Automatica, Vol. 12, pp. 601–611. Pergamon Press, 1976. Printed in Great Britain

A Survey of Design Methods for Failure Detection in Dynamic Systems*

ALAN S. WILLSKY†

Examination of statistical techniques for the detection of failures in dynamic systems reveals key concepts, similarities and differences in problem formulations, system structure, and performance.

Summary—In this paper we survey a number of methods for the detection of abrupt changes (such as failures) in stochastic dynamical systems. We concentrate on the class of linear systems, but the basic concepts, if not the detailed analyses, carry over to other classes of systems. The methods surveyed range from the design of specific failure-sensitive filters, to the use of statistical tests on filter innovations, to the development of jump process formulations. Tradeoffs in complexity vs performance are discussed.

I. INTRODUCTION

WITH the increasing availability and decreasing cost of digital hardware and software, there has developed a desire in several disciplines for the development of sophisticated digital system design techniques that can greatly improve overall system performance. A good example of this can be found in the field of digital aircraft control (see, for example, Doolin [45], Taylor [46], and Meyer and Cicolani [47]), where a great deal of effort is being put into the design of aircraft with reduced static stability, flexible wings, etc. Such vehicles can provide improved performance in terms of drag reduction and decreased fuel consumption, but they also require sophisticated control systems to deal with problems such as active control of unstable aircraft, suppression of flutter, the detection of system failures, and management of system redundancy. The demands on such a control system are beyond the capabilities of conventional aircraft control system design techniques, and the use of digital techniques is essential.

Another example can be found in the field of electrocardiography. In recent years a great deal of effort has been devoted to the development of digital techniques for the automatic diagnosis of electrocardiograms (ECG's; see, for example, [48]). Such systems can be used for preliminary screening of large numbers of ECG's, for the monitoring of patients in a hospital, etc.

In this paper we review some of the recent work in one area of system theory that is of importance in both of these examples, as well as in many other system design problems. Specifically, we will discuss the problem of the detection of abrupt changes in dynamical systems. In the aircraft control problem one is concerned with the detection of actuator and sensor failures, while in the ECG analysis problem one wants to detect arrhythmias—sudden changes in the rhythm of the heart. For the sake of simplicity in our discussion, we will refer to all such abrupt changes as 'failures', although, as in the ECG example, the abrupt change need not be a physical failure. Our aim in this survey is to provide an overview of a number of the basic concepts in failure detection.

The design of failure detection systems involves the consideration of several issues. One is usually interested in designing a system that will respond rapidly when a failure occurs; however, in high performance systems one often cannot tolerate significant degradation in performance during normal system operation. These two consideration are usually in conflict. That is, a system that is designed to respond quickly to certain abrupt changes must necessarily be sensitive to certain high frequency effects, and this in turn will tend to increase the sensitivity of the system to noise, via the occurrence of false alarms signaled by the failure detection system. The tradeoff between these design issues is best studied in the context of a specific example in which the costs of the various tradeoffs can be assessed. For example, one might be more willing to tolerate false alarms in a highly redundant system configuration than in a system without substantial back-up capabilities.

In general, one would like to design a failure detection system that takes system redundancy into account. For example, in a system containing several back-up subsystems we may be able to devise a simple detection algorithm that is easily implemented but yields only moderate false alarm rates. On the other hand, by implementing a more complex failure detection algorithm that takes careful account of system dynamics, one may be able to reduce requirements for costly hardware redundancy.

In addition to taking hardware issues into consideration, the designer of failure detection systems should consider the issue of computational complexity. One clearly needs a scheme that has reasonable storage and time requirements. It would also be useful to have a design methodology that admits a range of implementations, allowing a tradeoff study of system complexity vs performance. In addition, it would be desirable to have a design that takes advantage of new computer capabilities and structures, e.g. designs that are amenable to modular or parallel implementations.

In this paper we survey a variety of failure detection methods, and, keeping the issues mentioned above in mind, we will comment on the characteristics, advantages, disadvantages, and tradeoffs involved in the various techniques. In order to provide this survey with some organization and to point out some of the key concepts in failure detection system design, we have defined several categories of failure detection systems and have placed the designs we have collected into these groups. Clearly such a grouping can only be a rough approximation, and we caution the reader against drawing too much of an inference about individual designs based on our classification of them; several of the techniques could easily fall into a number of our classes. In addition, for the sake of brevity we have limited our detailed discussions to only a few of the many techniques. Our choice of those techniques has been motivated by a desire to span the range of available methods and by our familiarity with certain of these algorithms. Finally, we have attempted to collect all of those studies of the failure detection problem of which we are aware, and we apologize for any oversights.

We also note that problems of equal importance to that of failure detection are the issue of system reorganization

*Received 16 December 1975; revised 10 May 1976. The original version of this paper was not presented at any IFAC meeting. It was recommended for publication in revised form by associate editor A. Longmuir.

†Associate Professor of Electrical Engineering and Computer Science, Electronic Systems Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A. This research was supported in part by NASA Ames Research Center under Grant NGL-22-009-124 and in part by NASA Langley Research Center under Grant NSG-1112.

subsequent to the detection of a failure and the design of 'fault tolerant' control systems that retain their integrity in the presence of system failures or large changes in system operating conditions. Since the principle thrust of this paper is directed at the detection problem, we limit ourselves in these other areas to pointing out the work of several authors on the closed loop control problem. Specifically, Beard[4] introduces several notions of analytical sensor and actuator redundancy and discusses how one might restructure a control system following failure detection. The work of Sworder *et al.*[17]–[20], [37] and that of Ratner and Luenberger[21] is aimed primarily at the development of an adaptive control system for linear systems in which parameters may jump abruptly. Finally, several authors have devised control system designs that are 'fault tolerant' in that they remain stable in the presence of certain large changes in system characteristics. Belletrutti and MacFarlane[55] and Solheim[56] have used frequency domain methods to determine useful linear multivariable designs. Also, Wong and Athans[57] and Safonov and Athans[58] have determined conditions under which optimal linear control systems, with quadratic criteria and Gaussian, additive noise, remain stable under large system variations. The interested reader is referred to these references for details.

II. FORMULATIONS OF THE FAILURE DETECTION PROBLEM

In this paper we are mostly concerned with the analysis of linear stochastic models in the standard state space form.

System dynamics

$$x(k+1) = \Phi(k)x(k) + B(k)u(k) + w(k). \tag{1}$$

Sensor equation

$$z(k) = H(k)x(k) + J(k)u(k) + v(k) \tag{2}$$

where u is a known input, and w and v are zero-mean, independent, white Gaussian sequences with covariances defined by

$$E[w(k)w'(j)] = Q\delta_{kj}, \quad E[v(k)v'(j)] = R\delta_{kj} \tag{3}$$

where δ_{kj} is the Kronecker delta. We think of (1)–(3) as describing the 'normal operation' or 'no failure' model of the system of interest. If no failures occur, the optimal state estimator is given by the discrete Kalman filter equations[33]

$$\hat{x}(k+1|k) = \Phi(k)\hat{x}(k|k) + B(k)u(k) \tag{4}$$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)\gamma(k) \tag{5}$$

$$\gamma(k) = z(k) - H(k)\hat{x}(k|k-1) - J(k)u(k) \tag{6}$$

where γ is the zero-mean, Gaussian innovation process, and the gain K is calculated from the equations

$$P(k+1|k) = \Phi(k)P(k|k)\Phi'(k) + Q \tag{7}$$

$$V(k) = H(k)P(k|k-1)H'(k) + R \tag{8}$$

$$K(k) = P(k|k-1)H'(k)V^{-1}(k) \tag{9}$$

$$P(k|k) = P(k|k-1) - K(k)H(k)P(k|k-1). \tag{10}$$

Here $P(i|j)$ is the estimation error covariance of the estimate $\hat{x}(i|j)$, and $V(k)$ is the covariance of $\gamma(k)$. We refer to (4)–(10) as the 'normal mode filter' in the sequel.

In addition to the above estimator, one may also have a closed loop control law, such as the linear law

$$u(k) = G(k)\hat{x}(k|k). \tag{11}$$

We then obtain the normal operation configuration depicted in Fig. 1.

The problem of failure detection is concerned with the detection of abrupt changes in a system, as modeled by (1)–(3). Such abrupt changes can arise in a number of ways.

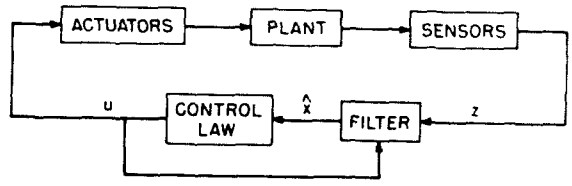

FIG. 1. No-failure system configuration.

