriminant analysis and conclude that MTS shows better performance. Cudney et al. [28] apply MTS in order to study the relationship between available vehicle level performance data for vehicle ride and the corresponding consumer satisfaction ratings for the purpose of improving customer-driven quality. They analyze six vehicle ride parameters from sixty-seven datasets using the MTS, determining the key parameters. The authors are able to show that the MTS methodology can be used in order to evaluate consumer satisfaction and manufacturers can utilize this knowledge to make the appropriate design decisions regarding attributes that are delivered in a product.

Cudney et al. [29] further investigate the performance data and customer satisfaction relationship using the Mahalanobis-Taguchi Gram-Schmidt (MTGS) technique and showed improved results. In their recent article, Foster et al. [30] develop an alternative search procedure to be used within the MTS. The adaptive One-Factor-AT-a-Time (aOFAT) procedure replaces the orthogonal arrays traditionally utilized in MTS, and features are individually added or removed from the classification system depending on their impact on the overall signal-to-noise ratio. The experimentation shows that the aOFAT procedure renders greater improvements over the median with the

same or fewer design alternatives being explored and also exhibits good ability to generalize to new instances after training.

MTS has also been used as a process control and improvement tool to aid manufacturing systems. Asada [31] use MTS to forecast the yield of a wafer production system, which is affected by variability of the electrical properties and environmental dust. MDs are used to compare the relationship between wafer yield and distance. Hayashi et al. [32] develop an MD based method in order to maximize productivity in a semi-conductor manufacturing facility. Their method consists of real-time monitoring of flow factor and in process wafers using MD values to identify the out-of-specification processes and/or tools and take the necessary actions.

Mohan et al. [33] propose a MTS-based diagnostic and root cause analysis scheme for online monitoring of grip-length of pull type fasteners. Their scheme utilizes data collected from an aerospace industry pull-type fastening tool instrumented with various sensors. Sensor data are analyzed using MTS and the quality of the current operation is communicated back to the user via a wireless network. The authors were able to obtain a 100% data collection on each fastener as opposed to other statistical process control techniques, which rely on sampling. Dasgupta [34] develop a unified framework for achieving process control and improvement during the implementation phase of Six Sigma. The proposed framework takes advantage of the adaptability of MTS as a classification, variable selection and monitoring tool. The framework is explained using a simulated example. Srinivasaraghavan and Allada [35] also propose a MTGS based methodology in order to support the implementation of lean processes. Their methodology assists contemporary lean assessment tools through providing a measure of

leanness by benchmarking exemplary lean industries along with suggestions for improvements based on cost considerations.

MTS is a data driven method that relies on multivariate statistical analysis. Therefore, it is important to compare MTS to well-established multivariate statistical techniques such as principal component analysis, discriminant analysis, cluster analysis, canonical component analysis, and multidimensional scaling, as well as artificial neural networks. For an in-depth comparison of these techniques to MTS, please refer to Kim et al. [36], Cudney et al. [37], Jugulum [38], and Woodall et al. [39]. Books relevant to this study are Harris [3], Isermann [40], and Wowk [41].

The contribution of this paper can be summarized as follows: (1) Applies MTS to the domain of rolling element bearing monitoring, diagnostics and prognostics for the first time to the best knowledge of the authors; (2) Builds upon MTS by introducing the use of MD-based fault clusters for fault isolation; and (3) Utilizes the progression of MD values over time in order to estimate the remaining useful life of components.

4.2. METHODOLOGY

In this paper, a Mahalanobis Taguchi System based fault detection, isolation, and prognostics scheme is presented. The proposed scheme utilizes a Mahalanobis Distance based fault clustering method in order to classify faults into different categories. Using these clusters, fault detection and root cause analysis is performed. The scheme also utilizes the progression of the MD values over time in order to facilitate prognosis of time to failure for components. The details of the scheme are presented in the rest of the section.

4.2.1. Mahalanobis Taguchi System. MTS starts with data collection on normal observations. Then, Mahalanobis Distance is calculated using certain characteristics to investigate whether MD has the ability to differentiate the normal group from an abnormal group. If MD cannot detect the normal group using those particular characteristics, then a new combination of characteristics need to be explored. When the right set of characteristics are found, Taguchi methods are employed to evaluate the contribution of each characteristic. If possible, dimensionality is reduced by eliminating those characteristics that do not add value to the analysis. MTS consists of four stages [23, 25]:

Stage 1: Construction of Mahalanobis Space

The first step in MTS is the construction of the Mahalanobis Space (MS) as the reference. In order to construct MS, the variables that represent "normal" conditions are defined and sampled. In this study, new bearings that are properly lubricated represent the normal case. The steps for the construction of MS are outlined below:

1. Calculate the mean for each characteristic in the normal data set as:

$$\bar{x}i = \frac{\sum_{j=1}^{n} X_{ij}}{n} \tag{1}$$

2. Then, calculate the standard deviation for each characteristic:

$$s_i = \sqrt{\frac{\sum_{j=1}^{n} (X_{ij} - \bar{x}_i)^2}{n-1}} \tag{2}$$

3. Next, normalize each characteristic, form the normalized data matrix (Z_{ij}) and take its transpose (Z_{ij}^T)

$$Z_{ij} = \frac{(X_{ij} - \overline{x}_i)}{s_i} \tag{3}$$

4. Then, verify that the mean of the normalized data is zero:

$$\overline{z}_i = \frac{\sum_{j=1}^{n} Z_{ij}}{n} = 0 \tag{4}$$

5. Verify that the standard deviation of the normalized data is one:

$$s_z = \sqrt{\frac{\sum_{j=1}^{n} (Z_{ij} - \overline{z}_i)^2}{n-1}} = 1 \tag{5}$$

6. Form the correlation matrix (C) for normalized data. Calculate the matrix elements (c_{ij}) as follows:

$$c_{ij} = \frac{\sum_{m=1}^{n} (Z_{im} Z_{jm})}{n-1} \tag{6}$$

7. Calculate the inverse of the correlation matrix (C^{-1}).

8. Finally, calculate MD as:

$$MD_j = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \tag{7}$$

where

x_{ij} is the i^{th} characteristic in the j^{th} observation,

n is the number of observations,

s_i is the standard deviation of the i^{th} characteristic,

Z_{ij} is the normalized value of the i^{th} characteristic in the j^{th} observation,

s_z is the standard deviation of the normalized values,

C is the correlation matrix,

C^{-1} is the inverse of the correlation matrix

MD_j is the Mahalanobis Distance for the j^{th} observation, and

k is the number of characteristics.

Stage 2: Validation of MS

In order to validate MS, observations that fall outside the normal group are identified and the corresponding MD values are calculated. The characteristics of the abnormal group are normalized using the mean and the standard deviation of the corresponding characteristics in the normal group. The correlation matrix corresponding to the normal group is used to compute the MDs of the abnormal cases. If MS has been constructed using the appropriate characteristics, then the MDs of the abnormal group should have higher values than that of the normal group.

Stage 3: Identification of the Useful Characteristics

In this stage, the useful characteristics are determined using orthogonal arrays (OA) and signal to noise (S/N) ratios. An orthogonal array is a table that lists the combination of characteristics, which enables testing the effects of the presence or absence of a characteristic. The size of the OA is determined by the number of characteristics and the levels they can take. In MTS, characteristics in the OA have two levels. Level-1 represents the presence of a characteristic while level-2 represents the absence of a characteristic. For the abnormal cases, MD values are calculated using the

combination of the characteristics dictated by the OA and the larger-the-better signal-to-noise ratio is calculated as follows [23, 25]:

$$\eta_q = -10\log\left[\frac{1}{t}\sum_{j=1}^{t}\frac{1}{MD_j}\right] \tag{8}$$

where

η_q is the signal-to-noise ratio for the q^{th} run of the OA, and

t is the number of abnormalities under consideration

Then, an average signal-to-noise ratio at level-1 and level-2 of each characteristic is obtained and the gain in signal-to-noise ratio values for each characteristic is calculated [23, 25]:

$$Gain = (Avg.S/N\,Ratio)_{Level-1} - (Avg.S/N\,Ratio)_{Level-2} \tag{9}$$

Stage 4: Decision Making

In the final stage, the application under investigation is monitored via data collection using the MS. MDs are calculated and if MD >> 1, it is concluded that the application is displaying abnormal behavior and appropriate corrective actions need to be taken. If $MD \leq 1$, then the conditions are normal.

4.2.2. MD-based Diagnostics. A fault is a change from the normal operation of a system. Fault detection is the first step in diagnostics, which indicates the occurrence of a fault in a monitored system. Diagnostics or also referred to as fault isolation, therefore identifies the root cause of a fault [40]. In this study, an MD-based data clustering technique is used in order to classify rolling element bearing failure data into different fault groups. The technique takes advantage of the fact that the Mahalanobis distance is

calculated using the inverse of the correlation matrix (Eqn. 6). The correlation matrix consists of the correlation coefficients, which is a normalized measure of the strength of the linear relationship between variables as shown below:

$$r_{X,\,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu X)(Y-\mu Y))}{\sigma_X \sigma_Y} \tag{10}$$

where

σ_X and σ_Y are the standard deviations of the random variables X and Y,

μX and μY are the expected values for X and Y,

E is the expected value operator, and

cov denotes the covariance.

The correlation between the MDs can be visualized by plotting the current MD value against the previous MD value. In this case, signatures obtained from faulty operating regimes are grouped into individual clusters as shown in Figure 4.2. In other words, data from the same failure are clustered together due to their correlation.

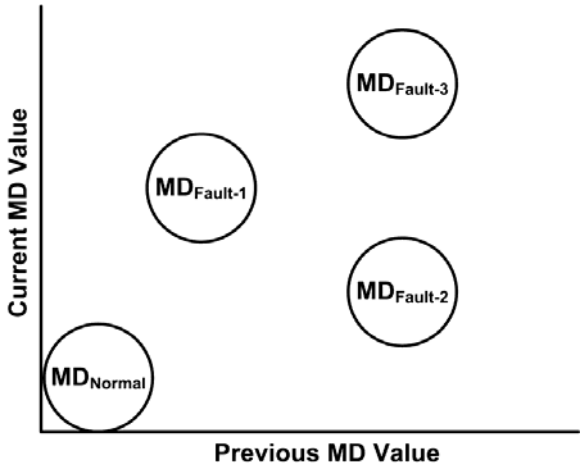

Figure 4.2. MD-based fault clustering

Upon completion of the clustering analysis, the designer selects the appropriate thresholds based on the mean and standard deviation of the MD values for each fault cluster. Consequently, the type of the fault is isolated by comparing the average MD value of the process with the thresholds set for various faulty conditions.

4.2.3. MD-based Prognostics. Prognostics involves a) detection of an abnormal condition or fault, b) identifying the root cause or fault isolation, and, c) estimating the remaining useful life or time to failure. In the prognostics case, the MD values are calculated using a predetermined time window as the process continues. Fault detection occurs once the MDs leave the normal regime and the tracking of the MD trend is initiated. By observing the direction of the transition of the MD trend between the normal case and one of the fault clusters, the possible root cause is identified. This is accomplished by calculating the angle θ between the unknown point U and the mean MD value of each of the known fault clusters assuming the fault progression will follow a roughly linear trajectory (Fig. 4.3). In this case, the smaller the angle, the more likely that the unknown point is progressing towards one of the fault clusters.

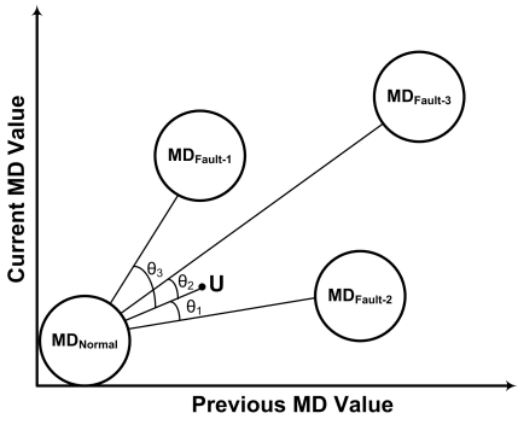

Figure 4.3. Illustration of angle based fault isolation

Once the MD data enters the root cause failure cluster, prognosis of the time to failure is initiated via linear approximation. First, the slope of the progression of the MD value is generated using the MD value of the current and the previous time windows. Then, using the slope and the current MD value, the time to failure is determined at each instant of time (11). The failure threshold is selected by the designer when the performance of the machine is unsatisfactory.

$$slope = \frac{MD_t - MD_{failure-threshold}}{time_t - time_{failure}} \tag{11}$$

In some cases, the fault clusters may lie on a linear path. The linear alignment of the fault clusters would make the classification of the fault progression difficult. For example, consider the fault clusters in Figure 4.4, and assume that the MD value of the current time window is being marked in the plot with an "x". As the process continues, the MD values leave the normal condition, following the direction of the arrow, but it is not clear whether the actual fault is progressing towards fault-1 or fault-2. In this case, a new set of MD values is calculated using one of the faulty data sets as a basis. Consequently, the MS is shifted and the identification of the MD progression becomes straightforward.

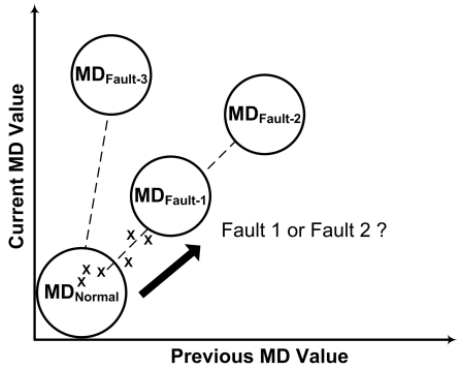

Figure 4.4. Linearly aligned fault clusters

4.3. DESIGN OF EXPERIMENTS

In order to facilitate the development of the proposed MTS-based prognostic scheme a CNC machine testbed has been constructed at Missouri University of Science and Technology. The testbed consists of a Taig MicroMill CNC machine, which houses two ORS (Ortadogu Rulman Sanayi) 6203 brand rolling element bearings in its spindle headstock. Each bearing has been instrumented with two dual vibration/temperature sensors (CTC TA-102-1A) along the x and y axes, while a separate vibration sensor is installed on the worktable in order to determine the noise floor and help in denoising the vibration signals. Figure 4.5 represents the placement of the sensors on the spindle headstock, while. Figure 4.6 shows the actual testbed setup.

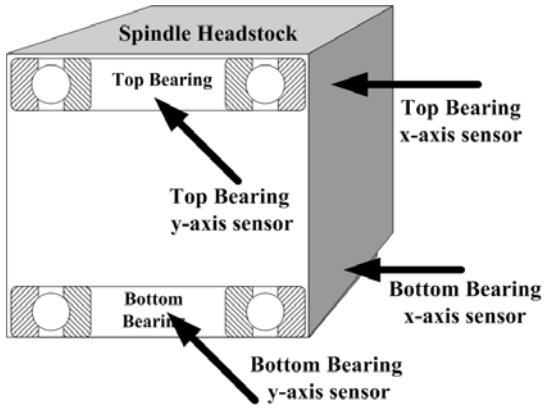

Figure 4.5. Sensor placement on the spindle headstock

The spindle of the CNC machine was operated at 1,600 rpm, while performing a facing operation on Al 6061 plates at a feed rate of 2 in/min. In order to eliminate the effects of tool wear and to achieve consistency between the tests, a brand new 3/8 inch 4-flute end mill was installed into the spindle at the start of each test.

Figure 4.6. Taig MicroMill CNC machine testbed

Vibration and temperature data from both bearings was sampled continuously at a rate of 10 kHz and the data was recorded to a hard drive every 30 seconds. It should be noted that acquisition of data while performing a cutting operation more closely represents the conditions that would be faced in a real-life manufacturing operation. In the normal operating condition case, new bearings were installed into the spindle headstock for each experiment. However, due to the long life of rolling element bearings, further modifications (i.e., seeded with faults) were made to the bearings in order to expedite the fault progression, which is explained below.

Out of the various failure types shown in Fig. 4.7, four different types of bearing operating conditions were investigated in the experiments:

1. Normal Condition: Brand new bearings were operated with no alterations.
2. Cage defect: Bearings were operated after all of the lubrication was removed by the use of solvents in order to accelerate the testing.

3. Inner race defect: A 0.005 inch groove was cut in the inner race of the bearings with the help of a wire EDM machine. Then the bearings were run with no lubrication.
4. Outer race defect: Same as the inner race defect, but this time the groove was placed on the outer race of the bearings.

In order to have adequate amount of data sets for the validation of the proposed scheme, three experiments with normal operating conditions, six experiments with cage defect fault, and, three experiments each of inner and outer race defect faults were performed until the bearings reached a complete failure state and stopped spinning.

It is important to note that due to its proximity to the milling cutter, the bottom bearing of the spindle headstock is exposed to higher level of forces and it was the one that always failed in the experiments. Therefore, the details of the experimental results and analysis, which are explained in the next section, are based on the data collected on the bottom bearings.

4.4. EXPERIMENTAL RESULTS

4.4.1. Feature Extraction. The construction of the Mahalanobis Space requires selection of an initial set of features. Therefore, kurtosis and root-mean-square (RMS) of vibration in the time domain, and the bearing characteristic frequencies in the frequency domain are extracted from the vibration signals of each bearing sensor. The temperature signal is also used as a feature without any further processing. . By assuming a fixed outer race, rotating inner race, and no slippage; equations (12), (13), and (14), provide

frequencies for ball pass inner race (BPFI), ball pass outer race (BPFO), and cage defect (CD), respectively [5].

$$BPFI = \frac{bf}{2}\left(1 + \frac{d}{e}\cos(\beta)\right) \tag{12}$$

$$BPFO = \frac{bf}{2}\left(1 - \frac{d}{e}\cos(\beta)\right) \tag{13}$$

$$CD = \frac{f}{2}\left(1 - \frac{d}{e}\cos(\beta)\right) \tag{14}$$

where

b is the number of rolling elements,

f is the bearing inner race rotational speed,

d is the diameter of the rolling elements,

e is the bearing pitch diameter, and

β is the rolling element contact angle.

The specifications for the experimental bearings are b = 8, f = 1600 rpm, d = 0.2362 in., e = 1.18 in., and β = 15°. Using equations (12-14), the cage defect frequency, ball pass inner race frequency and the ball pass outer race frequency are calculated as 15.91 Hz, 127.3 Hz, and 86.04 Hz, respectively. Upon completion of the feature extraction, each feature is analyzed individually. Across all of the experiments, a large variability in the bearing characteristics frequencies is observed despite the fact that the experiments were performed under identical and strictly controlled conditions.

