CHAPTER 1

INTRODUCTION

1.1 HISTORICAL PERSPECTIVE

Over the last several decades, there has been a wide range of approaches and implementation strategies for performing manual, semiautomated, or fully automated fault diagnosis and prognosis (i.e., health management) on critical systems in commercial and defense markets. Associated with these diagnostic and prognostic system designs are an equally diverse number of philosophies and associated architectures used to implement them for particular applications. For example, early-generation aircraft relied on manual detection and isolation of problems on the ground to meet their health management needs. These systems typically were analog and independent of one another, with only a schematic and voltmeter readings available to troubleshoot the problems. As aircraft systems became more complicated and integrated, built-in test (BIT) equipment, simple alarms, and trending analysis were implemented to warn the operators of safety-critical situations. However, the system maintainers still did not fully use this capability and often continued to rely on the voltmeter, schematics, and operator reports to solve problems or improve designs.

Continuing with aircraft diagnostic system evolution as an example (although other equipment has similar parallels), original equipment manufac-

turers (OEMs) and vehicle integrators began to realize that the output of the fault detection alarms or BIT equipment could be made available to support system troubleshooting. With this concept, fault indications were made available on the front of a line-replaceable unit (LRU) that indicated that a fault had been detected—they were originally mechanical but later were replaced with small light-emitting diodes (LEDs). These types of indicators progressed over the 1950s and 1960s. In many cases the LRU front panel contained a test switch to command the LRU to test itself, in a manner similar to how ground-support equipment could test the LRU. This capability became known as *built-in test equipment* (BITE). This capability began to decrease the need for some of the ground-support equipment previously used to test critical equipment. Depending on the system, the LRU fault indicators effectively could point the mechanic in the right direction, but schematics and voltmeters still were used in many conditions. However, the BITE of this era often was confusing, not reliable, and difficult to use. Mechanics often distrusted it. Many systems on airplanes such as the Boeing 707, 727, and early 737/747, and the McDonnell Douglas DC-8, DC-9, and DC-10 employed this type of diagnostic and maintenance practice.

In the 1970s, as computer systems began to be introduced, some of the increasingly complex systems began to use computers to perform their calculations. This was called *digital BITE*. With these computers came the ability to display fault detection and isolation information in digital form, normally via numeric codes, on the front panel of the LRU. The digital logic could produce codes that could better isolate the cause of the fault. The digital display offered the capability to display many different codes to identify each type of fault that was detected. These codes often pointed to some description in a manual that could be used to isolate and correct the fault. Many systems on the Boeing 757/767, Airbus A300/310, McDonnell Douglas DC-10, and Lockheed L-1011 employ this approach.

As the number of systems grew, use of separate front-panel displays to maintain the systems became less effective, particularly since each LRU often used a different technique to display its fault data. In addition, some of the systems had become increasingly integrated with each other. Digital data buses, such as ARINC 429, began to be used during this time period. Autopilot systems, for example, became among the first to use these digital data buses and depend on sensor data provided by many other systems. The success of these systems became a driving force in the definition of more sophisticated maintenance systems. The more sophisticated monitoring was necessary to meet the integrity and certification requirements of its automatic landing function. For example, the 767 maintenance control and display panel brought together the maintenance functions of many related systems. As the next step, ARINC 604, defined in 1986, provided a central fault display system (CFDS) that brought to one display the maintenance indications for all the systems on the airplane. This approach enabled more consistent access to maintenance data across the aircraft systems and a larger display than each

of the systems could contain individually and saved the cost of implementing front-panel displays on many of the associated system LRUs. In this approach, the CFDS were used to select the system for which maintenance data were desired and then routed the maintenance text from that system to the display. This approach was applied to many of the systems on later Boeing 737s and most systems on the Airbus A320/330/340 and the McDonnell Douglas MD11.

Systems continued to become more complex and integrated. A single fault on an airplane could cause fault indications for many systems, even when displayed using systems such as the CFDS. In many cases, the mechanic had little help in determining which indication identified the source fault and which was merely an effect. To solve this and related issues, the ARINC 624 was developed in the early 1990s. It defines a more integrated maintenance system that can consolidate the fault indications from multiple systems and provide additional functionality to support condition-based maintenance. With such an architecture, minimal ground-support equipment is needed to test airplane systems because most of this capability is included in the maintenance system. For example, most factory functional tests of airplane systems on the Boeing 747-400 and 777 airplanes consist of little more than execution of selected tests, monitoring fault displays, and monitoring certain bus data using the integrated maintenance system.