For example, in aerospace applications, one is often concerned with the failure of control actuators and surfaces. Such abrupt changes can manifest themselves as shifts in the control gain matrix B, increased process noise, or as a bias in equation (1), as might arise if a thruster developed a leak[31]. In addition, failures of sensors may take the form of abrupt changes in H, increases in measurement noise, or as biases in (2). For simplicity, we will refer to abrupt changes in (1) as 'actuator failures,' and shifts in (2) will be called 'sensor failures.' Again we point out that in many applications shifts in (1) or (2) may be used to model changes in observed system behavior that have nothing to do with actuators or sensors.

The main task of a failure detection and compensation design is to modify the normal mode configuration in order to include the capability of detecting abrupt changes and compensating for them by activating back-up systems, adjusting the feedback design appropriately, etc. Conceptually, we think of the detection-compensation system as part of the filtering portion of the feedback loop. As illustrated in Figs 2 and 3, the resulting filter design can take one of two forms. Either we perform a complete redesign of the filter, replacing (4)–(10) with a filter that is sensitive to failures, or we design a system that monitors the normal system configuration and adjusts the system accordingly. We will discuss examples of both of these structures.

As mentioned earlier, we will concentrate primarily on the problem of failure detection, which we consider to consist of three tasks—alarm, isolation, and estimation. The alarm task simply consists of making a binary decision—either that something has gone wrong or that everything is fine. The problem of isolation is that of determining the source of the failure—e.g. which sensor or actuator has failed, what type of arrhythmia has occurred, etc. Finally, the estimation problem involves the determination of the extent of failure. For example, a sensor may become completely non-operational (an 'off' or 'hard-over' failure), or it may simply suffer degradation in the form of a bias or increased inaccuracies, which may be modeled as an increase in the sensor noise covariance. In the latter case, estimates of the bias or the

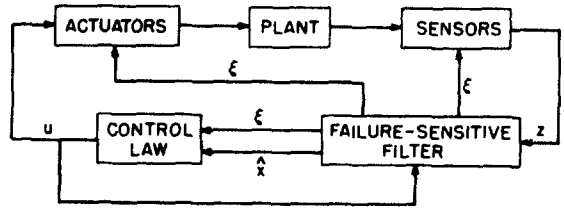

FIG. 2. Failure detection system involving failure-sensitive primary filter (here ξ denotes information concerning detected failures).

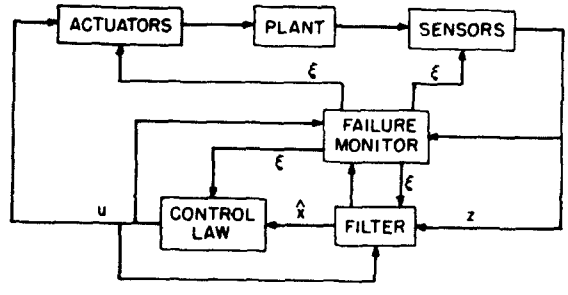

FIG. 3. Failure detection system involving a monitoring system for the no-failure configuration.

increase in noise may allow continued use of the sensor, albeit in a degraded mode. Clearly the extent to which we need to perform these various tasks depends upon the application. If a human operator is available, we may only be interested in generating an alarm that tells him to perform further tests. In other systems in which back-ups are available, we might settle for failure isolation without estimation. On the other hand, in the absence of hardware redundancy, we may be interested in using a degraded instrument and thus would need estimation information.

Intuitively we can associate increased software system complexity with the tasks—i.e. isolation requires more sophisticated data processing than an alarm, and estimation more than isolation. On the other side, as we increase failure detection capabilities, we may be able to *decrease* hardware redundancy. Also, in some applications we may be able to delay isolation and estimation until after an alarm has been sounded. In such a sequential structure, one increases detector complexity *after* a failure has been detected, thereby reducing the computational burden during normal operation. Again the details of such considerations depend upon the particular application.

Another tradeoff involving failure detection system complexity involves its relation to detection system performance. For example, one might expect that one could achieve better alarm performance by using *a priori* knowledge concerning likely failure modes. That is, by looking for specific forms of system behavior that are characteristic of certain failures, one should be able to improve detection performance. Thus, it seems likely that alarm performance, as measured by the tradeoff between false alarms and missed detections, will be improved if we attempt simultaneous detection, isolation, and estimation of failures. This tradeoff of complexity vs performance is extremely important in the design of failure detection systems.

In the following sections we will discuss several failure detection methods and will comment on their characteristics with respect to the issues mentioned in this and the preceding section. In addition, in our concluding remarks we briefly summarize some of the advantages and disadvantages of the various methods and point out several areas for future work.

In this section we have not attempted to define a general set of failure models to be considered, as the various techniques are based on quite different failure models. In addition, we will also see that there are differences between the hypothesized dynamic models for the different methods, i.e. modifications of the model (1), (2). These differences—both in dynamics and in failure models—are significant in that they indicate the basis for each of the methods and thus shed light on the relative merits and range of applicability of each technique.

III. 'FAILURE-SENSITIVE' FILTERS

Our first class of failure detection concepts is aimed at overcoming the problem of an 'oblivious filter'. As has been noted by many authors [1]–[3], [33], the optimal filter defined by (4)–(10) performs well if there are no modelling errors; however, it is possible for the filter estimate to diverge if there are substantial unmodeled phenomena. The problem occurs because the filter 'learns the state too well'—i.e. the precomputed error covariance P and filter gain K become small, and the filter relies on old measurements for its estimates and is oblivious to new measurements. Thus, if an abrupt change occurs, the filter will respond quite sluggishly, yielding poor performance. Consequently, one would like to devise filter designs that remain sensitive to new data so that abrupt changes will be reflected in the filter behavior.

Two well-known techniques for keeping the filter sensitive to new data are the exponentially age-weighted filter studied by Fagin [1] and Tarn and Zaborszky [2] and the limited memory filter proposed by Jazwinski [3]. Others, such as increasing noise covariances or simply fixing the filter gain are discussed by Jazwinski in [33]. These techniques yield only indirect failure information. That is, if an abrupt change occurs, these filters will respond faster than the normal filter, and one can base a failure detection decision on sudden changes of $\hat{x}$.

It is important to note a performance tradeoff evident in this method. As we increase our sensitivity to new data, by effectively increasing the bandwidth of the Kalman filter, our system becomes more sensitive to sensor noise, and the performance of the filter under no-failure conditions degrades. In some cases this can be rather severe, and one may not be able to tolerate the degradation in overall system performance under no-failure conditions. One might then consider a two filter system—the normal mode filter (4)–(10) as the primary filter, with this type of failure-sensitive filter as an auxiliary monitor, used only to detect abrupt changes. We remark that the tradeoff between detection performance and filter behavior under normal conditions is a characteristic of all failure detection systems and is analogous to the costs associated with false alarms and missed detections in standard detection problems [41].

The techniques mentioned so far in this section are rather indirect failure detection approaches. Several methods have been developed for the design of filters that are sensitive to specific failures. One method involves the inclusion of several 'failure states' in the dynamic model (1)–(3). Kerr [25] has considered a procedure in which failure modes, such as the onset of biases, are included as state variables. If the estimates of these variables vary markedly from their nominal values, a failure is declared. A two-confidence interval overlap decision rule for failure detection using such failure states is described and its performance is analyzed in [25]. Note that this approach provides failure isolation and estimation at the expense of increased dimensionality and some performance degradation under no-failure conditions since inclusion of the added states effectively opens up the bandwidth of the Kalman filter.

An alternative to the addition of failure states to the dynamic model is the class of detector filters developed by Beard [4] and Jones [5]. Their work has led to a systematic design procedure for the detection of a wide variety of abrupt changes in linear time-invariant systems. They consider the continuous-time, time-invariant, deterministic system model

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{11}$$

$$z(t) = Cx(t) \tag{12}$$

and design a filter of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{x}(t) = A\hat{x}(t) + D(z(t) - C\hat{x}(t)) + Bu(t). \tag{13}$$

The primary criterion in the choice of the gain matrix D is not that (13) provide a good estimate of x, as it is with observers or optimal estimators, but rather that the effects of certain failures are accentuated in the filter residual

$$\gamma(t) = z(t) - C\hat{x}(t). \tag{14}$$

The basic idea is to choose D so that particular failure modes manifest themselves as residuals which remain in a fixed direction or in a fixed plane.