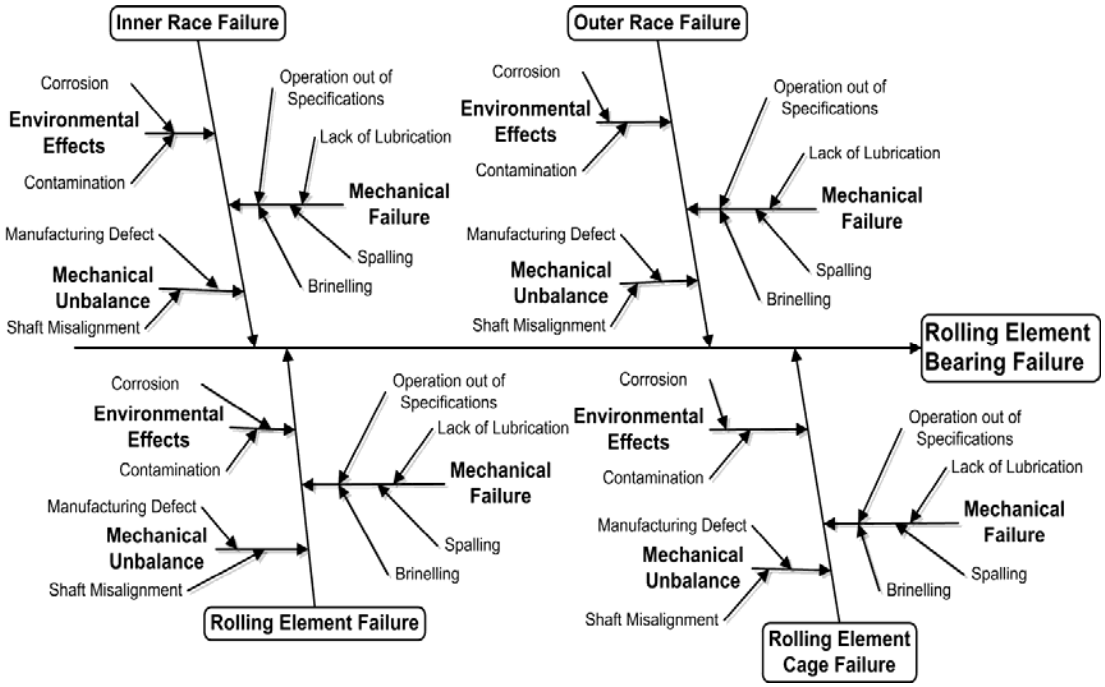

Figure 4.7. Rolling element bearing failure cause-effect diagram

Figures 4.8 through 4.9 highlight the variability of the cage defect frequencies and the bearing failure times. The failure times range between 4.3 hours to approximately 70 hours where as the amplitude of the cage defect frequency at the end of an experiment ranges from -60 dB to -30 dB. This confirms the challenge of setting thresholds based on individual features in order to perform diagnostics and prognostics. Therefore, the need for a multivariate tool that effectively combines multiple characteristics into a single metric, which is the Mahalanobis Distance, is validated.

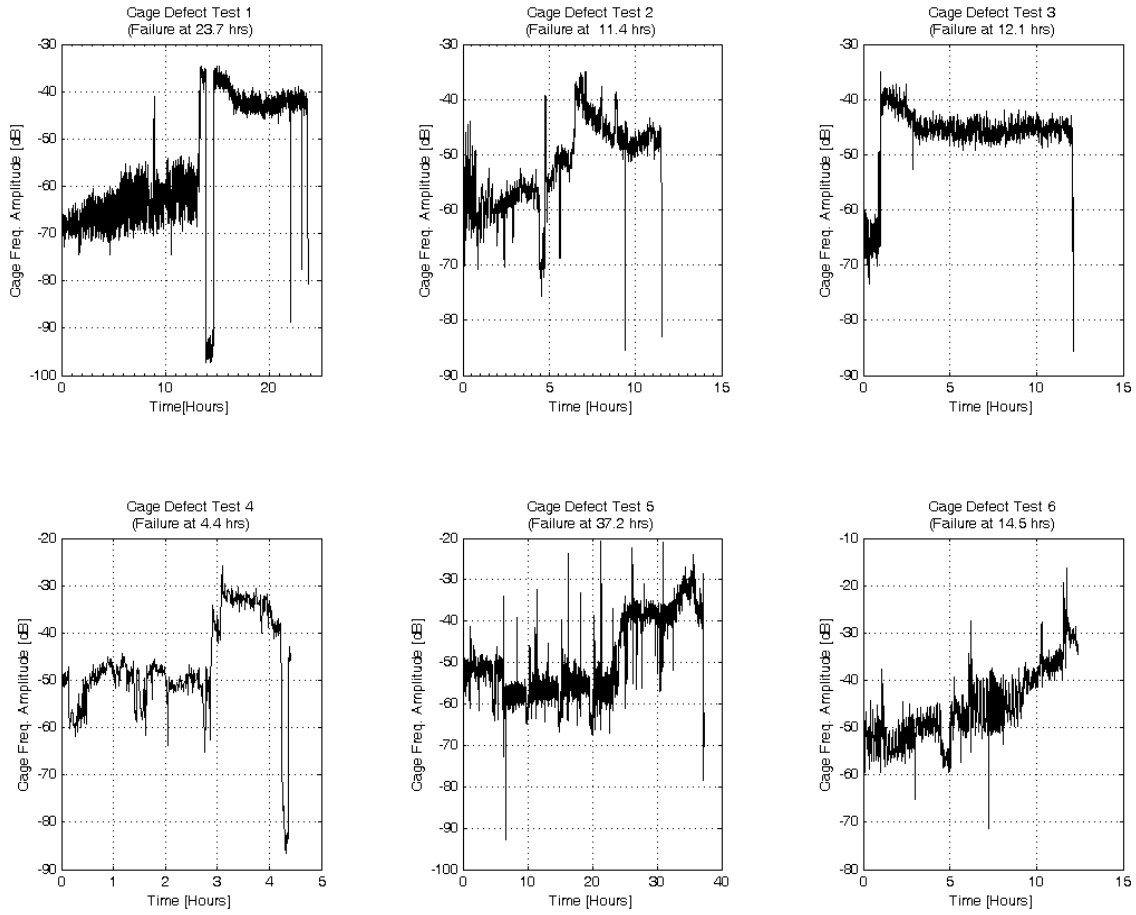

Figure 4.8. Progression of cage defect frequency over time for cage defect tests

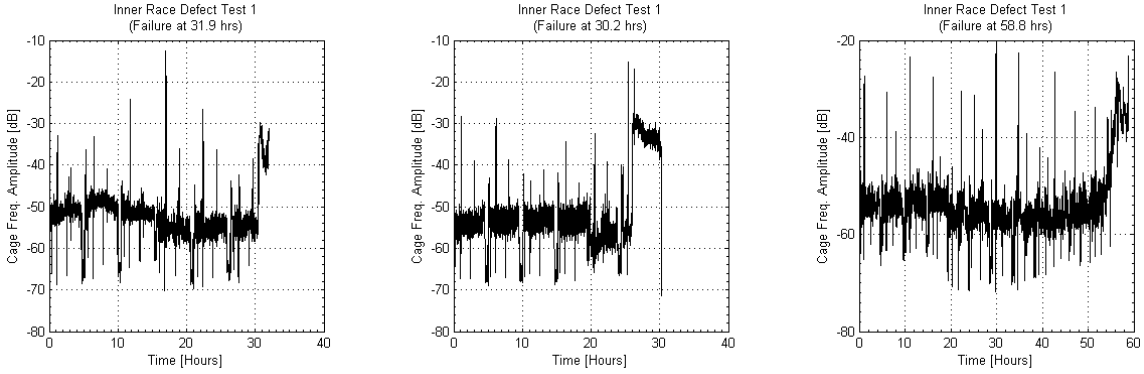

Figure 4.9. Progression of cage defect frequency over time for inner race defect tests

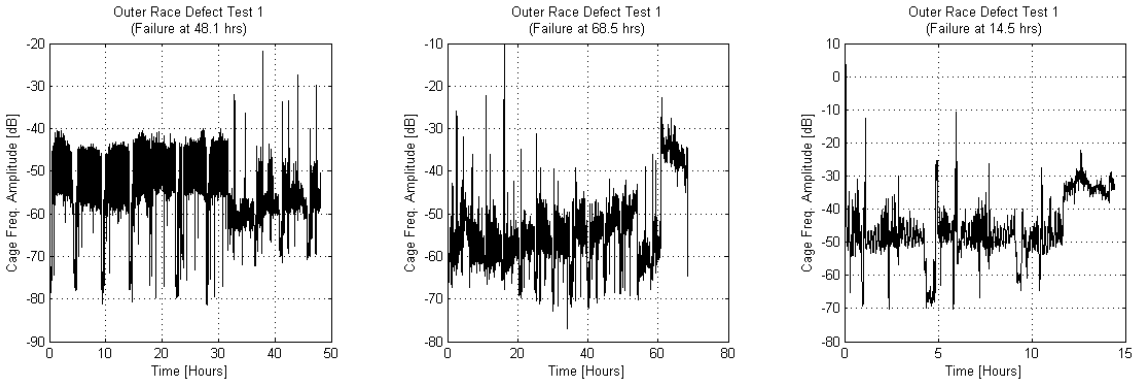

Figure 4.10. Progression of cage defect frequency over time for outer race defect tests

4.4.2. Identification of Key Features. The initial Mahalanobis Space is constructed and corroborated using ten features: cage defect frequencies (CD_x, CD_y), ball pass inner race frequency (BPFI_x, BPFI_y), ball pass outer race frequencies (BPFO_x, BPFO_y), root mean square of the vibration signal (RMS_x, RMS_y), and kurtosis of the vibration signal (KURT_x, KURT_y). The features are extracted by processing the vibration data collected from the sensors in the x and y axes. The raw temperature data (Temp) are also used as the eleventh feature.

Then, the $L_{12}(2^{11})$ orthogonal array is utilized for each of the twelve experiments. Using the larger-the-better type signal to noise ratios and the overall gains, the following eight features are selected as the final inputs to the MTS to be used in further analysis:

- Ball pass inner race frequencies in the x and y axes,
- Ball pass outer race frequencies in the x and y axes,
- Cage defect frequencies in x and y axes, and
- Vibration RMS in the x and y axes

Table 4.1, shows the results of the OA analysis for all of the experiments. The features that display a positive S/N are marked with an "X".

4.4.3. Fault Detection and Isolation. After the key features are identified, MD values for all of the experiments are recalculated. Then, the fault clusters are formed by plotting the current MDs against the previous MDs as described in Section 4.2.2 (Fig. 4.11).

Notice that there are no transition points between the fault clusters in Figure 4.11. This is because the bearings were seeded with faults in order to accelerate the testing (i.e., fault progression), as described in Section 4.3. Using the mean and three times the standard deviation of the MDs for each fault type, the thresholds for each fault cluster are identified. Table 4.2 shows the observed minimum and maximum mean and standard deviations of MDs during the experiments for each fault type. The differences in the values are due to the variability between the data sets as explained earlier.

Table 4.1. Results of the Orthogonal Array Analysis

Exp.	CD_x	BPFI_x	BPFO_x	CD_y	BPFI_y	BPFO_y	KURT_x*	KURT_y*	RMS_x	RMS_y	Temp. *
Cage Defect Test 1					X	X			X	X	X
Cage Defect Test2	X	X	X	X	X				X		
Cage Defect Test 3	X	X	X	X	X				X	X	
Cage Defect Test 4	X	X	X	X	X	X		X	X	X	
Cage Defect Test 5		X	X	X	X	X	X				
Cage Defect Test 6	X	X		X	X				X	X	
Inner Race Defect Test 1	X	X		X		X			X	X	
Inner Race Defect Test 2			X	X		X			X	X	
Inner Race Defect Test 3	X		X	X	X						X
Outer Race Defect Test 1	X	X			X	X			X	X	
Outer Race Defect Test 2	X	X	X	X	X		X		X	X	
Outer Race Defect Test 3	X	X	X	X		X		X	X	X	

Table 4.2. Mean and Standard Deviation of MDs

	Observed max. and min. mean of MDs	Observed max. and min. std. deviation of MDs
Normal Case	0.99 – 0.995	2.05 – 3.07
Cage Fault	2.2 – 15	0.75 – 1.56
Inner Race Fault	190 - 257	5.7 – 28.5
Outer Race Fault	273 - 423	10.4 – 32.8

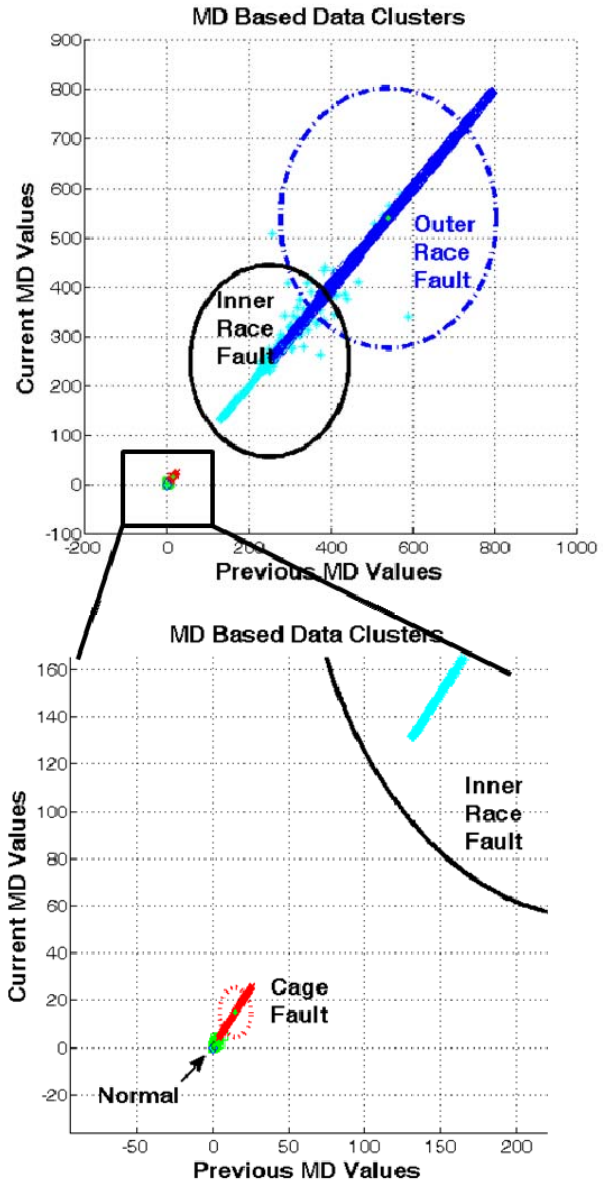

Figure 4.11. MD-based fault clusters

Figure 4.12 displays the fault detection and isolation scheme. The detection starts with the loading/acquisition of test data. Then, the mean of the MD values for the current time window is calculated. If the mean MD is larger than that of the threshold set for the normal case, a fault is detected. Next, the mean MD value is compared with the thresholds of the other fault cases. If the mean MD value stays within the thresholds of a cluster for a user determined time window, then the fault type is isolated.

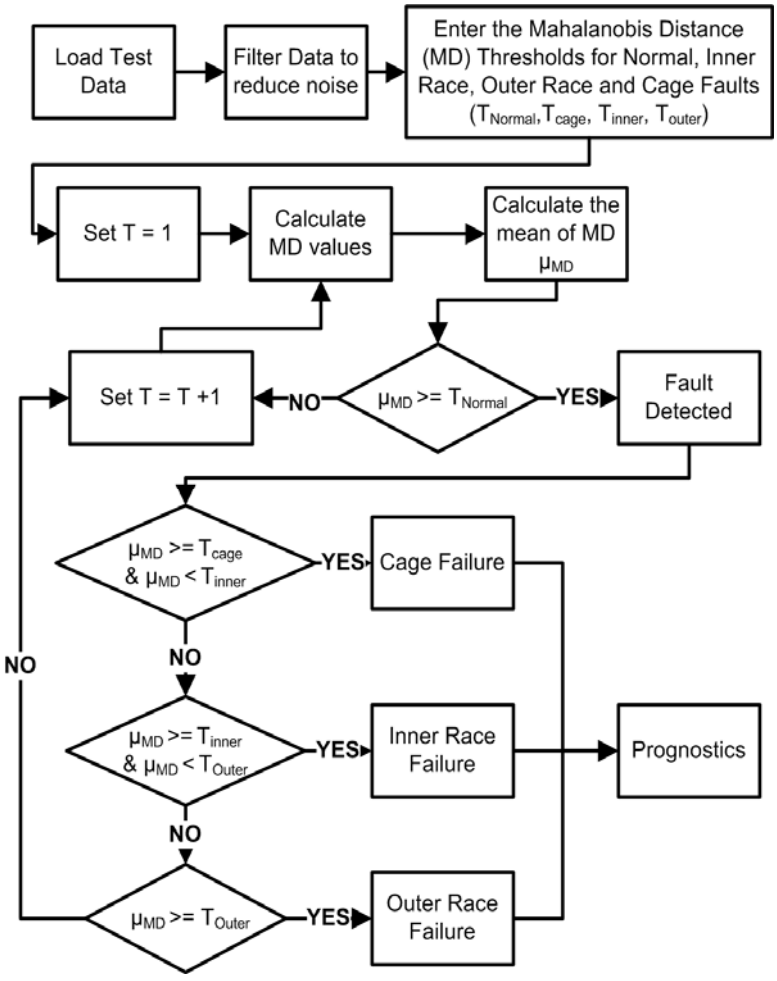

Figure 4.12. Fault detection and isolation scheme

Table 4.3 shows the fault detection and isolation results that were obtained using different combinations of thresholds for the fault clusters. The threshold values displayed in Table 4.3 were obtained by utilizing one third of the experimental data. These data were used for the training of the MTS. Upon analysis of the resulting MD values, the thresholds were selected. The remainder of the experimental data was then used for testing the diagnostic capability of the proposed methodology. The results show 100% success rate in correct detection and isolation of bearing faults.

