Diagnosing Multiple Persistent and Intermittent Faults

Johan de Kleer
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304 USA
dekleer@parc.com

Abstract

Almost all approaches to model-based diagnosis presume that the system being diagnosed behaves non-intermittently and analyze behavior over a small number (often only one) of time instants. In this paper we show how existing approaches to model-based diagnosis can be extended to diagnose intermittent failures as they manifest themselves over time. In addition, we show where to insert probe points to best distinguish among the intermittent faults those that best explain the symptoms and isolate the fault in minimum expected cost.

1 Introduction

Experience with diagnosis of automotive systems and reprographic machines [Fromherz *et al.*, 2003] shows that intermittent faults are among the most common and the most challenging kinds of faults to isolate. The notion of intermittency is a hard-to-define concept, so we first describe it intuitively before defining it more formally. A system consists of a set of components. A faulty component is one which is physically degraded such that it will not always function correctly. For example, a faulted resistor may no longer conduct the expected current when a non-zero voltage is applied across it. A worn roller in a printer may no longer grip the paper consistently thereby causing intermittent paper jams. In the case of a worn roller, it usually operates correctly but will infrequently slip and cause a paper jam. We therefore associate two probabilities with each potentially intermittent component: (1) the probability that the actual component deviates from its design such that it may exhibit a malfunction, and (2) the probability that the faulted component functions correctly when observed. For example, the probability that a roller being worn might be 10^{-5} while the probability of a worn roller actually malfunctioning might be 0.01.

We do not model the continuous behavior of a system over the time. Instead, the system is viewed over a sequence of observation events. For example, a piece of paper is fed into a printer and it may be observed to jam. A test-vector is applied to a combinational digital circuit and its output signals subsequently observed.

Intermittency can arise from at least two sources. If the system can be modeled at a more detailed (i.e., less abstract) level, apparent intermittency can disappear. For example, two wires may be shorted, making it appear as if a gate is intermittent, when in fact there is an unmodeled and unwanted bridge connection. The second type of intermittency arises from some stochastic physical process which is intermittent at any level of detail. This paper focuses on this second type of intermittency.

In this paper we show how to troubleshoot devices containing any combination of intermittent or persistent faults. Priors are assumed for all faults. The device may contain any number of faults. The approach both computes the posterior probabilities after observations are made as well as the additional probes needed to efficiently isolate the fault(s) in the device. Reprographic and automotive systems raise many modeling complexities, so we present our approach to isolating intermittent faults in the context of logic systems. See [de Kleer *et al.*, 2008] for the application of these concepts to reprographic engines. For benchmarks, we draw on the circuits in the widely available ISCAS85 benchmarks.

Diagnosis of intermittent faults is a broad challenge. In this paper we make the following presuppositions: (1) One observed variable per time instant, (2) All inputs are always known, but change over time, (3) Diagnosis starts with an observation in which a faulty output is observed, (4) No memory or energy storage components, (5)No measurement error, (6) Making measurements does not change the system, (7) Components fail independently (dependent failures may lead to misdiagnosis), (8) Underlying stochastic process is stationary. None of these limitations present a fundamental obstacle to our model-based approach, but are topics for future research.

2 GDE probability framework

We presume the consistency based framework as described in [de Kleer and Williams, 1987; de Kleer *et al.*, 1992]. For most of this paper we assume ‘weak’ fault models; or the IAB (ignorance of abnormal behavior) assumption. No fault models are presumed. Later in the paper we show how this assumption can be relaxed. Time is expressed easily in this formalism. The model of an inverter can be written as:

$$INVERTER(x) \rightarrow \Big[\neg AB(x) \rightarrow [in(x,t) = 0 \equiv out(x,t) = 1]\Big].$$

When ambiguous this paper represents the value v of variable x at time t as $T(x = v, t)$. Time is a sequence of instants t_0, t_1, ... The probability of X at time t is represented as $p_t(X)$.

