A Bayesian Approach to Diagnosis and Prognosis Using Built-In Test

John W. Sheppard, *Senior Member, IEEE,* and Mark A. Kaufman, *Member, IEEE*

***Abstract*—Accounting for the effects of test uncertainty is a significant problem in test and diagnosis, especially within the context of built-in test. Of interest here, how does one assess the level of uncertainty and then utilize that assessment to *improve* diagnostics? One approach, based on measurement science, is to treat the probability of a false indication [e.g., built-in-test (BIT) false alarm or missed detection] as the measure of uncertainty. Given the ability to determine such probabilities, a Bayesian approach to diagnosis, and by extension, prognosis suggests itself. In the following, we present a mathematical derivation for false indication and apply it to the specification of Bayesian diagnosis. We draw from measurement science, reliability theory, signal detection theory, and Bayesian decision theory to provide an end-to-end probabilistic treatment of the fault diagnosis and prognosis problem.**

***Index Terms*—Bayesian inference, built-in test (BIT), diagnosis, false indication, measurement uncertainty, prognosis.**

I. Introduction

TESTING at any level (e.g., production, field, or on-board) is about obtaining realistic results, or indications about a unit under test (UUT). Courses of action are selected based on test results. Ideally each test result would exactly reflect the condition of the UUT. Given the fact that there is no such thing as an ideal test, effective diagnosis requires understanding the relative probability of a false indication. But what is a false indication? How can false indications be predicted? How can one use such prediction to aid diagnosis rather than hinder diagnosis? These are the questions that concern us in this paper. Further, given the focus of the IEEE to standardize metrics for testability and diagnosability, answers to these questions are particularly relevant [17].

The ability to predict and track false indications in test and diagnosis is the long-time bane of embedded diagnostics (such as built-in test (BIT). Unfortunately, it is generally accepted that, even if we are able to predict false indications, verifying these predictions with field data is problematic. At best, we might be able to measure false-removal rates, cannot duplicate rates, or retest-OK rates, but these rates can be attributed to multiple causes that may or may not include false indications.

MIL-STD-2165 (now MIL-HDBK 2165A) [25] defines a false alarm as "a fault indicated by built-in test (BIT) or other monitoring circuitry where no fault exists." The former MIL-STD-1309C [24] defines false alarm the same way, limiting the definition to BIT. The IEEE standard dictionary is in agreement, defining false alarm more generally as "an indicated fault where no fault exists" [15]. A missed detection can be defined similarly as "an indication of no fault where a fault exists." Combining, we can then define a false indication simply as "an erroneous test result." This definition applies to all levels of test.

This paper does not segregate BIT from testing in general. BIT, after all, is a test; however, BIT does have some unique properties. Specifically, BIT is an intimate part of the UUT and is subjected to the same environment. BIT and the UUT may suffer degradation from the same causes. BIT is also toleranced closer to operational requirements than other levels of test. However, BIT results are still used to determine a course of action. These factors were considered during the development of the approach of this paper, and it was decided that BIT fits within the general model of test. In fact, we would further generalize our definition of BIT to include any observation recorded by the system about the performance of the system that can be used for test and diagnosis.

Work from the measurement science community provides a means for predicting uncertainty in measurement that can be used as the basis of predicting false indications. As long as we rely on calibration laboratories to validate these uncertainties and we focus on applying the uncertainties to the test results, we can mitigate the impact of our limitations in verifying false alarms. Specifically, we look to apply measurement uncertainty combined with probability of failure and probability of observing failure as components in an overall model for predicting false indications. We then use these probabilities in the context of a Bayesian belief network to perform diagnosis and, ultimately, prognosis. We recognize that there are other possible causes for false indications, including environmental factors, human error, and systems operating near their functional limits; however, we limit the discussion to measurement uncertainty as a starting point in our exploration.

The remainder of this paper is organized as follows. Section II provides our assumptions in addressing the false indication problem and Section III provides basic background material from measurement science. Sections IV–VII provide the main tools used in this paper—predicting probability of false indication, predicting probability of failure, and predicting instrument uncertainty, respectively. In Section VIII, we provide the foundations required for using a Bayesian approach by introducing the fundamentals of Bayes decision theory and detection theory, respectively. We then provide the main result in Section IX, showing how to combine these predictions in Bayesian diagnosis, a numerical example illustrating the application of the

Manuscript received October 15, 2003; revised March 7, 2005.

J. W. Sheppard is with ARINC Incorporated, Annapolis, MD 12401 USA (e-mail: jsheppar@arinc.com).

M. A. Kaufman is with the Naval Surface Warfare Center, Corona Division, Corona, CA 92878 USA (e-mail: mark.kaufman@navy.mil).

Digital Object Identifier 10.1109/TIM.2005.847351

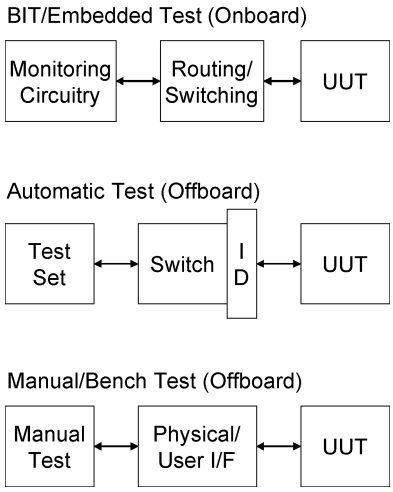

Fig. 1. Notional test architecture.

approach in Section X, followed by an interesting extension to handle prognosis in Section XI. We provide concluding remarks in Section XII.

II. Assumptions

As mentioned above, a false alarm has been defined traditionally as "an indication of a fault where no fault exists." Interpreting this in the context of testing a UUT, this has typically been limited to faults in the UUT alone. In the following, we will argue that, to facilitate creating a predictive model for false alarms, the definition must be applied to the entire test environment. Here, the test environment includes the test equipment, the hardware or software required to support testing (including switching matrices or interface devices as appropriate), as well as the UUT (Fig. 1). Further, the treatment for the predictive model will be extended to cover missed detection as well, thus, providing a model for overall false indication.

In this paper, we will consider potential sources of error in the test process. Under ideal conditions, with "accurate" measurement devices and full information about the UUT, diagnostic errors will still occur and cannot be eliminated completely. It is important to understand that some complicated elements of diagnosis involve chasing these errors [44]. We assume that the most likely sources of error are related to the following:

1) human error;
2) environmental conditions;
3) test equipment uncertainty;
   a) instrument uncertainty;
   b) power source uncertainty;
4) test design;
   a) transient effects;
   b) loading effects;
   c) threshold settings.

Human error is not something that can be predicted since it is tied to several factors including education level, training, fatigue, and other "human limitations." Therefore, we are not incorporating this factor into the proposed model. Further, while environmental effects account for a large percentage of error sources, we see no "generic" approach to modeling environmental error. Rather, we note that such error sources are determined *post facto*, and the results of such *post-mortem* analysis factored directly into test policies and procedures by incorporating additional environmental controls. We consider this issue one worthy of further research.

Test equipment uncertainty is believed to be a significant predictable source of uncertainty in our model. Test equipment in this case includes all of those elements required to perform a test, including BIT. We differentiate between the power source uncertainty (i.e., the probability that power levels applied as input to the UUT are at the intended levels) and instrument uncertainty (i.e., the probability that measurements and stimuli are within some specified tolerance). When considering power source uncertainty, we observe that most test procedures follow a sequence of events such as the following.

1) The UUT is evaluated to ensure it is safe to have power applied (i.e., safe-to-turn-on test).[1]
2) A power system integrity check is performed to determine if the power source indeed provides power at the required levels.[2]
3) The UUT is then tested.

Utilizing this sequence, we see that, in step "2," the power system itself is treated as a UUT and must adhere to the same type of prediction as we are proposing here. Given that, we can assume that if the power system integrity check passes (with some confidence level associated with it), then either we can factor power uncertainty out of the model, or the resulting confidence can be used as input to an overall probability of error. We will assume power uncertainty is negligible due to the fact the metrology process usually imposes a 4:1 accuracy ratio between the measurement device and the power source [3].

For instrument uncertainty, the calibration process is designed to 1) minimize error in the instrument, as well as 2) determine the level of uncertainty (by reading as opposed to by scale) of the instrument.[3] Thus, assuming the instrument is calibrated, probability of error can be determined based on this specified uncertainty. Assuming that the causes for power source uncertainty and instrument uncertainty are independent and the level of uncertainty due to power source integrity is *not* negligible, the resulting probabilities can simply be multiplied together.

The issue of test design (ignoring the effects of human error due to inadequate documentation, faulty simulation, inappropriate assumptions, etc.) can be limited to considering the three factors above—transient effects, loading effects, and threshold settings. Transient effects correspond to external sources of deviation due to, for example, changes in mode settings or environmental effects. As discussed earlier, we believe that the ability to devise general models of environmental uncertainty is limited

[1]Of course, with BIT, power is already applied since BIT is an integral part of the UUT.

[2]In the context of BIT, it is typical that power is assumed to be good; therefore, it could be a potential source for BIT false indication. This is especially the case given ac power, which is frequently "dirty" unless properly conditioned.

[3]BIT/BITE should also be regarded as "instrumentation" and should undergo the same calibration as off-board instrumentation. Otherwise, error introduced by BIT cannot be controlled or predicted.

at this time; therefore, they will be deferred to a later point for incorporation into the model.[4]

We also assume that, in the process of designing tests, settling time corresponding to mode changes (e.g., in the UUT, test set, or switching matrix) are set such that the corresponding transient effects do not impact the final measurement. In other words, we are assuming measurements are taken after the test environment has reached a stable state.

When considering loading effects, we first partition the test environment as in Fig. 1. As seen in the figure, embedded and offboard (automatic and manual) test environments fit the same general architecture. Specifically, the test set corresponds to the collection of instruments and power sources/supplies used to test the UUT. The switching matrix provides the means for routing the signals between the UUT and the appropriate instruments in the test set. The interface device (ID) provides the mechanical means to connect the UUT to the switching matrix. Finally, the UUT is what we are testing.

Each of the components in the test environment have their respective loading effects on the entire test environment. We are assuming that loading effects due to the test set and the switching matrix, up to the connectors on the test set side of the ID are factored out as result of metrology specifications relative to the test equipment. Specifically, the measurement uncertainty associated with a particular measurement devices, is specified at the interface rather than at the device itself. Thus, we only need to consider loading effects arising from the pathways between the ID and the point of measurement on the UUT. BIT may have a very short and simple connection to the UUT, but there is one nonetheless.

Since transient effects were factored out of the model, we will assume that any loading effects will result in a simple shift in the measurement (i.e., they will provide an offset). As with any other factor, there will be some level of uncertainty associated with this offset. We assume that the resulting probability of error is independent of error due to measurement, and that we can then include probability of error due to loading as another multiplier in the model.