Table 4.3. Fault Detection and Isolation Results

Fault Type	No. of Cases	Correct Detection	Missed Detection
Cage Fault	6	$T_{cage} = 2$ & $T_{inner} = 190$	
		6	0
		$T_{cage} = 3$ & $T_{inner} = 190$	
		6	0
		$T_{cage} = 5$ & $T_{inner} = 190$	
		6	0
Inner Race	3	$T_{inner} = 190$ & $T_{outer} = 270$	
		3	0
		$T_{inner} = 220$ & $T_{outer} = 270$	
		3	0
		$T_{inner} = 250$ & $T_{outer} = 270$	
		3	0
Outer Race	3	$T_{outer} = 300$	
		3	0

4.4.4. Prognostics. As described in Section 4.2.2, determining the direction of actual progression of a fault is difficult if the fault clusters are aligned linearly (Fig. 4.11). In general, transition data is utilized to obtain the direction of fault progression and to identify the root cause by finding the fault cluster the data leads to. Unfortunately, for the

experiments performed in this study, due to the nature of accelerated fault progression, the transition data is not available. If transition data were available, it could have displayed a gradual progression from the healthy state to the faulty state in a nonlinear fashion, which could be utilized for prognostics.

In order to determine the actual direction of the progression of a fault and to identify the root cause, a new set of MDs are calculated using the cage fault case as a basis for comparison. This realigns the fault clusters as seen in Figure 4.13, where the normal operating condition cluster now sits between the cage fault and inner race fault clusters. The realignment improves the probability of correctly identifying the root cause of a fault since the progression of a cage fault and an inner race fault would now proceed in opposite directions. By selecting the inner race fault and the outer race fault cases as the normal case in MD calculations, two other alignments of the fault clusters are obtained, which covers all of the possible fault progression scenarios.

Figure 4.14 shows the MD-value based prognostic scheme. The proposed scheme exploits the idea of using each of the fault cases as a reference in the generation of fault clusters. At time = 1, test data is acquired and the MD values are calculated using the normal case as the basis for comparison. Fault detection occurs when the MD of the current process crosses the predetermined threshold of the normal case. If a fault is detected, a flag is set to initiate the parallel calculation of MD values with the faulty cases as the references, and the time is advanced to the next time window. Then, by using the four references, the direction of progression of the fault to its fault cluster is determined and the root cause identified.

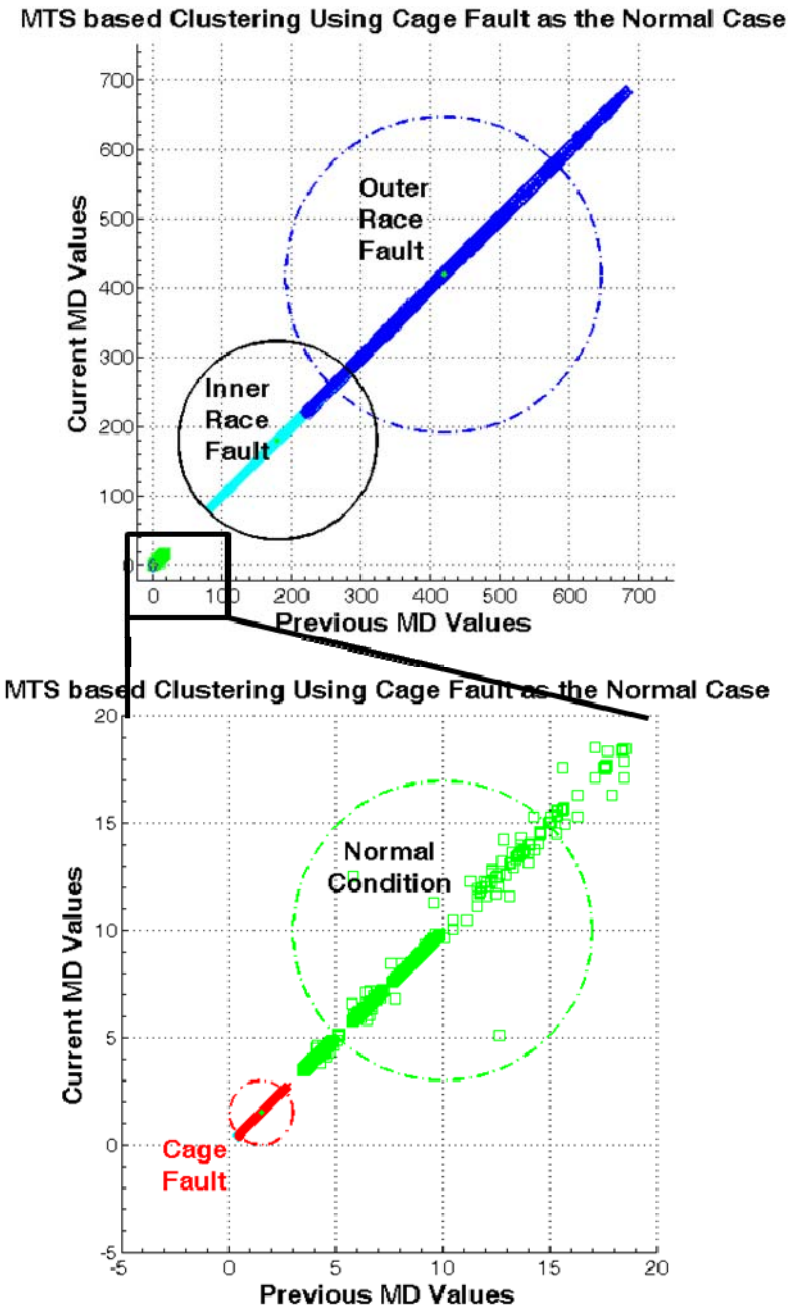

Figure 4.13. Fault clusters using cage fault as the normal case

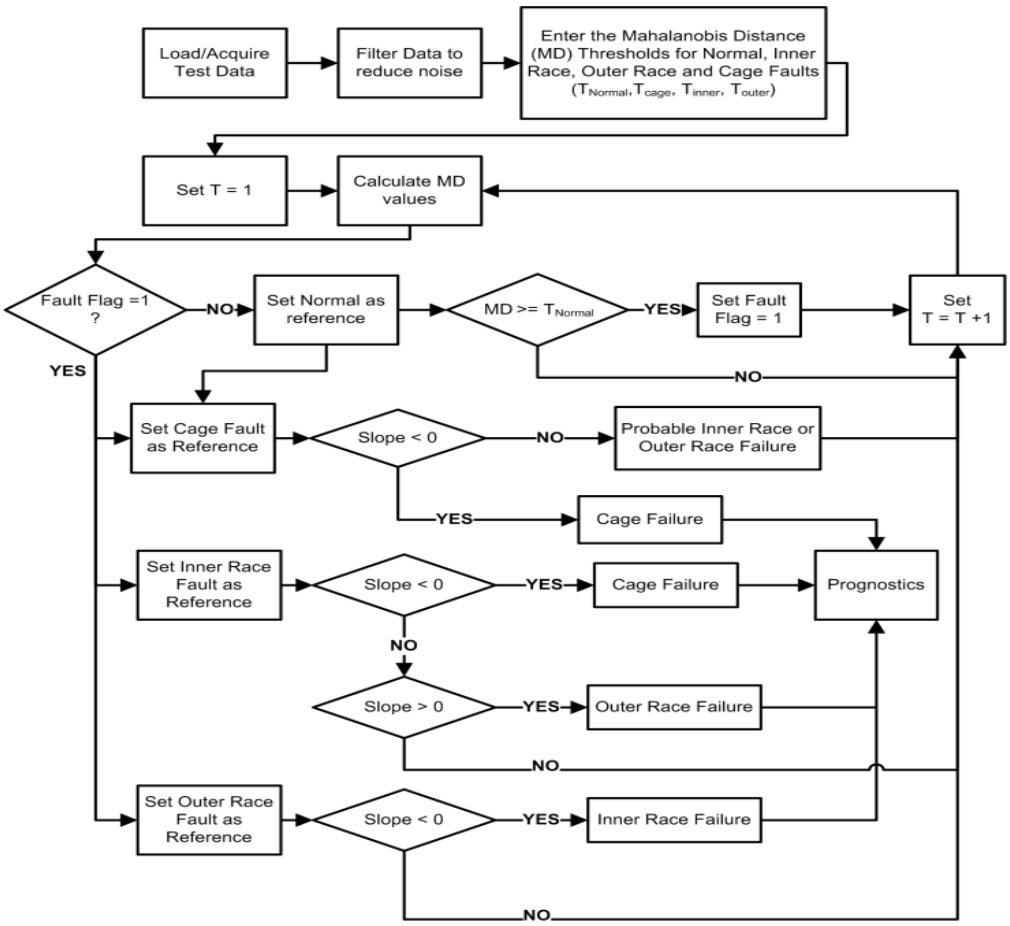

Figure 4.14. Detection of the fault progression

Figure 4.15 shows how the MD values evolve over time for cage defect test #6. Figure 4.16 shows the time-to-failure (TTF) estimation plot for the same experiment using a thirty minute time window and the failure detection threshold given in Table 4.1 using Eqn. (11). The estimation is triggered as soon as a particular operation enters a fault cluster and stays within the cluster for a period of two time windows in order to minimize false fault classification alarms. The failure MD thresholds used in the time to failure calculations was set to 6.3 for Figure 4.16. The prognostics scheme renders satisfactory results for all of the experiments.

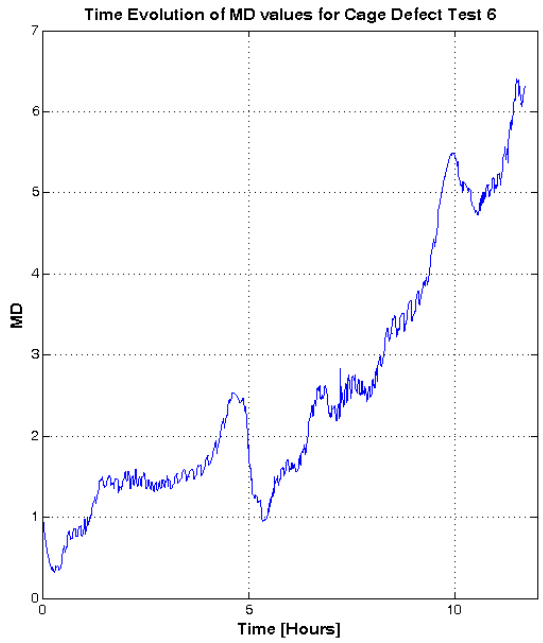

Figure 4.15. Time evolution of MD values for cage failure test 6

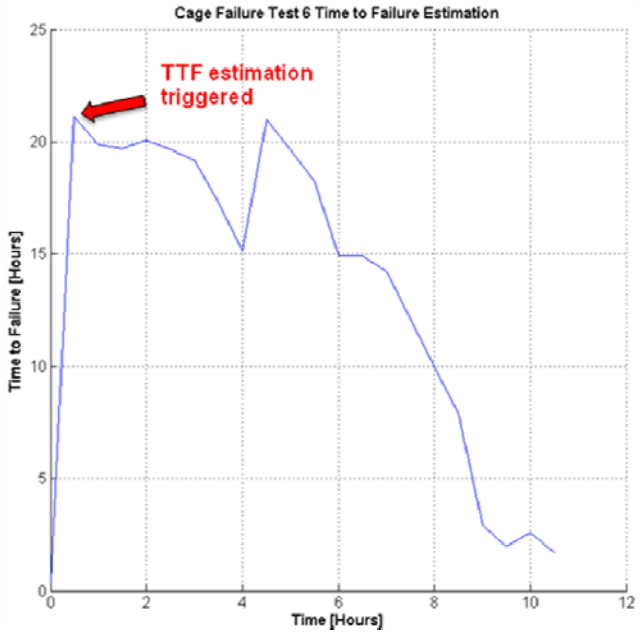

Figure 4.16. Time to failure estimation for cage failure test 6

4.5. CONCLUSIONS

In this paper, a novel Mahalanobis Taguchi System based fault detection, isolation and prognostics scheme is presented. The proposed scheme utilizes Mahalanobis Distance based fault clustering and the progression of MD values for diagnostics and prognostics. The performance of the scheme was validated through experiments performed on rolling element bearings installed in the spindle headstock of a micro-CNC machine testbed.

Advantages of the proposed approach can be summarized as follows:

- Experimental results show large variability between traditional bearing failure indicators, such as cage defect frequency, ball pass inner, outer race frequencies, and the overall failure times. The proposed approach overcomes the difficulty of setting individual thresholds for each fault indicator in such a high variability multivariable environment, by combining all of the pertinent information into a single metric, which is the Mahalanobis Distance.
- It provides a unique solution for rolling element bearing fault detection, isolation and prognostics, eliminating the need for developing a tool for each one separately.
- It is process independent. It can be applied to a wide variety of multi-variate problems provided that a priori normal and abnormal data are available.
- It provides a methodical way to identify the features critical for the process, reducing dimensionality of the problem. Therefore, the analysis overhead is reduced since efforts can be concentrated on the key features.

- Since the proposed scheme is not computationally exhaustive, it can easily be implemented onto wireless motes and deployed to perform online monitoring (i.e., real-time) and decision making in a variety of industrial environments.

The analysis and results presented in this paper, especially the thresholds, are valid for the particular experimental setup and operating conditions. However, the proposed approach can be applied to other processes.

4.6. REFERENCES

[1] H. Ocak, and K. A. Loparo, "A New Bearing Fault Detection and Diagnosis Scheme Based on Hidden Markov Modeling of Vibration Signals," Acoustics, Speech, and Signal Processing, Proceedings ICASSP, vol. 5, pp. 3141-3144 vol.5, 2001.

[2] C. J. Li, and S. M. Wu, "On-Line Detection of Localized Defects in Bearings by Pattern Recognition Analysis," ASME Journal of Engineering for Industry, vol. 111, no. 4, pp. 331-336, 1989.

[3] T. A. Harris, Rolling Bearing Analysis Fourth Edition, New York, NY 10158: Wiley-Interscience, 2001.

[4] P. D. McFadden, and J. D. Smith, "Model for the Vibration Produced by a Single Point Defect in a Rolling Element Bearing," Journal of Sound and Vibration, vol. 96, no. 1, pp. 69-82, 1984.

[5] D. Hochmann, and E. Bechhoefer, "Envelope Bearing Analysis: Theory and Practice," IEEE Aerospace Conference, pp. 3658-3666, 2005.

[6] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-Network-Based Motor Rolling Bearing Fault Diagnosis," Industrial Electronics, IEEE Transactions on, vol. 47, no. 5, pp. 1060-1069, 2000.

[7] S. Janjarasjitt, H. Ocak, and K. A. Loparo, "Bearing Condition Diagnosis and Prognosis Using Applied Nonlinear Dynamical Analysis of Machine Vibration Signal," Journal of Sound and Vibration, vol. 317, no. 1-2, pp. 112-126, 2008.

[8] B. Wu, M. Wang, S. Yu, and C. Feng, "An Approach of Bearing Fault Detection and Diagnosis at Varying Rotating Speed," IEEE ICCA, pp. 1634-1637, 2007.

[9] C. K. Mechefske, and J. Mathew, "Fault Detection and Diagnosis in Low Speed Rolling Element Bearings Using Inductive Inference Classification," Mechanical Systems and Signal Processing, vol. 9, pp. 275-286, 1995.

[10] G. Dalpiaz, A. Rivola, and R. Rubini, "Effectiveness and Sensitivity of Vibration Processing Techniques for Local Fault Detection in Gears," Mechanical Systems and Signal Processing, vol. 14, no. 3, pp. 387-412, 2000.

[11] R. Yan, and R. X. Gao, "An Efficient Approach to Machine Health Diagnosis Based on Harmonic Wavelet Packet Transform," Robotics and Computer-Integrated Manufacturing, vol. 21, no. 4-5, pp. 291-301, 2005.

[12] P. W. Tse, Y. H. Peng, and R. Yam, "Wavelet Analysis and Envelope Detection for Rolling Element Bearing Fault Diagnosis---Their Effectiveness and Flexibilities," Journal of Vibration and Acoustics, vol. 123, no. 3, pp. 303-310, 2001.

[13] X. Lou, and K. A. Loparo, "Bearing Fault Diagnosis Based on Wavelet Transform and Fuzzy Inference," Mechanical Systems and Signal Processing, vol. 18, no. 5, pp. 1077-1095, 2004.

[14] J. S. R. Jang, "Anfis: Adaptive-Network-Based Fuzzy Inference System," Systems, Man and Cybernetics, IEEE Transactions on, vol. 23, no. 3, pp. 665-685, 1993.

[15] A. K. S. Jardine, D. Lin, and D. Banjevic, "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance," Mechanical Systems and Signal Processing, vol. 20, no. 7, pp. 1483-1510, 2006.

[16] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S. Liang, "Adaptive Prognostics for Rolling Element Bearing Condition," Mechanical Systems and Signal Processing, vol. 13, pp. 103-113, 1999.

[17] Y. Li, T. R. Kurfess, and S. Y. Liang, "Stochastic Prognostics for Rolling Element Bearings," Mechanical Systems and Signal Processing, vol. 14, pp. 747-762, 2000.

[18] Y. Li, C. Zhang, T. R. Kurfess, S. Danyluk, and S. Y. Liang, "Diagnostics and Prognostics of a Single Surface Defect on Roller Bearings," Proceedings of the Institution of Mechanical Engineers -- Part C -- Journal of Mechanical Engineering Science, vol. 214, no. 9, pp. 1173-1185, 2000.

[19] J. Luo, B. A., P. K., L. Qiao, K. M., and C. S., "An Interacting Multiple Model Approach to Model-Based Prognostics," Systems, Man and Cybernetics, vol. 1, pp. 189-194 vol.1, 2003.

[20] P. Wang, and G. Vachtsevanos, "Fault Prognostics Using Dynamic Wavelet Neural Networks," AI in Equipment Service, 2000.

[21] H. Qiu, and J. Lee, "Feature Fusion and Degradation Using Self-Organizing Map," Proceedings Machine Learning and Applications, pp. 107-114, 2004.

[22] R. B. Chinnam, and P. Baruah, "A Neuro-Fuzzy Approach for Estimating Mean Residual Life in Condition-Based Maintenance Systems," International Journal of Materials and Product Technology, vol. 20, no. 1/2/3, pp. 166-179, 2004.

[23] G. Taguchi, S. Chowdury, and Y. Wu, The Mahalanobis Taguchi System, New York: McGraw Hill, 2001.