2.1 Updating diagnosis probabilities

Components are assumed to fail independently. Therefore, the prior probability a particular diagnosis $\mathcal{D}(Cp, Cn)$ is correct is:

$$p_1(\mathcal{D}) = \prod_{c \in C_p} p(c) \prod_{c \in C_n} (1 - p(c)), \tag{1}$$

where $p(c)$ is the prior probability that component c is faulted.

The posterior probability of a diagnosis $\mathcal{D}$ after an observation that x has value v at time t is given by Bayes' Rule:

$$p_t(\mathcal{D}|x = v) = \alpha p_t(x = v|\mathcal{D}) p_{t-1}(\mathcal{D}). \tag{2}$$

$p_{t-1}(\mathcal{D})$ is determined by the preceding measurements and prior probabilities of failure. α is a normalizing term that is identical for all $p(\mathcal{D})$ and thus need not be computed directly. The only term remaining to be evaluated in the equation is the observation function $p_t(x = v|\mathcal{D})$:

$p_t(x = v|\mathcal{D}) = 0$ if $\mathcal{D}, SD, OBS, T(x = v, t)$ are inconsistent,

else,

$p_t(x = v|\mathcal{D}) = 1$ if $T(x = v, t)$ follows from $\mathcal{D}, SD, OBS$

If neither holds, $p_t(x_i = v_k|\mathcal{D}) = \epsilon_{ik}$. Various ϵ-policies are possible [de Kleer, 2006] and a different ϵ can be chosen for each variable x_i and value v_k. Typically, $\epsilon_{ik} = \frac{1}{m}$. This corresponds to the intuition that if x ranges over m possible values, then each possible value is equally likely. In digital circuits $m = 2$ and thus $\epsilon = 0.5$.

In the conventional framework, observations that differ with predictions yield conflicts, which are then used to compute diagnoses. Consider the full adder digital circuit of Figure 1. Suppose all the inputs to the circuit are 0, and c_o is measured to be 1. This yields one minimal conflict:

$$AB(\text{A1}) \vee AB(\text{A2}) \vee AB(\text{O1}).$$

The single fault diagnoses are: [O1], [A1] and [A2] (for brevity sake we represent the diagnosis $\mathcal{D}(\{f_1, f_2, \ldots\}, C_n)$ by the faulty components: $[f_1, f_2, \ldots]$).

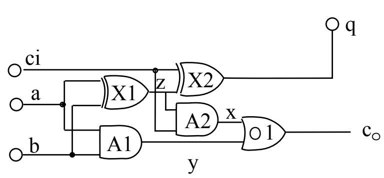

Figure 1: Full adder. This circuit computes the binary sum of c_i (carry in), a and b; q is the least significant bit of the result and c_o (carry out) the high order bit.

3 Extensions for support intermittent faults

In the conventional framework, $p(c)$ is the prior probability that component c is faulted. In the new framework, two probabilities are associated with each component: (1) $p(c)$ represents the prior probability that a component is intermittently faulted, and (2) $g(c)$ represents the conditional probability that component c is behaving correctly when it is faulted.

In the intermittent case, the same sequential diagnosis model and Bayes' rule update applies. However, a more sophisticated ϵ-policy is needed. Note that an ϵ-policy applies only when a particular diagnosis neither predicts $x_i = v_k$ nor is inconsistent with it. Consider the case where there is only a single fault. As we assume all inputs are given, the only reason that a diagnosis could not predict a value for x_i is if the faulted component causally affects x_i. Consider the single fault diagnosis $[c]$. If c is faulted, then $g(c)$ is the probability that it is producing a correct output. There are only two possible cases: (1) c is outputting a correct value with probability $g(c)$, (2) c is producing a faulty value with probability $1 - g(c)$. Therefore, if $x_i = v_k$ follows from $[], SD, OBS$ and ignoring conflicts:

$$p_t(x_i = v_k|[c]) = g(c).$$

Otherwise,

$$p_t(x_i = v_k|[c]) = 1 - g(c).$$

These two ϵ equations are equivalent to (no need to ignore conflicts): $p_t(x_i = v_k|\mathcal{D}) =$

$$\begin{array}{l} g(c) \quad\quad \text{if } [], OBS, SD \vdash T(x_i = v_k, t) \\ + \\ 1 - g(c) \text{ if } [c], \overline{o}(c), OBS, SD \vdash T(x_i = v_k, t) \end{array}$$

$\overline{o}(c)$ the incorrect (output negated) model for c.

