Monitoring, Prediction, and Fault Isolation in Dynamic Physical Systems

Pieter J. Mosterman and **Gautam Biswas**
Center for Intelligent Systems
Box 1679, Sta B
Vanderbilt University
Nashville, TN 37235.
pjm,biswas@vuse.vanderbilt.edu

Abstract

Diagnosis of dynamic physical systems is complex and requires close interaction of monitoring, fault generation and refinement, and prediction. We establish a methodology for model-based diagnosis of continuous systems in a qualitative reasoning framework. A temporal causal model capturing dynamic system behavior identifies faults from deviant measurements and predicts future system behavior expressed as signatures, i.e., qualitative magnitude changes and higher order time-derivative effects. A comparison of the *transient* characteristics of the observed variables with the predicted effects helps refine initial fault hypotheses. This allows for quick fault isolation, and circumvents difficulties that arise when interactions caused by feedback and dependent faults. This methodology has been successfully applied to the secondary cooling loop of fast breeder reactors.

Introduction

The complexity and sophistication of the new generation of physical systems along with the growing demand for their reliability and safety, is being met by automatic control and monitoring, and the use of *functional redundancy* techniques that exploit static and dynamic relations between observed variables in a system for fault detection and isolation. Functional relations among system parameters can be expressed as a set of mathematical constraints, and *filtering* and *observability* methods can be applied to generate *residual vectors* which can be processed by state estimation, parameter identification, and recognizing characteristic quantities[6]. *Topological methods,* on the other hand, characterize behavior relations as directed graphs constructed from system models under *normal* operating conditions and *faulty* situations[7]. Propagation of observed discrepancies in the graph help implicate system components.

An advantage of topological models is their compositional nature, the ability to dynamically partition the system into possible faulty subsystems given a set of observations and focus on specific constraints that are related to hypothesized faults. A problem with topological models is that they are often incomplete, underconstrained, and *ad hoc* (not derived from physical principles). This results in computationally intractable search spaces and the generation of spurious fault candidates. Models for diagnosis should describe normal and faulty system behavior, incorporate sufficient behavioral detail to map observed deviations to system components and parameters, and generate dynamic behavior caused by faults. Faults change system parameters, therefore, the assumption of constant parameters does not hold,[1] and the temporal effects of these causes have to be included.

The greater the number of relevant physical constraints included in a model, the lesser the search complexity and greater the accuracy in behavior generation. All this contributes to improved diagnostic accuracy. Bond graphs[9] provide a systematic framework for building well-constrained models of dynamic physical systems across multiple domains. Causality constraints derived from these models provide effective and efficient topology-based mechanisms for diagnosis. Analytic state-based representations derived from bond graphs express system behavior as a set of first order differential equations and form the basis for deriving a spectrum of qualitative to quantitative models. Compared to other qualitative reasoning approaches, bond graphs use energy and state conservation and continuity of power constraints to reduce the number of spurious behaviors generated. Methodologies exist for deriving bond graph models for diagnosis from physical system descriptions so that its elements can be directly related to system components and mechanisms under diagnostic scrutiny[2].

[1] Given the capacitor relation $V = \frac{q}{C}$, its dynamic behavior under faulty situations is governed by $\frac{dV}{dt} = \frac{1}{C}\frac{dq}{dt} + q\frac{d}{dt}(\frac{1}{C})$. The typical relation, $\dot{q} = i = C\frac{dV}{dt}$, assumes a correctly functioning component.

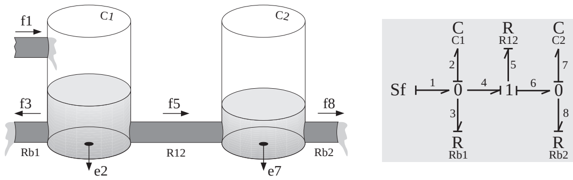

Figure 1: **The bi-tank system and its bond graph.**

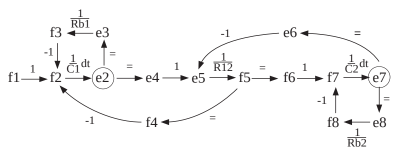

Figure 2: **Temporal causal graph.**

Our models for diagnosis are *temporal causal graphs* derived from bond graph models of the physical system. This graph, derived using the SCAP algorithm[9] expresses causal relations among system parameters extended with temporal behavior relations. The graphical structure represents effort and flow variables as vertices, and relations between the variables as directed edges[7]. Fig. 2 shows the temporal causal graph for the bi-tank system (Fig. 1). Junctions, transformers, and resistors introduce instantaneous magnitude relations, whereas capacitors and inductors introduce temporal effects. In general, these temporal effects are *integrating*, and their effect on the rate of change of an observed variable is determined by the path that links this variable to the point where a deviation occurs.