Finally, and perhaps most significantly, we consider the effects on setting the PASS/FAIL threshold. The characteristics of the underlying behavior distribution will have been fully determined by the previous factors. The effect of setting the threshold is one of optimizing risk, where frequently this is further subdivided into consumer and producer risk. Consumer risk refers to the potential loss to the consumer of a product due to error. Thus, this corresponds to the effects of missed detection because a consumer receiving and using a faulty item will have a direct negative impact on the consumer's ability to obtain full use of the product. Producer risk refers to the potential loss to the producer of a product due to error. This corresponds to the effects of false alarm because a producer repairing a fault-free item incurs unnecessary cost (e.g., lower yield, higher manufacturing cost, higher maintenance cost).

[4]It is also possible to regard human error and environmental effects as anomalous inputs; therefore, under the assumption that we apply the definition of false alarm to the entire test environment, one could argue that errors due to these causes would not be false alarms but would be indicative of their cause.

To summarize, our model for predicting probability of false indication depends upon the following factors:

1) instrument uncertainty following calibration;
2) loading effect offset and associated uncertainty;
3) threshold settings based on consumer-producer risk tradeoff.

III. Measurement Theory

A prominent issue in testing is the impact of *precision* and *accuracy* of test resources on the certainty in the resulting test outcome. Formalizing the impact of precision and accuracy on test confidence comes from work in measurement theory. Given the need to formalize these factors, we can consider each of them in terms of probability distributions. The precision of a resource characterizes the amount of "scatter" one can expect when repeated measurements are made of the same signal. Typically, precision is applied to the measurement process or resource rather than the measurement itself. Thus, the precision of a resource can be characterized by considering a probability distribution of measurements. A precise resource would yield a narrow distribution of measured values (i.e., a low variance), while an imprecise resource would yield a wide distribution of measured values (i.e., a high variance).

On the other hand, the accuracy of a measurement corresponds to the level of deviation from some *reference value*. As with precision, accuracy can be modeled as a probability distribution. Typically, accuracy is applied to the measured value rather than the process or instrument used in taking the measurement. Frequently, it is determined through taking several independent measurements and taking the mean deviation from the reference value.

Statistically, inaccuracy is evaluated in terms of two potential sources—referred to as random error and systematic error. Systematic error is also called *bias*. Many consider accuracy only in terms of random error in which bias is ignored or calibrated out (in which case the estimate becomes an unbiased estimate of error). Others consider total error in which the bias is included.

To determine the appropriate tolerances for a particular test, and thereby determine relevant test outcomes, one must consider the precision of the required resources. The precision will be used to determine the associated bias of the instrument. To determine the precision, typically, several independent measurements are taken for that instrument under known conditions and the resulting distribution determined. Then the width of the distribution is determined by using, for example, variance (a statistical measure defined as the average deviation from the mean of the distribution).

Next, measurement error is considered by examining the distribution characterizing accuracy. Given the system to be tested and a measurement to be made, a distribution of "nominal values" for that measurement can be determined. From this distribution, PASS/FAIL criteria are established (in the simplest case) based on the probability of a measurement occurring within some set of defined limits applied to that measurement.

Generally the PASS/FAIL criteria are determined by considering expected values for a fault-free unit. "Significant" deviation from these expected values results in the FAIL outcome for

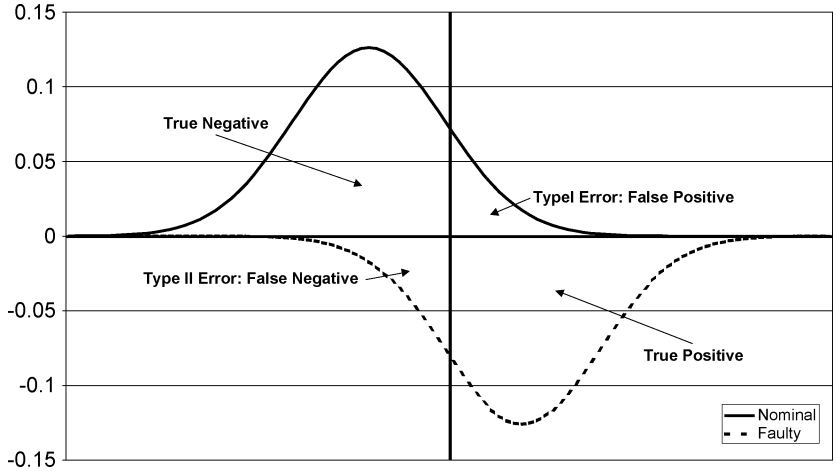

Fig. 2. Distributional sources of Type I and Type II errors, respectively.

that test. The limits define what is meant by "significant." Ideally, conditions are considered in which a similar distribution is determined in the presence of faulty conditions, and the limits are set such that maximum separation between the distributions (in terms of probability density) is obtained. During testing, a measurement value is typically mapped into a discrete outcome determined based upon which side of these limits the measurement falls.

Once the measured value is mapped into a discrete outcome, interpreting the test results becomes problematic. Specifically, when considering a single measurement, the probability of that measurement being within the set limits will be $\Pr(x)$, and the probability of it being out of the limits will be $1 - \Pr(x)$. (Actually, these are conditional probabilities since the actual probability depends on whether or not the unit has failed, as we will see below.)

In determining the base confidence (e.g., either the expected confidence or the maximum confidence) in the outcome, we want to consider both the probability of observing a PASS outcome given the test should pass and the probability of observing a FAIL outcome given the test should fail, i.e., $\Pr(o(P)|P)$ and $\Pr(o(F)|F)$ respectively. Problems occur when a measurement for a good unit is out of tolerance (referred to as a "Type I error") or a measurement for a bad unit is in tolerance (referred to as a "Type II error"). Given that we consider a test outcome of FAIL to be a "positive" indication of a problem, Type I errors correspond to false positives (i.e., false alarms) while Type II errors correspond to false negatives (i.e., missed detections). These two types of errors are depicted in Fig. 2.

Current BIT captures performance data about a particular subsystem or unit but frequently makes crisp decisions in terms of the expected cause (i.e., fault) generating the BIT indication. Given current trends for BIT to have high false alarm rates, it is particularly important that both Type I and Type II errors be accounted for within the underlying diagnostics. Specifically, ignoring these types of errors has the potential to generate inaccurate diagnosis with no guidance for how to detect or compensate for such inaccuracies. By directly modeling BIT uncertainty, the approach proposed in this paper propagates the BIT uncertainty during diagnostic inference to provide a more accurate prediction of the system or unit state of health.

IV. Determining Probability of False Indication

Based on the fact Type I and Type II errors occur in practice, we want to know the rate at which we will be faced with such errors. Formally, a false indication occurs whenever a test result is inconsistent with the underlying state of the unit being tested. When determining the probability of a false indication occurring, we must consider these joint effects. Consider the situation where we only have a single test, and it is possible for that test to either PASS or FAIL. Further, assume that we consider the possibility of a fault truly being present or not being present. Thus we need to consider four combinations as follows.

1) True pass: We observe the test passing $(o(P))$ and there is no fault (P).
2) Missed detection: We observe the test passing $(o(P))$ and there is a fault (F).
3) False alarm: We observe the test failing $(o(F))$ and there is no fault (P).
4) True fail: We observe the test failing $(o(F))$ and there is a fault (F).

Now we want to consider the probabilities of these joint events occurring, namely $\Pr(o(P), P)$, $\Pr(o(P), F)$, $\Pr(o(F), P)$, and $\Pr(o(F), F)$, respectively. Note that these four situations cover all possible combinations, so we can define the probability of false indication, $\Pr(FI^u)$, as follows:

$$\Pr(FI^u) = \frac{\Pr(o(P), F) + \Pr(o(F), P)}{\Pr(o(P), P) + \Pr(o(F), F) + \Pr(o(P), F) + \Pr(o(F), P)}. \quad (1)$$

Here, FI^u denotes fault indication in an "unconditional sense" where we consider only unconditional probabilities.

Note that, from a diagnostic perspective, this definition is not particularly useful. This is because the process of performing diagnosis corresponds to observing test results and then inferring diagnoses (i.e., possible faults) within the system. Thus, in reality, we are interested in the conditional probabilities of the

state of the unit given the test results. Recall the definition of conditional probability that says

$$\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)}.$$

Using this definition, we can rewrite (1) as shown in (2) at the bottom of the page. So

$$\begin{aligned}\Pr(FI^u) &= \frac{\Pr\left(o(P)|F\right)\Pr(F) + \Pr\left(o(F)|P\right)\Pr(P)}{\Pr(P)+\Pr(F)} \\ &= \Pr\left(o(P)|F\right)\Pr(F) + \Pr\left(o(F)|P\right)\Pr(P). \quad (3)\end{aligned}$$

This form will be more useful in diagnosis in that $\Pr(F)$ is derivable directly from the failure rate of the unit, and $\Pr(o(F)|F) = 1 - \Pr(o(P)|F)$ is a measure of the ability of a test to detect the associated failure which we will derive below.

Now we will consider these alternative conditional probabilities. Specifically:

1) $\Pr(o(F)|F)$ is the probability we observe a failure, given the unit is faulty;
2) $\Pr(o(P)|P)$ is the probability we do not observe a failure, given the unit is not faulty;
3) $\Pr(o(F)|P)$ is the probability we observe a failure, given the unit is not faulty;
4) $\Pr(o(P)|F)$ is the probability we do not observe a failure, given the unit is faulty.