Returning to the full adder example, if all components are equally likely to fail a priori, then $p([\text{A1}]), p([\text{A2}]), p([\text{O1}]) = \frac{1}{3}$. Suppose $y = 0$ is observed next. In the conventional framework this has no consequence, as the observation is the same as the prediction. However, in the intermittent case, such 'good' observations provide significant diagnostic information. Table 1 illustrates the results if c_o is first observed to be faulty, $g(c) = 0.9$ for all components, O1 is the actual fault, and y is observed continuously. As sampling continues, $p([\text{A1}])$ will continue to drop. An intelligent probing strategy would switch to observing x at time 4. As y is insensitive to any error at [O1], no erroneous value would ever be observed.

Failing components closer to the input are easier to isolate as there are fewer confounding effects of intermediate components through which the signals must propagate before reaching a suspect component. Table 2 presents a sequence of probes which isolates a fault in O1. Although O1 can never be identified as failing with probability 1, the table shows that repeated observations drive O1's probability asymptotically to 1.

It is important to note that the single fault assumption does not require an additional inference rule. The Bayes' rule update equation does all the necessary work. For example, when $c_o = 1$ is observed, the single fault diagnoses [X1] and [X2] both predict $c_o = 0$ and thus are both eliminated from consideration by the update equation. As an implementation detail, single faults can be computed efficiently as the intersection of all conflicts found in all the samples.

Table 1: Probabilities of component failure over time. Row 0 are the prior probabilities of the components intermittently failing, row 1 are the probabilities conditioned on knowing the system has a fault, and row 2 are the probabilities conditioned on $c_o = 1$ and so on.

i	A1	A2	O1	X1	X2	obs
0	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	
1	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	
2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	0	$c_o = 1$
3	0.310	0.345	0.345	0	0	$y = 0$
4	0.288	0.356	0.356	0	0	$y = 0$
5	0.267	0.366	0.366	0	0	$y = 0$
6	0.247	0.376	0.376	0	0	$y = 0$
7	0.228	0.386	0.386	0	0	$y = 0$
8	0.209	0.395	0.395	0	0	$y = 0$
9	0.192	0.403	0.403	0	0	$y = 0$
10	0.177	0.411	0.411	0	0	$y = 0$
11	0.162	0.419	0.419	0	0	$y = 0$
12	0.148	0.426	0.426	0	0	$y = 0$

4 Probing strategy

Consider again the example of Figure 1 and Table 1. Repeatedly measuring $y = 0$ will monotonically drive down the posterior probability of A1 being faulted. Choosing probes which drive down the posterior probability of component faults is at the heart of effective isolation of intermittent faults.

A probing strategy better for isolating some faults, might perform worse on others. To fairly compare proposed strategies we compare them by their expected diagnostic cost [de Kleer *et al.*, 1991]:

$$\text{DC}(D, A) = \sum_{s \in S} p(s) c(D, s, A),$$

where S is the set of diagnostic situations, $p(s)$ is the probability of a particular diagnostic situation, D is the device, A is the algorithm used to select probes, and $c(D, s, A)$ is the cost of diagnosing device D in diagnostic context s with algorithm A. In limited cases DC can be exactly calculated. Generally DC must be estimated by sampling, for example, with S consisting of pairs (f, v) where f is the fault and v is a set of device input-output pairs which cause D to manifest the symptom.

To successfully isolate the faulty intermittent component, the diagnoser must propose probes. In conventional model-based diagnosis, (following the myopic strategy) the best probe to make next is the one which minimizes,

$$H = -\sum_{\mathcal{D} \in DIAGNOSES} p(\mathcal{D}) \log p(\mathcal{D}). \qquad (3)$$

Even though the definition of $p(c)$ is different than in conventional model-based diagnosis, we still want to probe those variables which maximize the likelihood of distinguishing among the possible intermittent faults. The goal is the same — to lower the posterior probability of failure for all but one of the components.