We use the temporal causal graph to predict future behavior in response to hypothesized faults in terms of the qualitative values $(-, 0, +)$ of 0^{th} and higher order derivatives. These predictions form *signatures* of future behavior and are matched against actual observed 0^{th} and 1^{st} order system behavior. To determine these actual values requires monitoring of the system for a period of time after failure. For example, a signal with predicted normal magnitude and positive slope at the time of failure eventually exhibits a deviating magnitude as well. These effects are accounted for by a *progressive monitoring* scheme. With time, feedback and other interactions convolute fault behaviors, and this makes fault refutation unreliable. Overall, signature prediction, progressive monitoring, and suspension of fault transients analysis are three novel components of the diagnostic methodology.

Bond graph models allow the automatic derivation of steady state models[7], an added component for diagnostic analysis. However, in real applications steady state detection is hard, control actions are usually taken before steady state to avoid catastrophic situations, faults may be intermittent and not persist long enough, and cascading faults complicate steady state analysis. This paper does not discuss steady state analysis.

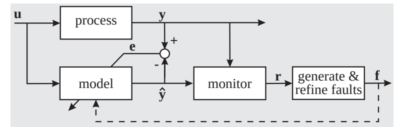

Figure 3: **Diagnosis of dynamic systems.**

Dynamic Continuous Systems Diagnosis

System models predict normal operating values, $\hat{\mathbf{y}}$, for a set of observed variables, $\mathbf{y}$, (Fig. 3). The variables monitored over time are matched against their predicted values in the desired normal mode of operation to compute *residuals* $\mathbf{r}$, as deviations from normal. An observer mechanism accounts for modeling errors, drift, and noise ($\mathbf{e}$) to ensure $\mathbf{r}$ is not spurious. When deviations occur, predictions for possible faults are generated by the diagnosis module using constraint analysis and propagation methods applied to the system model. To accurately isolate problems (identify the true fault) we predict future behaviors of the observed variables by introducing faults into the model and continuing to monitor the observables to check for *consistency* among the predictions and observations.

Our focus in this work is on abrupt faults that cause significant deviations from steady state operations called *transients*. If the goal is to quickly isolate faults and return the system to normalcy, it is essential to track and analyze system behavior at frequent intervals soon after the fault occurs, so that their unique transient effects are not lost. However, modeling, tracking, interpreting, and analyzing dynamic system transients is difficult. To keep model complexity and analysis computationally feasible while preserving some dynamic information, several methods perform diagnosis based on deviations from a static steady state model [1, 8] but their process models are invariably underconstrained and the number and size of fault candidate sets explode, making the diagnosis task impractical. Identifying and analyzing dynamic transients caused by faults lays great emphasis on the monitoring and prediction components of the overall diagnosis process, a feature that differentiates our efforts from a lot of the traditional work in model-based diagnosis (e.g., [4, 5, 8]).

Temporal Ordering When faults occur, their effects introduce changes in system behavior that propagate instantaneously to some parts of the system, but

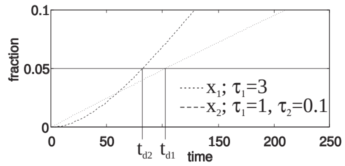

Figure 4: **Delay times for observing deviations.**

have delayed effects on other parts because of the *time constants* involved (fault effects on measured variables with larger time constants take longer to manifest than the measured values that have smaller time constants). Relations that do not embody time constants propagate abrupt changes instantaneously. These abrupt changes occur on a time-scale that is much smaller than the time-scale of observation, and, therefore, manifest as *discontinuities*. Because physical systems are inherently continuous, this is a sampling artifact associated with the time-scale of observation. The temporal properties associated with changes, especially discontinuities are exploited in analyzing faults. Other forms of temporal ordering are hard to exploit in a purely qualitative framework. When multiple phenomena with different time constants affect a measurement, the resultant time constant is hard to estimate without detailed computation. Moreover, faults change parameter values which affect time constants, so quantitative analysis is not always possible.