In considering test confidence alone, we can assume we do not know the failure probability of the UUT, i.e., in the worst case, that $\Pr(F) = \Pr(P)$. This leads to the $\Pr(F)$ and $\Pr(P)$ terms in (2) canceling out, so we can consider the lack of confidence (which we call suspicion) in our test, independent of the failure rate of the unit tested, to be the following:

$$\Pr(FI^c) = \frac{\Pr(o(F)|P)+\Pr(o(P)|F)}{\Pr(o(F)|P)+\Pr(o(F)|F)+\Pr(o(P)|P)+\Pr(o(P)|F)}. \quad (4)$$

Here, FI^c denotes fault indication in a "conditional sense" where we consider only conditional probabilities. Equation (6) can be simplified further by observing that the sum of the conditional probabilities $\Pr(o(F)|P) + \Pr(o(P)|P) = 1$ and $\Pr(o(P)|F) + \Pr(o(F)|F) = 1$. Thus

$$\Pr(FI^c) = \frac{1}{2}\left[\Pr\left(o(F)|P\right) + \Pr\left(o(P)|F\right)\right]. \quad (5)$$

To make sure that we have not violated the axioms of probability (specifically, the axiom that states that the sum of the probabilities over a domain must be 1), we note the following. First, $\Pr(o(P)|P) = 1 - \Pr(o(F)|P)$ and $\Pr(o(F)|F) = 1 - \Pr(o(P)|F)$. So, the probability of a true indication $(\Pr(TI^c))$, which reflects test confidence, can be computed similarly

$$\Pr(TI^c) = \frac{\Pr(o(F)|F)+\Pr(o(P)|P)}{\Pr(o(F)|P)+\Pr(o(F)|F)+\Pr(o(P)|P)+\Pr(o(P)|F)}. \quad (6)$$

Observe that, given the above axiom of probability, we must have $\Pr(TI^c) = 1 - \Pr(FI^c)$. Using the complementary of probabilities defined above we have

$$\begin{aligned}\Pr(TI^c) &= \frac{(1-\Pr(o(F)|P))+(1-\Pr(o(P)|F))}{\Pr(o(F)|P)+\Pr(o(F)|F)+\Pr(o(P)|P)+\Pr(o(P)|F)} \\ &= 1-\frac{1}{2}\left[\Pr(o(F)|P)+\Pr(o(P)|F)\right] \\ &= 1-\Pr(FI^c).\end{aligned}$$

So, we see that the axioms of probability hold as desired.

Thus, assuming independence in testing, we can derive corresponding false indication probabilities from instrument accuracy and reliability measures. Assuming T independent tests, this can be given as

$$\Pr(\mathbf{FI}) = \frac{\sum_{i=1}^{T}\Pr\left(FI_i^c\right)}{\sum_{i=1}^{T}\Pr\left(FI_i^c\right)+\sum_{i=1}^{T}\Pr\left(TI_i^c\right)} = \frac{1}{T}\sum_{i=1}^{T}\Pr\left(FI_i^c\right). \quad (7)$$

When wanting to predict $\Pr(FI^c)$, we can apply Baye's Rule to each of the respective terms where

$$\begin{aligned}\Pr\left(o(F)|P\right) &= \frac{\Pr\left(P|o(F)\right)\Pr\left(o(F)\right)}{\Pr(P)} \\ \Pr\left(o(P)|F\right) &= \frac{\Pr\left(F|o(P)\right)\Pr\left(o(P)\right)}{\Pr(F)}. \quad (8)\end{aligned}$$

Notice that the probabilities $(\Pr(F))$ are simply failure probabilities, which can be derived from reliability models. Further, $\Pr(P|o(P))$ can be derived from the instrument uncertainty, and $\Pr(o(P))$ corresponds to the prior probability of observing no fault (i.e., $1 - \Pr(o(F))$). As we will see, we will never have to actually provide or compute $\Pr(o(F))$ or $\Pr(o(P))$. This is significant in that our model provides a formal approach to specifying all of the numerical components of a Bayesian model for fault diagnosis and prognosis. In the end, we argue that these models more accurately represent available diagnostic information by incorporating better understanding of test measurement uncertainty. Thus, the models can be expected to yield better diagnostic results over models that do not capture sources of uncertainty in a formal way.

$$\Pr(FI^u) = \frac{\Pr\left(o(P)|F\right)\Pr(F) + \Pr\left(o(F)|P\right)\Pr(P)}{\Pr\left(o(P)|P\right)\Pr(P) + \Pr\left(o(F)|F\right)\Pr(F) + \Pr\left(o(P)|P\right)\Pr(F) + \Pr\left(o(F)|F\right)\Pr(P)} \quad (2)$$

V. Testing and Detection Theory

If we regard the source of error in testing to be the uncertainty inherent in the test environment, and we assume that this uncertainty arises from several independent random sources, then by the *Central Limit Theorem*, the combined uncertainty tends to be distributed according to a Gaussian distribution. Under this assumption, we can model both nominal behavior and anomalous behavior with respect to the measurement x as a Gaussian distribution, where the only difference between the two distributions is that the mean μ is shifted by some offset.

Without loss of generality, suppose we focus on false alarms. In particular, given the underlying behavior distribution of the UUT while faulty and fault free, we can evaluate various decision thresholds for a test by plotting the probability of false alarm against the probability of true alarm for each of these thresholds. The resulting curve is the so-called receiver operating characteristic or ROC curve (Fig. 3).

We would like to be able to characterize the ability of a test to discriminate noise from signal. Signal detection theory applies a measure called the "discriminability" of a detector[5] (i.e., test) and defines it as

$$d' = \frac{|\mu_1 - \mu_2|}{\sigma} \tag{9}$$

where μ_1 and μ_2 are the means of the two distributions and σ is the standard deviation. A different ROC curve can be plotted for every d'. As the distributions are more discriminable, then the more the bend in the curve will move to the upper left corner, as shown in the figure. If the curves are completely indiscriminable (i.e., signal cannot be separated from noise), then the curve will correspond to a 45° line from (0, 0) to (1, 1). Thus, this line represents a "worst case" for a test.

From the standpoint of test design or test selection, the discriminability of the test is an essential feature that should be evaluated. Specifically, highly discriminable tests will be better suited for managing decision risk (Section VIII) through the identification of clear decision thresholds with tighter associated tolerances. However, less discriminable tests will tend to require wider tolerances associated with those decision thresholds, thus, making the choice of thresholds to control risk more difficult.

VI. Determining Probability of Failure

Traditional reliability theory determines probability of failure based on the projected failure rate of an item. Failure rates are determined using combinations of historical data and physical models and are typically expressed in terms of numbers of failures per million hours (or some other unit). Traditionally, the failure rate is denoted λ, and the mean time between failures (MTBF) is $1/\lambda$.

Usually, failure probability is computed under the assumption of an exponential failure distribution, based on a further assumption of constant failure rate [22].[6] In other words, if we let D_i denote failure (or diagnosis) i, the probability of failure D_i is computed as

$$\Pr(D_i) = 1 - \exp[-\lambda_i t]. \tag{10}$$

In performing diagnosis under uncertainty, it is common to compute probability of failure as a relative measure over members of an ambiguity group directly from the failure rates.

Specifically, we define an ambiguity group to be a set of diagnoses for which no tests have been performed that differentiate the set (perhaps because they do not exist). Assume (for simplicity) that a diagnosis corresponds to a single fault mode. Assume also that each fault mode has an associated failure rate. Then, given a set of diagnoses D_i in an ambiguity group $\mathbf{A}$

$$\Pr(D_i|\mathbf{A}) = \frac{\lambda_{D_i}}{\sum_{D_j \in \mathbf{A}} \lambda_{D_j}}. \tag{11}$$

Thus, for a given diagnosis D_i, determining $\Pr(F)$ (i.e., $\Pr(D_i)$) for that diagnosis prior to testing (in which the initial ambiguity group is the entire set of possible diagnoses D) is simply

$$\Pr(D_i) = \frac{\lambda_{D_i}}{\sum_{D_j \in D} \lambda_{D_j}}. \tag{12}$$

However, note that the failure probability is time-dependent and Markov in nature. In other words, the actual probability of a fault, in addition to depending on its likelihood relative to the other members of an ambiguity group, also depends upon the time since the last evaluation of that unit. Since ambiguity group

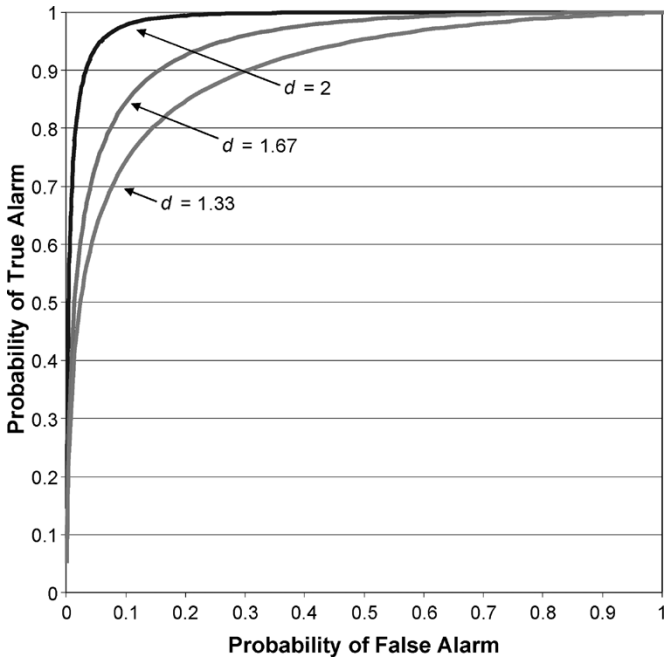

Fig. 3. Sample ROC curves.

[5]The discriminability of a test corresponds to the ability of the test to discern faulty states from fault-free states based on the underlying distribution of measurement uncertainty.

[6]Note that, when the failure rate is not constant, it is common for the Weibull distribution to be used instead. It is relevant to observe that, for the Weibull distribution, when the shape parameter β is set to 0, the distribution reduces to exponential.

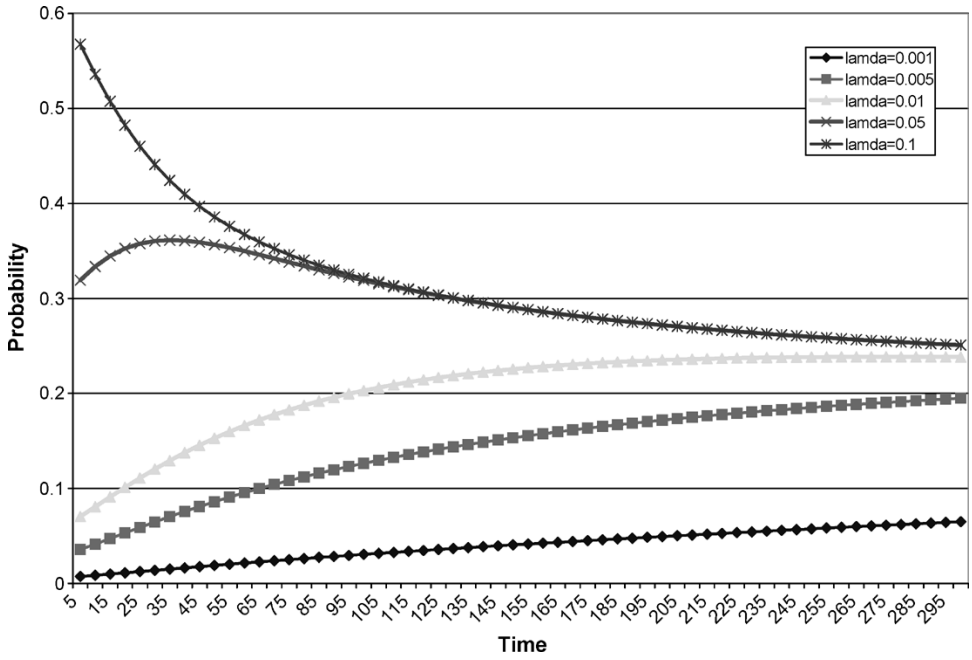

Fig. 4. Progression of relative probabilities with time.