The evolution of diagnostic and prognostic reasoning systems thus has strived to better identify the single LRU or component that is the source of the fault. This allows the maintainer to remove the failed component and correct the fault condition confidently. Although in many cases this is possible, there are many situations where it is not possible without the addition of sensors or wiring. Addition of these sensors increases the number of components that can fail and thus sometimes can worsen the maintenance effort. In addition, they add cost and weight to the airplane. There are clearly cases where the addition of such hardware can be beneficial, but the benefits of improved fault isolation must be weighed against the potential reduced reliability and increased cost and weight of the additional components.

1.2 DIAGNOSTIC AND PROGNOSTIC SYSTEM REQUIREMENTS

A critical part of developing and implementing effective diagnostic and prognostic technologies is based on the ability to detect faults in early enough stages to do something useful with the information. Fault isolation and diagnosis uses the detection events as the start of the process for classifying the fault within the system being monitored. Condition and/or failure prognosis then forecasts the remaining useful life (the operating time between detection and an unacceptable level of degradation). If the identified fault affects the life of a critical component, then the failure prognosis models also must reflect this diagnosis. Specific requirements in terms of confidence and

severity levels must be identified for diagnosis and prognosis of critical failure modes. In general, the fault diagnosis detection level and accuracy should be specified separately from prognostic accuracy.

As a minimum, the following probabilities should be used to specify fault detection and diagnostic accuracy:

1. The probability of anomaly detection, including false-alarm rate and real fault probability statistics.
2. The probability of specific fault diagnosis classifications using specific confidence bounds and severity predictions.

To specify prognostic accuracy requirements, the developer/end user must first define

1. The level of condition degradation beyond which operation is considered unsatisfactory or undesirable to the mission at hand.
2. A minimum amount of warning time to provide the operator and maintainer required information that can be acted on before the failure or condition is encountered.
3. A minimum probability level that remaining useful life will be equal to or greater than the minimum warning level.

1.3 DESIGNING IN FAULT DIAGNOSTIC AND PROGNOSTIC SYSTEMS

Following the evolution of diagnostic systems in the aircraft industry, prognostic initiatives started to be introduced in order to try to take advantage of the maintenance planning and logistics benefits. However, the early prognostic initiatives often were driven by in-field failures that resulted in critical safety or high-cost failures, and thus retrofitted technology was hard to implement and costly to develop. Hence diagnostic and prognostic system developers found the need to analyze and describe the benefits associated with reducing in-field failures and their positive impact on safety, reliability, and overall life-cycle-cost reduction. This has lead to many cost-benefit analyses and ensuing discussions and presentations to engineering management about why the diagnostic and prognostic technologies need to be included in the design process of the system and not simply an afterthought once field failures occur. This had lead us to the point where many complex vehicle/system designs such as the Joint Strike Fighter (F-35), DD(X), Expeditionary Fighting Vehicle (EFV), and various future combat system (FCS) vehicles are now developing ''designed in'' health management technologies that can be implemented within an integrated maintenance and logistics system that supports the equipment throughout its life time. This ''designed in'' approach to health man-

agement is performed with the hardware design itself and also acts as the process for system validation and managing inevitable changes from in-field experiences and evaluating system design tradeoffs, as shown in Fig. 1.1.

Realizing such an approach will involve synergistic deployments of component health monitoring technologies, as well as integrated reasoning capabilities for the interpretation of fault-detect outputs. Further, it will involve the introduction of learning technologies to support the continuous improvement of the knowledge enabling these reasoning capabilities. Finally, it will involve organizing these elements into a maintenance and logistics architecture that governs integration and interoperation within the system, between its on-board elements and their ground-based support functions, and between the health management system and external maintenance and operations functions. In this book we present and discuss the required functions of health management technologies that, if applied correctly, can directly affect the operations and maintenance of the equipment and positively affect the lifecycle costs.