Consider two variables, x_1 and x_2 related to the same fault. x_1 embodies a first order effect and x_2 a second order effect. Typically a measurement is considered *normal* if it is within a certain percentage (say 2%–5%) of its nominal value. Fig. 4 shows the delay times (t_{d1} and t_{d2}) associated with the two variables in crossing the error-threshold. At times between t_{d2} and t_{d1}, x_2 is reported deviant but x_1 is reported normal. A second order effect in continuous time should respond more slowly to the failure than a first order effect, but x_2 is reported to deviate first. In a qualitative reasoning framework, the temporal ordering of first and higher order effects of deviations from normal is, in general, impossible unless the sensor system is wired and calibrated to guarantee a temporal ordering in response times. A *normal observation* at a given time may be a slowly changing signal that has not reached its threshold, therefore, they are not used to refute faults in our consistency-based approach. The only situation where a normal observation reliably refutes a fault is when it is compared against a discontinuity (abrupt change).

Feature Detection Given the difficulties associated with temporal ordering between signals, the analysis of individual signal features becomes the primary discriminative factor in transient analysis. This analysis relies on *magnitude deviation, slopes, and discontinuity at time of failure*. Typically signal deviations are measured in terms of *magnitude changes* and estimated values of *first-order derivatives* only. Studies show that noise in signals and sensor errors make higher-order derivative estimation from measurements very unreliable [3]. In this work, we make the assumption that with appropriate filtering techniques qualitative measurement deviations above/below ($\pm$) normal and slopes ($\pm$) can be estimated reliably.

Using physical principles a first observed change in a measurement is considered discontinuous if magnitude and slope have opposite signs. Discontinuous changes due to parameter deviations are attributed to energy storage parameter changes. In steady state the stored energy in a system does not change, and energy storage parameters have no dynamic effects in steady state, therefore, after an initial discontinuous change, the system returns to its original point of operation. The discontinuity detection algorithm has been been successfully applied in hydraulic systems. Discontinuities in other domains may manifest differently, some may go undetected decreasing diagnosis resolution. Because their detection forms a necessary condition for refutation, this does not result in incorrect diagnosis.

Diagnosis System Implementation

Algorithms applied to the temporal causal graph process measurement deviations. This involves generating hypothesized faults from observed deviations and prediction of their future behavior. Monitoring of the predictions refines the set of possible faults.

Initial Component Parameter Implication

When discrepancies between measurements and nominal values are detected, a *backward propagation* algorithm operates on the temporal causal graph to implicate component parameters. Implicated component parameters are labeled $-$ (below normal) and $+$ (above normal). The algorithm propagates deviant values backward on the directed arcs and consistent $\pm$ deviation labels are assigned sequentially to vertices along the path if they do not have a previously assigned value. An example is shown in Fig. 5 for a deviant right tank pressure, ${e_7}^+$ system (Fig. 1). ${e_7}^+$ initiates backward propagation along $f_7 \stackrel{\frac{1}{C_2}dt}{\rightarrow} e_7^+$ and implicates C_2 below normal (C_2^-) or f_7 above normal (f_7^+). The next step along $f_6 \stackrel{1}{\rightarrow} f_7^+$ implicates f_6^+, and $f_8 \stackrel{-1}{\rightarrow} f_7^+$ implicates f_8^-, and so on. Propagation is terminated along a path when a conflicting assignment is reached. Because backward propagation does not explicitly take

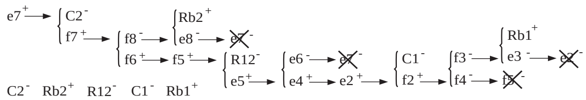

Figure 5: **Backward propagation to find faults.**

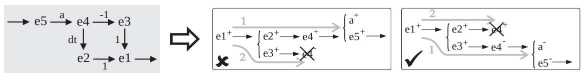

Figure 6: **Instantaneous edges propagate first.**

temporal effects into account, deviant values are propagated along edges with instantaneous relations first. This is to ensure that no faults are generated due to higher order effects which conflict with lower order effects. An example is shown in Fig. 6. If temporal effects are not taken into account, a^+ may be generated based on the observation, e_1^+. However, there is a path from a^+ to e_1 with lower order effect that is opposite. Because at the time of failure lower order effects always dominate, a^- should be generated, and this is achieved by traversing edges with instantaneous relations first. All component parameters along a propagation path are possible faults. As discussed, observed normal measurements do not terminate the backward propagation process.