Table 2: Probes required to isolate the failing O1.

i	A1	A2	O1	X1	X2	obs	H
2	0.345	0.310	0.345	0	0	$x = 0$	1.58
3	0.321	0.321	0.357	0	0	$y = 0$	1.58
4	0.332	0.299	0.369	0	0	$x = 0$	1.58
5	0.309	0.309	0.382	0	0	$y = 0$	1.58
6	0.319	0.287	0.394	0	0	$x = 0$	1.57
7	0.297	0.297	0.407	0	0	$y = 0$	1.57
8	0.306	0.275	0.419	0	0	$x = 0$	1.56
9	0.284	0.284	0.432	0	0	$y = 0$	1.55
10	0.292	0.263	0.445	0	0	$x = 0$	1.55
20	0.223	0.201	0.576	0	0	$x = 0$	1.41
40	0.107	0.097	0.796	0	0	$x = 0$	0.93
55	0.052	0.052	0.896	0	0	$y = 0$	0.59
56	0.052	0.047	0.901	0	0	$x = 0$	0.57
80	0.016	0.014	0.970	0	0	$x = 0$	0.23
90	0.010	0.009	0.982	0	0	$x = 0$	0.15
100	0.006	0.005	0.989	0	0	$x = 0$	0.10
120	0.002	0.002	0.996	0	0	$x = 0$	0.04
127	0.001	0.001	0.997	0	0	$y = 0$	0.03
128	0.001	0.001	0.998	0	0	$x = 0$	0.03

Table 1 suggests a very simple probing strategy from which we can compute an upper bound on optimal DC. The simplest strategy is to pick the lowest posterior probability component and repeatedly measure its output until either a fault is observed or its posterior probability drops below some desired threshold.

The number of samples needed depends on the acceptable misdiagnosis threshold e. Diagnosis stops when one diagnosis has been found with posterior probability $p > 1 - e$. To compute the upper bound on DC: (1) we take no advantage of the internal structure of the circuit; (2) we presume that every measurement can exonerate only one component; (3) we are maximally unlucky and never witness another incorrect output.

Let n be the number of components c_i and o_i the number of samples of the output of c_i. $p_1([c_i])$ is derived from the priors. The Bayes' update is:

$$p_t([c_i]) = \alpha p_t(x = v | [c_i]) p_{t-1}([c_i]).$$

Notice that $p_t(x = v | [c_i]) = g(c_i)$ when the output of component c_i is observed. After t samples:

$$p_t([c_i]) = \alpha g(c_i)^{o_i} p_1([c_i]),$$

where, $\Sigma o_i = t$, and,

$$\alpha = \frac{1}{\Sigma_i g(c_i)^{o_i} p_1([c_i])}.$$

Splitting the misdiagnosis threshold evenly across all components, we want to pick o_i such that:

$$\alpha g(c_i)^{o_i} p_1(c_i) < \frac{e}{n}.$$

We need to solve for o_i in:

$$\frac{g(c_i)^{o_i} p_1([c_i])}{\alpha} = \frac{e}{n}.$$

If o_i is sufficiently large, then $\alpha \approx 1 - p([c_i])$ so we can solve for o_i:

$$o_i = \frac{\log e - \log n - \log p_1([c_i]) + \log(1 - p_1([c_i]))}{\log g(c_i)}.$$

In the case of the full adder example all priors are equal, $g = 0.9$, and $e = 0.1$. So based on the preceding equation, $o_i = 28.1$. Hence, a worst case strategy is to measure each point 29 times. There are only 4 informative probe points. If we utilize a strategy of sequentially probing each such probe point 29 times this would yield an upper bound of 126. However, the last probe point need not be measured since the strategy would have failed to identify a fault earlier, and hence because the device has a fault, the probing at the final point is at the fault. There is no need to measure the non-symptom output because that cannot provide information on any single fault. Therefore, the upper bound is 58. This worst case is corroborated in Table 2 where it takes 56 probes to isolate O1 to with probability about 0.9. Faults in A1 and A2 would be detected with far fewer probes.

