Improving probability estimates to lower diagnostic costs

Johan de Kleer
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304 USA
Revision of July 25, 2006

Abstract

The core objective of model-based diagnosis is to identify candidate diagnoses which explain the observed symptoms. Usually there are multiple such candidate diagnoses and a model-based diagnostic engine proposes additional measurements to better isolate the true diagnosis. An objective of such an algorithm is to identify this diagnosis in minimum average expected cost. Minimizing this cost requires having accurate probability estimates for the candidate diagnoses. Most diagnostic engines utilize sequential diagnosis combined with Bayes Rule to determine the posterior probability of a candidate diagnosis given a measurement outcome. Unfortunately, one of the terms of Bayes rule, the conditional probability of a measurement outcome given a candidate diagnosis, must often be estimated (noted as ϵ in most formulations). This paper presents a reformulation of the sequential diagnosis process used in diagnostic engines and shows how different ϵ policies lead to varying results.

1 Introduction

Model-based diagnosis has been applied to a wide range of applications including automobiles [Struss and Price, 2004], spacecraft [Williams and Nayak, 1996], mobile robots [Steinbauer and Wotawa, 2005] and software [Köb and Wotawa, 2004] to mention just a few. The core objective of model-based diagnosis is to identify candidate diagnoses which explain the observed symptoms. Usually there are multiple such diagnoses and a model-based diagnostic engine proposes additional measurements to better isolate the true diagnosis. An objective of such an algorithm is to identify this diagnosis in minimum average expected cost (i.e., the sum of the costs of the measurements). Minimizing this cost requires having accurate probability estimates for the candidate diagnoses. Most diagnostic engines utilize a greedy sequential diagnosis combined with Bayes Rule to determine the posterior probability of a candidate diagnosis given a measurement outcome. Unfortunately, one of the terms of Bayes rule, the conditional probability of an measurement outcome given a candidate diagnosis, must often be estimated (noted as ϵ in most formulations). This paper presents a reformulation of the sequential diagnosis process used in most diagnostic engines and shows the results of various ϵ-policies. In order to minimize possible confounding of different domain models and to have easy access to many examples we draw all our examples from a widely available combinatorial logic test suite from ISCAS-85 [Brglez and Fujiwara, 1985].

In order to focus on the impact of varying ϵ-policies we make the following assumptions. (All the assumptions can be relaxed, but would confound the results.): (1) All measurements have equal cost, (2) No intermittent faults, (3) No multi-step lookahead, (4) The inference engine used to derive the consequences of observations is complete, (5) All the system's inputs are known, (6) One symptomatic output is given, (7) Time is not modeled, (8) The system has at most two faults, (9) The behavioral model for each component is completely described, (10) The system is well-designed (no unattached inputs or outputs or cycles).

2 GDE Probability Framework

This basic framework is described in [de Kleer and Williams, 1987; de Kleer *et al.*, 1992].

Definition 1 *A system is a triple (SD,COMPS, OBS) where:*

1. *SD, the system description, is a set of first-order sentences.*
2. *COMPS, the system components, is a finite set of constants.*
3. *OBS, a set of observations, is a set of first-order sentences.*

Definition 2 *Given two sets of components Cp and Cn define $\mathcal{D}(Cp, Cn)$ to be the conjunction:*

$$\Big[\bigwedge_{c\in Cp} AB(c)\Big] \wedge \Big[\bigwedge_{c\in Cn} \neg AB(c)\Big].$$

Where $AB(x)$ represents that the component x is ABnormal (faulted).

A diagnosis is a sentence describing one possible state of the system, where this state is an assignment of the status normal or abnormal to each system component.

Definition 3 *Let $\Delta \subseteq$COMPS. A diagnosis for (SD,COMPS,OBS) is $\mathcal{D}(\Delta, COMPS - \Delta)$ such that the following is satisfiable:*

$$SD \cup OBS \cup \{\mathcal{D}(\Delta, COMPS - \Delta)\}$$

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

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

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 is given by Bayes Rule:

$$p(\mathcal{D}|x = v) = \frac{p(x = v|\mathcal{D})p(\mathcal{D})}{p(x = v)}. \qquad (2)$$

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

$$p(x = v|\mathcal{D}) = 1 \quad \text{if} \quad x = v \text{ follows from } \mathcal{D}, SD,$$

$$p(x = v|\mathcal{D}) = 0 \quad \text{if} \quad \mathcal{D}, SD, (x = v) \text{ are inconsistent.}$$

If neither holds,

$$p(x = v|\mathcal{D}) = \epsilon_{ik}, \qquad (3)$$

where $\epsilon_{ik} = \frac{1}{m}$. 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 $\epsilon = .5$.