D corresponds (as defined above) to a group of possible diagnoses that have not been differentiated, they have not yet been evaluated; therefore, the more accurate method for computing the relative probabilities of the diagnoses in D (given a specified time t since last testing) is

$$\Pr(D_i) = \frac{1 - \exp\left[-\lambda_{D_i} t\right]}{\sum_{D_j \in D} \left(1 - \exp\left[-\lambda_{D_j} t\right]\right)}. \tag{13}$$

Suppose we have three possible diagnoses with failure rates (assuming common units such as per million hours) set at 0.001, 0.005, 0.01, 0.05, and 0.1, respectively. Using (12), we find the relative failure probabilities for these are approximately 0.006, 0.03, 0.06, 0.3, and 0.6, respectively. However, if we use (13), we find that, due to the dependence on time, these relative probabilities drastically change, eventually converging with time to a uniform distribution (Fig. 4). We can then use this time-dependent view in the calculation of probability of false indication, as defined above.

VII. Determining Instrument Uncertainty

Consider now the task of determining uncertainty in a measurement from some instrument or built-in test circuitry. Typically, a measurement μ is mapped to a PASS or FAIL outcome when testing; however, based on the uncertainty of the measurement, u_μ, the probability of observing the PASS or FAIL outcome can be determined.

Note that when measuring the state of a UUT, we are not measuring that state directly but are measuring some signal that depends on a number of independent variables (inputs, internal state, fault conditions, environmental conditions, etc.). Thus, we can consider the measurement process as one of estimating a function $y = f(x_1, \ldots, x_n)$. Uncertainty in the dependent variable y ultimately depends on the uncertainty in the independent variables x_i, (denoted u_i, respectively).

In the metrology and calibration community, the uncertainty of an instrument is determined through one of two methods: Type A uncertainty is determined based on statistical sampling, and Type B uncertainty is determined based on the assumption of some underlying distribution [48]. If we assume a Gaussian distribution, we find that, frequently, the Type A and Type B uncertainties are approximately the same. Either way, our objective is to identify *standard uncertainty*, which is typically treated as the standard deviation of the underlying distribution.

More precisely, consider some measurement, x_m that is intended to observe some "true value" x. The uncertainty associated with this measurement is then u_x, and we say that $x = x_m \pm u_x$. Recognizing that the estimation process involves combining multiple measurements, we are seeking some $y = f(x_1, \ldots, x_n)$. Given that the uncertainty u_i of independent variable x_i can be determined either statistically or analytically, the goal is to determine the uncertainty in the dependent variable y. The "combined" standard uncertainty of y, denoted $u_c(y)$, represents the estimated standard deviation (or standard error) in the result, and is computed as

$$u_c(y) = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i}\right)^2 u^2(x_i)}. \tag{14}$$

Note that this assumes independence among the variables x_i. If we wish to consider correlated effects among the variables [1], then our estimate (based on the *law of uncertainty propagation* [50]) becomes

$$u_c(y) = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i}\right)^2 u^2(x_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} u(x_i, x_j)}. \tag{15}$$

Here, $u^2(x_i)$ denotes the estimated variance of x_i, and $u(x_i, x_j)$ denotes the estimated covariance associated with x_i and x_j. This law is derived from a first-order Taylor series approximation of

the function $y = f(x_1, \ldots, x_n)$. For many functional forms, (15) reduces to simple linear forms. For complex cases where there is significant nonlinearity, higher order approximations are required.

Interpreting the result of (15), since we are assuming a Gaussian (i.e., normal) distribution and treating standard uncertainty as the standard error of the estimate y, we note that the actual value $Y = y \pm u_c$ approximately 68% of the time and $Y = y \pm 2u_c$ approximately 95% of the time. Consider now a test that has been "normalized" such that its distribution is centered on zero with a nominal range of $\pm z$. This distribution has particular standard deviation, σ, which is determined through the process of designing the test.

Our interest is in determining the probability of some observation, given the underlying state of the system. For example, $\Pr(o(P)|P)$ and $\Pr(o(F)|F)$ would tell us the probability of observing PASS given the unit is fault free and the probability of observing a FAIL given the unit is faulty, respectively. Note that these probabilities work from the discrete random variables of unit state and test outcome; however, the test outcome, ultimately, depends on the underlying measurement. So consider the case where a measurement y is taken and $y \in \pm z$ (i.e., the measurement indicates a PASS observation). The probability of the test actually PASSing given the measurement y is given as

$$\Pr(P|y) = \Pr(P|o(P))\Pr(o(P)|y) + \Pr(P|o(F))\Pr(o(F)|y). \tag{16}$$

However, when conditioning on observing a PASS, we note that the second term drops out since $\Pr(o(F)|y) = 0$. Also, since $\Pr(o(P)|y) = 1$, we have

$$\Pr(P|y) = \Pr\left(P|o(P)\right). \tag{17}$$

Similarly

$$\Pr(F|y) = \Pr(F|o(F))\Pr(o(F)|y) + \Pr(F|o(P))\Pr(o(P)|y). \tag{18}$$

So, by the same argument, when conditioning only on observing a FAIL, we find that

$$\Pr(F|y) = \Pr\left(F|o(F)\right). \tag{19}$$

Continuing in this line of reasoning, observe that the probability of taking a particular measurement y given the unit passes can be computed by considering the distribution around y limited to the nominal range.[7] In other words

$$\Pr(y|P) = \frac{1}{u_c\sqrt{2\pi}} \int_{-z}^{z} \exp\left[-\frac{1}{2}\left(\frac{x-y}{u_c}\right)^2\right] dx. \tag{20}$$

Similarly, consider the case where a measurement is taken and the system is faulty. The probability of receiving the actual measurement y conditioned on this faulty state is then given as

$$\Pr(y|F) = \frac{1}{u_c\sqrt{2\pi}} \int_{-\infty}^{-z} \exp\left[-\frac{1}{2}\left(\frac{x-y}{u_c}\right)^2\right] dx + \frac{1}{u_c\sqrt{2\pi}} \int_{z}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-y}{u_c}\right)^2\right] dx. \tag{21}$$

Thus, observing that the limits of integration between (20) and (21) cover the entire possible range and that the integral over this range must equal 1, we see that $\Pr(y|P) = 1 - \Pr(y|F)$.

Without loss of generality, focus on $\Pr(P|y) = \Pr(P|o(P)$ [from (17)]. If we apply Bayes rule, we get

$$\Pr(P|y) = \frac{\Pr(y|P)\Pr(P)}{\Pr(y)}.$$

Substituting the various integrals, we get

$$\Pr(P|y) = \frac{\left[\frac{1}{u_c\sqrt{2\pi}} \int_{-z}^{z} \exp\left[-\frac{1}{2}\left(\frac{x-y}{u_c}\right)^2\right] dx\right]\Pr(P)}{\left[\frac{1}{u_c\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-y}{u_c}\right)^2\right] dx\right]}. \tag{22}$$

Observe that the denominator reduces to 1.0 and that the second term in the numerator is a constant defined by the failure probability of the unit tested (i.e., $\Pr(P) = 1 - \Pr(F)$). Thus

$$\Pr\left(P|o(P)\right) = (1 - \Pr(F)) \times \left[\frac{1}{u_c\sqrt{2\pi}} \int_{-z}^{z} \exp\left[-\frac{1}{2}\left(\frac{x-y}{u_c}\right)^2\right] dx\right]. \tag{23}$$

Following a similar line of reasoning, we can then derive $\Pr(F|o(F))$ as

$$\Pr\left(F|o(F)\right) = \Pr(F)\left[\frac{1}{u_c\sqrt{2\pi}} \int_{-\infty}^{-z} \exp\left[-\frac{1}{2}\left(\frac{x-y}{u_c}\right)^2\right] dx + \frac{1}{u_c\sqrt{2\pi}} \int_{z}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-y}{u_c}\right)^2\right] dx\right]. \tag{24}$$

Note that these equations assume the availability of a specific measured value (i.e., y). In a dynamic diagnostic environment, this may be interesting; however, fixed probabilities are usually required for a Bayesian model. One approach is to use the expected value for the appropriate range. Another (and the one we adopt) is to select a "worst case" value. In this case, such a value would correspond to one that is coincident with one of the limits,[8] z (since this would maximize probability of false alarm or nondetection).[9] Given this, we now have all of the pieces required to determine test confidence for Bayesian diagnosis.

[7]Usually, this would be restricted to the dynamic range of the instrument; however, we assume the tests have been defined properly such that the measurements of a particular test are of consistent accuracy. Therefore, the number of in-tolerance measurements having different accuracy would be insignificant. This allows us it "ignore" the tails of the distribution as being negligible in cumulative probability.

[8]Technically, since we do not know *a priori* which limit would be exceeded, we should take the expected value, i.e., the average in this case, of the two limit values. With symmetric distributions, however, this is not necessary, and we can work with one of the limits chosen *without loss of generality*.

[9]Should historical data be available, it might be possible to refine these probabilities based on that history, but such issues of diagnostic "maturation" are beyond the scope of this paper.

Assume, without loss of generality, that we use a worst case of $y = +z$ for our measurement value. Then, (23) and (24) reduce to

$$\Pr(P|o(P)) = (1-\Pr(F))\left[\frac{1}{u_c\sqrt{2\pi}}\int_{-z}^{z}\exp\left[-\frac{1}{2}\left(\frac{x-z}{u_c}\right)^2\right]dx\right]. \quad (25)$$

$$\Pr(F|o(F)) = \Pr(F)\left[1-\frac{1}{u_c\sqrt{2\pi}}\int_{-z}^{z}\exp\left[-\frac{1}{2}\left(\frac{x-z}{u_c}\right)^2\right]dx\right]. \quad (26)$$

An alternative approach to deriving these probabilities is to note that the probabilities of the state of the UUT and the measurement of the test both follow normal distributions that can be modeled jointly using the bivariate normal distribution [20]. Specifically, the bivariate normal distributed is defined as

$$P(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho}}\exp\left[-\frac{z}{2(1-\rho)^2}\right]$$

$$\text{where} \quad z = \frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}.$$

Here, ρ is the correlation coefficient $\sigma_{12}/\sigma_1\,\sigma_2$.