To illustrate the Beard–Jones approach, let us consider a simple example from [4]. Suppose we wish to detect a failure of the ith actuator, i.e. in the actuator driven by the ith component of u. If we assume the failure takes the form of a constant bias, our state equation becomes

$$\begin{aligned}\dot{x}(t) &= Ax(t) + B[u(t) + \nu e_i] \\ &= Ax(t) + Bu(t) + \nu b_i, \quad t \geq t_0\end{aligned} \tag{15}$$

where e_i is the ith standard basis vector, b_i is the ith column of B, and t_0 is the (unknown) time of failure. Suppose we consider the case of full state measurement—i.e. let $C = I$. In this case we obtain a differential equation for the residual

$$\dot{\gamma}(t) = [A - D]\gamma(t) + \nu b_i. \tag{16}$$

If we choose $D = \sigma I + A$, we obtain

$$\dot{\gamma}(t) = -\sigma\gamma(t) + \nu b_i$$

$$\gamma(t) = e^{-\sigma(t-t_0)}\gamma(t_0) + \frac{\nu[1-e^{-\sigma t}]}{\sigma} b_i. \quad (17)$$

Thus, as the effect of the initial condition dies out, $\gamma(t)$ maintains a fixed direction (b_i) with magnitude proportional to failure size (ν). Note that as we increase σ, thus increasing filter gain, the initial condition dies out faster, but the magnitude of the steady-state value of γ decreases. Thus, if there is any noise in the system, we cannot make σ arbitrarily large.

In their work Beard and Jones consider the design of such filters for an extremely wide variety of failure modes, including actuator and sensor shifts and shifts in A and B. The initial deterministic analysis for all of these cases was considered by Beard[4], while a systematic design procedure is given by Jones[5] for the design of the gain D to allow detection of several failures modes. Jones' approach is quite geometric in nature, and his formulation allows one to gain considerable insight into the detection problem. As pointed out in [5], the gain selection problem is quite similar to the output decoupling problem and requires the introduction of the important concept of 'mutually detectable failure modes' in order to answer the question of whether or not one can simultaneously distinguish between several types of failures. Thus the question of failure isolation is of central importance in the design methodology derived in [5].

The results in [4], [5] represent perhaps the most thorough study of the basic concepts underlying failure detection. The tradeoff between detection and filter performance is discussed in depth in [5] and an attempt is made in [4] to introduce the concept of the level of redundancy in a dynamical system.

As mentioned in the example, the basic design procedure is deterministic. However, in this simple example we can see how one can take noise into account. If the system (11), (12) contains noise, we have seen that one may not wish to make the scalar σ as large as possible. In fact, one could choose σ so as to minimize the mean-square estimation error in the detector filter when there is no failure. In his thesis[5], Jones describes a procedure in which one first chooses the structure of D for failure detection purposes and then chooses the remaining free parameters in order to minimize the estimation error covariance. Although this yields a suboptimal filter design, it may work quite well, as it did in the problem reported in [5].

In summary, the Jones–Beard design methodology is extremely useful conceptually, can be used to detect a wide variety of failures, and provides detailed failure isolation information. It is suboptimal as an estimator, and if this presents a serious problem, one might wish to use the detector filter as an auxiliary monitoring system. This appears to be only a minor drawback, and the major limitation of the approach is its applicability only to time-invariant systems.

IV. VOTING SYSTEMS

Voting techniques are often useful in systems that possess a high degree of parallel hardware redundancy. Memoryless voting methods can work quite well for the detection of 'hard' or large failures, and the papers of Gilmore and McKern[6], Pejsa[7], and Ephgrave[8] discuss the successful application of voting techniques to the detection of hard gyro failures in inertial navigation systems.

In standard voting schemes, one has (at least) three identical instruments. Simple logic is then developed to detect failures and eliminate faulty instruments, for example, if one of the three redundant signals differs markedly from the other two, the differing signal is eliminated. Recently, Broen[9] has developed a class of voter–estimators that possesses advantages relative to standard voting techniques. Consider the dynamical system

$$x(k+1) = \Phi x(k) \quad (18)$$

with a triply redundant set of sensors

$$y_i(k) = Hx(k) + v_i(k), \quad i = 1, 2, 3. \quad (19)$$

Broen develops a set of recursive filter equations for computing the estimate $\hat{x}(k)$ that minimizes

$$J_k = \sum_{i=0}^{k}\sum_{j=1}^{3} w_{ji}\gamma_j'(i)R_j^{-1}\gamma_j(i) \quad (20)$$

where R_j is the covariance of the measurement noise v_j, and γ_j is the innovations sequence

$$\gamma_j(i) = y_j(i) - H\Phi^{i-k}\hat{x}(k). \quad (21)$$

Here w_{ji} is a function of $y_1(i)$, $y_2(i)$, $y_3(i)$ which is large if $y_j(i)$ is close to the other two $y_m(i)$ and is small if $y_j(i)$ deviates greatly from the other two. In this way, one obtains a 'soft' voting procedure in which faulty sensors are smoothly removed from considerations. This greatly alleviates the cost of false alarms, but the price is the on-line computation of the filter gain, which is a function of the w_{ji}. In addition, the limitation of the dynamic model (18) with respect to the general model (1) is quite apparent.

Voting schemes are in general relatively easy to implement and usually provide fast detection of hard failures, but they are only applicable in systems possessing a high level of parallel redundancy. They do not in general take advantage of redundant information provided by unlike sensors, and thus cannot detect failures in single or even doubly redundant sensors. In addition, voting techniques can have difficulties in detecting 'soft' failures, such as a small bias shift.

V. MULTIPLE HYPOTHESIS FILTER-DETECTORS

A rather large class of adaptive estimation and failure detection schemes involves the use of a 'bank' of linear filters based on different hypotheses concerning the underlying system behavior. In the work of Athans and Willner[10] and Lainiotis[11], several different sets of system matrices are hypothesized. Filters for each of the models are constructed, and the innovations from the various filters are used to compute the conditional probability that each system model is the correct one. In this manner, one can do simultaneous system identification and state estimation. In addition, an abrupt change in the probabilities can be used to detect changes in true system behavior. This technique has been investigated in the context of the adaptive control of the F-8C digital fly-by-wire aircraft by Athans *et al.*[35] and also has been applied to the problem of classifying rhythms and detecting rhythm shifts in electrocardiograms. Extremely good results in the latter case are reported by Gustafson *et al.* in [36].

Techniques involving multiple hypotheses have also been used to design failure detection systems. Montgomery, Caglayan and Price[12], [13] have used such a technique for digital flight control systems and have studied its robustness in the presence of nonlinearities via simulations. Recently a technique involving a bank of observers has been devised, and a successful application to a hydrofoil sensor failure problem is reported by Clark *et al.* in [34]. Also, Willsky, Deyst, and Crawford[15], [16] have applied the methodology devised by Buxbaum and Haddad in [14] to study failure detection for an inertial navigation problem. We will briefly describe this technique to illustrate some of the concepts underlying the bank of filters approach. We also refer the reader to Wernersson[42] for a technique that is similar to that discussed in [16].

Consider the system

$$x(k+1) = \Phi(k)x(k) + w(k) \quad (22)$$

$$z(k) = H(k)x(k) + v(k). \quad (23)$$

We are interested in detecting sudden shifts in certain of the components of x, e.g. bias states. We model these shifts by choosing the distribution of w appropriately. Let $\{f_1, \ldots, f_r\}$ be the set of hypothesized failure directions. We then assume that w has a high probability of being the usual process noise and a small probability of including a burst of noise in each of

the failure directions. Thus the density for $w(k)$ is

$$p_0 N(0, Q) + \sum_{i=1}^{r} p_i N(0, Q + \sigma_i f_i f_i') \tag{24}$$

$$\sum_{i=0}^{r} p_i = 1, \quad p_0 \gg p_i \quad i = 1, \ldots, r. \tag{25}$$

Here $N(m, P)$ is a normal density with mean m and covariance P.

If we hypothesize such a density at each point in time and if we assume that $x(0)$ is normally distributed, we have the following expression for the conditional density of $x(k)$ given $z(1), \ldots, z(k)$

$$p(x, k) = \sum_{i_0=0}^{r} \cdots \sum_{i_{k-1}=0}^{r} p_i N(\eta_i, p_i). \tag{26}$$

Here $\mathbf{i} = (i_0, \ldots, i_{k-1})$ and the density has the following interpretation. Let $\mathbf{j} = (j_0, \ldots, j_{k-1})$ be a random k-tuple where $j_s = i$ if there is a shift in the f_i direction at time s ($i = 0$ is used to denote no shift). Then

$$p_i = \text{Prob}\,(\mathbf{j} = \mathbf{i} | z(1), \ldots, z(k)) \tag{27}$$

and η_i and p_i are the mean and covariance of the Kalman filter designed assuming $\mathbf{j} = \mathbf{i}$, i.e. assuming $w(s)$ has covariance $Q + \sigma_{i_s} f_{i_s} f_{i_s}'$. The $\mathbf{p_i}$ can be computed in a sequential manner as a function of the various filter innovations. We refer the reader to [14]–[16] for the details of the calculations.