[24] P. C. Mahalanobis, "On the Generalized Distance in Statistics," Proceedings, National Institute of Science of India, vol. 2, no. 1, pp. 49-55, 1936.

[25] G. Taguchi, and R. Jugulum, The Mahalanobis-Taguchi Strategy - a Pattern Technology System, New York: John Wiley & Sons, 2002.

[26] R. B. Chinnam, B. Rai, and N. Singh, "Tool-Condition Monitoring from Degradation Signals Using Mahalanobis Taguchi System," in Robust Engineering, ASI's 20th Annual Symposium, 2004, pp. 343-351.

[27] H. Wang, C. Chiu, and C. Su, "Data Classification Using Mahalanobis Taguchi System," Journal of Chinese Institute of Industrial Engineers, vol. 21, no. 6, pp. 606-618, 2004.

[28] E. A. Cudney, K. Paryani, and K. M. Ragsdell, "Applying the Mahalanobis-Taguchi System to Vehicle Ride," Journal of Industrial and Systems Engineering, vol. 1, no. 3, pp. 251-259, 2007.

[29] E. A. Cudney, R. Jugulum, and K. Paryani, "Forecasting Consumer Satisfaction for Vehicle Ride Using a Multivariate Measurement System," International Journal of Industrial and Systems Engineering, vol. 4, pp. 683-696, 2009.
[30] C. R. Foster, R. Jugulum, and D. D. Frey, "Evaluating an Adaptive One-Factor-at-a-Time Search Procedure within the Mahalanobis-Taguchi System," International Journal of Industrial and Systems Engineering, vol. 4, pp. 600-614, 2009.
[31] M. Asada, "Wafer Yield Prediction by the Mahalanobis-Taguchi System," Statistical Methodology, IEEE International Workshop on, pp. 25-28, 2001.
[32] S. Hayashi, Y. Tanaka, and E. Kodama, "A New Manufacturing Control System Using Mahalanobis Distance for Maximizing Productivity," Semiconductor Manufacturing, IEEE Transactions on, vol. 15, no. 4, pp. 442-446, 2002.
[33] D. Mohan, C. Saygin, and J. Sarangapani, "Real-Time Detection of Grip Length Deviation During Pull-Type Fastening: A Mahalanobis–Taguchi System (Mts)-Based Approach," The International Journal of Advanced Manufacturing Technology, vol. 39, no. 9, pp. 995-1008, 2008.
[34] T. Dasgupta, "Integrating the Improvement and the Control Phase of Six Sigma for Categorical Responses through Application of Mahalanobis-Taguchi System MTS," International Journal of Industrial and Systems Engineering, vol. 4, pp. 615-630, 2009.
[35] J. Srinivasaraghavan, and V. Allada, "Application of Mahalanobis Distance as a Lean Assessment Metric," The International Journal of Advanced Manufacturing Technology, vol. 29, no. 11, pp. 1159-1168, 2006.
[36] S. B. Kim, K.-L. Tsui, T. Sukchotrat, and V. C. P. Chen, "A Comparison Study and Discussion of the Mahalanobis-Taguchi System," International Journal of Industrial and Systems Engineering, vol. 4, pp. 631-644, 2009.
[37] E. A. Cudney, D. Drain, K. Paryani, and S. Naresh, "A Comparison of the Mahalanobis-Taguchi System to a Standard Statistical Method for Defect Detection," Journal of Industrial and Systems Engineering, vol. 2, no. 4, pp. 250-258, 2009.
[38] R. Jugulum, "Comparison between Mahalanobis-Taguchi System and Artificial Neural Networks," Journal of Quality Engineering Forum, vol. 10, no. 1, pp. 60-73, 2002.
[39] W. H. Woodall, R. Koudelik, K.-L. Tsui, S. B. Kim, Z. G. Stoumbos, and C. P. Carvounis, "A Review and Analysis of the Mahalanobis-Taguchi System," Technometrics, vol. 45, pp. 1-15, 2003.
[40] R. Isermann, Fault Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance, Berlin, Germany: Springer, 2005.
[41] V. Wowk, Machinery Vibrations, McGraw Hill, 1991.

5. MAHALANOBIS TAGUCHI SYSTEM (MTS) AS A MULTI-SENSOR BASED DECISION MAKING PROGNOSTICS TOOL FOR CENTRIFUGAL PUMP FAILURES

5.1. INTRODUCTION

Centrifugal pumps are extensively used in a wide variety of process industries such as chemical processing plants, food service, electric utility and automotive companies. There are many types of centrifugal pumps that are specialized for the type of application. Drum centrifugal pumps are generally utilized to transport the contents of drums or barrels, ballast pumps are used on ships in order to load or remove water from a ship's ballast tanks, grinder centrifugal pumps are used so as to grind the waste material in a sewage system for easier transportation of the liquids, and magnetic drive centrifugal pumps are generally used in laboratories and chemical processing. Centrifugal pumps are often critical components in the entire production/process chain [1].

According to a report prepared under the SAVE program for the European Commission, pumps annually consume 117 Terawatt-hours of electricity in Europe alone [2]. Unexpected failure of centrifugal pumps can become costly due to the interruption in processes and further costs are incurred in terms of repair/replacement. In addition, the failure of a pump can also cause potential damage in the equipment elsewhere in the process chain. In a centrifugal pump, bearing, seal and impeller are critical components that directly affect the desired performance of the pump.

Bearing, seal and impeller defects, and cavitation can result in problems such as abnormal noise, leakage, drop in hydraulic performance, structural vibration and damage on the pump components by corrosion and pitting. Therefore, a variety of techniques that

strive to maximize the availability of pumps have been developed and presented in the literature. Based on the literature survey regarding centrifugal pumps presented below, failures of the seal and the impeller are found to be the predominant failure modes. Figures 5.1 and 5.2 show the cause-effect diagram of for seal and impeller failures in centrifugal pumps respectively. In general, researchers have approached the centrifugal pump performance monitoring and fault diagnosis problem in two ways: (1) utilizing qualitative or quantitative model based techniques, (2) using process history based models. Table 5.1 summarizes the various types of failure modes and the sensors that can be utilized to identify these failures, where as Table 6.2 shows the detection techniques.

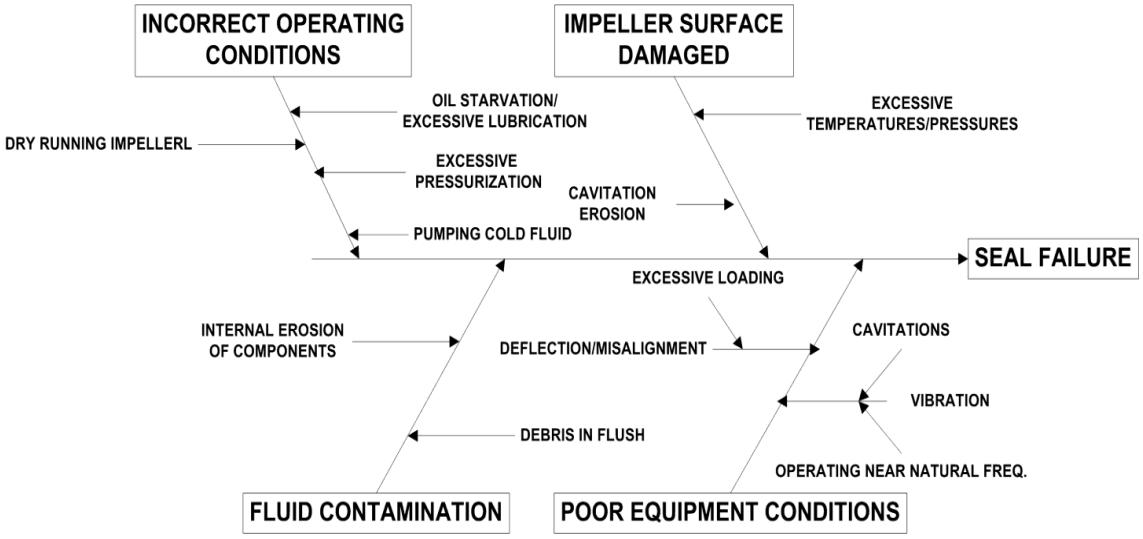

Figure 5.1. Cause-effect diagram for seal failure

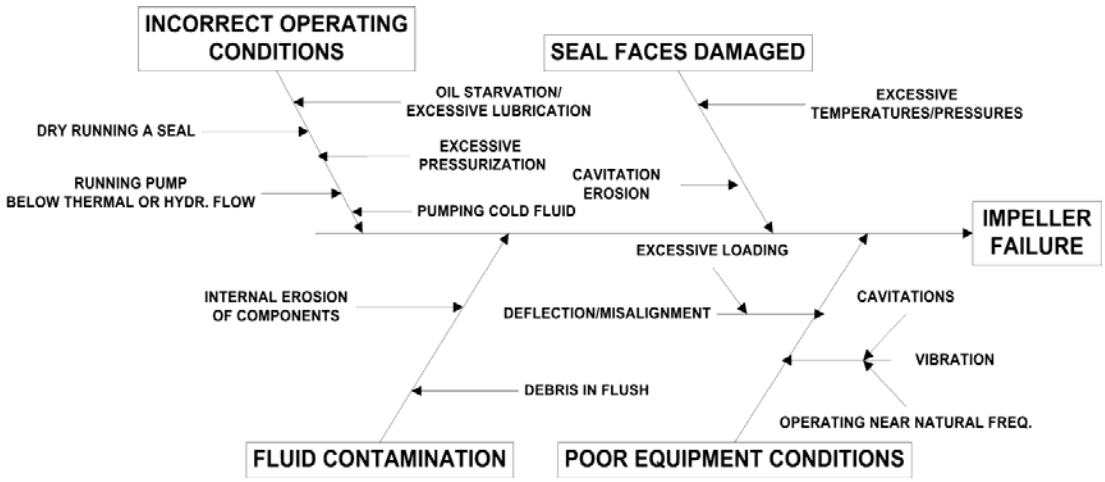

Figure 5.2. Cause-effect diagram for impeller failure

In their paper, Byskov et al. [3] investigate the flow field in a centrifugal pump impeller at design load and quarter-load using large eddy simulation (LES). They compare velocities predicted by the LES and steady-state Reynolds averaged Navier-Stokes (RANS) simulations based on the Baldwin-Lomax and Chien k-ε turbulence models with experimental data obtained from PIV. They report that using LES for analyzing the flow field in centrifugal pumps provides an improved insight into the basic fluid dynamics with a satisfactory accuracy compared to experiments.

Coutier-Delgosha et al. [4] study a pump with two-dimensional curvature blade geometry in cavitating and non-cavitating conditions using different experimental techniques and a three-dimensional numerical model. They compare experimental and numerical results under various flow conditions. Their results show the ability of the developed model to simulate complex three dimensional cavitation in rotating machinery and the associated effects of pump performance, which can be utilized in future pump designs.

Table 5.1. Summary of the Various Failure Modes and Sensors Used for Detection

Failure Modes	Eddy Current Proximity Probes	Optical Pickup	Current Transformer	Thermo-couples	Accelero-meters	Strain Gauge	Piezo electric Transducer	Oil Debris Sensor	Proximity Probes
Cavitation			X [12]		X [1, 8]		X [13, 14]		
Axial Thrust					X [1]				
Radial Thrust					X [1]				
Pressure Pulsations							X [9]		
Impeller Failure	X [3]		X [12]		X [1, 8, 10]		X [13, 14]		X [10]
Seal Failures			X [12]				X [13, 14]		
Bearing Failures					X [15]		X [13, 14]	X [1]	
Vibration					X [1, 8 ,10]				X [1, 8, 10]
Fatigue Failures		X [11]				X [11]			X [11]
Shaft Cracking Failures		X [11]							X [11]

Table 5.2. Techniques Used for Fault Detection

Failure Modes	Acoustic Emission Tech.	Vibration Analysis	Frequency Analysis	High BW Dynamic Analysis	Low BW Performance Analysis	Pit Count	Torsional Vibration Analysis
Cavitation	X [6]		X [12 - 14]	X [8]	X [8]	X [8]	
Axial Thrust		X [1]					
Radial Thrust		X [1]					
Pressure Pulsations		X [9]	X [9]				
Impeller Failure	X [6]	X [1]	X [12 - 14]	X [8]	X [8]		
Seal Failure			X [12 - 14]				
Bearing Failure		X [1]	X [13 - 15]	X [1]	X [1]		
Vibration		X [1]	X [10, 15]				
Fatigue Failure		X [1]	X [11]				X [11]
Shaft Cracking Failure			X [11]				X [11]
Shaft Misalignment		X [1]	X [11]				X [11]

Hirschi et al. [5] propose a three-dimensional numerical method which can predict the cavitation behavior of centrifugal pump. The proposed method, which allows performance drop prediction, consists of assuming the cavity interface as a free surface boundary of the computation domain and computing the single phase flow. The unknown shape of the interface is determined using an iterative procedure matching the cavity surface to a constant pressure boundary. Using the proposed method, the main cavitation characteristics of the pump impeller can predicted and it compares well with Navier-Stokes computations.

Langthjem and Olhoff develop a computationally simple and fast numerical model in order to calculate the flow induced noise in a two-dimensional centrifugal pump [6]. They use a discrete vortex model in order to simulate the flow within the pump and their results show that the proposed method is one of the simplest methods of capturing the essential features of rotational flow. Medvitz et al. [7] use a multi-phase computational fluid dynamics (CFD) method to analyze centrifugal pump flow under developing cavitation conditions. They use a quasi-three dimensional analysis method to model a 7-blade pump impeller across wide range of flow coefficients and cavitation numbers. Their results show that performance trends associated with off-design flow and blade-cavitation, including breakdowns compare qualitatively well with experimental results.

Byington et al. [8] develop an inline intelligent pump monitor for critical hydraulic pumps and motors used in aircraft systems. They acquire pressure, flow and temperature data from the pump system and reduce the raw data by processing and extracting salient features using high and low bandwidth analysis. Their overall approach includes performance modeling, signal processing and feature extraction, feature level fusion, automated classification using neuro-fuzzy classifiers, and knowledge fusion for estimating degradation through the collection of in-line pump data. They demonstrate the success of their approach by running five different pumps to failure.

Dong et al. [9] investigate the effect of the modification of impeller geometry on unsteady flow, pressure fluctuations, and noise in a centrifugal pump. They utilize particle image velocimetry (PIV), pressure and noise measurements after performing various modifications to the impeller and try to determine the overall effects on the pump

performance. Koo and Kim [10] develop a Wigner Distribution based vibration monitoring system in order to analyze vibration signals collected from nuclear reactor coolant pump. They use a rotor kit in order to simulate abnormal conditions such as bearing rubbing, shaft bending and shaft misalignment and within the pump and develop a neural network based online diagnostic method which analyzes the vibration data by utilizing Fourier Transforms and a special software developed with the WD. They achieve a classification accuracy of %81.25 using their proposed method. The study shows that the WD makes it easy to analyze the vibration signals and it helps operators better grasp the cause of vibrations in the pump.

Lebold et al. [11] develop a non-intrusive torsional vibration method in order to monitor and track small changes in crack growth in the shaft of a reactor coolant pump. They initiate and grow fatigue cracks in a laboratory scale shaft under a controlled process. The shaft is equipped with a photosensitive optical encoder and the encoder data is processed with a customized torsional vibration algorithm in order to produce a torsional vibration spectrum. They show that the torsional vibration measurement method has the ability to detect and track natural frequency shifts, which potentially allows online diagnostics and prevention of shaft failure due to cracks.

Perovic et al. [12] develop a fuzzy logic system in order to cavitation, blockage and impeller damage in centrifugal pumps. They establish fault signatures from the pump motor current data by relating spectral features to individual faults using fuzzy logic inference. They then build a fuzzy logic system for the final diagnostic decision making. They test their method by performing experiments on two centrifugal pumps and state

that the analysis of fault signatures using fuzzy logic inference achieves good success in pump diagnostics.

Sakhtivel et al. [13] perform vibration based fault diagnosis of a mono-block centrifugal pump studying normal, bearing fault, impeller fault, seal fault impeller and bearing fault together, and cavitation. They measure vibrations at the pump inlet using a piezo-electric accelerometer and use clear acrylic pipes at the inlet and outlet of the pump in order to visualize the cavitation. They extract standard statistical features such as standard deviation, kurtosis, skewness, and sample variance from the vibration data and use a C4.5 decision tree algorithm in order the classify the features into the operating conditions mentioned above. They report good experimental results which are applicable to practical applications of fault diagnosis of mono-block centrifugal pumps.

In a later paper, Sakhtivel et al. [14] present the use of decision tree and rough sets to generate classification rules from statistical features extracted from vibration signals under normal and faulty operating conditions of centrifugal pump. They build a fuzzy classifier using decision tree and rough set rules using experimental data. The performance of the decision tree and the rough set are compared. They also evaluate the classification accuracy of a principal component analysis based decision tree-fuzzy system. The results show that the performance of the decision tree-fuzzy hybrid system performs better than the rough set-fuzzy hybrid system.

Zhang et al. [15] propose a fuzzy neural network for fault diagnosis of rotary machines. The fault diagnosis system is based on a series of standard fault pattern pairings between fault symptoms and fault and the fuzzy neural network is trained to memorize these pairings. Since fuzzy neural networks adopt bi-directional association,

they make use of information from both the fault symptoms and the faults, which improve the recognition rates. They verify their results through experiments performed on water pump sets of an oil plant and report a well distinguished ability to perform diagnosis of rotary machines.

As it can be seen from the literature survey, most of the authors report success in condition monitoring and fault diagnosis of pumps using various quantitative model-based, qualitative model-based and process history based methods. However, to the best knowledge of the authors, a method which can also successfully predict the remaining useful life of a pump or its components in addition to fault diagnosis is not presented. In this paper, a Mahalanobis-Taguchi System (MTS) based prognostics tool for centrifugal pumps is developed.

MTS has been widely used in various diagnostic applications that deal with data classification. In MTS, Mahalanobis Distance (MD) is used to identify the degree of abnormality of data, whereas the Taguchi methods, which utilize orthogonal arrays (OA) and signal-to-noise ratio (S/N), are used to evaluate the accuracy of predictions and to optimize the system [16]. MD, introduced by P.C. Mahalanobis in 1936 [17], is a multivariate generalized measure used to determine the distance of a data point to the mean of a group. MD is measured in terms of the standard deviations from the mean of the samples and provides a statistical measure of how well the unknown data set matches with the ideal one. The advantage of the MD is that it is sensitive to the intervariable changes in the reference data. Therefore, it has traditionally been used to classify observations into different groups and diagnosis [18].