If we consider the distribution around x_1 to be the expected value of the state of the UUT under nominal conditions, and the distribution around x_2 to be the expected value of the instrument regardless of which UUT is being measured, then we note that these distributions are independent. Thus, we can argue that they should not be correlated and set $\rho = 0$. In this case, the bivariate normal distribution reduces to

$$P(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right].$$

Note that marginalizing out x_1 or x_2 yield the corresponding univariate normal distributions.

$$P(x_1) = \int_{-\infty}^{\infty} P(x_1,x_2)dx_2 = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right]$$

$$P(x_2) = \int_{-\infty}^{\infty} P(x_1,x_2)dx_1 = \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left[-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right].$$

This makes sense, because if x_1 and x_2 are independent, we have $P(x_1,x_2) = P(x_1)P(x_2)$, giving us

$$P(x_1,x_2) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right] \times \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left[-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right] = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right].$$

Consider now the corresponding cumulative probabilities, defined as

$$\Pr(x_1,x_2) = \iint \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right] \times \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left[-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right]dx_1dx_2.$$

Notice that each of the terms in the integrand are independent of each other (from the standpoint of the variables of integration), which allows us to rewrite the integral as follows:

$$\begin{aligned}\Pr(x_1,x_2) &= \int \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left[-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right] \times \int \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right]dx_1dx_2 \\ &= \left(\frac{1}{\sigma_1\sqrt{2\pi}}\int \exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right]dx_1\right) \times \left(\frac{1}{\sigma_2\sqrt{2\pi}}\int \exp\left[-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right]dx_2\right) \\ &= \Pr(x_1)\Pr(x_2).\end{aligned}$$

Finally, observe that this is exactly the form of (25) and (26).

It is interesting to note that repeated measurements enable revision of uncertainty and can tend to reduce uncertainty. The idea of "repeat polling" has been proposed as a technique for reducing the probability of a false indication by allowing transient signals and other sources of measurement uncertainty to "dampen out" [36]. Specifically, using this idea of repeat polling, Phillips *et al.* show that uncertainty can be revised as

$$u_c(t) = \left(\frac{1}{u_c^2(0)}+\frac{1}{u_c^2(t-1)}\right)^{-\frac{1}{2}}. \quad (27)$$

Given constant uncertainty (for a particular measurement type), we see that

$$u_c(t) = \left(\sum_{i=1}^{t}\frac{1}{u_c^2}\right)^{-\frac{1}{2}} = \frac{u_c}{\sqrt{t}}. \quad (28)$$

Further, the best estimate of the measurand based on repeated measurements can be determined by computing the weighted sum

$$y = x(t)\left(\frac{\gamma^2}{1+\gamma^2}\right) + y(t-1)\left(\frac{1}{1+\gamma^2}\right) \quad (29)$$

where $x(t)$ is the current measurement, $y(t-1)$ is the best estimate through $t-1$ measurements, and $\gamma = u_c(t-1)/u_c(0)$ represents the ratio between uncertainty at the previous time step and initial uncertainty (thus showing the amount of relative reduction at this time step).

VIII. Applying Bayes Decision Theory

Assuming we are able to generate the probability distributions for nominal and faulty behavior, we return to the diagram provided in Fig. 2 to consider the effects of locating the decision boundaries. For this discussion, we will draw on results from Bayes decision theory and its derivative, signal detection theory

[11]. For this discussion, we will assume we are performing a two-category classification problem where, given a test result, we want to find a decision threshold such that if the value is less than or equal to that threshold, we assign a PASS outcome; otherwise, we assign a FAIL outcome. Further, given that we assign this outcome based on an observation of the measured signal with some level of uncertainty, we want to select the threshold that minimizes misclassification error.

For now, to keep things simple, assume that the UUT can only fail one way, and we have a single test designed to detect that failure. Let ω_i represent the "actual state" of the unit being tested, where ω_1 corresponds to a nominal unit, and ω_2 corresponds to a faulty unit. Further, let x_j represent the measured value of the test performed on the UUT. Finally, assume the action of deciding on a test outcome is designated by α_j, where α_1 corresponds to deciding to assign a PASS outcome and α_2 corresponds to deciding to assign a FAIL outcome.

As mentioned above, the concepts of consumer risk and producer risk are fundamental drivers for determining where to set the decision threshold. To make this more formal, we will define a "risk" function as follows. Let $R(\alpha_i|x)$ be the *conditional risk* associated with deciding on α_i given the measurement is x. We will determine this risk as an expected value of "loss" computed over the set of all decisions. Specifically

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) \Pr(\omega_j|x) \tag{30}$$

where $\lambda(\alpha_i|\omega_j)$ is the "loss" associated with choosing α_i when the true state was ω_j and c is the number of categories (in this case two). From this, we can compute the overall risk associated with a decision rule as

$$R = \int R\left(\alpha(x)|x\right) p(x) dx. \tag{31}$$

Recall that we are focusing on the two-category case. To do this, we will apply some shorthand. Namely, let λ_{ij} denote the loss of deciding α_i when the true state is ω_j.[10] Then we can rewrite the conditional risk equation (30) as

$$R(\alpha_i|x) = \sum_{j=1}^{2} \lambda_{ij} \Pr(\omega_j|x). \tag{32}$$

Expanding, we get the two equations that provide the risk for each of the decision categories

$$\begin{aligned} R(\alpha_1|x) &= \lambda_{11} \Pr(\omega_1|x) + \lambda_{12} \Pr(\omega_2|x) \\ R(\alpha_2|x) &= \lambda_{21} \Pr(\omega_1|x) + \lambda_{22} \Pr(\omega_2|x). \end{aligned}$$

In this case, the "rational" decision is to select a choice that minimizes conditional risk. In other words, decide α_1 if $R(\alpha_1|x) < R(\alpha_2|x)$. We can represent this in a way that determines the decision in terms of probabilities. Specifically

$$\lambda_{11} \Pr(\omega_1|x) + \lambda_{12} \Pr(\omega_2|x) < \lambda_{21} \Pr(\omega_1|x) + \lambda_{22} \Pr(\omega_2|x).$$

[10]Note that λ_{ij} is distinct from the failure rate. The overloading of symbols comes from the fact different communities have defined these concepts. We will continue to use the traditional symbols since their interpretation should be clear based on context.

Rearranging terms, we get

$$\lambda_{11} \Pr(\omega_1|x) - \lambda_{21} \Pr(\omega_1|x) < \lambda_{22} \Pr(\omega_2|x) - \lambda_{12} \Pr(\omega_2|x)$$

and simplifying, we find

$$(\lambda_{11} - \lambda_{21}) \Pr(\omega_1|x) < (\lambda_{22} - \lambda_{12}) \Pr(\omega_2|x).$$

Suppose, for the sake of discussion, that $\lambda_{11} = \lambda_{22} = 0$ (i.e., there is no loss associated with a correct classification), and $\lambda_{12} = \lambda_{21} > 0$. Then the decision rule reduces to deciding ω_1 if $\Pr(\omega_1|x) > \Pr(\omega_2|x)$, which once again appears to be "rational."

Now assume the more general case where $\lambda_{11} = \lambda_{22} = 0$, but $(\lambda_{12} \neq \lambda_{21}) > 0$. If we define category 1 to be PASS and category 2 to be FAIL, then λ_{12} represents the *producer risk* and λ_{21} represents the *consumer risk*. Then the decision rule becomes deciding ω_1 if $\lambda_{21} \Pr(\omega_1|x) > \lambda_{12} \Pr(\omega_2|x)$. Applying Bayes' Rule, this can be expanded to deciding ω_1 if

$$\lambda_{21} p(x|\omega_1) \Pr(\omega_1) > \lambda_{12} p(x|\omega_2) \Pr(\omega_2).$$

Observe that we can rewrite the decision rule using the "likelihood ratio" as follows. Decide ω_1 if

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{\lambda_{12}}{\lambda_{21}} \times \frac{\Pr(\omega_2)}{\Pr(\omega_1)}. \tag{33}$$

Given this rule, let $\Re_1$ denote the region over x where we decide α_1 and $\Re_2$ denote the region over x where we decide α_2. Recall that our goal is to minimize overall risk R. Let

$$\begin{aligned} R_{mm} &= \lambda_{12} \int_{\Re_1} p(x|\omega_2) dx \\ \Phi &= \lambda_{21} \int_{\Re_2} p(x|\omega_1) dx - \lambda_{12} \int_{\Re_1} p(x|\omega_2) dx. \end{aligned}$$

Then we can redefine risk as $R = R_{mm} + \Pr(\omega_1)\Phi$. In other words, we find that risk is *linear* in $\Pr(\omega_1)$. We would like to find a boundary that makes R independent of the prior $\Pr(\omega_1)$, but that only occurs if $\Phi = 0$. If we treat Φ as if it was the "slope" of the risk line, then we would be looking for regions that set the slope to zero. Then risk reduces to R_{mm}, which is referred to as the *minimax risk*. This occurs when

$$\lambda_{21} \int_{\Re_2} p(x|\omega_1) dx = \lambda_{12} \int_{\Re_1} p(x|\omega_2) dx. \tag{34}$$

In other words, recalling the likelihood ratio, we can set the decision boundary independent of the priors such that

$$\frac{\int_{\Re_2} p(x|\omega_1) dx}{\int_{\Re_1} p(x|\omega_2) dx} = \frac{\lambda_{12}}{\lambda_{21}}. \tag{35}$$

Note that we can extend this idea to the full diagnostic problem where, rather than deciding between two categories corresponding to PASS and FAIL, we consider each diagnosis to be a separate category. In this case, the above result can be extended such that the classification assigned corresponds to

the category with the maximum *a posteriori* probability. More formally

$$\begin{aligned} h_{MAP} &= \arg\max_{\omega\in\Omega} [\lambda_\omega \Pr(\omega|x)] \\ &= \arg\max_{\omega\in\Omega} [\alpha\lambda_\omega \Pr(x|\omega)\Pr(\omega)] \\ &= \arg\max_{\omega\in\Omega} [\lambda_\omega \Pr(x|\omega)\Pr(\omega)] \end{aligned} \tag{36}$$

where α is a normalizer, and λ_ω is the loss associated with ω being the incorrect classification. The resulting classification is referred to as the *maximum a posteriori hypothesis*. In other words, the diagnosis yielding the highest posterior probability is proposed as the most probable fault.

IX. Diagnosis With Bayesian Networks

Given the previous discussion, we see a direct application of Bayesian decision theory to diagnosis with BIT under the assumption we can estimate BIT uncertainty (see Section VII). From these estimates, it can be proven [11] that the MAP hypothesis, in fact, minimizes overall risk as defined in (30). In other words, diagnosis is improved over traditional diagnosis that ignores measurement uncertainty by the fact that overall risk associated with the resultant diagnosis has been minimized.