Rarely can we compute a bound on a particular diagnostic algorithm so neatly. In most cases the diagnostic cost of a particular algorithm can only be evaluated empirically. Computing DC using the minimum entropy criterion (equation 3) on a randomly chosen set of vectors, and assuming all faults are equally likely, $g = 0.9$, $e = 0.1$, results in an expected cost of $DC = 22.2$ with observed error 0.05. This DC is far smaller than the 58 bound calculated earlier, because faults in A1 and A2 evaluated in far fewer probes and DC is an average. As the manifestation of intermittent faults is inherently random, there will always be a chance of misdiagnosis. The observed error of 0.05 is better than our desired 0.1. In general, the observed error will be close to the theoretical error, but it will vary because of specific characteristics of the circuit being analyzed and the sequence of random numbers generated in the simulation.

5 Statistical nature of probing strategy

Fault isolation in the intermittent case is more subtle than in the deterministic case. It will always make mistakes. As an example, consider the very simple two buffer sequence of Figure 2. Assume A and B are equally likely to fault and $g = 0.9$. As we assume the input is given, and that a faulty ob-

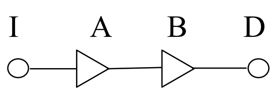

Figure 2: Two buffers in sequence. One is intermittent.

servation was observed, there is only one measurement point, the output of A that can provide any useful information.

In this simple case there is no choice of measurement point. First, consider the case where B is intermittent. The output of A will always be correct. Following the same line of development as earlier, after n measurements:

$$p_{n+1}([A]) = \frac{g^n}{g^n+1}, \quad p_{n+1}([B]) = \frac{1}{g^n+1}.$$

To achieve the probability of misdiagnosis to be less than e:

$$\frac{g^n}{g^n+1} < e. \quad n = \frac{\log e}{\log g},$$

As $g^n \approx 0$. With $g = 0.9$ and $e = 0.01$, $n = 43.7$. Now consider the case where A is faulted. Until the fault is manifest, A will be producing a good value and the results will look the same as the case where B is faulted. Roughly, the fault will be manifest in, $\frac{1}{2-2g} = 4.5$ samples. Thus, the expected diagnostic cost of the circuit of Figure 2 is roughly 24.1. Our algorithm obtains a similar result. Again we see that it is considerably more expensive to isolate intermittents further from the input(s).

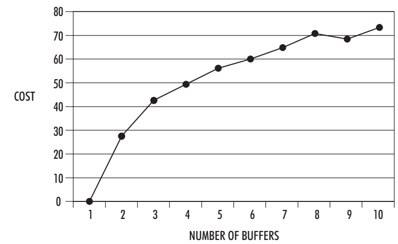

Figure 3: DC of isolating an intermittent buffer in a row of n.

Figure 3 plots diagnostic cost (obtained from one simulation) of isolating a single intermittent buffer in a sequence of buffers using the minimum entropy strategy. The graph illustrates that, as is the case with persistent faults, groups of components are simultaneously exonerated so that DC grows roughly logarithmic with circuit size. Note that the minimum entropy strategy is much better than the simple approach presented earlier.

If inputs vary, a slightly different approach is needed. It is computationally impractical to find the variable to measure next which provides the maximum expected information gain over all possible system inputs. Instead, we employ a heuristic to ensure measuring the best variable which is directly adjacent to some remaining suspect component. This avoids the possible suboptimal situation where the particular new inputs make the proposed measurement insensitive to any of the faults (e.g., suppose the symptomatic value propagated through a known good and-gate, and if the other input of this and-gate were 1 when the fault was first observed and was 0 in later inputs, measuring the output of that and-gate would be useless for those input vectors, and it would take a many samples to isolate the faulty component).