1.4 DIAGNOSTIC AND PROGNOSTIC FUNCTIONAL LAYERS

A comprehensive health management system philosophy integrates the results from the monitoring sensors all the way through to the reasoning software that provides decision support for optimal use of maintenance resources. A core component of this strategy is based on the ability to (1) accurately predict the onset of impending faults/failures or remaining useful life of critical components and (2) quickly and efficiently isolate the root cause of failures once failure effects have been observed. In this sense, if fault/failure predictions can be made, the allocation of replacement parts or refurbishment actions can

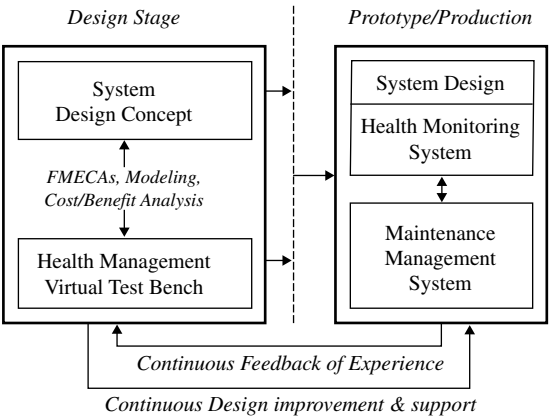

Figure 1.1 The "designed in" approach to health management.

be scheduled in an optimal fashion to reduce the overall operational and maintenance logistic footprints. From the fault isolation perspective, maximizing system availability and minimizing downtime through more efficient troubleshooting efforts is the primary objective.

In addition, the diagnostic and prognostic technologies being developed under a number of new acquisition programs implementing an autonomous logistics system concept require an integrated maturation environment for assessing and validating prognostics and health management (PHM) system accuracy at all levels in the system hierarchy. Developing and maintaining such an environment will allow for inaccuracies to be quantified at every level in the system hierarchy and then be assessed automatically up through the health management system architecture. The final results reported from the system-level reasoners and decision support is a direct result of the individual results reported from these various levels when propagated through. Hence an approach for assessing the overall PHM system accuracy is to quantify the associated uncertainties at each of the individual levels, as illustrated in Fig. 1.2, and build up the accumulated inaccuracies as information is passed up the vehicle architecture.

This type of hierarchical verification and validation (V&V) and maturation process will be able to provide the capability to assess diagnostic and prognostic technologies in terms of their ability to detect subsystem faults, diagnose the root cause of the faults, predict the remaining useful life of the faulty component, and assess the decision-support reasoner algorithms. Specific metrics to be discussed in this book will include accuracy, false-alarm rates, reliability, sensitivity, stability, economic cost/benefit, and robustness, just to name a few. The technical performance and accuracy of the diagnostic and prognostic algorithms also will need to be evaluated with performance met-

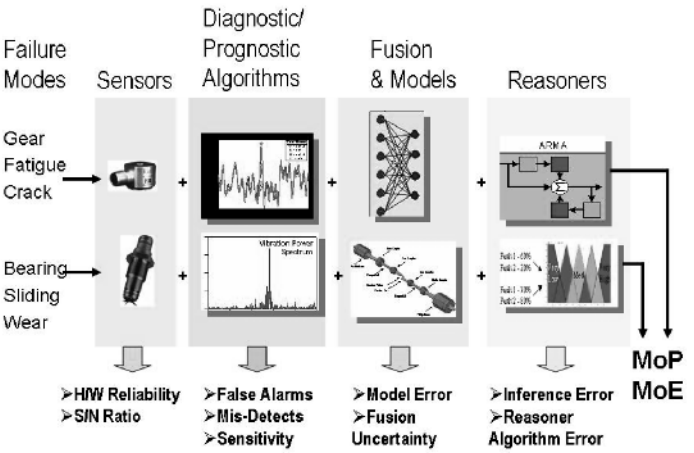

Figure 1.2 Functional layers in the health management system architecture.

rics, whereas system-level capabilities in terms of achieving the overall operational goals, economic cost/benefit, and assessment of the business case will be evaluated with effectiveness measures (also discussed in detail in Chap. 7).

Cost-effective implementation of a diagnostic or prognostic system will vary depending on the design maturity and operational/logistics environment of the monitored equipment. However, one common element to successful implementation is feedback. As components or LRUs are removed from service, disassembly inspections must be performed to assess the accuracy of the diagnostic and prognostic system decisions. Based on this feedback, system software and warning/alarm limits should be optimized until desired system accuracy and warning intervals are achieved. In addition, selected examples of degraded component parts should be retained for testing that can better define failure progression intervals.