Recall that $\Pr(TI) = 1 - \Pr(FI)$. We will use this as a measure of confidence in our test results. Note that, if we have measures for $\Pr(o(F)|P)$ and $\Pr(o(P)|F)$ for each of the measured tests, then we can use Bayes Rule and break the confidence into its respective parts for PASS or FAIL. Specifically, our confidence in a PASS result becomes $\Pr(P|o(P)) = 1 - \Pr(F|o(P))$. Similarly, our confidence in a FAIL result becomes $\Pr(F|o(F)) = 1 - \Pr(P|o(F))$.

Once we have formulated probabilities such as these, we are in a position to use these probabilities as confidence values in a diagnostic system. There are many diagnostic systems that allow for reasoning under uncertainty; however, the natural approach that uses probabilities directly is the Bayesian belief network (BBN) [35]. Formally, a $\mathrm{BBN} = \langle V, \mathbf{E}, \mathbf{C}\rangle$ is a *Bayesian Belief Network*, where

1) V is a set of vertices corresponding to random variables $V_i \in V$;
2) $\mathbf{E}$ is a set of directed edges $e_{ij} \in \mathbf{E}$ where the source of the edge corresponds to V_i, the destination of the edge corresponds to V_j, and the edge represents a conditional dependence relationship of V_j on V_i;
3) $\mathbf{C}$ is a set of conditional probability tables $C_{\mathrm{Pa}(i)} \in \mathbf{C}$, where each entry provides the probability of V_i given the set of parents of $V_i(\mathrm{Pa}(i))$.

The structure of the BBN depends on the concept of conditional independence. Given two random variables, V_1 and V_2, V_1 is *conditionally independent* of V_2 iff $\Pr(V_1|V_2) = \Pr(V_1)$. Further, given random variables V_1, V_2, and V_3, V_1 is *conditionally independent* of V_2 given V_3 iff $\Pr(V_1|V_2, V_3) = \Pr(V_1|V_3)$ and $\Pr(V_2|V_1, V_3) = \Pr(V_2|V_3)$. Within the BBN, we say that a random variable V_i is conditionally independent of all other variables not connected to it given its parents $\mathrm{Pa}(V_i)$.

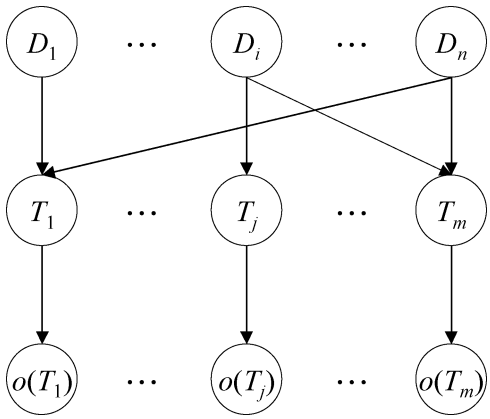

Fig. 5. Diagnostic Bayesian network structure.

If we continue to assume that tests are performed independently from one another, then we connect tests (as random variables) to possible diagnoses (as random variables). Usually, we can also assume that the diagnoses are independent from one another (i.e., the existence of one fault does not cause another fault to occur). Thus, the only dependence relationships modeled are between tests and diagnoses. Note that these two assumptions are not necessarily true, and in general such dependence relationships, when known, can be modeled directly by the BBN by inserting appropriate edges between pairs of tests or between pairs of diagnoses. In addition, we must find a way to handle the relationships between the intended states of the tests and the observations of those tests. Thus, we can define a causal model such as the one illustrated in Fig. 5.

The diagnostic problem consists of inferring the probability of each of the diagnoses in the BBN given the test results. Note that the joint probability distribution over all of the variables in the BBN $\Pr(V)$ is given as the product of the probability distributions of each over each of the vertices (random variables) conditioned on their parents, i.e.,

$$\Pr(V) = \prod_{V_i \in V} \Pr\left(V_i|\mathrm{Pa}(V_i)\right). \tag{37}$$

Assume we subdivide the set of random variables V into two subsets, T and D corresponding to tests (e.g., BIT indications) and diagnoses, respectively, as discussed previously. Further, assume that $T \cup D = V$ and $T \cap D = \varnothing$ (i.e., T and D are disjoint but define the entire set of random variables V). Finally, assume T contains the set of observations (e.g., the evidence from BIT) and D contains everything else (including, if needed, random variables representing the "true" states of the tests as if we were able to know the underlying state perfectly). Given a target set of test results, τ, we calculate $\Pr(T' = \tau)$ (where $T \subseteq T$) by marginalizing out the remaining variables given by $V \setminus T'$. Marginalization is carried out by summing over all $\Pr(V \setminus T', T = \tau)$, where "$\setminus$" denotes set difference

$$\Pr(T' = \tau) = \sum_{V\setminus T} \Pr(V \setminus T', T' = \tau). \tag{38}$$

In the case of Fig. 5, we note the prior probabilities for the diagnoses D_i, are given by the probabilities derived from (13).

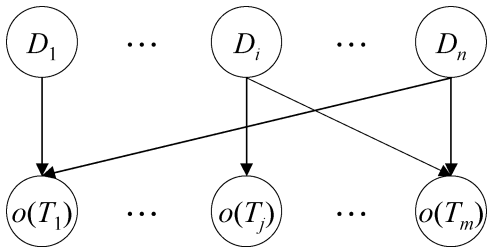

Fig. 6. Simplified diagnostic Bayesian network.

The probabilities for $\Pr(T_j|D_i)$ arise from constructing the appropriate diagnostic model and reflect the causal nature of the faults. The probabilities of observing the appropriate state corresponds to the conditional probabilities defined in (23) and (24). Note that, if one can assume the probabilities $\Pr(T_j|D_i)$ are either 1 or 0 (i.e., there is no uncertainty), then the network can be simplified to take the form in Fig. 6 with the associated conditional probability tables being derived as follows:

$$\Pr\left(o(T_i)|\mathrm{Pa}\left(o(T_i)\right)\right) = \Pr(o(T_i)|D_{T_i}) = \prod_{D_j \in D_{T_i}} \Pr(o(T_i)|D_j) \tag{39}$$

where D_{T_i} is the specific set of diagnoses (i.e., a subset of all diagnoses in the model), all of which must "pass" to observe the dependent test passing (i.e., $D_{T_i} = \mathrm{Pa}(o(T_i))$).

When evaluating BBNs, there are several techniques for deriving probabilities based on evidence values. For the restricted set of BBNs corresponding to polytrees, exact solutions exist. For general BBNs, computing the joint probability distribution is NP-hard; therefore, a variety of techniques exist ranging from constructing join-trees to applying various Monte Carlo sampling techniques [42].

Applying the method described above for determining probability of false alarm or probability of missed detection, we can use these probabilities directly without the need for stochastic sampling methods. By assuming independence among the random variables in D (i.e., the diagnoses) as well as independence among the random variables in T (i.e., the tests), the characteristics of direction-dependent separation (i.e., d-separation) allow for simple propagation of the probabilities from the tests to the diagnoses.[11] Specifically, we say that a set of evidence nodes E in a BBN (i.e., nodes that can be observed) d-separates two sets of nodes X and Y if every undirected path from a node $x \in X$ to a node $y \in Y$ is "blocked" given E [42]. A path is blocked given E if there exists some node z on the path for which one of the following three conditions holds:

1) if $z \in E$, z has one parent that is on the path, and z has one child that is also on the path;
2) if $z \in E$ and z has two children, both of which are on the path;
3) if $z \notin E, \forall x \mathrm{pa}(x) = z \notin E$, and z has two parents, both of which are on the path.

[11]In the event additional dependence relationships need to be defined, then propagation of probabilities along these dependence links must also be accounted for. The theory of Bayesian networks allows for this; however, such additional dependencies may lead to large internal cycles, which are conditions under which the NP-hardness of Bayesian inference becomes problematic.

Now, we can assert the diagnoses are conditionally independent of each other given the test results because of the third case, even though the diagnoses have no parents. Specifically, taking the graph transpose of the network (i.e., reversing the directions of the arcs) results in the evidence variables being the parents of the diagnoses, and none of the diagnoses being evidence variables.

Given the conditional independence of the diagnoses, we can then compute the posterior probabilities of each of the diagnoses given the test results as follows. First, we will assume that we are using the network form presented in Fig. 5 and partition the random variables into three sets: D (the diagnoses), T (the true test states), and O (the test observations). The evidence variables will be restricted to O.

$$\begin{aligned}\Pr(D_i|O) &= \alpha \Pr(O|D_i) \Pr(D_i) \\ &= \alpha \Pr(D_i) \sum_{T_j \in T} \Pr\left(o(T_j)|T_j\right) \Pr(T_j|D_i). \end{aligned} \tag{40}$$

Here, α is a normalizer over the set D, equal to

$$\alpha = \sum_{D_i \in D} \Pr(D_i) \sum_{T_j \in T} \Pr\left(o(T_j)|T_j\right) \Pr(T_j|D_i).$$

Observe that $\Pr(T_j|D_i) \in \{0, 1\}$ as described earlier, so the members of the sum are restricted only to those tests that observe D_i. Then we only need to consider $\Pr(D_i)$, which corresponds to the prior probability for D_i based on failure rate, and $\Pr(o(T_j)|T_j)$, which corresponds to the confidence value assigned to the observed test result. Using the Baye's maximum *a posteriori* hypothesis from (36), we determine the most likely diagnosis simply as

$$D_{MAP} = \arg\max_{D_i \in D} \left\{\Pr(D_i|O)\right\}. \tag{41}$$

In other words, we provide the most probable diagnosis as a means of minimizing expected error (i.e., risk) in the diagnostic process.

X. Numerical Example

To illustrate the concepts described in the previous sections, we present a small numerical example. Suppose, for the sake of discussion, that we are considering the BIT from the stability augmentation system (SAS) of a helicopter. Stability augmentation systems provide stability control for the three axes of the aircraft, namely roll, pitch, and yaw. Without loss of generality, we will consider just the roll axis. In evaluating the performance of the roll stability control in the SAS, we consider the health of at least three components: the roll control unit, the roll gyro, and an accelerometer. For our test scenario, we note that if the expected output of the control unit agrees with the actual, derived roll outputs from the accelerometer and roll gyro, then the system is functioning properly. On the other hand, if any two of these three elements disagree, a fault exists in one of the two units involved in the disagreement. This scenario can be represented with the Bayesian network shown in Fig. 7. Note that we are using the simplified form of the network (Fig. 6) based on the assumption the tests are designed correctly to evaluate the indicated faults.