Note that the implementation of (26) requires an exponentially growing bank of filters: there are $(r+1)^k$ terms in (26). To avoid this problem a number of approximation techniques have been proposed [14]–[16]. The one used in [16] involves hypothesizing shifts only once every N steps. At the end of each N step period we "fuse" the $(r+1)$ densities into a single density and begin the procedure again. In this way we implement only $(r+1)$ filters at any time. We note that the techniques devised in [10]–[12] do not involve growing banks of filters, since the number of hypothesized models do not grow in time. However, it is possible for all of the filters in the bank to become oblivious, and thus shifts between the hypotheses may go undetected (see [16], [36] for examples). The technique of periodic fusing of the densities and initiation of new bank effectively avoids this problem, as would designing the original bank using age-weighted filtering techniques.

The technique described above was applied to the problem of detecting gyro and accelerometer bias shifts in a time-varying inertial calibration and alignment system. The results of these tests are extremely impressive. This is not surprising, as the multiple hypothesis method computes precisely the quantities of interest—the probabilities of all types of failures under consideration. The cost associated with such a high level of performance is an extremely complex failure detection system. Note, however, that the parallel structure of the system allows one to consider highly efficient parallel processing computer implementations. In addition, the use of reduced-order filters for the various failure hypotheses may increase the practicality of such a scheme, or one might consider the use of a simpler detection-only system to detect failures, with a switch to a multiple hypothesis procedure for failure isolation and estimation after a failure has been detected.

However, even if such a failure detection scheme cannot be implemented in a particular application, it provides a useful benchmark for comparison with simpler techniques. In addition, by studying the simulation of a multiple hypothesis method, one can gain useful insight into the dynamics of failure propagation and detection (see the discussion in [16]).

One of the earliest techniques for detecting a switch between different models was derived by Newbold and Ho[52]. They considered the problem of detecting a *single* switch between two dynamic models ('failed' and 'unfailed') with different process noise covariances. By eliminating the possibility of multiple switches, one immediately reduces the exponential growth in the required number of filters to linear growth, since we implement one new filter at each point in time based on the hypothesis that the process has just changed. Newbold and Ho proposed an implementation that eliminated this linear growth by using a decision rule based on the *sequential probability ratio test* (SPRT)[61]. This test proceeds as follows: given two hypotheses (for example, the process switched at time n or the process did not), we compute the *a posteriori* probabilities of the two models (in the same manner as in the other multiple model methods; e.g. see [35], [36] and the calculation discussed above in connection with equations (26) and (27)). We compare the logarithm of the ratio of these two probabilities to two thresholds. If it exceeds one threshold or falls below the other we terminate the test with a decision corresponding to the threshold that is crossed. Until the log likelihood ratio exceeds the threshold, we defer decision. The SPRT is an extremely powerful test, and we refer the reader to [61] for a detailed description of it. We note here only that the SPRT minimizes the time to reach a decision for given probabilities of making the wrong decisions, e.g. declaring hypothesis 1 when it is really hypothesis 2 and vice versa. Given this property, Newbold and Ho were motivated to propose that one perform 'occasional' tests—i.e. we run a single SPRT. Once it reaches a decision, we again hypothesize that a shift has occurred and start again. We note that this has the same flavor as the method proposed earlier in which one hypothesizes changes only every N steps. The advantage of the Newbold–Ho algorithm is that the size N can vary, depending upon actual data. That is we do not start a new test until the previous SPRT tells us that nothing happened in the preceding interval. As with all such 'occasional' or 'every N step' tests, these algorithms may not respond optimally if a change occurs in the middle of a test. That is, we may have to wait until the next test to detect the change. We refer the reader to [52] for details of the above method, for another suboptimal approximation, and for the application of this technique to a gyro failure detection problem.

McGarty[23] has developed a method for rejecting bad measurements that bears some similarity to the approaches just discussed. Each measurement has a binary random variable $g(k)$ associated with it. If $g(k) = 1$ the measurement is 'good', i.e. the measurement contains the signal of interest, while $g(k) = 0$ denotes a bad data point, the measurement is pure noise. McGarty devises a maximum likelihood approach for estimating the values of the exponentially growing set of possibilities ($g(i) = 1$ or 0, $i = 1, \ldots, k$). He also allows these variables to have a sequential correlation, i.e. knowing that the present measurement is good or bad says something about the next observation. A computationally feasible approximation method is devised and simulation results are described. We refer the reader to [23] for details. We also note that Nahi[54] considered the same problem formulation as McGarty, but he considered the problem of finding the best *linear* estimator of the state of the system. Thus, he obtained a linear filter, similar to a Kalman filter, but with dynamics appropriately modified to take into account the *a priori* probability of any measurement containing only noise. Although the solution is far simple than McGarty's, it does not remove bad measurements or provide any adaptivity in the filter structure.

Recently, Athans *et al.*[51] have also considered the problem of designing an estimator that can detect and remove bad or false measurements. Their approach is Bayesian in nature—i.e. an estimate is generated of the *a posteriori* probability that a given measurement is false. The method of calculation of these pseudo-probabilities is quite similar to that used in the other multiple hypothesis methods (see [10]–[14]). The reader is referred to [51] for details of the analysis and for a discussion of some successful simulation results.

VI. JUMP PROCESS FORMULATIONS

The problem of the detection of abrupt changes in dynamical systems suggests the use of jump process techniques in devising system design methodologies (see [39],

[49]–[50] for general results on jump processes). One models potential failures as jumps, characterized by *a priori* distributions which reflect initial information concerning failure rates. The size of the possible failures are usually taken to be known. One could, however, model failure magnitude as a random variable. This leads to a compound jump process formulation which greatly complicates the desired analysis. In any event, taking such a jump process formulation, one can devise failure-sensitive control laws and methods for computing the conditional probability of failure. Control problems of this type have received a great deal of attention in the literature. Sworder and Robinson[17]–[20], [37] and Ratner and Luenberger[21] have considered the design of control laws which take into account the possibility of sudden shifts in system matrices. The results they have obtained are for the full-state feedback problem with no system randomness other than the jumping of the system matrices among a finite set of possible matrices.

Davis[22] has utilized nonlinear estimation techniques to solve a fault detection problem. His formulation is as follows: consider the scalar stochastic equations

$$dx(t) = a(t)x(t)\,dt + g(t)\,dv(t) \tag{28}$$

$$dy = h(t)x(t)\,dt + dw(t) \tag{29}$$

where w and v are independent Brownian motion processes and

$$a(t) = a_0(t)[1 - \xi(t)] + a_1(t)\xi(t) \tag{30}$$

where

$$\xi(t) = \begin{cases} 0 & t < T \\ 1 & t \geq T \end{cases} \tag{31}$$

and T is a random variable. Here we interpret a_0 as the unfailed dynamics, and a_1 represents the failure mode. Davis derives the optimal, infinite-dimensional equations for the computation of the conditional mean of x and the conditional probability

$$\hat{\xi}(t|t) = Pr[t \geq T | y(s), \quad 0 \leq s < t]. \tag{32}$$

An implementable approximation is described in [22], but evaluation of its performance has not as yet been made.

Note that Davis' method leads to an estimate of x that is suboptimal under no-failure conditions. Chien[24] has devised a jump process formulation that avoids this difficulty for the problem of the detection of a jump or a ramp in a gyro bias. He considers the dynamical model.

$$\dot{x}(t) = \omega x(t) + w(t) \tag{33}$$

where w is a white noise process. Three hypotheses are conjectured for the form of the gyro output.

Normal Mode H_0:

$$z(t) = x(t) + v(t) \qquad \forall t \tag{34}$$

Bias Mode H_1:

$$z(t) = x(t) + m\xi(t) + v(t) \qquad t > T \tag{35}$$

Ramp Mode H_2:

$$z(t) = x(t) + n(t - T)\xi(t) + v(t) \quad t > T \tag{36}$$

where n and m are unknown constants, v is white noise, T is the time of failure, and $\xi(t)$ is as in (31).

Chien's approach is as follows: design a filter based on H_0 (which will thus yield the optimal estimate for $t < T$, assuming no false alarms occur), and determine the *steady-state* effect of the degradations H_1 and H_2 on the filter residuals. If one then hypothesizes a failure rate q—i.e.

$$\text{Prob}\,(T > t) = e^{-qt} \tag{37}$$

and if one further assumes a nominal size for the bias m, one can then compute an approximate stochastic differential equation for $Pr(H_1|z(s), s \leq t)$, in which the input to this equation is the residual γ of the H_0 filter. The details of the analysis are described in [24].

For his problem Chien is able to demonstrate that his detection procedure—based on the assumption of a nominal value for the bias failure m—has the capability of detecting biases larger than m and also can be used to detect ramps (mode H_2). Of course, the delay times until detection in these cases are greater than if one implemented a filter based on the proper bias size or if one were looking for a ramp, indicating the potential usefulness of estimating the failure magnitude. The major advantages of Chien's approach are the simplicity of the detector, implementation of a scalar stochastic equation, and the fact that one obtains an estimate of precisely the quantity of interest—the conditional probability of failure. The simplicity of the scheme may, in fact, make it a great deal more robust in the face of system modelling errors, such as the use of an extremely simplified gyro error model, than more sophisticated approaches. Also, this approach leads to no degradation in performance prior to detection of the failure. In addition, the use of a probabilistic description of the time of failure allows one to avoid the problem of the oblivious filter—i.e. the fact that a failure can occur at any time has been incorporated in the design, which therefore will remain sensitive to new data.