6 Benchmarks

The approach has been fully implemented on GDE/HTMS architecture [de Kleer and Williams, 1987; de Kleer, 1992]. Table 3 lists the diagnostic cost of diagnosing circuits in the ISCAS85 benchmark. The first column is the ISCAS85 circuit name. The second column is the number of distinct gates of the circuit. The remaining columns are computed by hypothesizing all possible single intermittent faults. For each fault, a single (constant) test-vector is found which gives rise to an observable symptom for one of the outputs. For each row, for each component, two simulations are performed (with 1 switched to 0, and a 0 switched to a 1). The next column is the DC to isolate the intermittent component with misdiagnosis probability $e = 0.1$, $g = 0.9$ and all faults equally likely apriori. The last column is the observed error rate. Not surprisingly, the number of samples needed to isolate intermittent faults is far more than is needed for persistent faults.

Table 3: Summary of results on the ISCAS85 benchmarks.

circuit	components	noni	i cost	error	i cost	error
		cost	e=0.1		e=0.01	
c17	6	2.7	29	.07	45	.005
c432	160	5.5	94	.08	119	.03
c499	201	5.4	106	.09	126	.04
c880	384	5.3	79	.09	104	.02

7 Learning $g(c)$ for single faults

Although the prior probability of component failure can be estimated by the manufacturer or previous experience with that component in similar systems, $g(c)$ typically varies widely. Therefore, we estimate the $g(c)$'s instead of presuming their values. Estimating $g(c)$ requires significant additional machinery which we only describe here in the single fault case. The generalization to multiple faults is analogous to the one used for reprographic modules [de Kleer *et al.*, 2008] and requires a maximization step. Fortunately, the single fault case is easy to analyze.

Our approach does not depend on isolating c for any observation. We define $G(c)$ to be the number of times c could have been working correctly, and $B(c)$ as the number of times c has been observed working incorrectly when faulted. Corroboratory measurements which c cannot influence are ignored (conflicting measurements exonerate any component which cannot influence it, in which case $g(c)$ is no longer relevant). We estimate $g(c)$:

$$g(c) = \frac{G(c)}{G(c) + B(c)},$$

if either $G(c)$ or $B(c)$ is 0, then we define $g(c) = 0.5$. The situations in which c is working incorrectly are straightforward to detect — they are simply the situations in which the Bayes' update equation utilizes $p_t(x_i = v_k|\mathcal{D}) = 1 - g(c)$. The cases in which c is working correctly requires additional inferential machinery. Consider again the example of Figure 1 where all the inputs are 0, and the expected $c_o = 0$ is observed. Or-gate O1 cannot be behaving improperly because its inputs are both 0, and its observed output is 0. And-gate A1 cannot be behaving improperly because its inputs are both 0, and its output must be 0, as O1 is behaving correctly and its output was observed to be 0. Analogously, and-gate A2 cannot be faulted. However, we have no evidence as to whether X1 is behaving improperly or not, because if X1 were behaving improperly, its output would be 1, but that cannot affect the observation because and-gate A2's other input is 0. X2 cannot causally affect c_o. If we observe $c_o = 0$ at an instant in which all the inputs are 0, we have learned that A1, A2 and O1 cannot be misbehaving alone, and we learn nothing about the faultedness of X1 and X2. Hence, G(A1), G(A2) and G(O1) are incremented. All other counters are left unchanged. In summary, as a consequence of observing $c_o = 0$, G(A1), G(A2), and G(O1) are incremented. The net consequence will be that the diagnostician may have to take (slightly) more samples to eliminate A1, A2 or O1 as faulted.

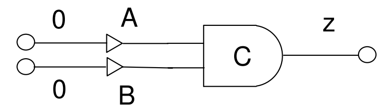

Figure 4: A three component circuit that computes the conjunction of its inputs.