The benefits of MTS as a pattern recognition and data classification tool can be summarized as follows:

- A robust methodology that is insensitive to variations in multidimensional systems.
- Can handle many different types of data sets and effectively consolidates the data into a useful metric.
- Implementation of MTS requires limited knowledge of statistics.
- Relies typically on simple arithmetic, contextual knowledge, and intuition.
- Its success has been demonstrated in various practical applications.

In their paper, Chinnam et al. [19], use ten features derived from two degradation signals (thrust force and torque) in a drilling operation in order to determine the condition of the drilling tool. The authors utilize MD in order to generate a measurement scale and then they use the Taguchi methods to reduce the number of features down to six. They use the 99^{th} percentile of the MDs from the normal group as a threshold, which shows a superior performance when compared with the any of the individual features.

Wang et al. [20] use the iris data, which is often used in statistical textbooks, and credit card data from a Taiwanese bank in order to display the effectiveness of MTS in data classification. They compare the performance of MTS with discriminant and stepwise discriminant analysis and conclude that MTS shows better performance. Cudney et al. [21] apply MTS in order to study the relationship between available vehicle level performance data for vehicle ride and the corresponding consumer satisfaction ratings for the purpose of improving customer-driven quality. They analyze six vehicle ride parameters from sixty-seven datasets using the MTS, determining the key parameters.

The authors are able to show that the MTS methodology can be used in order to evaluate consumer satisfaction and manufacturers can utilize this knowledge to make the appropriate design decisions regarding attributes that are delivered in a product.

Cudney et al. [22] further investigate the performance data and customer satisfaction relationship using the Mahalanobis-Taguchi Gram-Schmidt (MTGS) technique and showed improved results. In their recent article, Foster et al. [23] develop an alternative search procedure to be used within the MTS. The adaptive One-Factor-AT-a-Time (aOFAT) procedure replaces the orthogonal arrays traditionally utilized in MTS, and features are individually added or removed from the classification system depending on their impact on the overall signal-to-noise ratio. The experimentation shows that the aOFAT procedure renders greater improvements over the median with the same or fewer design alternatives being explored and also exhibits good ability to generalize to new instances after training.

MTS has also been used as a process control and improvement tool to aid manufacturing systems. Asada [24] use MTS to forecast the yield of a wafer production system, which is affected by variability of the electrical properties and environmental dust. MDs are used to compare the relationship between wafer yield and distance. Hayashi et al. [25] develop an MD based method in order to maximize productivity in a semi-conductor manufacturing facility. Their method consists of real-time monitoring of flow factor and in process wafers using MD values to identify the out-of-specification processes and/or tools and take the necessary actions.

Mohan et al. [26] propose a MTS-based diagnostic and root cause analysis scheme for online monitoring of grip-length of pull type fasteners. Their scheme utilizes

data collected from an aerospace industry pull-type fastening tool instrumented with various sensors. Sensor data are analyzed using MTS and the quality of the current operation is communicated back to the user via a wireless network. The authors were able to obtain a 100% data collection on each fastener as opposed to other statistical process control techniques, which rely on sampling.

Dasgupta [27] develop a unified framework for achieving process control and improvement during the implementation phase of Six Sigma. The proposed framework takes advantage of the adaptability of MTS as a classification, variable selection and monitoring tool. The framework is explained using a simulated example. Srinivasaraghavan and Allada [28] also propose a MTGS based methodology in order to support the implementation of lean processes. Their methodology assists contemporary lean assessment tools through providing a measure of leanness by benchmarking exemplary lean industries along with suggestions for improvements based on cost considerations.

MTS is a data driven method that relies on multivariate statistical analysis. Therefore, it is important to compare MTS to well-established multivariate statistical techniques such as principal component analysis, discriminant analysis, cluster analysis, canonical component analysis, and multidimensional scaling, as well as artificial neural networks. For an in-depth comparison of these techniques to MTS, please refer to Kim et al. [29], Cudney et al. [30], Jugulum [31], and Woodall et al. [32]. Books relevant to this study are Harris [3], Isermann [40], and Wowk [41].

The contribution of this paper can be summarized as follows: (1) Uses MTS in order to fuse multi-sensor information into a single system performance metric;

(2) Builds upon MTS by introducing the use of MD-based fault clusters for fault isolation; (3) Utilizes the progression of MD values over time in order to estimate the remaining useful life of components; and (4) Applies MTS to the domain of centrifugal pump monitoring, diagnostics and prognostics for the first time to the best knowledge of the authors.

5.2. METHODOLOGY

In this paper, a Mahalanobis Taguchi System based fault detection, isolation, and prognostics scheme is presented. The proposed scheme utilizes a Mahalanobis Distance based fault clustering method in order to classify faults into different categories. Using these clusters, fault detection and root cause analysis is performed. The scheme also utilizes the progression of the MD values over time in order to facilitate prognosis of time to failure for components. The details of the scheme are presented in the rest of the section.

5.2.1. Mahalanobis Taguchi System. MTS starts with data collection on normal observations. Then, Mahalanobis Distance is calculated using certain characteristics to investigate whether MD has the ability to differentiate the normal group from an abnormal group. If MD cannot detect the normal group using those particular characteristics, then a new combination of characteristics need to be explored. When the right set of characteristics are found, Taguchi methods are employed to evaluate the contribution of each characteristic. If possible, dimensionality is reduced by eliminating those characteristics that do not add value to the analysis. MTS consists of four stages [23, 25]:

Stage 1: Construction of Mahalanobis Space

The first step in MTS is the construction of the Mahalanobis Space (MS) as the reference. In order to construct MS, the variables that represent "normal" conditions are defined and sampled. In this study, new bearings that are properly lubricated represent the normal case. The steps for the construction of MS are outlined below:

1. Calculate the mean for each characteristic in the normal data set as:

$$\bar{x}i = \frac{\sum_{j=1}^{n} X_{ij}}{n} \tag{1}$$

2. Then, calculate the standard deviation for each characteristic:

$$s_i = \sqrt{\frac{\sum_{j=1}^{n} (X_{ij} - \bar{x}_i)^2}{n-1}} \tag{2}$$

3. Next, normalize each characteristic, form the normalized data matrix (Z_{ij}) and take its transpose (Z_{ij}^T)

$$Z_{ij} = \frac{(X_{ij} - \bar{x}_i)}{s_i} \tag{3}$$

4. Then, verify that the mean of the normalized data is zero:

$$\bar{z}_i = \frac{\sum_{j=1}^{n} Z_{ij}}{n} = 0 \tag{4}$$

5. Verify that the standard deviation of the normalized data is one:

$$s_z = \sqrt{\frac{\sum_{j=1}^{n} (Z_{ij} - \bar{z}_i)^2}{n-1}} = 1 \tag{5}$$

6. Form the correlation matrix (C) for normalized data. Calculate the matrix elements (c_{ij}) as follows:

$$c_{ij} = \frac{\sum_{m=1}^{n}(Z_{im}Z_{jm})}{n-1} \tag{6}$$

7. Calculate the inverse of the correlation matrix (C^{-1}).

8. Finally, calculate MD as:

$$MD_j = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \tag{7}$$

where

x_{ij} is the i^{th} characteristic in the j^{th} observation,

n is the number of observations,

s_i is the standard deviation of the i^{th} characteristic,

Z_{ij} is the normalized value of the i^{th} characteristic in the j^{th} observation,

s_z is the standard deviation of the normalized values,

C is the correlation matrix,

C^{-1} is the inverse of the correlation matrix

MD_j is the Mahalanobis Distance for the j^{th} observation, and

k is the number of characteristics.

Stage 2: Validation of MS

In order to validate MS, observations that fall outside the normal group are identified and the corresponding MD values are calculated. The characteristics of the abnormal group are normalized using the mean and the standard deviation of the

corresponding characteristics in the normal group. The correlation matrix corresponding to the normal group is used to compute the MDs of the abnormal cases. If MS has been constructed using the appropriate characteristics, then the MDs of the abnormal group should have higher values than that of the normal group.

Stage 3: Identification of the Useful Characteristics

In this stage, the useful characteristics are determined using orthogonal arrays (OA) and signal to noise (S/N) ratios. An orthogonal array is a table that lists the combination of characteristics, which enables testing the effects of the presence or absence of a characteristic. The size of the OA is determined by the number of characteristics and the levels they can take. In MTS, characteristics in the OA have two levels. Level-1 represents the presence of a characteristic while level-2 represents the absence of a characteristic. For the abnormal cases, MD values are calculated using the combination of the characteristics dictated by the OA and the larger-the-better signal-to-noise ratio is calculated as follows [23, 25]:

$$\eta_q = -10\log\left[\frac{1}{t}\sum_{j=1}^{t}\frac{1}{MD_j}\right] \tag{8}$$

where

η_q is the signal-to-noise ratio for the q^{th} run of the OA, and

t is the number of abnormalities under consideration

Then, an average signal-to-noise ratio at level-1 and level-2 of each characteristic is obtained and the gain in signal-to-noise ratio values for each characteristic is calculated [23, 25]:

$$Gain = (Avg.S/N\,Ratio)_{Level-1} - (Avg.S/N\,Ratio)_{Level-2} \tag{9}$$

Stage 4: Decision Making

In the final stage, the application under investigation is monitored via data collection using the MS. MDs are calculated and if MD >> 1, it is concluded that the application is displaying abnormal behavior and appropriate corrective actions need to be taken. If $MD \leq 1$, then the conditions are normal.

5.2.2. MD-based Diagnostics. A fault is a change from the normal operation of a system. Fault detection is the first step in diagnostics, which indicates the occurrence of a fault in a monitored system. Diagnostics or also referred to as fault isolation, therefore identifies the root cause of a fault [40]. In this study, an MD-based data clustering technique is used in order to classify rolling element bearing failure data into different fault groups. The technique takes advantage of the fact that the Mahalanobis distance is calculated using the inverse of the correlation matrix (Eqn. 6). The correlation matrix consists of the correlation coefficients, which is a normalized measure of the strength of the linear relationship between variables as shown below:

$$r_{X,\,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu X)(Y-\mu Y))}{\sigma_X \sigma_Y} \tag{10}$$

where

σ_X and σ_Y are the standard deviations of the random variables X and Y,

μX and μY are the expected values for X and Y,

E is the expected value operator, and

cov denotes the covariance.

The correlation between the MDs can be visualized by plotting the current MD value against the previous MD value. In this case, signatures obtained from faulty

operating regimes are grouped into individual clusters as shown in Figure 5.3. In other words, data from the same failure are clustered together due to their correlation.

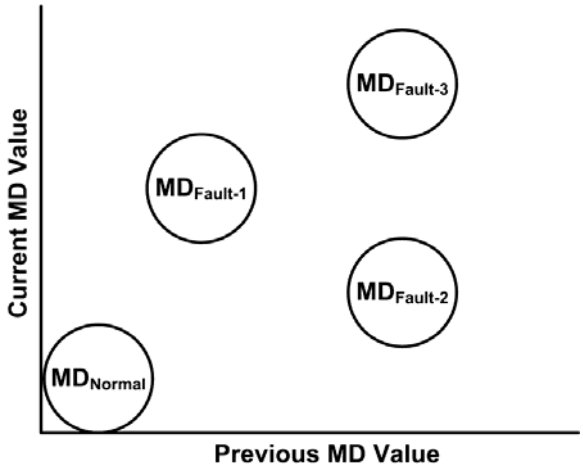

Figure 5.3. MD-based fault clustering

Upon completion of the clustering analysis, the designer selects the appropriate thresholds based on the mean and standard deviation of the MD values for each fault cluster. Consequently, the type of the fault is isolated by comparing the average MD value of the process with the thresholds set for various faulty conditions.

5.2.3. MD-based Prognostics. Prognostics involves a) detection of an abnormal condition or fault, b) identifying the root cause or fault isolation, and, c) estimating the remaining useful life or time to failure. In the prognostics case, the MD values are calculated using a predetermined time window as the process continues. Fault detection occurs once the MDs leave the normal regime and the tracking of the MD trend is initiated. By observing the direction of the transition of the MD trend between the normal

case and one of the fault clusters, the possible root cause is identified. This is accomplished by calculating the angle θ between the unknown point U and the mean MD value of each of the known fault clusters assuming the fault progression will follow a roughly linear trajectory (Fig. 5.4). In this case, the smaller the angle, the more likely that the unknown point is progressing towards one of the fault clusters.

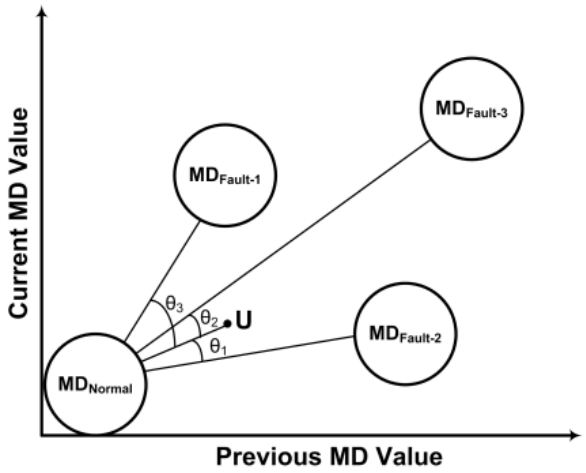

Figure 5.4. Illustration of angle based fault isolation

Once the MD data enters the root cause failure cluster, prognosis of the time to failure is initiated via linear approximation. First, the slope of the progression of the MD value is generated using the MD value of the current and the previous time windows. Then, using the slope and the current MD value, the time to failure is determined at each instant of time (11). The failure threshold is selected by the designer when the performance of the machine is unsatisfactory.

$$slope = \frac{MD_t - MD_{failure-threshold}}{time_t - time_{failure}} \tag{11}$$

In some cases, the fault clusters may lie on a linear path. The linear alignment of the fault clusters would make the classification of the fault progression difficult. For example, consider the fault clusters in Figure 5.5, and assume that the MD value of the current time window is being marked in the plot with an "x". As the process continues, the MD values leave the normal condition, following the direction of the arrow, but it is not clear whether the actual fault is progressing towards fault-1 or fault-2. In this case, a new set of MD values is calculated using one of the faulty data sets as a basis. Consequently, the MS is shifted and the identification of the MD progression becomes straightforward.

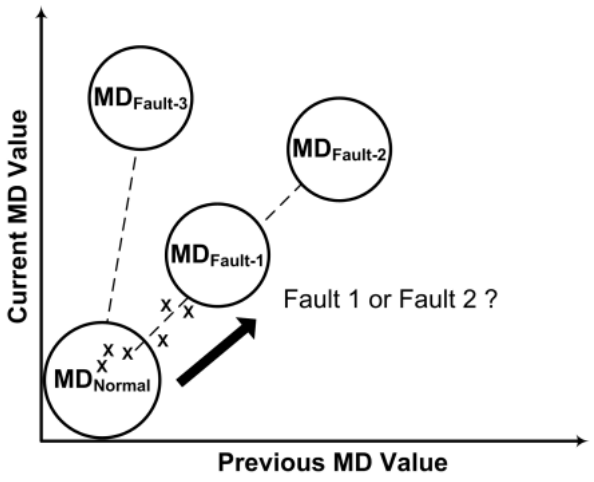

Figure 5.5. Linearly aligned fault clusters

5.3. DESIGN OF EXPERIMENTS

In order to facilitate the development of the proposed multi-sensor based MTS prognostics tool, a centrifugal water pump testbed has been constructed at the Missouri University of Science and Technology. The 1/2 HP centrifugal pump has been instrumented with the following sensors at the inlet and the outlet of the pump:

- Inlet: pressure, temperature
- Outlet: pressure, temperature and flow

In addition to these sensors, two accelerometers were installed in the pump casing in order to measure the lateral and vertical vibration. Figure 5.6 shows a schematic representation of the testbed setup.

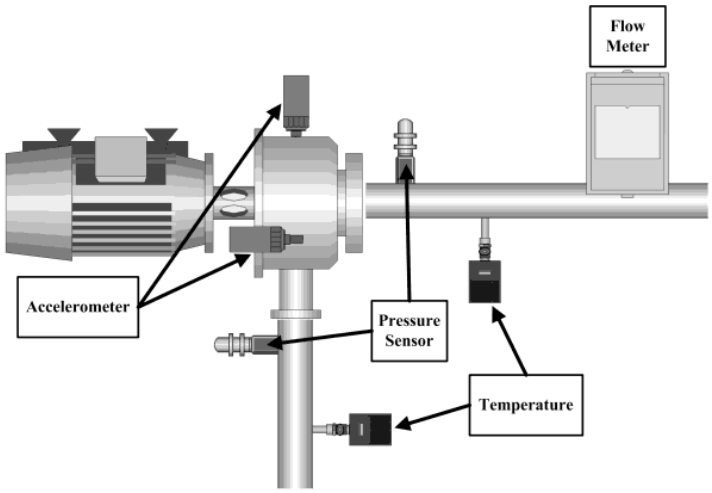

Figure 5.6. Centrifugal water pump testbed

The water pump testbed was operated for 150 hours in order to collect the signature for normal operating conditions. After the initial data collection was complete,

three types of pump failures one at a time were seeded in order to accelerate the testing. In one of the tests, the seal was mechanically defaced to create an artificial seal failure. The pump was operated with the damaged seal and sensor signatures were collected. During the experiment, a drop in the inlet pressure was noticed.

In another experiment, impeller failure was investigated. The failure was seeded by intermittent dry running of the pump, which caused erosion on the impeller surface. Since the experiments were taking a long time, the blades on the impeller were mechanically filed down in order to further accelerate the testing. The experiment was carried out until a large change from the normal sensor signature was noticed. As the damage on the impeller progressed, corresponding drop in flow values was observed. The accelerometer signatures were also observed to deviate from the normal case. Figure 5.7 shows the damage on the impeller surface at the end of the experiment.

Figure 5.7. Erosion damage on the impeller surface

In the final experiment, the filter clog failure was studied. The pump was operated while contaminants such as sand were added in to the pump's water reservoir. Since the

filter is located right before the main body of the pump, a clog in the filter resulted in a drop in the inlet and outlet pressures, and the outlet flow.