The drawbacks of the scheme are the use of a fixed bias size and the use of the steady-state effect of the failure on the filter residual. The first of these may not be too much of a problem, as Chien as pointed out, but the second may cause difficulties. Specifically, this limits the approach to time-invariant systems and filters. In addition, as the transient effect of the failure has been ignored, it may be difficult to make quick detections of certain changes, i.e. we may have to wait until the transient dies out. In the next section we will discuss an approach (the GLR method) which has several concepts in common with Chien's approach and which allows one to overcome these two drawbacks, at the cost of added computational complexity, of course.

In summary, jump process formulations appear to be quite natural for failure detection problems. One usually makes approximations in the analysis in order to obtain implementable solutions. These simplifications impose some limitations on the capabilities of the designs, but there is at present no systematic analytical procedure for evaluating these limitations or for studying tradeoffs between design complexity and system performance.

VII. INNOVATIONS-BASED DETECTION SYSTEMS

Chien's failure detection technique can also be placed in the class of failure detection methods that involve the monitoring of the innovations of a filter based on the hypothesis of normal system operation. In such a configuration the overall system uses the normal filter until the innovations monitoring system detects some form of aberrant behavior. The fact that the monitoring system can be attached to a filter–controller feedback system is particularly appealing, since overall system behavior is not disturbed until after the monitor signals a failure and since the monitoring system can be designed to be added to an existing system.

Mehra and Peschon[26] have suggested a number of possible statistical tests to be performed on the innovations. One of these is a chi-squared test which was applied in [15], [16] by Willsky, Deyst and Crawford. Let $\gamma(k)$ be the p-dimensional innovations for the filter defined by (4)–(10). If the system is operating normally, the innovations is zero-mean and white with known covariance $V(k)$. In this case the quantity

$$l(k) = \sum_{j=k-N+1}^{k} \gamma'(j)V^{-1}(j)\gamma(j) \tag{38}$$

is a chi-squared random variable with Np degrees of freedom[26], [15], [16]. If a system abnormality occurs, the

statistics of γ change, and one can consider a detection rule of the form

$$l(k) > \epsilon \Rightarrow \text{FAILURE}$$
$$l(k) \leq \epsilon \Rightarrow \text{NO FAILURE}. \tag{39}$$

With the aid of chi-squared tables, one can compute the probability P_F of false alarm as a function of the innovations window length N and the decision threshold ϵ. The probability P_D of correct detection depends upon the particular failure mode as indicated in [16] and the discussion of the GLR approach to follow. We note that for a given failure mode, as N increases the probability of correct detection may decrease—i.e. by averaging a larger number of residuals we smooth out the effect of a failure on γ, and the detector may become somewhat oblivious, or at the very best respond quite slowly, to new data. On the other hand, too small a value of N may yield an unacceptably high value of P_F.

The implementation of the chi-squared test (38), (39) is quite simple, but, as one might expect, one pays for this simplicity with rather severe limitations on performance. As described in [15], [16] this method was applied to the same inertial calibration and alignment problem to which the Buxbaum–Hadded multiple hypothesis approach[14]–[16], described in Section V was applied. The performance of the chi-squared test was mixed. The method is basically an alarm method—i.e. the system (38), (39) makes no attempt to isolate failures—and one finds that those failure modes that have dramatic effects on γ are detectable by this method; however more subtle failures are more difficult to detect with this simple scheme. Comparing the performance of the multiple hypothesis and chi-squared systems, we see that in some cases we can obtain superior alarm capabilities if we simultaneously attempt to do failure isolation and estimation. One can obtain some failure isolation information by considering the components of γ separately (this may be especially useful for sensor failures), and we refer the reader to [15], [16] for a detailed discussion of this and other aspects of the chi-squared method.

Another innovations-based approach, developed by Merrill[27], is motivated by a desire to suppress bad sensor data. Merrill devises a modification of the least squares criterion in order to suppress extremely large residuals, which are given a very large weighting in the usual least squares framework, and he applies his methodology to a power system application. The use of weighted residuals tests (as in [38]), combined with Merrill's nonquadratic criterion were used by Schweppe, Hanschin *et al.*[62], [64] for bad data analysis in static power system estimation problems. Dynamic versions of these concepts were briefly considered by Merrill[27], and Peterson[63] has extended the Schweppe–Handschin technique to the case of bad data suppression for dynamic systems. This method essentially involves performing a static test at each point in time, incorporating the new measurement and the predicted estimate of this measurement based on previous data, regarded as an additional measurement of the present state. We refer the reader to [63] for details.

A final technique in this category has been studied by several researchers—Willsky and Jones[28], [29], McAulay and Denlinzer[30], Deyst and Deckert[31], Sanyal and Shen[32], and Chow, Dunn and Willsky[38]—and we will describe the most general formulation of the approach, developed in [28], [29]. This technique, which we call the generalized likelihood ratio (GLR) approach, was in part motivated by the shortcomings of the simpler chi-squared procedure. The GLR approach, which can be applied to a wide range of actuator and sensor failures, makes an attempt to isolate different failures by using knowledge of the different effects such failures have on the system innovations. The method provides an optimum decision rule for failure detection and provides useful failure identification information for use in system reorganization subsequent to the detection of a failure. In addition, one can devise a number of simplifications of the technique and can study analytically the tradeoff between GLR complexity and GLR performance.

Consider the basic dynamical model (1)–(3). The following are 4 possible modifications of these equations that incorporate certain sudden system changes (see Willsky and Jones[28], [29] and Gustafson, Willsky and Wang[36] for physical motivation for these and other failure modes of the same general type)

Dynamics Jump

$$x(k+1) = \Phi(k)x(k) + B(k)u(k) + w(k) + \nu\delta_{k+1,\theta}. \tag{40}$$

Here ν is an unknown n-vector, θ is the unknown time of failure, and δ_{ij} is the Kronecker delta. Such a model can be used to model sudden shifts in bias states, as in the inertial problem studied in [15], [16].

Dynamic Step

$$x(k+1) = \Phi(k)x(k) + B(k)u(k) + w(k) + \nu\sigma_{k+1,\theta}. \tag{41}$$

Here σ_{ij} is the unit step

$$\sigma_{ij} = \begin{cases} 1 & i \geq j \\ 0 & i < j. \end{cases} \tag{42}$$

This model can be used to model certain actuator failures (compare to the Beard–Jones example in Section III; see equation (15)).

Sensor Jump

$$z(k) = Hx(k) + Ju(k) + v(k) + \nu\delta_{k,\theta}. \tag{43}$$

We can use this to model bad data points.

Sensor Step

$$z(k) = Hx(k) + Ju(k) + v(k) + \nu\sigma_{k,\theta}. \tag{44}$$

Sudden changes in sensor biases fit into this model.

By the linearity of the system (1)–(3) and the filter (4)–(10), one can determine the effect of each of the failure modes on the innovations. The general form is

$$\gamma(k) = G(k;\theta)\nu + \tilde{\gamma}(k) \tag{45}$$

where $\tilde{\gamma}(k)$ is the filter innovations if no failure occurs, and the matrix G can be precomputed (see [29], [38]). This matrix, which is different for each of the four cases (40)–(44), is called the *failure signature matrix* and provides us with an explicit description of how various failures propagate through the system and filter.

The full-blown GLR method involves the following: we assume we are looking for one of the four classes of failures and have computed the appropriate signature matrix. Given the residuals, we compute the maximum likelihood estimates of ν and θ, and, assuming that these estimates are correct, we compute the log-likelihood ratio for failure vs no failure (see Van Trees[41] for a general discussion of GLR methods and see Middleton and Esposito[53] for some detailed analysis of the utility of GLR). The implementation of the full GLR requires a linearly growing bank of matched filters, computing the best estimates of ν assuming a particular value of $\theta \in \{1, \ldots, k\}$.

A number of remarks can be made concerning the GLR system. We note that, as with other methods such as Buxbaum–Haddad or Chien, the inclusion of the variable θ to indicate our uncertainty as to the time of failure keeps the detection system sensitive to new data. However, it is the estimation of θ that causes the growing complexity problem. On the other hand, even if the full GLR is not implementable, it can serve as a benchmark for other schemes and can in fact be used as a starting point for the design of simpler systems. One simplification that eliminates the growing complexity is the restriction of the estimate of θ to a window

$$k - N \leq \theta \leq k - M$$

where the lower bound is included to limit complexity, and the upper bound is set by failure observability and false alarm considerations. Successful simulation runs with $N = M$ (i.e. when we don't optimize θ at all and have only one matched filter for $\hat{\nu}$) are reported by Willsky and Jones in [29]. We remark only that the price one pays for 'windowing' the estimate'of θ is in a reduction in the accuracy of the estimate of ν. For example, in the case of $N = M$, we often are able to detect failures extremely quickly, but if $\hat{\theta} = k - N$ is not the correct time of failure, the estimate of ν may be severely degraded, e.g. our estimate of the slope of a ramp changes as we change our estimate of the time at which it started. We note that the estimation of θ is similar to time-of-arrival estimation problems that arise in various applications, and refer the reader to Van Trees [44] for a general discussion of several techniques.