Prediction

The main task is to derive the signature of dynamic qualitative deviations in magnitude and derivatives of the observed variables under the fault conditions.

Dynamic Behavior The forward propagation algorithm propagates the effect of faulty parameters along instantaneous and temporal links in the temporal causal graph to establish a qualitative value for all measured system variables. Temporal links imply integrating edges, and, therefore, affect the derivative of the variable on the other side of the link. Initially, all deviation propagations are 0^{th} order magnitude values. When an integrating link is traversed, the magnitude change becomes a 1^{st}-order (derivative) change, shown by an $\uparrow$ ($\downarrow$) in Fig. 7. Similarly, a first order change propagating across an integrating link creates a second-order (derivative) change ($\uparrow\uparrow$ ($\downarrow\downarrow$) in Fig. 7), and so on.

Forward propagation with increasing derivatives is

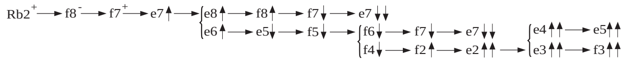

Figure 7: **Forward propagation yields a signature.**

terminated when a signature of sufficient order is derived. The *sufficient* order of the signature for a component fault is determined by a measurement selection algorithm. It is important to keep the sufficient order low because higher order effects typically involve larger time periods allowing more interactions. As a result, feedback effects can modify the transient signal features, making fault detection difficult. On the other hand, in general this requires more measurement points, and a trade-off has to be made. A *complete* signature contains derivatives specified to its sufficient order. When the complete signature of an observed variable has a deviant value, monitoring should report a non normal value for this variable.

When assigning values to vertices, situations may occur where the variable has an assigned deviation for the higher order derivatives but the lower order derivatives are not assigned values. Under the single-fault assumption in prediction, this implies that the lower order derivatives of the prediction for the fault under scrutiny are non-deviating.

Monitoring

This module compares reported signatures and actual observations as they change dynamically after faults have occurred. A number of heuristic mechanisms governed by the dynamic characteristics of the specific system previously discussed improve monitoring quality.

Sensitivity A heuristic parameter is the sensitivity or the time step employed in the monitoring process, which is a function of the different rates of response the system exhibits. Too large a time step may not distinguish between discontinuities and smooth changes. Too small a time step may unnecessarily burden the real-time diagnosis processor. To ensure reliability, we heuristically estimate the time step as a significant fraction of the smallest time constant in the system. The upperbound on this fraction is chosen $\frac{1}{3}$, based on its convergence characteristic in forward Euler numerical simulation.

Progressive Monitoring Transients generated by failures are dynamic, therefore, the signatures of the observed variables change over time. For example, a variable may have a magnitude reported normal and a 1^{st} derivative which is above normal. Over time the variable value will go above normal. Incorporating effects of higher order derivatives in the comparison process is referred to as progressive monitoring. It replaces derivatives that do not match the observed value with the value of derivatives of one order higher in the signature. Fig. 8 shows time stamps marked 1, 2, and 3, where a lower order effect is replaced by man-

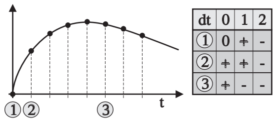

Figure 8: **Progressive monitoring.**

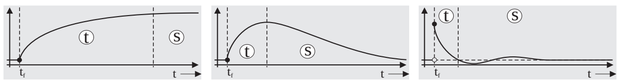

Figure 9: **Qualitative signal transients.**

ifested higher order effects. If the predicted deviation of higher order derivatives do not match the observed value, the fault is rejected.

Progressive monitoring is activated when there is a discrepancy between a predicted value and a monitored value that deviates (this applies to 0^{th} and higher order derivatives). At every time point, it is checked whether the next higher derivative could make the prediction consistent with the observation. If this next higher derivative value is normal the next higher derivative value is considered, until there is either a conflict in prediction and observation, a confirmation, or an unknown value is found.