Table 4: Result of probing strategy using learning, with A1 failing at $i = 11$.

i	A1	A2	O1	X1	X2	obs	H
2	0.40	0.20	0.40	0	0	x=0	1.52
3	0.25	0.25	0.50	0	0	y=0	1.50
4	0.27	0.18	0.55	0	0	x=0	1.44
5	0.20	0.20	0.60	0	0	y=0	1.37
6	0.21	0.16	0.63	0	0	x=0	1.31
7	0.17	0.17	0.67	0	0	y=0	1.25
8	0.17	0.14	0.69	0	0	x=0	1.20
9	0.14	0.14	0.71	0	0	y=0	1.15
10	0.15	0.12	0.73	0	0	x=0	1.11
11	1.00	0	0	0	0	y=1	0

Consider the example of Figure 4. At $t = 0$, the inputs are both 0, and the output is observed to be $z = 0$. Neither the single fault A nor B can influence $z = 0$, therefore the counters for A and B are left unchanged. However, C can influence the output and therefore $G(C)$ is incremented.

More formally, in response to an observation $x_i = v_k$, $G(c)$ is incremented if a different value for x_i logically follows from the negated model $\overline{o}(c)$; c causally influences the observation and $G(C)$ is incremented.

Table 4 presents the sequence of probes to isolate a fault in A1 of the full adder. A1's fault is manifest at $i = 11$. Notice that the conditional probabilities shift differently as compared to the non-learning case (Table 1). As $g(c) = 0.5$ is assumed without any evidence, measuring $x = 0$ immediately drops A2's probability to 0.5 compared to the other two diagnoses. When using learning initial $g(c)$'s need to be chosen with some care as they significantly influence DC. Learning significantly increases DC. For example, DC for the full adder rises from 22 to 39 for the same error rate (with no knowledge of initial $g(c)$'s).

Table 5: 4 combinations that need consideration for intermittent fault $[X1, X2]$

X2:X1	incorrect behavior	correct behavior
incorrect behavior	$(1 - g(X1))(1 - g(X2))$	$g(X1)(1 - g(X2)$
correct behavior	$(1 - g(X1))g(X2)$	$g(X1)g(X2)$

8 Multiple faults

Bayes' rule applies as before, but evaluating the observation function is more complex. Consider multiple intermittent faults. For every diagnosis we have to consider every possible combination of faulty/correct behavior. If $|C_p| = n$, then 2^n cases need to be considered. (The single-fault case $n =$ requires 2 terms as described earlier.) Consider our full adder example and the candidate diagnosis $[X1, X2]$. Table 5 lists the 4 combinations to be analyzed. We see that: $p(c_0 = 1|[X1, X2]) = g(X1)(1 - g(X2)) + g(X1)(1 - g(X2))$.

More formally: Let $M(C_p, C_n, S, r, t)$ be the predicate where $S \subset C_p \subset$ COMPS which holds iff r holds at time t if all components in S are producing an incorrect value. More formally,

$$M(C_p, C_n, S, r, t) = \\ [C_n], OBS, SD, \bigwedge_{c \in S} \overline{o}(c), \bigwedge_{c \in C_p - S} o(c) \vdash T(r, t).$$

We can now define the observation function:

$$p_t(r|\mathcal{D}(C_p, C_n)) = \\ \sum_{[S \subset C_p] \wedge M(C_p, C_n, S, r, t)} \prod_{c \in S} (1 - g(c)) \prod_{c \in C_p - S} g(c).$$

r can represent any observation, but in all the examples in this paper r is $x_i = v_k$. The computation required is exponential in the number of simultaneous single faults considered, not the number of components, so $p_t(x_i = v_k|\mathcal{D}(C_p, C_n))$ can be evaluated relatively efficiently. These can be evaluated directly with the ATMS/HTMS framework.

A persistent fault in component c can be modeled as an intermittent fault with $g(c) = 0$. As the probing strategy is the same for persistent and intermittent faults, no change is needed to the algorithms (except $g(c)$ is more difficult to learn) to diagnose multiple simultaneous faults of both types.

Modeling persistent faults as $g(c) = 0$ has two drawbacks: (1) it is a 'strong' fault model (i.e., the component always behaves incorrectly if it is faulted), and (2) it cannot distinguish between intermittent and persistent faults in a single component. Therefore we introduce fault modes as in [de Kleer and Williams, 1989]. An inverter might fail with its output stuck at 1, or its output stuck at 0, or it might intermittently produce the wrong output:

$$INVERTER(x) \rightarrow OK(x) \vee SA0(x) \vee SA1(x) \vee U(x),$$

where $OK(x)$ corresponds to the earlier $\neg AB(x)$,

$$OK(x) \rightarrow [in(x, t) = 0 \equiv out(x, t) = 1],$$

and,

$$SA0(x) \rightarrow [out(x) = 0], SA1(x) \rightarrow [out(x) = 1],$$

and $U(x)$ is treated as intermittent fault (just as $AB(x)$ was earlier).