Also, we note that even if the physical system and filter are time-invariant, the GLR monitoring system is time-varying, as the failure signature G includes transient effects. In some cases one may be able to neglect these and utilize a simpler steady state signature. This is quite similar to Chien's use [24] of the steady-state effect of the failure on the residuals, and the criticisms of that approach, given in Section VI, apply here as well.

One can also simplify the implementation by either partially or completely specifying the failure magnitude ν. Constrained GLR (CGLR) is based on the assumption that

$$\nu = \alpha f_i \tag{46}$$

where α is an unknown scalar and f_i is one of r possible failure directions. This technique is described in [29]. If we completely specify ν

$$\nu = \nu_0 \tag{47}$$

we obtain the simplified GLR (SGLR) algorithm which is extremely simple to implement, as we have completely eliminated the need for the matched filters to estimate ν. The use of specified failure sizes is similar to that proposed by Chien [24], although in SGLR one can use the time-varying failure signature, which should aid in failure detection. As initial results for the detection of electrocardiogram arrhythmias indicate (see Gustafson *et al.* [36]), the estimation of ν is not nearly as important for detection as the matching of failure signatures. Also, by the use of several values of ν_0 (i.e. by implementing several parallel SGLR's), one can achieve a high level of failure isolation without a great deal of additional software complexity. In addition, one could consider a 'dual-mode' procedure in which SGLR is used for alarm and isolation, with full GLR used only afterward in order to estimate the magnitude of the failure.

We note that the use of assumed failure sizes combined with SPRT tests on the residuals, i.e. essentially SGLR with a two threshold test, was used by Deckert *et al.* in [59], [60] to design a sensor failure detection system for the NASA F8 Digital-Fly-By-Wire aircraft. In this system, the determination of the failure time θ was greatly simplified, since dual-redundant sensors were available. Thus, a discrepancy between the two like instruments was used to 'trigger' the initiation of SPRT's that used filter residuals, as well as other differences among the sensed quantities, to decide which instrument actually failed. This combination of part-voting, part-estimation/detection leads to major computational simplifications. We refer the reader to [59], [60] for details.

The various simplifications of GLR, as well as full GLR, are amenable to certain analysis, such as the calculation of P_F, P_D and, at least for SGLR, the expected time delay in detection. By performing such analyses, one can study in detail the tradeoff between complexity and performance. A methodology for such comparisons is presently being developed and is being applied to an aircraft failure detection problem. Initial results are reported by Chow *et al.*, in [38], and a description of a detailed methodology will be reported in the near future, (see Bueno *et al.* [43]). In addition to the calculation of P_F and P_D, the comparison methodology reported in [43] includes the computation of cross-detection probabilities—i.e. the probability of detecting a failure of type A when a failure of type B has occurred. Such information can be useful in designing failure isolation procedures and also in determining if failure detector A can be successfully utilized as an alarm for failures of type B. This can lead to substantial simplifications in a failure alarm system. Also, we refer the reader to [29], [36] and [38] for successful simulations of the GLR method.

Presently the GLR method is being extended to other failure modes, such as:

Hard-Over Actuator Failure

$$x(k+1) = \Phi(k)x(k) + [B + M\sigma_{k+1,\theta}]u(k) + w(k). \tag{48}$$

With this model we can take into account complete (or 'off') failures of certain actuators. For example an off failure of the ith actuator can be modeled by choosing M all zero except for the ith column, which is taken to be the negative of the ith column of B. The GLR detector for (48) is presently under development [38], [43], and we note that this model is more difficult than the others as the effect of the failure is modulated by the input values $u(k)$.

Increased Process Noise Failures

$$x(k+1) = \Phi(k)x(k) + B(k)u(k) + w(k) + \xi(k)\sigma_{k+1,\theta}. \tag{49}$$

Here ξ is additional white process noise.

Hard-Over Sensor Failures

$$z(k) = Hx(k) + Ju(k) + v(k) + [Mx(k) + su(k)]\sigma_{k,\theta}. \tag{50}$$

Here the failures are modulated by u and x, and a failure of the ith sensor is modeled by choosing the ith *rows* of M and S appropriately.

Added Sensor Noise Failures

$$z(k) = Hx(k) + Ju(k) + v(k) + \xi(k)\sigma_{k+1,\theta}. \tag{51}$$

The analysis of these failure modes is presently being performed [38], [43], and it is anticipated that SGLR algorithms will also be developed.

In addition to these failure modes, one can develop additional models along these lines for particular applications. In particular, we have developed several additional models similar to those described by equations (40)–(44) for our work on the detection and classification of arrhythmias in electrocardiograms. The results reported in [36] are rather striking, as in all the tests performed we observed no false alarms, detected all rhythm changes immediately, with no incorrect estimates of θ, and classified all rhythm changes correctly. These tests utilized the full GLR approach and have provided useful insight into the characteristics of the method. For example, the use of maximum likelihood estimates of ν and θ precludes the use of *a priori* statistics on these variables. In the ECG problem, one is quite interested in accurate estimates of ν, and one also can come up with reasonable *a priori* statistics on ν based on physical arguments. Thus, it may pay to incorporate such *a priori* statistics into the GLR system, and this can be done rather easily by proper initialization of the matched filters estimating ν. On the other hand, for the ECG problem one does not want to look for abrupt changes at one point in the record more than at another, and thus it does not make sense to include *a priori* statistics on θ. In fact, one can argue that inclusion of *a priori* failure information tends to discount the observed data in order to avoid false alarms, unless failures are extremely likely, and one should probably avoid the inclusion of such information unless one is especially worried about false alarms. However, if one wishes to use such data, one can utilize the interpretation of the likelihood ratios as ratios of conditional probabilities of failure times in order to determine the appropriate modification of GLR [29].

Finally, we note that the GLR system provides extremely useful information for system compensation subsequent to the detection of a failure. For example, one can utilize the

GLR-produced estimates of ν and θ to determine an optimal update procedure for the filter estimate and covariance [29]. Once this update has been performed, the GLR system can be used to detect further failures, thus allowing the detection of multiple events. We refer the reader to [29], [38] for further discussions of the use of GLR-produced information in the design of failure compensation systems.

VIII. CONCLUSIONS

In this paper we have discussed a number of the issues involved in the design of failure detection systems. We have also reviewed a variety of existing failure detection methods and have discussed their characteristics and design tradeoffs. The failure detection problem is an extremely complex one, and the choice of an appropriate design depends heavily on the particular application. Issues such as available computational facilities and level of hardware redundancy enter in a crucial way in the design decision. For example, as we have mentioned, the use of a sophisticated failure detection–compensation system may allow one to reduce the level of hardware redundancy without much of a loss in overall system reliability.

Let us say a few words about the relative merits of the various failure detection methods described. Of course the most reliable means of failure detection is straight voting among like instruments. In this case one never runs into errors introduced by comparing the outputs of dissimilar devices. However, one pays the price of hardware redundancy in order to implement a voting scheme. In addition, voting has its built-in limitations. Like instruments may not be exactly alike, and one may have to use other instruments to compensate for these discrepancies. For example, redundant accelerometers may be mounted in different places in an aircraft and one must use rate gyro information to subtract out rotational effects. In addition, voting cannot be used to detect failures that affect both instruments in the same way, e.g. common power supply failures, common thermal effects, etc., and once a single instrument in a triad has failed, we can no longer perform failure detection by voting. Finally, voting techniques may have difficulties in detecting subtle degradations in instrument behavior.

The remaining techniques discussed in this paper are aimed at some or all of the above drawbacks of voting techniques. Of course, they all have their own limitations. In the first place, the use of indirect information necessarily leads to higher false alarm and missed detection probabilities, and the price one pays in order to keep these values small is computational complexity. Much of this complexity arises from the fact that one must detect *when* a change occurs. If one uses a dual-redundant system—as proposed by Deckert *et al.*, in [59], [60]—one can use a direct comparison to trigger a more sophisticated identification technique. This can greatly reduce the computational burden. For example, the growing bank of filters for the multiple hypothesis and GLR methods can be eliminated. This tradeoff between hardware and software complexity is certainly deserving of further study.

There are a number of key issues that must be considered in comparing the merits of the various failure detection methods. Among these, the following fall out as being of obvious importance

Types of failure modes that can be considered
Complexity in implementation
Performance, as measured by false alarms, delays in detection, etc.
Robustness in the presence of modelling errors.