Temporal Behavior When feedback effects begin changing transient characteristics, fault isolation should be suspended by the monitoring process. Signals may exhibit *compensatory* or *inverse responses* (Fig. 9) [8]. For compensatory responses the `slope becomes` 0. For inverse responses, the `magnitude and slope deviations have opposing sign` assuming there was `no discontinuous change at` t_f. If a discontinuous magnitude change were present, the transient at t_f could manifest as a decrease of this magnitude resulting in a slope with opposite sign. However, this is not an inverse response since the transient effects are exactly those as exhibited at t_f. A *reverse response* occurs for `discontinuous changes at` t_f, and signal overshoot causes the `magnitude deviation to reverse sign`. Qualitative observations of magnitude and slope detect these behaviors from an initial magnitude deviation. When these situations are detected, transient verification (stage **t**, Fig. 9) for that particular signal is suspended and steady state detection is activated (stage **s**, Fig. 9).

Monitoring and Diagnosis Example

We demonstrate the use of the monitoring scheme on a particular fault, a sudden increase in outflow resistance R_{b2}. Fig. 10 shows the results of progressive monitoring, where at times the signatures of the observed vari-

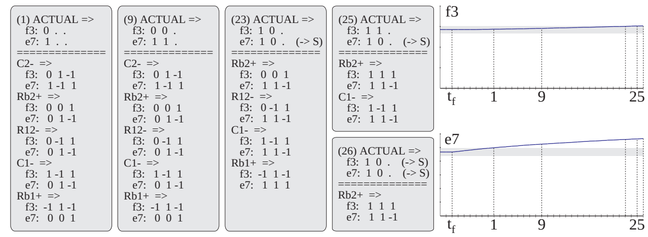

Figure 10: **Progressive monitoring for fault R_{b2}^{+}.**

ables are modified because of higher-order effects. For example, the signature of R_{b1}^{+} for e_7 changes from $0, 0, 1$ in step 9 to $1, 1, 1$ in step 23. The 2^{nd} order derivative, which is positive, is assumed to have affected the magnitude to make the candidate consistent with the observation $1, 1, .$ in step 9. Discontinuity detection was not employed. When discontinuity detection was used, the same result was obtained in three steps[7]. The diagnosis engine as described correctly detected and isolated all single-fault parameter deviations, if pressure in one tank and outflow of the other were measured. Similar results were obtained on a three-tank system[7].

Discussion and Conclusions

Our transient-based diagnosis scheme has been successfully applied to a number of different hydraulic systems[7]. As an example, we illustrate its application to the secondary cooling loop of a fast breeder reactor. The need for a qualitative approach in this system is motivated by its high-order (six), nonlinearity, and the non availability of precise numerical simulation models. The precision of flow sensors is limited and excessive expense is a deterrent to excessive hardware redundancy.

Heat from the reactor core is transported to the turbine by a primary and secondary cooling system. Liquid sodium is pumped through an intermediate heat exchanger to transport heat from the primary cooling loop to the feed water loop by means of a superheater and evaporator vessel (Fig. 11). Pump losses are modeled by R_1. The coil in the intermediate heat exchanger that accounts for flow momentum build-up is represented by a fluid inertia, I_{IHX}. The two sodium vessels are capacitances, C_{EV} and C_{SH}. An overflow column, C_{OFC}, maintains a desired sodium level in the main motor. All connecting pipes are modeled as resistances.

Design documents were used to choose relative parameter values so the behaviors generated would be qualitatively accurate. The monitoring sample rate was set at 20 seconds. Component failures were modeled by changing model parameters by a factor of