TABLE I
SAS CONDITIONAL PROBABILITY TABLE

Pr(o(AC) | Accel, Ctrl)	0.975	Pr(o(AC) | ¬Accel, Ctrl)	0.867
Pr(o(AC) | Accel, ¬Ctrl)	0.867	Pr(o(AC) | ¬Accel, ¬Ctrl)	0.292
Pr(o(AG) | Accel, Gyro)	0.823	Pr(o(AG) | ¬Accel, Gyro)	0.756
Pr(o(AG) | Accel, ¬Gyro)	0.756	Pr(o(AG) | ¬Accel, ¬Gyro)	0.664
Pr(o(CG) | Ctrl, Gyro)	0.905	Pr(o(CG) | ¬Ctrl, Gyro)	0.787
Pr(o(CG) | Ctrl, ¬Gyro)	0.787	Pr(o(CG) | ¬Ctrl, ¬Gyro)	0.521

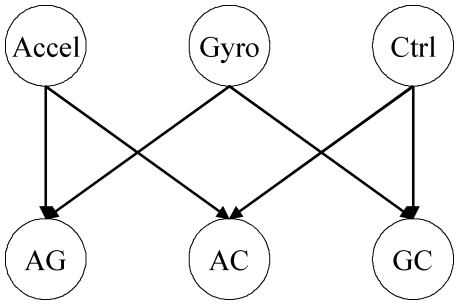

Fig. 7. Simple stability augmentation system BBN.

To interpret the elements of this network, Accel, Gyro, and Ctrl correspond to the diagnoses of whether the accelerometer, gyro, or control unit are faulty respectively. AG represents the observation associated with comparing the accelerometer output with the gyro output. AC compares the accelerometer output with the control output, and GC compares the gyro output with the control output.

Given this network structure, the next step is to define the conditional probability tables on the dependence links and the prior probabilities on the diagnoses. The prior probabilities are based upon the failure probabilities of the corresponding units. Suppose the SAS has been operating without failure for 250 h and the failure rates (assuming per thousand hours) for the accelerometer, gyro, and control unit are 30, 10, and 2, respectively. Since we must also include the absence of a fault in our probability calculations, we assign a "failure rate" of no-fault to be 958. Then, at this particular time, the relative failure probabilities (from (13) would be 0.301872, 0.277246, 0.118843, and 0.302039, respectively.[12]

To determine the conditional probabilities, we must first consider the instrument uncertainty. For the sake of simplicity, we will only derive the conditional probability table for AG. Without loss of generality, let o(AG) represent an observation that AG fails and ¬o(AG) represent an observation that AG passes. Similarly, let Accel and Gyro represent the logic states that the accelerometer or gyro have failed, respectively. Thus, ¬Accel would correspond to the accelerometer being fault free (likewise for ¬Gyro).

Given the absence of a measurement when first constructing the model, we might consider assuming the expected measurement [nom(y_f)] occurs at a decision limit for the test (i.e., z). Thus, we might set $y = +z$ (since we "normalized" the distribution to have zero-mean). What is interesting about this assumption, however, is that the mean of the distribution is shifted to the limit, so all probabilities of failure given the corresponding single faults reduce to 0.5. As mentioned, this is not meaningful for diagnosis because the test now has no discriminability. Instead, we will shift into the appropriate range for each of the outcomes considered. For example, for the test, AG, let's assume that the expected faulty measurement is at $z + 0.1$. Then, when considering the combined fault, we have

$$\Pr(o(\text{AG})|\text{Accel}) = \Pr(o(\text{AG})|\text{Gyro}) = 0.579$$
$$\Pr(o(\text{AG}|\text{Accel, Gyro}) = 1 - (1 - \Pr(o(\text{AG}|\text{Accel}))$$
$$(1 - \Pr(o\,(\text{AG})\,|\text{Gyro})) = 0.823$$

These probabilities would need to be adjusted based on the measurement uncertainty (defining the shape of the distribution) if assumptions other than the limits were chosen. Using the above approach, we can construct the conditional probability table given in Table I.

Using this approach of setting an expected measured value on the relevant side of the test limit, it is interesting to note that the probabilities of failure for each of the units (given no evidence) is {Accel: 0.302; Ctrl: 0.277, Gyro: 0.119; NF: 0.216}, which correspond to the failure probabilities (except for No Fault-for the Bayesian network, the probability for No Fault is given as Pr (NF|Accel, Ctrl, Gyro) and is derived from the fact that this probability is zero if any fault exists. Thus, the probability derived from the "failure rate" for No Fault is not required, except to determine the prior probabilities for the faults in the system. Suppose, we indicate that AC and AG both fail but CG passes. Logically, we would expect Accel to be faulty, and indeed, we find revised probabilities of {Accel: 0.539; Ctrl: 0.069, Gyro: 0.230; NF: 0.000}. Thus, we would conclude from the tests that Accel is the most likely to have failed.

XI. PROGNOSIS WITH DYNAMIC BAYESIAN NETWORKS

The traditional approach to fault diagnosis assumes tests are applied at a specific point in time from which one can infer the condition of the system under test and make a diagnosis. The problem of prognosis, while essentially an extension of diagnosis, is complicated by the fact that time becomes a significant factor in the analysis. In fact, one can represent the prognosis problem as a time series prediction problem in which one attempts to infer a future state from some sequence of past states.

It turns out that the Bayesian approach to diagnosis discussed in Section IX can be generalized in a straightforward way to address prognosis as well. In the most basic case, consider the

[12]The high number of significant digits is provided to demonstrate the subtle differences in probabilities at this point in the analysis.

state of the system as if it can be represented at some time t as a single random variable s_t. Assume, further, that the state at time $t+1$ depends only upon the system state at time t. Then we can represent the time series corresponding to the system state progression as a first-order Markov chain.

Several techniques exist for solving Markov chains (i.e., determining the probability vectors for being in the various states after some number of iterations) and can be classified as "steady-state" methods and "transient" methods. For steady-state methods, it is assumed that the underlying Markov chains are ergodic (i.e., for discrete-time Markov chains they are irreducible and aperiodic; for continuous-time Markov chains they are irreducible and a steady-state probability vector π exists). In the case of prognosis, ergodicity is a reasonable assumption. Steady-state methods can be classified further as direct or iterative. Examples of direct methods include Gaussian elimination and the Grassman algorithm, while iterative methods include the power method, Jacobi's method, and the Gauss–Seidel method [2].

Missing in our model is the fact that we do not have direct knowledge of the underlying state of the system. Specifically, we perform tests to observe conditions of the system, from which we infer the system state. Consequently, the basic Markov chain is not sufficient for our purpose—we need to differentiate between observable random variables and "hidden" (or unobservable) random variables. This leads to the concept of a hidden Markov model (HMM) [38].

Formally, an HMM $= \langle N, M, A, B_j, \pi \rangle$, where N is the number of states in the model (denote the states as $S = \{s_1, \ldots, s_N\}$), M is the number of distinct observation symbols per state (denote the symbols as $V = \{v_1, \ldots, v_M\}$), A is the state transition probability distribution $A = \{a_{ij}\} = \Pr(q_{t+1} = s_j | q_t = s_i)$, B_j is the observation probability distribution in state s_j, $B_j = b_j(k) = \Pr(v_k \text{ at } t | q_t = s_j)$, and π is the initial state distribution, $\pi = \{\pi_I\} = \Pr(q_0 = s_i)$ [38].

Several tools and algorithms exist for building and processing HMMs [38]. Even with the tools available for processing HMMs, we are still faced with a problem when focusing on prognosis. Specifically, HMMs assume a progression of single random variables (just like a Markov chain). One approach to handling this problem is to create separate HMMs for each test and couple the set of HMMs to a diagnostic model (such as the Bayesian network) to infer underlying state. There is, however, a better way.

The Markov chain and the HMM can be formulated as special cases of a graphical model first formalized by Dean and Kanazawa called the "dynamic Bayesian network" (DBN) [10]. DBNs have been studied further by Murphy who provided alternatives for representation, inference, and learning [27]. The purpose of a DBN is to model probability distributions over semi-infinite collections of random variables, Z_i, that progress according to some temporal model. Typically, the random variables are partitioned into three subsets indexed by time—$Z_t = (U_t, X_t, Y_t)$ where U_t is the set of inputs at time t, X_t is the set of hidden (i.e., unobservable) variables at time t, and Y_t is the set of outputs at time t. Then, given the set Z, a DBN is defined to be a pair $\langle B_1, B_{\rightarrow} \rangle$, where B_1 is a Bayesian network defining

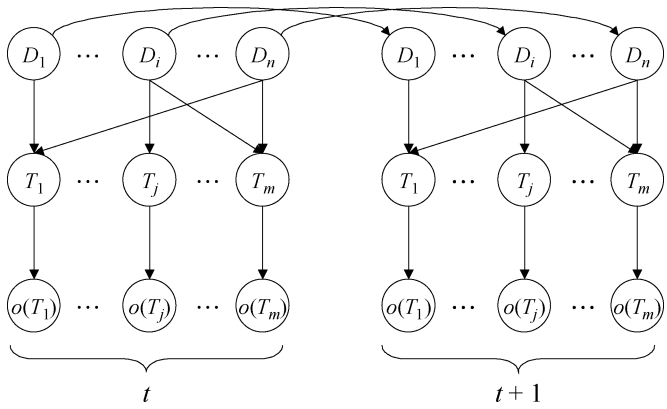

Fig. 8. Dynamic Bayesian network for prognosis.

the prior distribution $\Pr(Z_1)$, and $B_{\rightarrow}$ is a "two-slice" temporal Bayesian network defining the distribution $\Pr(Z_t | Z_{t-1})$ such that

$$\Pr(Z_t | Z_{t-1}) = \prod_{i=1}^{N} \Pr\left(Z_t^i | \mathrm{Pa}\left(Z_t^i\right)\right) \tag{42}$$

where Z_t^i is the ith node at time t, which could be a component of any of the partitions, and $\mathrm{Pa}(Z_t^i)$ are the parents of Z_t^i in the network. Of interest is the fact that the parents of a node $\mathrm{Pa}(Z_t^i)$ can either be from the same time slice or from the previous time slice (i.e., the resulting model is restricted to being a first-order Markov model). Of course, the general formulation of DBNs allows for higher order models simply by expanding the allowable set of parents to previous time slices.

To put this definition in the context of prognosis, we can construct a DBN for prognosis by "chaining" successive BBNs together. Under the first-order Markov assumption, we only need to represent two slices of the DBN and then "unroll" as necessary in processing the model. For example, Fig. 8 shows how to link the BBNs in sequence. Note that only the diagnoses are linked through time since they change state directly. Changes in observation state are derived from the underlying state changes in the system. This approach is distinct from the HMM that links observations together. Further, the DBN supports multiple random variables, so it can represent the different diagnoses, tests, and observations directly.