5.4. EXPERIMENTAL RESULTS

5.4.1. Identification of Key Parameters. The construction of the Mahalanobis Space requires the selection of an initial set of parameters [18]. In some cases, raw sensor data may provide sufficient information in order to differentiate between various failure modes. When the raw sensor data is not sufficient, it is customary to search for higher level features that can be extracted by further processing sensor data e.g., the bearing characteristic frequencies obtained from vibration data for rolling element bearings. In this study, the raw sensor measurements are used as the input parameters for the MTS since they provide adequate information. The input parameters are:

- Lateral acceleration
- Vertical acceleration
- Inlet pressure
- Outlet pressure
- Outlet flow

Then, the $L_8(2^7)$ orthogonal array is utilized for the S/N analysis. Using the larger-the-better type signal to noise ratios and the overall gains, all of the parameters mentioned above are selected as the final input parameters for further analysis since none of the inputs could be eliminated. Figures 5.8 through 5.10 show the results for the OA analysis for the experiments.

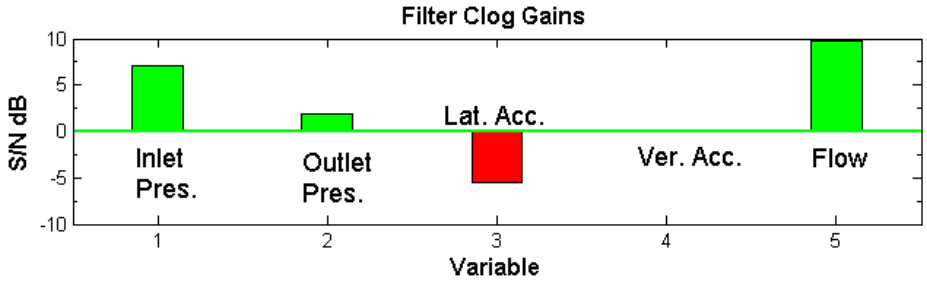

Figure 5.8. OA analysis gains for filter clog failure

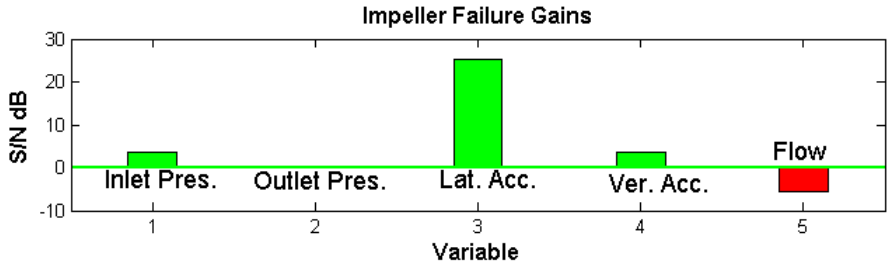

Figure 5.9. OA analysis gains for impeller failure

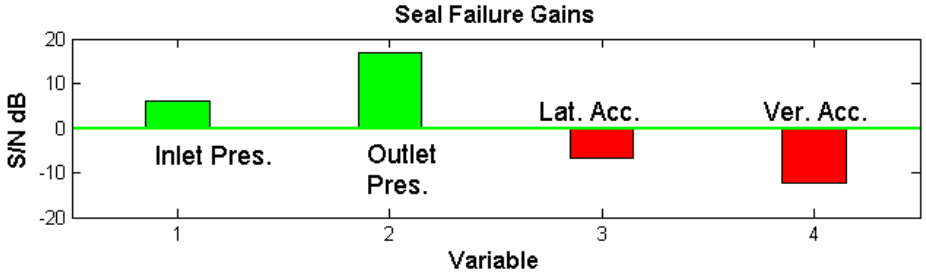

Figure 5.10. OA analysis gains for seal failure

5.4.2. Analysis on the Number and Type of Sensors Required. In an ideal case, one would like to use as few as possible sensors to perform diagnostics and prognostics on a system. The implementation of multiple sensors may be become costly, or may altogether be unfeasible due to the mechanical limitations of the system. In order

to determine the number and type of sensors required for identifying different centrifugal pump failures, a series of analyses using only the data from one sensor at a time were performed for the filter clog, seal and impeller failure cases.

5.4.2.1. Pressure sensor installed on the pump outlet. The outlet pressure of a centrifugal pump is one of the main performance characteristics. Therefore, it would be ideal to be able to determine the health state of the entire pump components by measuring only the outlet pressure. In order determine the adequacy of using only a pressure sensor on the outlet of the pump, MD fault clusters are formed as mentioned in Section 5.2.2. Figure 5.11 shows the fault clusters.

As it can be seen from Figure 5.11, the normal and the impeller failure cases can clearly be identified using the outlet pressure. However, the seal failure and the filter clog failure fault clusters overlap with each other

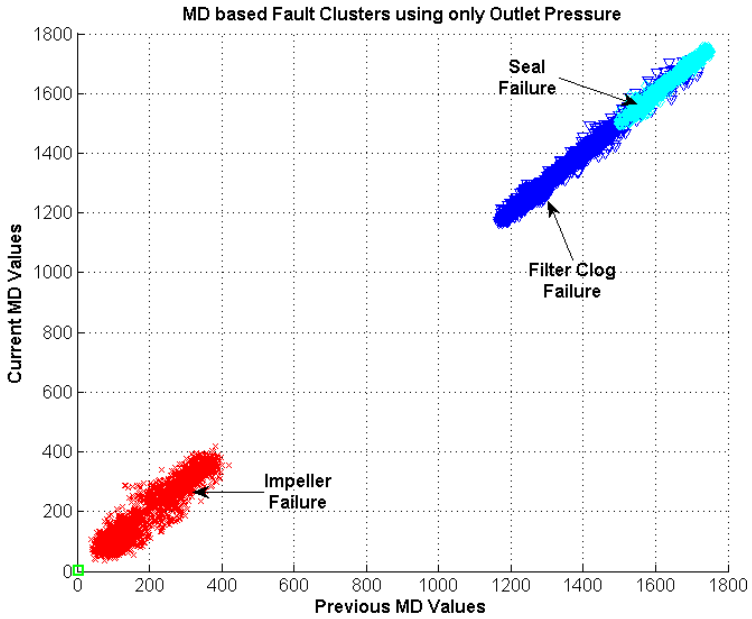

Figure 5.11. MD based fault clusters using only the outlet pressure

5.4.2.2. Accelerometer installed on the pump casing in the lateral axis. Another good indicator of the health state of a centrifugal pump is the amount of vibration that it generates. In a normal case the pump should operate within a specified range of vibrations, therefore the amount of vibration experienced on the pump could be good indicator of potential problems. Figure 5.12 depicts the MD based fault clusters using the data obtained from the lateral accelerometer on the pump casing. Figure 5.13 shows a zoomed-in plot of failure clusters.

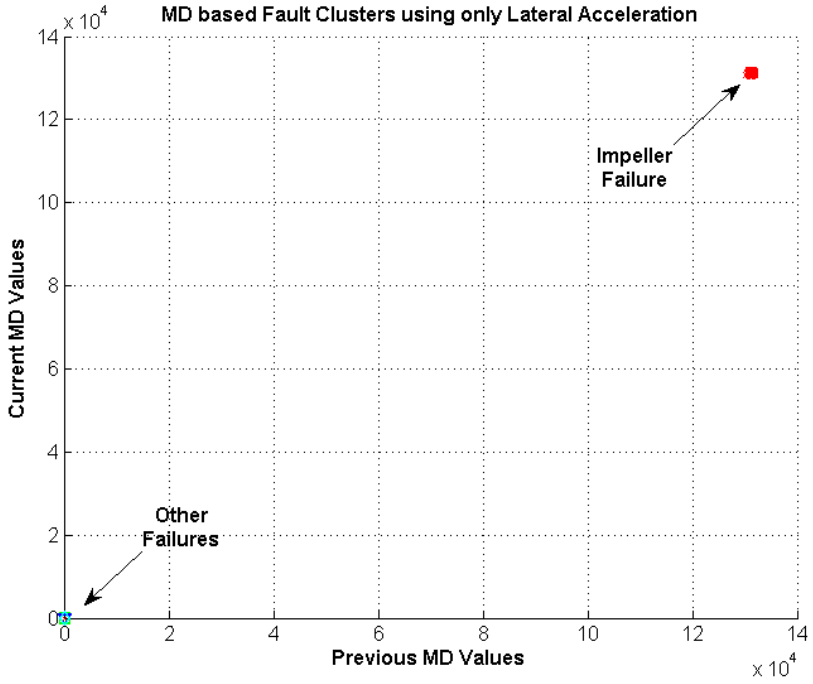

Figure 5.12. MD based fault clusters using only lateral acceleration

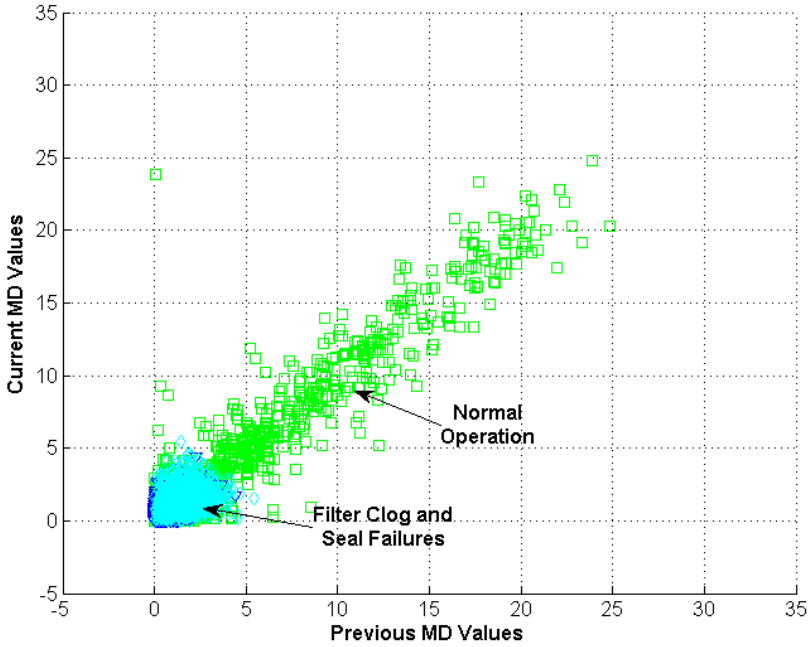

Figure 5.13. Zoomed-in version of Fig. 5.12

As it can be seen from Figures 5.12 and 5.13, the impeller failure can clearly be identified using only the lateral acceleration information. However, the normal, seal failure and filter clog failure conditions completely overlap with each other and cannot be separated.

5.4.2.3. Fault detection and isolation using multiple sensors. The analyses in Section 5.3.2.1 and 5.3.2.2 show that using a single type of sensor is not sufficient for successful fault isolation on a centrifugal pump. Therefore multiple types of sensors are required.

Using all of the sensors mentioned in Section 5.3, MD values for all of the experiments are calculated. Table 5.3 shows the observed minimum and maximum values of MDs as well as the standard deviations during the experiments for each fault type. Using the mean and the standard deviation of the MDs for each fault type, the thresholds

for the each cluster is identified. Figure 5.14 shows the fault clusters that are used in the final analysis. Notice that there are no transition points between the fault clusters in Figure 5.14. This is because the experiments were performed in an accelerated manner. Figure 5.15 zooms-in on the normal operation, filter clog, and the impeller failures.

Table 5.3. Mean and Standard Deviation of MDs

	Observed min. and max. values of MDs	Observed std. deviation of MDs
Normal Case	0.02 – 10.11	0.89
Filter Clog Failure	482 - 582.17	38.16
Seal Failure	561.33 – 827.02	59.21
Impeller Failure	28430 - 29179	155.59

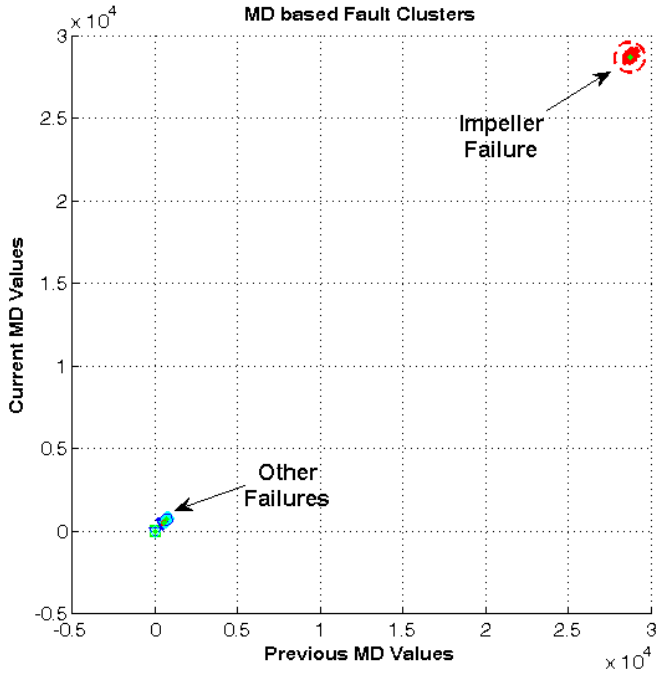

Figure 5.14. MD based fault clusters using all sensors

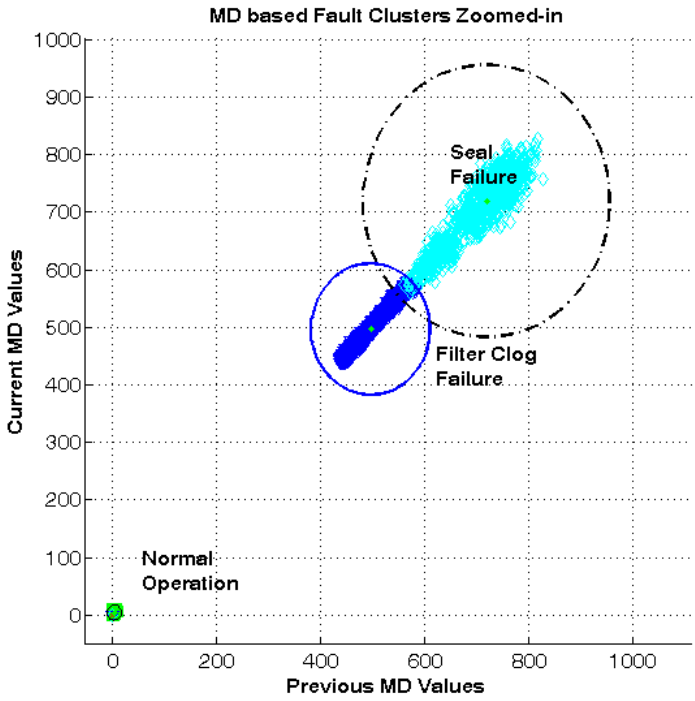

Figure 5.15. MD based fault clusters zoomed-in

Figure 5.16 displays the fault detection and isolation scheme for the centrifugal pump. The detection starts with the loading/acquisition of test data. Then, the mean of the MD values for the current time window is calculated. If the mean MD is larger than that of the threshold set for the normal case, a fault is detected. Next, the mean MD value is compared with the thresholds of the other fault cases. If the mean MD value stays within the thresholds of a cluster for a user determined time window, then the fault type is isolated.

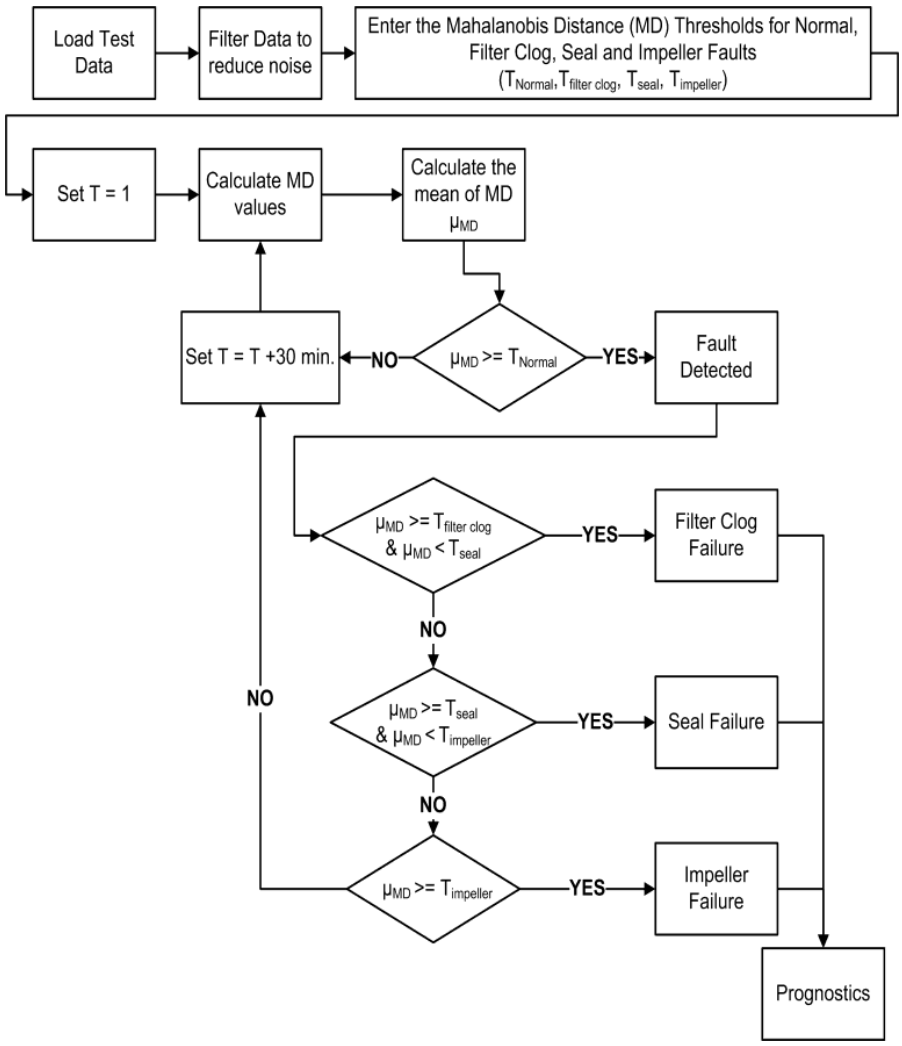

Figure 5.16. Fault detection and isolation scheme

Table 5.4 shows the fault detection and isolation results that were obtained using different combinations of thresholds for the fault clusters. The results show a high success rate in correct detection and isolation of centrifugal pump faults. In the case where the filter clog failure threshold is selected as 570 and the seal failure threshold is selected as 562, the proposed scheme misses the detection. This is due to the slight overlapping between the fault clusters as can be seen in Figure 5.15.