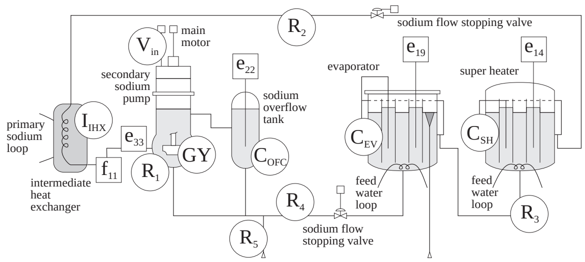

Figure 11: **Secondary sodium cooling loop.**

Fault	Diagnosis	k	Fault	Diagnosis	k
R_1^+	R_1^+	58	R_1^-	R_1^-	43
R_2^+	R_2^+	27	R_2^-	R_2^-	46
R_3^+	R_3^+	1255	R_3^-	$\emptyset$ (R_3^-)	699 (699)
R_4^+	R_5^+ (R_4^+, R_5^+)	3429 (378)	R_4^-	R_4^-	43
R_5^+	R_1^+, R_2^-, R_3^+ R_4^+, R_5^+	2	R_5^-	$R_3^-, R_4^-,$ R_5^-	687
C_{SH}^+	C_{SH}^+	73	C_{SH}^-	C_{SH}^-	16
C_{EV}^+	C_{EV}^+	45	C_{EV}^-	-	-
C_{OFC}^+	C_{OFC}^+	9	C_{OFC}^-	C_{OFC}^-	3
I_{IHX}^+	I_{IHX}^+	16	I_{IHX}^-	I_{IHX}^-	2

Table 1: **Fault detection and identification.**

5. Capacitive and inductive failures produced abrupt change of flow and pressure, respectively. The margin of error was set at 2% for practical reasons, and signatures were generated up to 3^{rd} order derivatives. Steady state was difficult to detect and not used.

The boxed variables in Fig. 11 are typical measurements. Simulation results (Table 1) showed that most faults could be accurately diagnosed in a reasonable number of time steps. R_3^-, R_4^+, and C_{EV}^- were the exceptions. To detect C_{EV}^-, flow of sodium through the overflow mechanism would have to be monitored. This is a configuration change that we will deal with in future work. R_3^- and R_4^+ were detected but not isolated because the overflow mechanism was not modeled in the temporal causal graph (the values in parentheses are results if the overflow mechanism is modeled). Precision in diagnosis may improve by considering predicted effects of order higher than 3, but as noted before, they take longer to manifest which may then cause cascading faults to appear. In real situations, cascading multiple faults are more likely than independent multiple faults. Cascading faults are best handled by quick analysis of transients to establish root-causes and then suspend diagnosis when other faults begin to influence the measured transients. In spite of the loss of precision, the results are more practical from a computational viewpoint.

In summary, our results indicate that the qualitative topology-based diagnosis scheme that integrates monitoring, prediction, and fault isolation works well for complex dynamic systems. Successful diagnosis was achieved by setting monitoring parameters in keeping with the dynamic characteristics of the system, and by tracking transients effectively soon after failures occurred. Future work will require the development of more sophisticated feature detection schemes, and a stronger focus on measurement selection to establish *distinguishability* among possible faults. We are presently working on a measurement selection algorithm that relies on minimal graph coverage techniques.

Acknowledgments We acknowledge the partial financial support of PNC, Japan in conducting this project. Dr. Takashi Washio provided details for the secondary sodium cooling loop system. N. Sriram implemented the system in C++.

References

[1] G. Biswas, R. Kapadia, and X.W. Yu. Combined qualitative quantitative steady state diagnosis of continuous-valued systems. *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 27A, pp. 167-185, 1997.

[2] G. Biswas and X. Yu. A formal modeling scheme for continuous systems: Focus on diagnosis. *Proc. 13th IJCAI*, pp. 1474–1479, Chambery, France, Aug. 1993.

[3] M.J. Chantler et al. The Use of Quantitative Dynamic Models and Dependency Recording for Diagnosis. *7th Intl. Principles of Diagnosis Workshop*, pp. 59-68, Val Morin, Canada, Oct. 1996.

[4] J. deKleer and B.C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.

[5] D. Dvorak and B. Kuipers. Model-based Monitoring of Dynamic Systems. *Proc. 11th IJCAI*, pp. 1238-1243, Detroit, MI, 1989.

[6] P. Frank. Fault diagnosis: A survey and some new results. *Automatica*, 26:459–474, 1990.

[7] P.J. Mosterman and G. Biswas. An integrated architecture for model-based diagnosis. *7th Intl. Principles of Diagnosis Workshop*, pp. 167-174, Val Morin, Canada, Oct. 1996.

[8] B.L. Palowitch. *Fault Diagnosis of Process Plants using Causal Models.* PhD dissertation, Massachusetts Institute of Technology, August 1987.

[9] R.C. Rosenberg and D. Karnopp. *Introduction to Physical System Dynamics*. McGraw-Hill, New York, New York, 1983.