To perform inference with the DBN (and thereby predict future states), one approach is fairly straightforward. First, infer the current state (i.e., the state in the current time slice) from the test observations using the algorithm described in Section IX. Next, "unroll" the DBN to the desired number of time slices (assuming the state progressions occur in discrete time steps—DBNs can handle continuous time, but the computation is more complex). Then, propagate beliefs through time by observing that

$$\Pr\left(D_i^{t+1}\right) = \Pr\left(D_i^{t+1} | D_i^t\right) \Pr\left(D_i^t\right) + \Pr\left(D_i^{t+1} | \neg D_i^t\right) \Pr\left(\neg D_i^t\right). \tag{43}$$

In fact, given the assumption that only diagnoses progress in state through time and that a diagnosis only depends upon itself in the previous time step, this part of the model reduces to a

simple Markov chain, which can be either discrete time or continuous time.

Key to constructing the DBN is defining the temporal transition probabilities. In the simplest case, failure probabilities based on (9) can be used. When better information is available (e.g,. based on historical data), probabilities derived from this information can be used. The point is that the DBN is fully general and can be adapted to available knowledge about the system being analyzed. Theoretically, causal relationships between faults (i.e., a fault at time step t causes another fault to occur at time step $t + 1$) can be represented directly with the DBN as well (even though such models are rarely useful).

XII. Summary

In this paper, we discussed a formal approach to predicting probability of false indication based on instrument uncertainty and using this information as a means of setting test uncertainty in diagnosis and prognosis with BIT. The diagnostic approach is based on Bayesian belief networks and incorporates information on failure probability, instrument uncertainty, and the predictions for false indication. Prognosis is performed using an extension of the Bayesian belief network, called a dynamic Bayesian network to model changes over time. This is the first time such a unified Bayesian view of the test, diagnosis, and prognosis has been presented, especially in the context of BIT. The advantage to the discussed method is that it provides a formally consistent and theoretically sound approach to diagnosis and prognosis that can be adapted and matured as better estimates of the associated probabilities become available.

By applying Bayesian inference and estimating probabilities from estimated or predicted reliability information and expected instrument uncertainty, resulting diagnoses accurately reflect the current understanding of test and failure uncertainty, thereby providing a more accurate picture of the state of the underlying system. Recalling our assertion that BIT is just a set of tests that happen to be embedded in the system, improving the accuracy of BIT diagnostics is particularly important given the history of BIT systems for having high false alarm rates.

Acknowledgment

The authors would like to thank the members of the IEEE SCC20 Diagnostic and Maintenance Control Subcommittee for reviewing several critical parts of this paper. Specifically, we thank T. Wilmering, G. Bowman, L. Orlidge, and Li Pi Su for their comments as the ideas were formulated. This work was a direct outgrowth of the standardization work being performed under the DMC for testability/diagnosability prediction and diagnostic maturation. We also thank the anonymous reviewers of this paper for several comments that helped tighten the presentation significantly.

References

[1] B. Baygün, "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 688–703, May 1995.

[2] G. Bolch, S. Greiner, H. de Meer, and K. Trivedi, *Queuing Networks and Markov Chains*. New York: Wiley , 1998.

[3] *Calibration Philosophy in Practice*, 2nd ed. Everett, WA: Fluke Corporation, 1994.

[4] H. Castrup, "Risk-based control limits," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2001, 4B-b.

[5] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*. New York: Wiley, 1998.

[6] A. R. Colclough, "Foundational positions in statistics and their relationship to measures of uncertainty," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 1992, pp. 394–404.

[7] S. F. Cook, "The application of Taguchi design in the calibration laboratory," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 1993, pp. 486–499.

[8] G. D'Agostini, "Probability and Measurement Uncertainty in Physics—A Bayesian Primer," Deutsches Electronen Synchnotron (DESY), DESY 95-242, December 1995.

[9] D. Deaver, "An application of the guide to measurement uncertainty," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2000, pp. 325–334.

[10] T. Dean and K. Kanazawa, "A model for reasoning about persistence and causation," *Artif. Intell.*, vol. 93, no. 1–2, pp. 1–27, 1989.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.

[12] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA: Academic, 1990.

[13] A. Hero, "Signal detection and classification," in *The Digital Signal Processing Handbook*, V. Madisetti and D. B. Williams, Eds. Boca Raton, FL: CRC, 1997.

[14] E. J. Horvitz, J. S. Breese, and M. Henrion, "Decision theory in expert systems and artificial intelligence," *Int. J. Approximate Reasoning*, vol. 2, pp. 247–302, 1988.

[15] *The Authoritative Dictionary of IEEE Standard Terms*, IEEE Standard 100-2000, 7th Edition, IEEE Standards Association Press, Piscataway, NJ, 2000.

[16] *IEEE Standard for Artificial Intelligence Exchange and Service Tie to All Test Environments*, IEEE Standard 1232-2002, IEEE Standards Association Press, Piscataway, NJ, 2002.

[17] *Draft Standard Testability and Diagnosability Characteristics and Metrics*, IEEE Standard P1522/D6.0, IEEE Standards Association Press, Piscataway, NJ, Oct. 2002.

[18] *Industrial Automation Systems and Integration—Product Data Representation and Exchange—Part 11: The EXPRESS Language Reference Manual*, ISO 10303-11:1994, Int. Organization for Standardization, Geneva, Switzerland, 1994.

[19] D. Jackson, "Calibration intervals—new models and techniques," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2002, 2D-a.

[20] ——, "Measurement risk analysis methods," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2005, 3B-3.

[21] M. Kaveh, H. Wang, and H. Hung, "On the theoretical performance of a class of estimators of the number of narrow-band sources," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-35, no. 9, pp. 1350–1352, Sep. 1987.

[22] J. Klion, *Practical Electronic Reliability Engineering*. New York: Van Nostrand Reinhold, 1992.

[23] D. Maupin, "Missile test confidence via test systems certification and associated data analysis," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2003, 6B-a.

[24] *Definition of Terms*, MIL-STD 1309C, Nov. 18, 1989.

[25] *Testability Program for Electronic Systems and Equipments*, MIL-STD 2165, Jan. 26, 1985.

[26] T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[27] K. Murphy, "Dynamic bayesian networks: representation, inference, and learning," Ph.D. Dissertation, Department of Computer Science, Univ. of California, Berkeley, 2002.

[28] R. Nicholas, "Consumer risk for non-Gaussian measurand distributions," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2002, 3D-b.

[29] ——, "Measurement decision risk simplified," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 1999, pp. 175–205.

[30] ——, "Measurement uncertainty misconceptions: 'MUMS' the problem," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2000, pp. 343–351.

[31] ——, "Non-linear propagation of uncertainty: part 1. Independent Gaussian densities," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2001, 4B-a.

[32] M. Nyberg, "Model based fault diagnosis: methods, theory and automotive engine applications," Ph.D. Dissertation, Department of Electrical Engineering, Linköping Univ., Linköping, Sweden, 1999.

[33] M. Nyberg and L. Nielsen, "Design of a complete FDI system based on a performance index with application to an automotive engine," in *Proc. IFAC Fault Detection, Supervision, and Safety for Technical Processes*, Hull, UK, pp. 812–817.
[34] B. Parry, "Applications of statistics and uncertainty to dimensional metrology," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 1995, pp. 43–58.
[35] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1991.
[36] S. D. Phillips, W. T. Estler, M. S. Levenson, and K. R. Eberhardt, "Calculation of measurement uncertainty using prior information," *J. Research of the National Institute of Standards and Technology*, vol. 103, no. 6, pp. 625–632, Nov.–Dec. 1998.
[37] F. Provost and T. Fawcett, "Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions," in *Proc. Third Int. Conf. Knowledge Discovery and Data Mining (KDD-97)*, Huntington Beach, CA, 1997, pp. 43–58.
[38] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[39] S. G. Rabinovich, "Methods for accounting for dependencies between components of measurement uncertainty," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 1994, pp. 370–378.
[40] L. M. Rudner, "The classification accuracy of measurement detection theory," in *Proc. National Council Measurement in Education*, Chicago, IL, Apr. 2003.
[41] ——, "An examination of decision-theory adaptive testing procedures," in *Proc. American Educational Research Association*, New Orleans, LA, Apr. 2002.
[42] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Upper Saddle River, NJ: Pearson Education, 2002.
[43] J. W. Sheppard and M. Kaufman, "IEEE 1232 and P1522 standards," in *AUTOTESTCON 2000 Conf. Record*, Sep. 2000, IEEE Press, pp. 388–397.
[44] J. W. Sheppard and W. R. Simpson, "Managing conflict in system diagnosis," in *Research Perspectives and Case Studies in System Test and Diagnosis*. Norwell, MA: Kluwer, 1998.
[45] H. M. Shertukde and Y. Bar-Shalom, "Detection and estimation for multiple targets with two omnidirectional sensors in the presence of false measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 749–763, May 1990.
[46] W. R. Simpson and J. W. Sheppard, *System Test and Diagnosis*. Norwell, MA: Kluwer, 1994.
[47] K.-D. Sommer and M. Kochsiek, "Uncertainty evaluation and modeling of measurements," in *Proc. Measurement Science Conf.*, Anaheim, CA, Jan. 2003, 4C-b.
[48] B. N. Taylor and C. E. Kuyatt, "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results," NIST, Technical Rep. 1297, 1994.
[49] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. San Diego, CA: Academic, 1999.
[50] J. S. Tripp and P. Tcheng, "Uncertainty Analysis of Instrument Calibration and Application,", NASA/TP-1999-209545, Oct. 1999.

John W. Sheppard (M'86–SM'97) received the B.S. degree in computer science from Southern Methodist University, Dallas, TX, in 1983 and the M.S. and Ph.D. degrees in computer science from Johns Hopkins University, Baltimore, MD, in 1990 and 1997, respectively.

He is a Research Fellow with ARINC Incorporated, Annapolis, MD, and a Lecturer in the Department of Computer Science, Johns Hopkins University. He has written over 100 publications in artificial intelligence and diagnostics, including two books. He has been active in the IEEE Computer Society and the IEEE Standards Association, being Secretary and past Chair of the Diagnostic and Maintenance Control subcommittee of IEEE Standards Coordinating Committee 20 on Test and Diagnosis for Electronic Systems.

Mark A. Kaufman (M'87) received the B.S.E.E. degree from California State Polytechnic University, Pomona, and the M.S. degree in scientific and engineering management from West Coast University, Los Angeles, CA.

He is a Testability and Diagnostics Engineer with the Naval Surface Warfare Center, Corona Division, Corona, CA. He has been working in the metrology and test assessment areas for the Navy for over 20 years. He has held several offices within the IEEE, including being the current Chair of the Diagnostic and Maintenance Control subcommittee of IEEE Standards Coordinating Committee 20 on Test and Diagnosis for Electronic System, past Vice Chair of SCC20, and Vice Chair of the AUTOTESTCON conference. In addition, he has served as President and Chairman of the Board of the Measurement Science Conference.