9 Related Work

There has been relatively little work in the model-based diagnosis community on intermittent faults. The approach of this paper exploits the notion of exoneration — ruling out components as failing when observed to be behaving correctly. Previous work along this line is the alibi notion of [Raiman, 1990] and of corroboration in [Brown *et al.*, 1982]. [Koren and Kohavi, 1977] converts single fault intermittent diagnosis tasks of combinational logic to dynamic programming. [Contant *et al.*, 2002] presents an intermittent diagnosis approach applicable to Discrete Event Systems. [Abreu *et al.*, 2008] applies methods similar to those presented in this paper to spectrum-based program debugging.

References

[Abreu *et al.*, 2008] R. Abreu, P. Zoeteweij, and A.J.C. van Gemund. A dynamic modeling approach to software multiple-fault localization. In *19th International Workshop on Principles of Diagnosis*, pages 7–14, Blue Mountains, Australia, 2008.

[Brown *et al.*, 1982] J. S. Brown, R. Burton, and J. de Kleer. Pedagogical, natural language, and knowledge engineering issues in SOPHIE I, II, and III. In D. Sleeman and J. S. Brown, editors, *Intelligent Tutoring Systems*, pages 227–282. Academic Press, New York, 1982.

[Contant *et al.*, 2002] O. Contant, S. Lafortune, and D. Teneketzis. Failure diagnosis of discrete event systems: The case of intermittent faults. In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 4006–4011, 2002.

[de Kleer and Williams, 1987] J. de Kleer and B. C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32(1):97–130, April 1987. Also in: *Readings in NonMonotonic Reasoning*, edited by Matthew L. Ginsberg, (Morgan Kaufmann, 1987), 280–297.

[de Kleer and Williams, 1989] J. de Kleer and B.C. Williams. Diagnosis with behavioral modes. In *Proc. 11th IJCAI*, pages 1324–1330, Detroit, 1989.

[de Kleer *et al.*, 1991] J. de Kleer, O. Raiman, and M. Shirley. One step lookahead is pretty good. In *Working Notes of the 2nd Int. Workshop on Principles of Diagnosis*, pages 136–142. Torino, Italy, 1991.

[de Kleer *et al.*, 1992] J. de Kleer, A. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56(2-3):197–222, 1992.

[de Kleer *et al.*, 2008] J. de Kleer, Bob Price, Lukas Kuhn, Minh Do, and Rong Zhou. A framework for continuously estimating persistent and intermittent failure probabilities. In *19th International Workshop on Principles of Diagnosis*, pages 63–70, Blue Mountains, Australia, 2008.

[de Kleer, 1992] J. de Kleer. A hybrid truth maintenance system. PARC Technical Report, January 1992.

[de Kleer, 2006] J. de Kleer. Getting the probabilities right for measurement selection. In *17th International Workshop on Principles of Diagnosis*, pages 141–146, Burgos, Spain, 2006.

[Fromherz *et al.*, 2003] M.P.J. Fromherz, D.G. Bobrow, and J. de Kleer. Model-based computing for design and control of reconfigurable systems. *The AI Magazine*, 24(4):120–130, 2003.

[Koren and Kohavi, 1977] Israel Koren and Zvi Kohavi. Diagnosis of intermittent faults in combinational networks. *IEEE Trans. Computers*, 26(11):1154–1158, 1977.

[Raiman, 1990] O. Raiman. A circumscribed diagnosis engine. In G. Gottlob and W. Nejdl, editors, *Proc. Int. Workshop on Expert Systems in Engineering*, pages 90–101. Springer Verlag LNCS 462, 1990.