Consider other possible values for ϵ_{ik}. As ϵ approaches 0, some diagnoses would be assigned far smaller posterior probabilities which would lead to inaccurate conclusions and excessive measurement cost. For example, multiple faults would be assigned far smaller probability than is actually the case. So long as $\epsilon > 0$ the GDE algorithm will identify the correct diagnosis after sufficient measurements ($\epsilon = 0$ would assign 0 probability to correct diagnoses). As ϵ approaches 1, there would be little need to use Bayes Rule and the relative likelihoods of any two diagnoses would always be a constant. This would force GDE to needlessly consider very unlikely candidate diagnoses. Looked at differently, as ϵ varies from 0 to 1 approximates the spectrum of abductive-based to consistency-based diagnostic frameworks [Brusoni *et al.*, 1998]. ϵ clearly must lie between 0 and 1, but should it be $\epsilon = \frac{1}{m}$?

There are a number of reasons that a candidate diagnosis might fail to predict a value for a measured variable.

- Incompleteness in the inference engine used (e.g., GDE's).
- Incompleteness in the component models.
- The model may predict a disjunction of values (as can be the case in qualitative models).
- Lack of knowledge of the actual faulty behavior of a component.

As rule 3 is applied whenever a candidate fails to predict a measurement, any incompleteness will result in incorrectly reducing the posterior of the candidate and increase overall diagnostic cost. Although the lack of inferential completeness is common in model-based diagnosis engines, in this paper we focus on the last source and use complete models and a complete inference procedure.

3 Using an ϵ-policy

In order to avoid excessive computational cost, many diagnostic algorithms utilize a greedy minimum entropy approach to select the best next measurement (*i.e.*, the one which, on average, minimizes the cost of identifying the actual diagnosis). [de Kleer and Williams, 1987] shows how expected entropy outcomes for hypothetical measurements can be determined without additional inference. The outcome can be calculated directly from the current probability distribution of measurement outcomes given $\epsilon_{ik} = \frac{1}{m}$. We now generalize this approach to allow an arbitrary ϵ-policy. Fortunately, the outcomes can be evaluated directly in the general case as well. Given a set of diagnoses, $DIAGNOSES$, and assuming all measurements are of unit cost,

$$H = - \sum_{\mathcal{D} \in DIAGNOSES} p(\mathcal{D}) \log p(\mathcal{D}), \qquad (4)$$

estimates the number of measurements needed to complete a diagnosis. We define,

$$S_{ik} = \{\mathcal{D} \in DIAGNOSES | \ \mathcal{D} \cup SD \cup OBS \vdash x_i = v_{ik}\},$$

$$U_i = \{\mathcal{D} \in DIAGNOSES | \mathcal{D} \notin S_{ik} \text{for any } k\}.$$

$$p(S_{ik}) = \sum_{C_j \in S_{ik}} p_j,$$

$$p(U_i) = \sum_{C_j \in U_i} p_j,$$

$$p(x_i = v_{ik}) = p(S_{ik}) + \epsilon_{ik} p(U_i).$$

ϵ_{ik} is determined by the diagnostic policy, under the restriction that $\Sigma_{k=1}^{m} \epsilon_{ik} = 1$ for all i. The expected entropy after measuring $x_i = v_{ik}$ is:

$$H_e(x_i) = \sum_{k=1}^{m} p(x_i = v_{ik}) H(x_i = v_{ik}). \qquad (5)$$

Let p' be the probability after making the measurement. Substituting equation 2 into equation 4 gives:

$$\begin{aligned} H(x_i = v_{ik}) &= - \sum_{l \in S_{ik} \cup U_i} p'_l \, log \, p'_l \\ &= - \sum_{l \in S_{ik}} \frac{p_l}{p(x_i = v_{ik})} log \frac{p_l}{p(x_i = v_{ik})} \\ &\quad - \sum_{l \in U_i} \frac{\epsilon_{ik} p_l}{p(x_i = v_{ik})} log \frac{\epsilon_{ik} p_l}{p(x_i = v_{ik})} \end{aligned}$$