As discussed by Beard [4] and Jones [5], the detector filters described in Section III can be used to detect an extremely wide variety of system failures. In addition, Jones has considered the problem of the distinguishability of failure modes and has provided an analytical framework in which one can answer this question. This is essentially an observability issue, and Jones' work represents the first major effort in this area. As mentioned earlier, a minor drawback with this method is its suboptimality in producing a state estimate, while a major limitation is its applicability only to time-invariant systems.

Of all of the methods, the multiple hypothesis techniques of Section V are the most complex and, if implemented in full, will yield the best performance for the widest class of failures. This is true, since all of the failure models considered in this paper can be modelled as abrupt switches among several hypothesized models. As the results of Athans *et al.* [35] and Gustafson *et al.* [36] indicate, these methods are particularly well-suited to the detection of changes that manifest themselves via switches in system dynamics or parameters. The potential of multi-filter methods is sufficiently great that serious consideration should be given to the development of useful approximate methods, such as those used by Gustafson *et al.* [36], Willsky *et al.* [15], [16], and Newbold and Ho [52].

As mentioned in Section VI, jump process formulations are particularly natural for the failure detection problem. The results of Chien [24] indicate the power of this framework, but the tractability and performance of methods such as that of Davis [22] for complex systems remains to be demonstrated.

The innovations- based detection systems offer several advantages. In the first place, they can be adapted to utilize the residuals of an existing filter with relative ease, and a wide range of procedures—from the simple statistical tests of Mehra and Peschon [26] to the more complex GLR method of Willsky *et al.* [28], [29], [38], [43]—are available, offering various tradeoffs between performance and complexity. In addition, since we implement a Kalman filter based on no failures, we suffer no performance degradation prior to detection, unlike, for example, the Beard–Jones approach. The GLR method is particularly amenable to detailed analysis (see [38], [43], [65]) of performance and allows the study of failure mode distinguishability in an informational sense, much as in the work of Jones [5]. As discussed in Section VII, GLR performance is quite outstanding for failure that can be modelled as additive effects. This, combined with the range of implementations available and its analytical tractability, make it appealing. The method is being extended to the case of parametric failures, as in (48)–(51). but in these cases the multi-filter methods of Section V should prove to be theoretically superior. The advantage of GLR in this case may be in its implementability.

Finally, we note that the issue of robustness of these methods has not been discussed in this paper. This remains an extremely important open question that must be answered before one can completely assess the various methods. It is fair to say that the more complex the system model and the more model-dependent the technique, the more danger there is that one will run into sensitivity problems. This is not to say that sophisticated techniques cannot be used in practice, but rather it does indicate that in using such techniques one should take care in defining the dynamic model to be used and, as a rule, should use as simple a model as possible. The results of Gustafson *et al.* [36] using GLR and multifilter methods on real ECG data and the work Deckert *et al.* [59], [60] using a SPRT-based test for failure detection for the F8 aircraft indicate that such methods are most certainly feasible in practice.

The development of failure detection methods is still a relatively new subject. At this time most of the work has been at a theoretical level with only a few real applications of techniques [6]–[9], [13], [31], [36], [59]–[60]. Much work is yet to be done in the development of implementable systems complete with a variety of design tradeoffs. Work is needed in the development of efficient techniques for failure compensation and system reorganization. In addition, as mentioned above, there is a great need for the analysis of the robustness of various failure detection systems in the presence of variations in system parameters and in the presence of modeling errors and system nonlinearities. These and other issues, such as the incorporation of fault-tolerant computer concepts into an overall reliable design methodology, see Deyst [40], await investigation in the future.

Acknowledgements—The author is in the debt of many colleagues for their comments during numerous discussions on the subject of failure detection. In particular, special thanks must go to Dr. R. C. Montgomery, Dr. A. K. Caglayan, Dr. D. B. Price, and Mr. J. L. Elliott of NASA Langley Research Center, Mr. B. F. Doolin of NASA Ames Research Center, Dr. H. L. Jones, Dr. B. S. Crawford, Dr. T. H. Kerr, Mr. J. H. Fagan, and Mr. W. O'Halloran of The Analytic Sciences Corporation, Dr. J. J. Deyst, Dr. D. E. Gustafson, Mr. J. C. Deckert, Dr. M. Desai, Dr. J. L. Speyer, and Mr. J.-Y. Wang of The Charles Stark Draper Laboratory, Dr. K. P. Dunn of the M.I.T. Lincoln Laboratory, and Prof. M. Athans, Dr. S. B. Gershwin, Mr. E. Chow, and Mr. R. Bueno of M.I.T. The author also wishes to thank the reviewers and editors for their extremely helpful comments.

REFERENCES

[1] S. L. FAGIN: Recursive linear regression theory, optimal filter theory, and error analyses of optimal systems. *IEEE Int. Conv. Record*, 216–240 March (1964).

[2] T. J. TARN and J. ZABORSZKY: A practical nondiverging filter. *AIAA Jl* **8**, 1127–1133 (1970).

[3] A. H. JAZWINSKI: Limited memory optimal filtering. *IEEE Trans. Aut. Control* **13**, 558–563 (1968).

[4] R. V. BEARD: *Failure Accommodation in Linear Systems Through Self-Reorganization*, Rept. MVT-71-1, Man Vehicle Laboratory, Cambridge, Massachusetts, February (1971).

[5] H. L. JONES: Failure Detection in linear systems. Ph.D. Thesis, Dept. of Aeronautics and Astronautics, M.I.T., Cambridge, Mass., Sept. (1973).

[6] J. GILMORE and R. MCKERN: A redundant strapdown inertial system mechanization—SIRU presented at the *AIAA Guidance, Control and Flight Mechanics Conf.* Santa Barbara, CA, Aug., 17–19 (1970).

[7] A. J. PEJSA: *Optimum orientation and accuracy of redundant sensor arrays*. Honeywell Aerospace Div., Minneapolis, MN (1971).

[8] J. T. EPHGRAVE: *Redundant Adaptive Strapdown Navigation Systems*, Aerospace Report No. TOR-0066 (5306)-10, The Aerospace Corp., Oct., 31 (1969).

[9] R. B. BROEN: A nonlinear voter-estimator for redundant systems. *Proc. 1974 IEEE Conf. on Decision and Control*, Phoenix, Arizona pp. 743–748.

[10] M. ATHANS and D. WILLNER: A practical scheme for adaptive aircraft flight control systems. *Symp. on Parameter Estimation Techniques and Applications in Aircraft Flight Testing*, NASA Flt. Res. Ctr., Edwards AFB, April 24, 25 (1973).

[11] D. G. LAINIOTIS: Joint detection, estimation, and system identification. *Inform. Control* **19**, 75–92, Aug. (1971).

[12] R. C. MONTGOMERY and A. K. CAGLAYAN: A self-reorganizing digital flight control system for aircraft. *AIAA 12th Aerospace Sciences Meeting*, Washington, D.C., Jan. 30–Feb. 1 (1974).

[13] R. C. MONTGOMERY and D. B. PRICE: Management of analytical redundancy in digital flight control systems for aircraft. *AIAA Mechanics and Control of Flight Conference*, Anaheim, CA, August 5–9 (1974).

[14] P. J. BUXBAUM and R. A. HADDAD: Recursive optimal estimation for a class of nongaussian processes. *Proc. Symp. on Computer Processing in Communications*, Polytech Inst. of Brooklyn, April 8–10 (1969).

[15] A. S. WILLSKY, J. J. DEYST and B. S. CRAWFORD: Adaptive filtering and self-test methods for failure detection and compensation. *Proc. of the 1974 JACC*, Austin, Texas, June 19–21, 1974.

[16] A. S. WILLSKY, J. J. DEYST, Jr. and B. S. CRAWFORD: Two self-test methods applied to an inertial system problem. *J. Spacecr. Rockets* **12**, No. 7, 434–437 July (1975).

[17] B. D. PIERCE and D. D. SWORDER: Bayes and Minimax controllers for a linear system with stochastic jump parameters. *IEEE Trans. Aut. Control* **AC-16**, No. 4, 300–307 Aug. (1971).

[18] D. D. SWORDER: Bayes' controllers with memory for a linear system with jump parameters. *IEEE Trans. Aut. Control* **AC-17**, 118–121 Feb. (1972).

[19] D. D. SWORDER and V. G. ROBINSON: Feedback regulators for jump parameter systems with state and control dependent transition rates. *IEEE Trans. Aut. Control* **AC-18**, No. 4, 355–360 Aug. (1973).

[20] V. G. ROBINSON and D. D. SWORDER: A computational algorithm for design of regulators for linear jump parameter systems. *IEEE Trans. Aut. Control* **AC-19**, 47–49 Feb. (1974).