1.5 PREFACE TO BOOK CHAPTERS

This book details the technologies in condition-based maintenance (CBM) and prognosis and health management (PHM) that have been introduced over the recent past by researchers and practitioners and are making significant inroads in such application domains as mechanical, thermal, electromechanical, and more recently, electrical and electronics systems. Thus our target applications are exclusively hardware-related. It is well recognized, though, that modern dynamical systems are a tightly coupled composite of both hardware and software. Software reliability is undoubtedly a serious concern and a challenge. Although we recognize its importance within the general area of reliability and maintainability, we are not addressing issues of software reliability in this book. The interested reader is referred to research publications on this topic stemming from the computer science community.

This book is structured as follows: Chapter 2 introduces those fundamental system concepts that set the stage for the effective design of fault diagnostic and prognostic technologies. We review systems-based methodologies that have a direct and significant impact on the design of CBM/PHM systems. The CBM/PHM designer must be thoroughly familiar with the physics of failure mechanisms and must posses an understanding of methods for the optimal selection of monitoring strategies, algorithms to detect and isolate faults and predict their time evolution, and systems approaches to design of experiments and testing protocols, performances metrics, and means to verify and validate the effectiveness and performance of selected models. A formal framework is established to conduct trade studies aimed at comparing alternative options and assisting in the selection of the ‘‘best’’ technologies that meet customer requirements. Failure modes and effects criticality analysis forms the foundation for good CBM/PHM design. We detail how it assists in deciding on the severity of failure modes, their frequency of occurrence,

and their testability. It considers fault symptoms and the required sensor suite to monitor their behavioral patterns. It also may list the candidate diagnostic and prognostic algorithms that are best suited to address specific failure modes.

Chapter 3 discusses sensor systems employed for monitoring and interrogating critical system components/subsystems. Our focus here is not only on conventional sensors used primarily for monitoring and control purposes but also on sensor systems and sensing strategies that are specifically designed to respond to CBM/PHM design requirements. Advances in the latter category over the recent past have made available to the CBM/PHM community new sensors that are promising to improve substantially the reliability, cost-effectiveness, coverage, and sensitivity to fault signatures of monitoring and fault tracking devices. If CBM/PHM systems are to penetrate the reliability/maintainability domain for critical military and industrial assets successfully, sensor technology must provide the assurance that measurements are accurate and reliable. Hardware and software sensor systems must perform ''better'' than the components/subsystems they are intended to monitor. We discuss in this chapter basic transducing principles and signal conditioning issues. We review sensor systems used to monitor fundamental physical processes of heat transfer, pressure, speed, vibration, etc. Of special concern are devices that respond to structural damage and anomalies encountered in rotating equipment, electrical systems, etc. Thus we elaborate on properties and discuss ultrasonic systems, proximity devices, strain gauges, acoustic emission sensors, etc. Recent advances in wireless sensing, Micro-Electro-Mechanical Systems (MEMS), and smart sensors are introduced. An important consideration in sensing strategies refers to the optimal placement of sensors so that metrics of fault detectability and distinguishability are maximized. Wireless communication protocols, sensor interfaces, and standards for sensor systems are surfacing lately and are establishing a systematic framework for advanced CBM/PHM system design.

Chapter 4 addresses key issues regarding data processing and database management methodologies. Whereas the preceding chapter introduced the hardware technologies and sensing strategies required to acquire CBM/PHM-relevant data, this chapter focuses on and provides answers to such important questions as: How do we relate fault (failure) mechanisms to the fundamental ''physics'' of complex dynamic systems? Where is the information? How do we extract it from massive raw databases? What is an optimal database configuration and database management schema? Answers to these questions set the groundwork and constitute the underpinnings for fault diagnosis and prognosis. Thus we begin by discussing essential signal-processing tools for virtual sensing, data validation, and data analysis. The signal-processing toolbox addresses a variety of time-domain and frequency-domain methods for data analysis and fault feature extraction. The sections here are aimed specifically at familiarizing the reader with tools employed for the analysis of signals encountered in CBM/PHM applications. Examples from vibration monitoring are used to illustrate the basic concepts. Case studies from aircraft engine and

gearbox fault signal analysis are presented. Feature selection and extraction techniques constitute the cornerstone for accurate and reliable fault diagnosis. We present a systematic approach and metrics of performance for feature selection and extraction. Tracking system usage patterns and factoring their potential impact into fault initiation and propagation studies are of great concern when designing CBM/PHM algorithms for such complex systems as aircraft, among many others. We conclude this chapter with a brief introduction to database management schemes. The preceding three chapters have layed the foundation for the main objective of this book: fault diagnosis and prognosis.