Table 5.4. Fault Detection and Isolation Results

Fault Type	Correct Detection	Missed Detection
Filter Clog Failure	T_{filter} = 482 & T_{seal} = 562	
	1	0
	T_{filter} = 500 & T_{seal} = 562	
	1	0
	T_{filter} =570 & T_{seal} = 562	
	0	1
Seal Failure	T_{seal} = 562 & $T_{impeller}$ =28430	
	1	0
	T_{seal} = 600 & $T_{impeller}$ = 28430	
	1	0
	T_{seal} = 700 & $T_{impeller}$ = 28430	
	1	0
Impeller Failure	$T_{impeller}$ = 28430	
	1	0

5.4.3. Prognostics. As described in Section 5.2.2, determining the direction of actual progression of a fault is difficult if the fault clusters are aligned linearly (Figures 5.14 and 5.15). In general, transition data is utilized to obtain the direction of fault progression and to identify the root cause by finding the fault cluster the data leads to. Unfortunately, for the experiments performed in this study, due to the nature of accelerated fault progression, the transition data is not available. If transition data were available, it could have displayed a gradual progression from the healthy state to the faulty state in a nonlinear fashion, which could be utilized for prognostics.

In order to determine the actual direction of the progression of a fault and to identify the root cause, a new set of MDs are calculated using the filter clog case as a basis for comparison. This realigns the fault clusters as seen in Figure 5.17, where the normal operating condition cluster now sits between the filter clog failure and the seal

failure clusters. The realignment improves the probability of correctly identifying the root cause of a fault since the progression of a filter clog fault and a seal fault would now proceed in opposite directions. By selecting the seal failure and the impeller failure cases as the normal case in MD calculations, two other alignments of the fault clusters are obtained, which covers all of the possible fault progression scenarios.

Fig. 5.18 shows the MD-value based prognostic scheme. The proposed scheme exploits the idea of using each of the fault cases as a reference in the generation of fault clusters. At time = 1, test data is acquired and the MD values are calculated using the normal case as the basis for comparison. Fault detection occurs when the MD of the current process crosses the predetermined threshold of the normal case. If a fault is detected, a flag is set to initiate the parallel calculation of MD values with the faulty cases as the references, and the time is advanced to the next time window. Then, by using the four references, the direction of progression of the fault to its fault cluster is determined and the root cause identified.

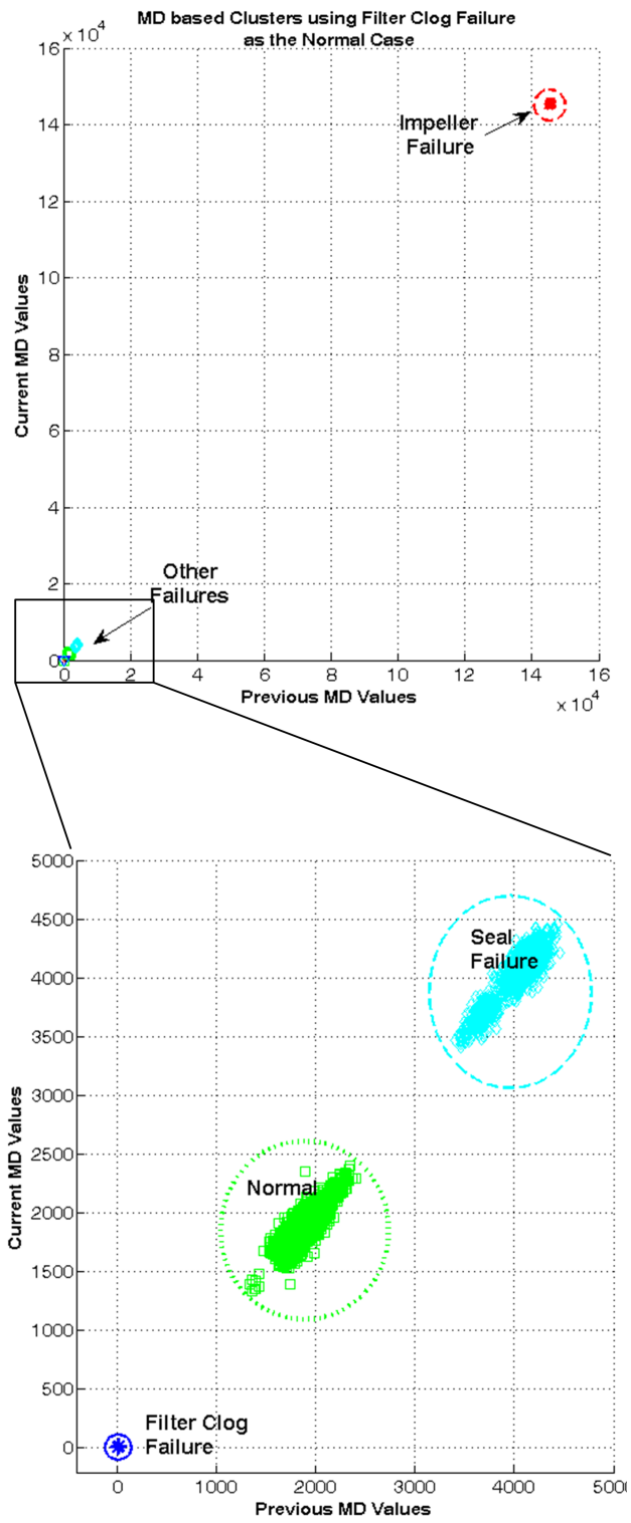

Figure 5.17. Realigned fault clusters using filter clog as the normal case

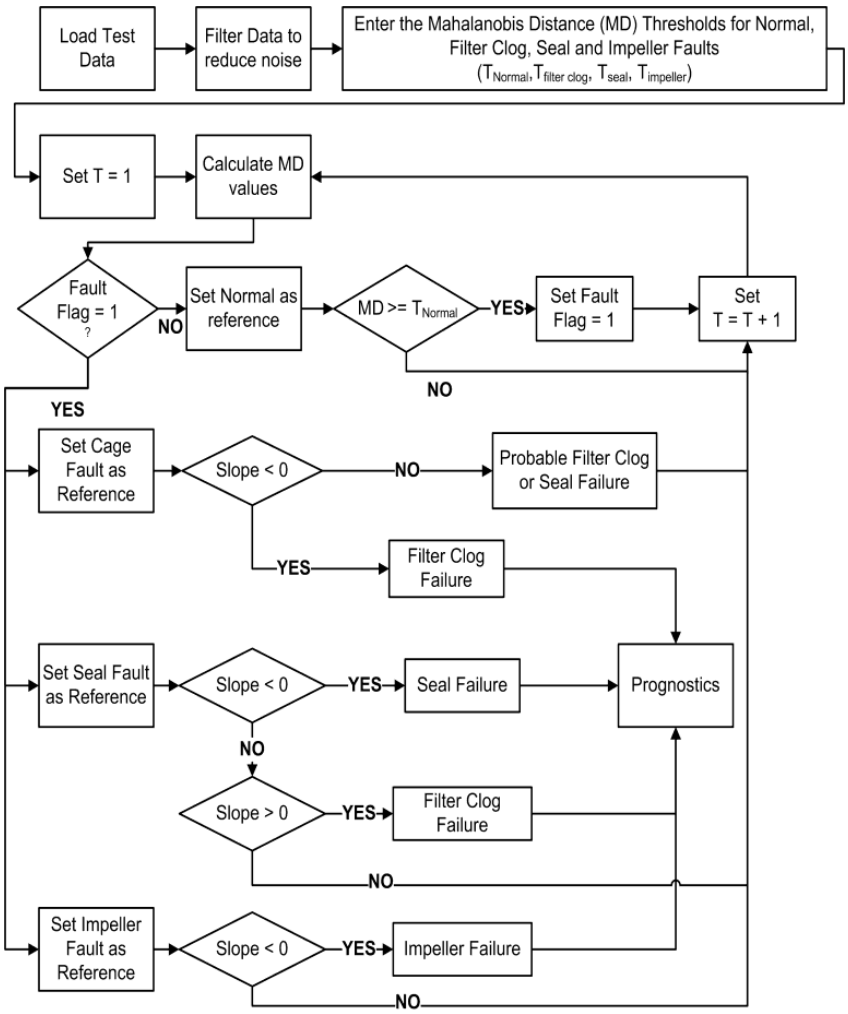

Figure 5.18. MD value based prognostic scheme

Figure 5.19 shows the time evolution of MD values for the seal failure experiment. Figure 6.20 shows the time-to-failure (TTF) estimation plot for the same experiment using a thirty minute time window using Eqn. (11). The failure threshold value selected for the experiment is 827. The estimation is triggered as soon as a particular operation enters a fault cluster and stays within the cluster for a period of two time windows in order to minimize false fault classification alarms

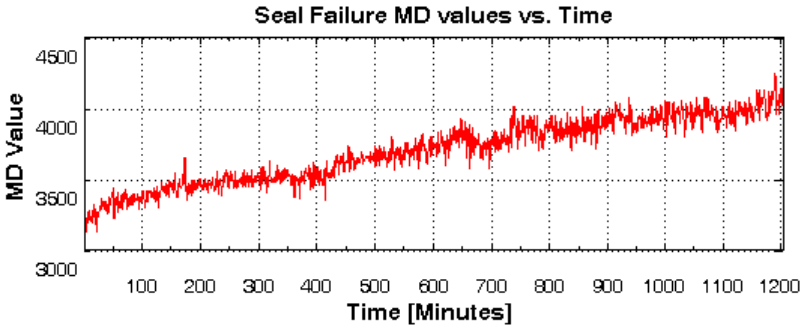

Figure 5.19. Time evolution of MD values for the seal failure experiment

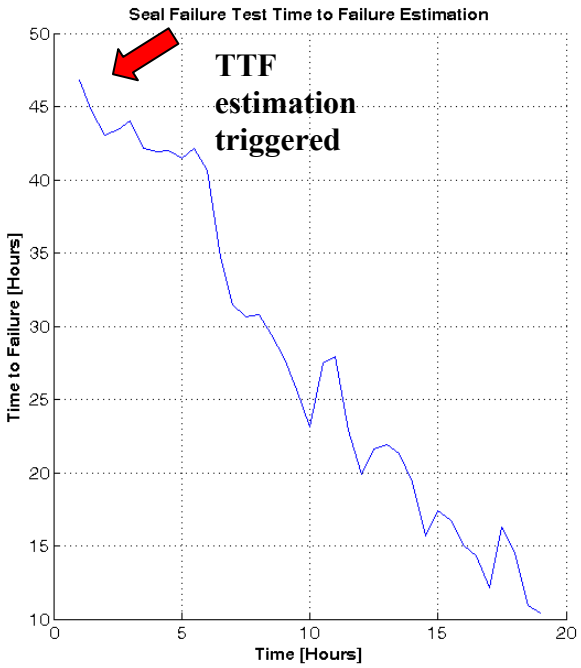

Figure 5.20. Time to failure estimation for the seal failure experiment

Figure 5.21 displays the time evolution of MD values for the filter clog failure experiment, while Figure 5.22 shows the time to failure estimation for the same experiment using a 30 minute time window and a failure threshold value of 582.

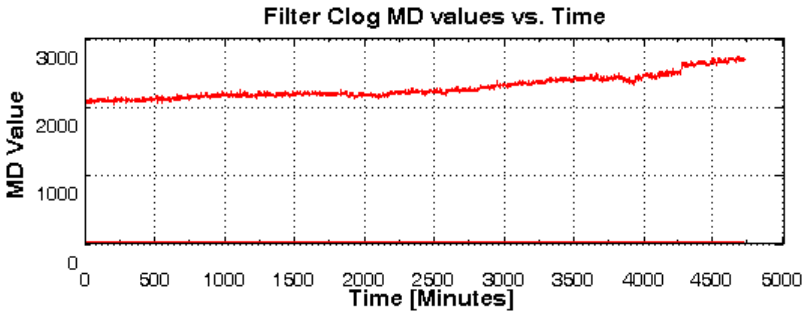

Figure 5.21. Time evolution of MD values for the filter clog failure experiment

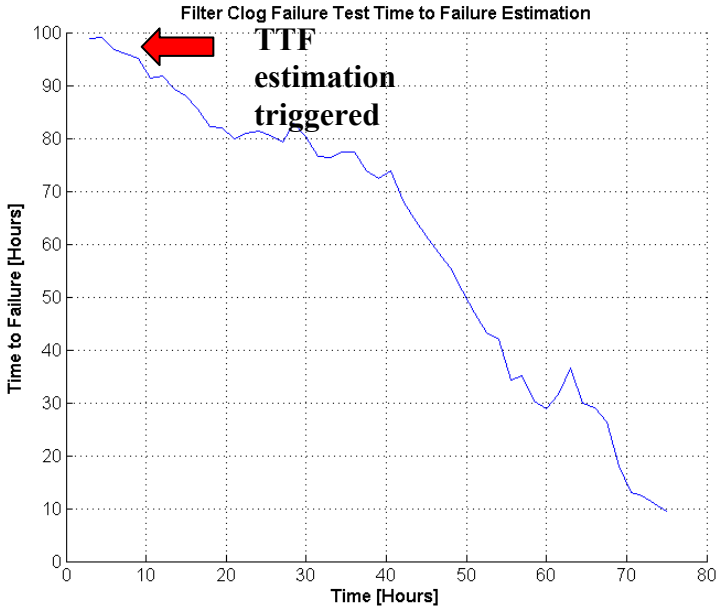

Figure 5.22. Time to failure estimation for the filter clog failure experiment

5.5. CONCLUSIONS

In this paper, a novel multi-sensor based Mahalanobis Taguchi System fault detection, isolation and prognostics scheme is presented. The proposed scheme utilizes Mahalanobis Distance based fault clustering and the progression of MD values for

diagnostics and prognostics. The performance of the scheme was validated through experiments performed on a centrifugal water pump testbed.

Advantages of the proposed approach can be summarized as follows:

- The proposed approach fuses all of the pertinent information obtained from a multi-sensor environment into a single performance metric, which is the Mahalanobis Distance.
- The proposed approach can function equally well in both the data and feature domains. In the case when the raw data collected from sensors is adequate to differentiate between various system health states, the data can directly be used as inputs to the system. If the raw data is not sufficient, then higher level features can be extracted from the data by further analysis, and these features can be utilized as inputs.
- It provides a unique solution for centrifugal pump fault detection, isolation and prognostics, eliminating the need for developing a tool for each one separately.
- It is process independent. It can be applied to a wide variety of multi-variate problems provided that a priori normal and abnormal data are available.
- It provides a methodical way to identify the features critical for the process, reducing dimensionality of the problem. Therefore, the analysis overhead is reduced since efforts can be concentrated on the key features.
- Since the proposed scheme is not computationally exhaustive, it can easily be implemented onto wireless motes and deployed to perform online monitoring (i.e., real-time) and decision making in a variety of industrial environments.

The analysis and results presented in this paper, especially the thresholds, are valid for the particular experimental setup and operating conditions. However, the proposed approach can be applied to other processes.

5.6. REFERENCES

[1]E. Lihovd, T. I. Johannessen, C. Steinebach, and M. Rasmussen, “Intelligent Diagnosis and Maintenance Management,” Journal of Intelligent Manufacturing, vol. 9, no. 6, pp. 523-537, 1998.

[2] H. Falkner, and D. Reeves, Study on Improving the Energy Efficiency of Pumps, ETSU, AEAT PLC. SAVE Study, United Kingdom, 2001.

[3] R. K. Byskov, C. B. Jacobsen, and N. Pedersen, “Flow in a Centrifugal Pump Impeller at Design and Off-Design Conditions---Part Ii: Large Eddy Simulations,” Journal of Fluids Engineering, vol. 125, no. 1, pp. 73-83, 2003.

[4] O. Coutier-Delgosha, R. Fortes-Patella, J. L. Reboud, M. Hofmann, and B. Stoffel, “Experimental and Numerical Studies in a Centrifugal Pump with Two-Dimensional Curved Blades in Cavitating Condition,” Journal of Fluids Engineering, vol. 125, no. 6, pp. 970-978, 2003.

[5] R. Hirschi, P. Dupont, F. Avellan, J. N. Favre, J. F. Guelich, and E. Parkinson, “Centrifugal Pump Performance Drop Due to Leading Edge Cavitation: Numerical Predictions Compared with Model Tests,” Journal of Fluids Engineering, vol. 120, no. 4, pp. 705-711, 1998.

[6] M. A. Langthjem, and N. Olhoff, “A Numerical Study of Flow-Induced Noise in a Two-Dimensional Centrifugal Pump. Part I. Hydrodynamics,” Journal of Fluids and Structures, vol. 19, no. 3, pp. 349-368, 2004.

[7] R. B. Medvitz, R. F. Kunz, D. A. Boger, J. W. Lindau, A. M. Yocum, and L. L. Pauley, “Performance Analysis of Cavitating Flow in Centrifugal Pumps Using Multiphase Cfd,” Journal of Fluids Engineering, vol. 124, no. 2, pp. 377-383, 2002.

[8] C. S. Byington, M. Watson, D. Edwards, and B. Dunkin, "In-Line Health Monitoring System for Hydraulic Pumps and Motors." pp. 3279-3287.

[9] R. Dong, S. Chu, and J. Katz, “Effect of Modification to Tongue and Impeller Geometry on Unsteady Flow, Pressure Fluctuations, and Noise in a Centrifugal Pump,” Journal of Turbomachinery, vol. 119, no. 3, pp. 506-515, 1997.

[10] I. S. Koo, and W. W. Kim, “The Development of Reactor Coolant Pump Vibration Monitoring and a Diagnostic System in the Nuclear Power Plant,” ISA Transactions, vol. 39, no. 3, pp. 309-316, 2000.

[11] M. S. Lebold, K. Maynard, K. Reichard, M. Trethewey, J. Hasker, C. Lissenden, and D. Dobbins, "A Non-Intrusive Technique F or on-Line Shaft Crack Detection a Nd Tracking." pp. 1-11.

[12]S. Perovic, P. J. Unsworth, and E. H. Higham, "Fuzzy Logic System to Detect Pump Faults from Motor Current Spectra." pp. 274-280 vol.1.