[21] R. S. RATNER and D. G. LUENBERGER: Performance-adaptive renewal policies for linear systems. *IEEE Trans. Aut. Control* **AC-14**, No. 4, 344–351 Aug. (1969).

[22] M. H. A. DAVIS: The application of nonlinear filtering to fault detection in linear systems. *IEEE Trans. Aut. Control* **AC-20**, No. 2, 257–259 April (1975).

[23] T. P. MCGARTY: State estimation with faulty measurements: an application of Bayesian outlier rejection. *Proc. Fifth Symposium on Nonlinear Estimation and Its Applications*, San Diego, CA, Sept. (1974).

[24] T. T. CHIEN: *An Adaptive Technique for a Redundant-Sensor Navigation System*. Rept. T-560, Draper Labs., Cambridge, MA, February (1972).

[25] T. H. KERR: A two ellipsoid overlap test for real time failure detection and isolation by confidence regions. *Pittsburgh Conf. on Modelling and Simulation*, April 24–26 (1974).

[26] R. K. MEHRA and J. PESCHON: An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica* **7**, 637–640.

[27] H. M. MERRILL: Bad data suppression in state estimation, with applications to problems in power. Ph.D. Thesis, Dept. of Electrical Engineering, M.I.T., Cambridge, Mass., June 1972.

[28] A. S. WILLSKY and H. L. JONES: A generalized likelihood ratio approach to state estimation in linear systems subject to abrupt changes. *Proc. of 1974 IEEE Conf. on Decision and Control* Phoenix, Arizona, Nov. (1974).

[29] A. S. WILLSKY and H. L. JONES: A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Trans. Aut. Control* **AC-21**, 108–112 Feb. (1976).

[30] R. J. MCAULAY and E. DENLINGER: A decision-directed adaptive tracker. *IEEE Trans. Aero. and Elec. Sys.* **AES-9**, 229–236 March (1973).

[31] J. J. DEYST and J. C. DECKERT: RCS jet failure identification for the space shuttle. *Proc. IFAC* 75, Cambridge, MA, August (1975).

[32] P. SANYAL and C. N. SHEN: Bayes' decision rule for rapid detection and adaptive estimation scheme with space applications. *IEEE Trans. Aut. Control* **AC-19**, 228–231 June (1974).

[33] A. H. JAZWINSKI: *Stochastic Processes and Filtering Theory*. Academic Press, New York (1970).

[34] R. N. CLARK, D. C. FOSTH and V. M. WALTON: Detecting instrument malfunctions in control systems. *IEEE Trnas. Aerospace Electronic Systems* **AES-11**, No. 4, 465–473 July (1975).

[35] M. ATHANS, K.-P. DUNN, C. S. GREENE, W. H. LEE, N. R. SANDELL JR., I. SEGALL and A. S. WILLSKY: The stochastic control of the F-8C aircraft using the multiple model adaptive control (MMAC) method. *Proc. 1975 IEEE Conf. on Decision and Control*, Houston, Texas, December (1975).

[36] D. E. GUSTAFSON, A. S. WILLSKY and J.-Y. WANG: *Final Report: Cardiac Arrhythmia Detection and Classification Through Signal Analysis*. The Charles Stark Draper Laboratory, Cambridge, Mass., Rept. No. R-920 July (1975).

[37] D. D. SWORDER: Feedback control of a class of linear systems with jump parameters. *IEEE Trans. Aut. Control* **AC-14**, No. 1, 9–14, Feb. (1969).

[38] E. CHOW, K.-P. DUNN and A. S. WILLSKY: Research status report to NASA Langley research center: a dual-mode generalized likelihood ratio approach to self-reorganizing digital flight control system design.

M.I.T. Electronic Systems Laboratory, Cambridge, MA, April (1975).
[39] R. BOEL, R. VARAIYA and E. WONG: Martingales on jump processes I: representation results, and II: applications, *SIAM J. Control* **13**, 999–1061 August (1975).
[40] J. J. DEYST: *A design approach to highly reliable flight control systems*. The Charles Stark Draper Laboratory, Inc., Cambridge, MA (1975).
[41] H. L. VAN TREES: *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. Wiley, New York (1971).
[42] ÅKE WERNERSSON: On Bayesian estimators for discrete time linear systems with Markovian parameters. *Sixth Symp. on Nonlinear Estimation and Its Applications*, San Diego, CA, Sept. 15–17 (1975).
[43] R. BUENO, E. CHOW, S. B. GERSHWIN and A. S. WILLSKY: Research status report to NASA Langley research center: *a dual-mode generalized likelihood ratio approach to self-reorganizing digital flight control system design*. Rept. ESL-IR-642, M.I.T. Electronic Systems Laboratory, Cambridge, MA Nov. (1975).
[44] H. L. VAN TREES: *Detection, Estimation, and Modulation Theory, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Wiley, New York (1971).
[45] B. F. DOOLIN: Reliability issues for future aircraft. MIT–NASA/Ames workshop on system reliability issues for future aircraft. M.I.T. Cambridge, MA. August 18–20 (1975).
[46] L. TAYLOR: Active control aircraft problems. MIT–NASA/Ames workshop on system reliability issues for future aircraft. M.I.T., Cambridge, MA, August 18–20 (1975).
[47] G. MEYER and L. CICOLANI: A formal structure for advanced automatic flight-control systems. NASA TN D-7940, May (1975).
[48] CHR. ZYWEITZ and B. SCHNEIDER: *Computer Application on ECG and VCG Analysis*, North Holland (1973).
[49] A. SEGALL: A Martingale approach to modeling, estimation and detection of jump processes. Ph.D. Dissertation, Stanford University, Stanford, CA, August (1973).
[50] D. L. SNYDER: *Random Point Processes*. Wiley, New York (1975).
[51] M. ATHANS, R. H. WHITING and M. GRUBER: A suboptimal estimation algorithm with probabilistic editing for false measurements with applications to target tracking with wake phenomena. *Sixth Symp. on Nonlinear Estimation and Its Applications*, San Diego, CA, Sept. 15–17 (1975).
[52] P. M. NEWBOLD and Y.-C. HO: Detection of changes in the characteristics of a Gauss–Markov process. *IEEE Trans. Aerospace Elec. Sys.* **AES-4**, No. 5, 707–718 Sept. (1968).
[53] D. MIDDLETON and R. ESPOSITO: Simultaneous optimum detection and estimation of signals in noise. *IEEE Trans. Inf. Th.*, Vol. IT-14, No. 3, 434–444 May (1968).
[54] N. E. NAHI: Optimal recursive estimation with uncertain observation. *IEEE Trans. Inf. Th.* **IT-15**, No. 4, 457–462 July (1969).
[55] J. J. BELLETRUTTI and A. G. J. MACFARLANE: Characteristic loci techniques in multivariable-control-system design. *Proc. IEE* **118**, No. 9, 1291–1297 Sept. (1971).
[56] O. A. SOLHEIM: Some integrity problems in optimal control systems. *AGARD Conference Proceedings*, No. 137.
[57] P. K. WONG and M. ATHANS: Closed-loop structural stability for linear-quadratic optimal systems. 1976 *IEEE Conf. on Decision and Control*, Clearwater, FL Dec. (1976).
[58] M. G. SAFONOV and M. ATHANS: Gain and phase margin for multiloop LQG regulators. 1976 *IEEE Conf. on Decision and Control*, Clearwater, FL Dec. (1976).
[59] J. C. DECKERT, M. N. DESAI, J. J. DEYST and A. S. WILLSKY: Dual redundant sensor FDI techniques applied to the NASA F8C DFBW aircraft. 1976 *AIAA Guidance and Control Conf.* San Diego, CA, August (1976).
[60] J. C. DECKERT, M. N. DESAI, J. J. DEYST and A. S. WILLSKY: A dual-redundant sensor failure detection algorithm for the F8 aircraft. 1975 *IEEE Conf. Decision and Control*; submitted to *IEEE Trans. Aut. Control*.
[61] J. C. HANCOCK and P. A. WINTZ: *Signal Detection Theory*. McGraw-Hill, New York (1966).
[62] F. C. SCHWEPPE and E. J. HANDSCHIN: Static state estimation in electric power systems. *Proc. IEEE* **62**, No. 7, 972–982 July (1974).
[63] D. W. PETERSON: Hypothesis, estimation, and validation of dynamic social models—energy demand modeling. Ph.D. Dissertation, Dept. of Electrical Engineering, M.I.T., June (1975).
[64] E. HANDSCHIN, F. C. SCHWEPPE, J. KOHLAS and A. FIECHTER: Bad Data analysis for power system state estimation. *Proc. of IEEE Summer Power Meeting*, Anaheim, CA (1974).
[65] E. Y. CHOW: Analytical studies of the generalized likelihood ratio technique for failure detection. S.M. Thesis, Dept. of Elec. Eng. and Comp. Sci., M.I.T., Feb. (1976).