Chapter 5 is an extensive treatment of fault diagnostic technologies. We begin with a general framework for fault diagnosis, and then we introduce relevant definitions and a brief review of fault diagnosis requirements and performance metrics. A more detailed discussion of performance metrics is deferred to Chapter 7. Diagnostic methods historically have been divided into two major categories: methods that rely primarily on data and employ data-driven algorithms or tools and those that exploit physics-based models and measurements to classify fault conditions. Data-driven methods emphasize historical data diagnostic tools, statistical legacy data classification, and clustering algorithms based on fuzzy logic and neural network constructs. We introduce in this category reasoning and decision-making methods—fuzzy logic. Dempster-Shafer evidential theory is an effective means to combine fault evidence, resolve conflicting information, and address incomplete data sets, novel wavelet neural networks, and multimodel filter classifiers—whose primary objective is to detect and isolate incipient failures accurately and reliably while minimizing false alarms. We present next a concise summary of basic dynamic system models and system identification methods because they form the theoretical foundation for model-based diagnostic routines and assist in characterizing fault behavioral patterns. We continue in this section with a discussion of virtual (soft or analytical) sensors—software models that augment typically scarce fault databases—and filtering/estimation methodologies borrowed from the classic engineering disciplines—Kalman filtering, recursive least-squares estimation, and particle filtering—that take advantage of both state models that describe the fault or fatigue initiation and evolution and sensor measurements. We call on a variety of techniques to address the fault diagnosis problem: fatigue and crack-propagation models, finite-element analysis, and other modeling tools that are finding increased popularity because they provide valuable insight into the physics of failure mechanisms. Examples and case studies are used to illustrate the algorithmic developments throughout this chapter.

Failure prognosis is the Achilles' heel of CBM/PHM systems. Once an incipient failure or fault is detected and isolated, as discussed in the preceding chapter, the task of the prognostic module is to estimate as accurately and precisely as possible the remaining useful life of the failing component/subsystem. Long-term prediction entails large-grain uncertainty that must be rep-

resented faithfully and managed appropriately so that reliable, timely, and useful information can be provided to the user. Chapter 6 introduces recent advances in the development of prognostic algorithms. Such developments are still in their early stages, although a number of techniques have been proposed whose validity has not yet been confirmed via ground-truth data. The research community is approaching the prognosis problem not only recognizing the need for robust and viable algorithms but also understanding that sufficient and complete ground-truth failure data from seeded fault testing or actual operating conditions are lacking. This is a major impediment to the training and validation of the algorithmic developments. The failure prognosis problem basically has been addressed via two fundamental approaches. The first one builds on model-based techniques, where physics-based, statistical probabilistic, and Bayesian estimation methods are used to design fatigue or fault growth models. The second approach relies primarily on the availability of failure data and draws on techniques from the area of computational intelligence, where neural-network, neuro-fuzzy, and other similar constructs are employed to map measurements into fault growth parameters. We begin this chapter by suggesting the challenge faced by the designer, and we present a notional framework for prognosis. In the model-based category, we detail contributions in physics-based fatigue models. We introduce concepts from the dynamical model identification area that form the basis for such well-known statistical modeling techniques as auto regressive moving average (ARMA) and ARMA with exogenous input (ARMAX). The treatment here is intended to introduce the reader to fundamental concepts in prediction. Bayesian probability theory is discussed as the foundational framework for estimation and prediction, and two estimation methods—Kalman filtering and particle filtering—are detailed as two methods that are finding utility in failure prognosis. The methods are illustrated with examples from the area of aircraft failure mode prediction. We conclude this chapter with a discussion of data-driven prediction techniques, highlighting the utility of neural-network constructs employed as the mapping tools. Learning aspects are emphasized as essential ingredients in the improvement process, where the prediction algorithms continue to ''learn'' and adapt their parameters as new data become available. Tools for managing uncertainty are also suggested, aimed at reducing the uncertainty bounds that are unavoidably growing with the prediction horizon. Recent results from seeded fault testing on aircraft components complement the algorithmic modules. Probabilistic and particle-filtering methods seem to hold the greatest promise because they combine physics-based failure models with available measurements acquired online in real time. An adaptive framework allows for continuous adjustment of model parameters for improved prediction accuracy and precision. Research currently is under way to develop tools for material microstructural characterization that lead to meta-models capable of capturing more accurately the dynamics of failure mechanisms. It is anticipated that these developments, coupled with the availability of adequate seeded fault data, will drive the design and implementation of