[13] N. R. Sakthivel, V. Sugumaran, and S. Babudevasenapati, "Vibration Based Fault Diagnosis of Monoblock Centrifugal Pump Using Decision Tree," Expert Systems with Applications, vol. 37, no. 6, pp. 4040-4049, 2010.
[14] N. R. Sakthivel, V. Sugumaran, and B. B. Nair, "Comparison of Decision Tree-Fuzzy and Rough Set-Fuzzy Methods for Fault Categorization of Mono-Block Centrifugal Pump," Mechanical Systems and Signal Processing, vol. In Press, Corrected Proof, 2010.
[15] S. Zhang, T. Asakura, X. Xu, and B. Xu, "Fault Diagnosis System for Rotary Machine Based on Fuzzy Neural Networks," JSME, vol. 46, no. 3, pp. 1035-1041, 2003.
[16] G. Taguchi, S. Chowdury, and Y. Wu, The Mahalanobis Taguchi System, New York: McGraw Hill, 2001.
[17] P. C. Mahalanobis, "On the Generalized Distance in Statistics," Proceedings, National Institute of Science of India, vol. 2, no. 1, pp. 49-55, 1936.
[18] G. Taguchi, and R. Jugulum, The Mahalanobis-Taguchi Strategy - a Pattern Technology System, New York: John Wiley & Sons, 2002.
[19] R. B. Chinnam, B. Rai, and N. Singh, "Tool-Condition Monitoring from Degradation Signals Using Mahalanobis Taguchi System," in Robust Engineering, ASI's 20th Annual Symposium, 2004, pp. 343-351.
[20] H. Wang, C. Chiu, and C. Su, "Data Classification Using Mahalanobis Taguchi System," Journal of Chinese Institute of Industrial Engineers, vol. 21, no. 6, pp. 606-618, 2004.
[21] E. A. Cudney, K. Paryani, and K. M. Ragsdell, "Applying the Mahalanobis-Taguchi System to Vehicle Ride," Journal of Industrial and Systems Engineering, vol. 1, no. 3, pp. 251-259, 2007.
[22] E. A. Cudney, R. Jugulum, and K. Paryani, "Forecasting Consumer Satisfaction for Vehicle Ride Using a Multivariate Measurement System," International Journal of Industrial and Systems Engineering, vol. 4, pp. 683-696, 2009.
[23] C. R. Foster, R. Jugulum, and D. D. Frey, "Evaluating an Adaptive One-Factor-at-a-Time Search Procedure within the Mahalanobis-Taguchi System," International Journal of Industrial and Systems Engineering, vol. 4, pp. 600-614, 2009.
[24] M. Asada, "Wafer Yield Prediction by the Mahalanobis-Taguchi System," Statistical Methodology, IEEE International Workshop on, pp. 25-28, 2001.
[25] S. Hayashi, Y. Tanaka, and E. Kodama, "A New Manufacturing Control System Using Mahalanobis Distance for Maximizing Productivity," Semiconductor Manufacturing, IEEE Transactions on, vol. 15, no. 4, pp. 442-446, 2002.
[26] D. Mohan, C. Saygin, and J. Sarangapani, "Real-Time Detection of Grip Length Deviation During Pull-Type Fastening: A Mahalanobis–Taguchi System (Mts)-Based Approach," The International Journal of Advanced Manufacturing Technology, vol. 39, no. 9, pp. 995-1008, 2008.
[27] T. Dasgupta, "Integrating the Improvement and the Control Phase of Six Sigma for Categorical Responses through Application of Mahalanobis-Taguchi System Mts," International Journal of Industrial and Systems Engineering, vol. 4, pp. 615-630, 2009.

[28] J. Srinivasaraghavan, and V. Allada, "Application of Mahalanobis Distance as a Lean Assessment Metric," The International Journal of Advanced Manufacturing Technology, vol. 29, no. 11, pp. 1159-1168, 2006.
[29] S. B. Kim, K.-L. Tsui, T. Sukchotrat, and V. C. P. Chen, "A Comparison Study and Discussion of the Mahalanobis-Taguchi System," International Journal of Industrial and Systems Engineering, vol. 4, pp. 631-644, 2009.
[30] E. A. Cudney, D. Drain, K. Paryani, and S. Naresh, "A Comparison of the Mahalanobis-Taguchi System to a Standard Statistical Method for Defect Detection," Journal of Industrial and Systems Engineering, vol. 2, no. 4, pp. 250-258, 2009.
[31] R. Jugulum, "Comparison between Mahalanobis-Taguchi System and Artificial Neural Networks," Journal of Quality Engineering Forum, vol. 10, no. 1, pp. 60-73, 2002.
[32] W. H. Woodall, R. Koudelik, K.-L. Tsui, S. B. Kim, Z. G. Stoumbos, and C. P. Carvounis, "A Review and Analysis of the Mahalanobis-Taguchi System," Technometrics, vol. 45, pp. 1-15, 2003.
[33] R. Isermann, Fault Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance, Berlin, Germany: Springer, 2005.

SECTION

2. CONCLUSIONS AND FUTURE WORK

2.1. CONCLUSIONS

Two of the main reasons for the loss of performance in a manufacturing system are problems regarding efficient and effective management of inventory and difficulties that arise due to the physical health condition of the components of the system. The overall objective of this dissertation is to develop sensor data-based decision making models in order to address these problems.

Technological developments in the recent years have introduced highly capable sensors such as RFID and other related Auto-ID tools, which provide non-contact object identification and inventory visibility. Such technological capabilities provide real-time visibility of each single entity in the supply chain; presenting many opportunities for process improvement and re-engineering. On the other hand, the technology also presents various challenges due to the lack of established industrial standards and application roadmaps and the difficulties that arise while dealing with voluminous data in a timely fashion.

A typical Auto-ID application requires effective integration of business processes and networking topologies and protocols. First, business decision-making models must be re-engineered to incorporate Auto-ID data. Business process re-engineering should involve the integration of real-time data into the critical processes such as production planning, scheduling and execution. This integration requires the development of

intelligent decision making models that can effectively process the sensor data into information and suggest the appropriate actions. On the other hand, in order for the decision-making component to be realistic, the network must facilitate effective and efficient routing of Auto-ID data packets. Academic studies usually focus either on manufacturing-specific decision-making, or on networking. Due to the gap between these two sub-components, the solutions provided to the industry are not directly applicable; further testing on Auto-ID technologies within the proposed solution is usually necessary in order to fine-tune it to the production environment.

In Paper I, an Auto-ID testbed architecture that aims to bridge this gap has been developed. The testbed provides a flexible test environment that can be used for evaluating Auto-ID technologies and solutions for industry. The testing environment, as well as the testing conditions, can be varied easily in order to match the needs of the industry. The testbed can be used as a low-cost technology development and assessment platform (proof-of-concept) due to the availability of reconfigurable models, which facilitate rapid decision making model development, hardware/software benchmarking, and implementation where hardware, software, and networking issues are considered simultaneously.

Additionally collection of RFID data in a timely manner, processing of such voluminous data, and making timely decisions that are tied into manufacturing execution systems presents a further challenge. Benefits provide by RFID such as waste elimination, inventory reduction, automatic replenishment, stock-out reduction, and overall cost savings can only realized if the challenge is conquered. Therefore, there is a

need for RFID data-based effective decision making algorithms that can lead to such benefits.

In Paper II, two static inventory models that rely on fixed baseline inventory levels, and a dynamic, forecast-integrated inventory model, which utilize RFID data was proposed. The results showed that the forecast-integrated model can adapt to the system dynamics more effectively than the current practice. The paper established that there are opportunities for RFID technology to provide significant benefits on the shop floor, well beyond the automation-oriented advantages, such as labor savings. The results displayed the potential of integrating RFID generated data with appropriate decision making models in a manufacturing environment

A further misapprehension regarding RFID technology is the assumption that a straightforward deployment will yield 100% satisfactory results in terms of item visibility. However, most of the reported studies have been performed under "controlled" laboratory settings and upon technology transfer to a real industrial environment, unforeseen problems, such as readability may arise from radio frequency (RF) interference, RF absorbing materials, and environmental conditions. Therefore, additional design and development need to be undertaken in areas in such as antenna configuration, antenna power control, etc. in order to provide feasible solutions for the intended application.

In Paper III, a complete solution that involves hardware considerations and decision making model development for such a real industry application was presented. The additional challenges that come up due to an extremely low temperature environments were addressed by the development of a novel RFID-based smart freezer.

The proposed solution utilized backpressure inventory control, systematic selection of antenna configuration, and antenna power control. The proposed RFID antenna configuration design methodology coupled with locally asymptotically stable distributed power control (LASDPC) ensured a 99% read rate of items while minimizing the required number of RFID antennas in the confined cold chain environments with non-RF friendly materials, which resulted in high inventory visibility. The proposed RFID-based Smart Freezer performance was verified through simulations of supply chain and experiments on an industrial freezer testbed operating at -100°F.

It is important to note that RFID is only one of many possible sensors that can be "embedded" in business processes in order to improve system performance. Even if the RFID/Auto-ID implementation for inventory management is successful, the performance of the overall system still depends on the physical health condition of its components. Therefore, it is necessary to be able to monitor and diagnose problems that can arise due to the wearing out of components. RFID sensors can usually provide satisfactory results in order to prevent performance loss due to inventory management problems. However, in the component health management case, a single sensor type usually cannot provide sufficient information complex decision making. Therefore, a combination of sensors should be used in an integrated manner in order to achieve desired performance levels.

The papers in the literature present several sensor signals; and features that are derived from these signals, which need to be monitored for component health management. If these signals/features show a large variability between experiments, it becomes difficult to select suitable thresholds to be used in monitoring diagnostics and prognostics. In addition, the existing body of work does not present a systematic method

which can identify the most pertinent variables and eliminate the redundant ones in order to reduce analysis overhead. Therefore, a new multivariable method is necessary to identify the input variables necessary to be monitored for fault detection, root cause analysis, and fault prognosis.

To address these issues, Paper IV introduced a novel Mahalanobis Taguchi System (MTS) based fault detection, isolation, and prognostics scheme. The proposed data-driven scheme built upon MTS by utilizing Mahalanobis Distance (MD) based fault clustering for diagnostics and by the use of the progression of MD values over time for prognostics. MD thresholds derived from the clustering analysis were used for fault detection and isolation. The performance of the scheme was validated via experiments performed on rolling element bearings inside the spindle headstock of a micro computer numerical control (CNC) machine testbed. The analysis was performed in the feature domain by extracting useful features from the raw sensor data. The experiments showed that the proposed approach provides a reliable multivariate analysis and real-time decision making tool that; presents a single tool for fault detection, isolation and prognosis, eliminating the need to develop each separately and (2) offers a systematic way to determine the key features, thus reducing analysis overhead.

In Paper V, the Mahalanobis Taguchi System based prognostics tool was expanded into a multi-sensor/multi-component/multi-failure-mode environment, which added on to the data collection and processing complexity. In this case, the analysis was performed in the data domain, i.e. directly on the raw data collected from the sensors. The need for multiple sensors in order to be able to effectively monitor the health of components was also demonstrated. Overall the paper showed that MTS can be

successfully used to fuse multi-sensor information into a single system performance metric for monitoring, diagnostics and prognostics purposes, and the proposed MTS-based prognostic tool is applicable to various industrial systems.

2.2. FUTURE WORK

The inventory management decision making models developed in Papers II and III are specific to the needs of the application. As part of the future work, other well-known inventory management models could be contrasted to the proposed models in terms of performance.

Due to the physical limitations of the laboratory environment and time constraints, the verification of the proposed MTS-based prognostics tool were performed via accelerated testing of rolling element bearings and a centrifugal pump. This resulted in the lack of transition data, which shows the gradual progression of a component from a healthy state into a faulty state. The ideas presented in Papers IV and IV could be better tested if it were possible to collect sensor data from a real industrial production environment.

Since the proposed scheme is not computationally exhaustive, it can easily be implemented onto wireless motes and deployed to perform real-time monitoring and decision making in a variety of industrial environments. In addition, with the use of RFID read/write tags, the current condition of a component can be stored on the component itself, effectively turning components into mobile databases. This would have a positive impact on the maintenance policies and component inventory management of manufacturing organizations.

APPENDIX

PROOF OF THEOREM 3

Let the Lyapunov candidate be

$$V = e_i^T(k)e_i(k) + \frac{1}{\alpha}tr[\tilde{\theta}^T k)\tilde{\theta}(k)]$$

whose first difference is given by

$$\Delta V = \underbrace{e_i^T(k+1)e_i(k+1) - e_i^T(k)e_i(k)}_{\Delta V_1} + \underbrace{\frac{1}{\alpha}tr[\tilde{\theta}^T(k+1)\tilde{\theta}(k+1) - \tilde{\theta}^T(k)\tilde{\theta}(k)]}_{\Delta V_2}$$

$$\Delta V_1 = (A_0 e_i(k) + \Psi_1(k) + \frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c} + \varepsilon(k))^T.(A_0 e_i(k) + \Psi_1(k) + \frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c} + \varepsilon(k)) - e_i^T(k)e_i(k)$$

after mathematical manipulation, ΔV_1 becomes

$$\Delta V_1 = e_i^T(k)A_0^T A_0 e_i(k) + 2e_i^T(k)A_0^T\Psi_1 + 2e_i^T(k)A_0^T\frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c} + 2e_i^T(k)A_0^T\varepsilon + \Psi_1^T\Psi_1 + 2\frac{\lambda e_i^T(k)\varepsilon}{e_i^T(k)e_i(k)+c} + \varepsilon^T\varepsilon - e_i^T(k)e_i(k)$$

Next, substituting the parameter update law in ΔV_2 becomes

$$\Delta V_2 = \frac{1}{\alpha}tr[(\Delta\tilde{\theta}(k) + \tilde{\theta}(k))^T.(\Delta\tilde{\theta}(k) + \tilde{\theta}(k)) - \tilde{\theta}^T(k)\tilde{\theta}(k)]$$

$$= \frac{1}{\alpha}tr[\alpha^2\varphi(e_i^T(k+1))(e_i(k+1)\varphi^T) - 2\alpha e_i(k+1)\varphi^T\tilde{\theta}(k)]$$

$$= \frac{1}{\alpha}tr[\alpha^2\varphi(A_0 e_i(k) + \Psi_1(k) + \frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c} + \varepsilon(k))^T.(A_0 e_i(k) + \Psi_1(k) + \frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c}$$

$$+\varepsilon(k))\varphi^T - 2\alpha(A_0 e_i(k) + \Psi_1(k) + \frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c} + \varepsilon(k))\varphi^T\tilde{\theta}(k)]$$

Applying the Cauchy-Schwartz inequality

$$((a_1 + a_2 + ... + a_n)^T.(a_1 + a_2 + ... + a_n) \le n.(a_1^T a_1 + a_2^T a_2 + ... + a_n^T a_n))$$

for the first term in the above equation, and applying the trace operator (given a vector $x \in \Re^n$, $tr(xx^T) = x^T x$),

$$\le 4\alpha\varphi^T\varphi(e_i^T(k)A_0^T A_0 e_i(k) + \Psi_1^T\Psi_1 + \frac{\lambda^2 e_i^T(k)e_i(k)}{(e_i^T(k)e_i(k)+c)^2} + \varepsilon^T\varepsilon)$$
$$-2\Psi_1^T(A_0 e_i(k) + \Psi_1 + \varepsilon + \frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c})$$

Next, applying $\Delta V = \Delta V_1 + \Delta V_2$, and cancelling similar terms, the first difference is given by

$$\Delta V \le e_i^T(k)A_0^T A_0 e_i(k) + \underbrace{2e_i^T(k)A_0^T \frac{\lambda e_i(k)}{e_i^T(k)e_i(k)+c}}_{1} + \underbrace{2e_i^T(k)A_0^T\varepsilon}_{1} - \Psi_1^T\Psi_1$$

$$+\frac{\lambda^2 e_i^T(k)e_i(k)}{(e_i^T(k)e_i(k)+c)^2} + \underbrace{2\frac{\lambda e_i^T(k)\varepsilon}{e_i^T(k)e_i(k)+c}}_{1} + \varepsilon^T\varepsilon - e_i^T(k)e_i(k) + 4\alpha\varphi^T\varphi e_i^T(k)A_0^T A_0 e_i(k)$$

$$+4\alpha\varphi^T\varphi\Psi_1^T\Psi_1 + 4\alpha\varphi^T\varphi\frac{\lambda^2 e_i^T(k)e_i(k)}{(e_i^T(k)e_i(k)+c)^2} + 4\alpha\varphi^T\varphi\varepsilon^T\varepsilon$$

Now apply Cauchy Schwartz inequality ($2ab \le a^2 + b^2$) to terms numbered as 1 in the above equation, and using the facts

$$\varepsilon^T\varepsilon \le \lambda^2 e_i^T(k)e_i(k) \quad , \quad \frac{e_i^T(k)e_i(k)}{(e_i^T(k)e_i(k)+c)^2} < e_i^T(k)e_i(k)$$

The norm results in:

$$\Delta V \le -(1 - 3A_{0\max}^2 - 6\lambda^2 - 4\alpha\varphi_{\max}^2 A_{0\max}^2 - 8\alpha\varphi_{\max}^2\lambda^2)\|e_i\|^2 - (1 - 4\alpha\varphi_{\max}^2)\|\Psi_1\|^2$$

Hence $\Delta V < 0$ if the gains are taken as:

$$\alpha \le 1/4\psi_{i\max}^2, \ \alpha \le 1/4\psi_{i\max}^2 \ and \ \lambda = \sqrt{\frac{1-3A_{0\max}^2 - 4\alpha\varphi_{\max}^2 A_{0\max}^2}{(6+8\alpha\varphi_{\max}^2)}}$$

This shows that the SNR error, $e_i(k)$, and the channel estimation error $\tilde{\theta}(k)$ are locally asymptotically stable.

VITA

Ahmet Soylemezoglu was born on September 5, 1976. He received his B.S. in Mechanical Engineering from Eastern Mediterranean University in July, 2000. He received his M.S. in Engineering Management in December 2001 and his M.S. in Manufacturing Engineering in May, 2004 from University of Missouri-Rolla. He earned his Ph.D. in Engineering Management in May 2010 from Missouri University of Science and Technology (formerly University of Missouri-Rolla). Dr. Soylemezoglu's research interests include discrete event systems, flexible manufacturing systems, shop floor control, computer integrated manufacturing, preventive maintenance and shop floor prognostics.