reliable and robust prognostic technologies. Performance metrics for prognostic algorithms are discussed in the next chapter.

Chapter 7 addresses important issues relating to CBM/PHM system requirements and performance/effectiveness metrics. Assessing the technical and economic feasibility of CBM/PHM systems constitutes an essential component of the overall design process. Researchers and CBM/PHM practitioners have been attempting to define such metrics for specific application domains without reaching as yet a consensus. We introduce in this chapter a set of basic metrics developed over the past years by researchers in this field with the objective of bringing to the reader an initial attempt at this difficult topic and to motivate further developments. Recent mandates to implement CBM/PHM technologies on critical military and commercial assets are dictating the need for the definition and development of fault diagnosis and prognosis requirements. Defining a candidate list of requirements is predicated on the availability of performance metrics for major modules of the CBM/PHM system. We begin by introducing fault feature evaluation metrics. Since features or condition indicators are the foundational elements of diagnosis, it is important that the selected feature vector meet certain criteria that will enable the realization of accurate and robust diagnostic routines. Information, distance, and dependence measures form the theoretical basis for feature evaluation. They lead to a two-sample z-test that assesses whether the means of two groups or feature distributions are statistically different from each other. Performance metrics for fault diagnostic systems are discussed next. Among them, such typical metrics as false positives and false negatives are defined. The receiver operating characteristic offers the designer a means to evaluate how well a fault is diagnosed and to trade off between false positives and false negatives. An architecture suitable for diagnostic metrics assessment is suggested, and examples are used to illustrate its utility. Defining metrics for prognostic systems presents unique challenges. The probabilistic nature of the prediction model and the inherent uncertainty dictate that measures of performance must be defined and assessed statistically. Accuracy, precision, and confidence are notions closely associated with our ability to predict the occurrence of an event such as the time to failure of a particular component. They are used to define a set of performance metrics for prognostic systems. Diagnostic and prognostic effectiveness metrics are discussed next, and cost-benefit analysis concepts are reviewed briefly as applied to CBM/PHM systems. We conclude this chapter by defining a candidate list of CBM/PHM system requirements based on specific performance metrics.

The eighth and final chapter answers the question: Who is the customer, and how does he or she benefit from CBM/PHM practices? The potential CBM/PHM customer list started initially with the maintainer. The maintainer is the recipient of the diagnostic and prognostic modules output; he or she exploits it to schedule maintenance so that equipment uptime is maximized; resources for repair, maintenance, and overhaul are optimized; and major failure events are avoided. The list, though, is growing, and it is generally

recognized that additional benefits may be derived through CBM/PHM impacting the system designer's objectives, the operations manager, and the military field commander, among others. This chapter focuses on logistics support and enumerates potential benefits and enabling technologies needed to integrate CBM/PHM into the logistics culture. We describe how accurate prediction of the system's conditions affects the operational planning and scheduling tasks, the procurement of spare parts and material needed to restore the system to normal operation, financial performance, personnel training and system redesign considerations given the evidence provided by the CBM/PHM on the health of system components, the performance of sensors or lack thereof, and required sensing strategies. We introduce possible methodologies for logistics support and optimal scheduling strategies and present case studies to illustrate the concepts and techniques introduced in this chapter.

1.6 REFERENCES

G. Bloor et al., "An Evolvable Tri-Reasoner IVHM System," Proceedings of *IEEE Aerospace Conference,* Big Sky, MT, 2000.

J. Gray and G. Bloor, "Application Integration," Boeing's 3rd Annual Data Exchange Conference, Mesa, AZ, 2000.

G. Karsai and J. Gray, "Component Generation Technology for Semantic Tool Integration," *Proceedings of the IEEE Aerospace,* Big Sky, MT, 2000.