Substituting H into this equation gives:

$$\begin{aligned} H_e(x_i) &= - \sum_{k=1}^{m} \sum_{l \in S_{ik}} p_l log \frac{p_l}{p(x_i = v_{ik})} \\ &\quad - \sum_{k=1}^{m} \sum_{l \in U_i} (\epsilon_{ik} p_l) log \frac{\epsilon_{ik} p_l}{p(x_i = v_{ik})}. \end{aligned}$$

Expanding the logarithms:

$$
\begin{aligned}
H_e(x_i) \quad = \quad & -\sum_{k=1}^{m}\sum_{l\in S_{ik}} p_l log\, p_l \\
& +\sum_{k=1}^{m}\sum_{l\in S_{ik}} p_l log\, p(x_i = v_{ik}) \\
& -\sum_{k=1}^{m}\sum_{l\in U_i} \epsilon_{ik} p_l log\, p_l - \sum_{k=1}^{m}\sum_{l\in U_i} \epsilon_{ik} p_l log\, \epsilon_{ik} \\
& +\sum_{k=1}^{m}\sum_{l\in U_i} \epsilon_{ik} p_l log\, p(x_i = v_{ik}).
\end{aligned}
$$

The first and third terms are simply the current entropy H and is necessarily constant. The second and fifth terms are the negative entropy (*i.e.*, of the probability density distribution of x_i). The expected entropy $H_e(x_i)$ to minimize has the following form:

$$H + \sum_{k=1}^{m} p(x_i{=}v_{ik}) log\, p(x_i{=}v_{ik}) - p(U)\sum_{k=1}^{m} \epsilon_{ik} log \epsilon_{ik}$$

The best proposed measurement is the one which maximizes information gain:

$$-\sum_{k=1}^{m} p(x_i{=}v_{ik}) log\, p(x_i{=}v_{ik}) + p(U)\sum_{k=1}^{m} \epsilon_{ik} log \epsilon_{ik}.$$

Expected information gain is the expected reduction in number of additional measurements needed to isolate the true diagnosis and always lies between 0 and 1 (where m, the number of values a variable can have is the base of the logarithm). There is thus no need to utilize additional inferential machinery to hypothesize the results of possible measurement outcomes. Given a policy for distributing $p(U_i)$, the proposed measurement can be evaluated directly from known probabilities. Furthermore, if the ϵ-policy is fixed, then $\sum_{k=1}^{m} \epsilon_{ik} log \epsilon_{ik}$ is constant throughout the diagnosis task.

4 Advantages of the GDE Framework

One of the fundamental advantages of the GDE framework is that it is unnecessary to enumerate all the possible fault modes beforehand. Thus a diagnostic algorithm can successfully diagnose a system having never-before-seen faults. These fault modes are a challenge to more conventional diagnostic approaches which require far more prior knowledge of all the system's fault modes.

GDE's probabilistic framework allows it to identify the best measurement to make next to localize the system's fault. Consider the simple four inverter circuit of Figure 1 and $a = 0$.

To see the effects of different ϵ_{ik}'s consider the some simplistic policies.Table 1 lists expected costs of measuring all the variables, first after $a = 1$, and then after $e = 0$. The costs are given for three values of ϵ_{ik}. Note that $\epsilon = 1$ is equivalent to using no probabilistic information at all, and as a consequence the resulting costs cannot be used to rank proposed

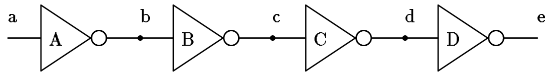

Figure 1: Four sequential inverters.

Table 1: Expected information gains for cascaded inverters after measuring a and e (with $p = .01$).

	$a = 0$			$a = 0, e = 0$		
	$\epsilon_{i1} = .01$	$\epsilon_{i1} = .5$	$\epsilon_{i1} = 1$	$\epsilon_{i1} = .01$	$\epsilon_{i1} = .5$	$\epsilon_{i1} = 1$
a	0	0	0	0	0	0
b	.07	.03	0	.99	.85	0
c	.14	.06	0	1	.98	0
d	.19	.09	0	.99	.85	0
e	.23	.10	0	0	0	0

measurements. As long as $0 < \epsilon < 1$, GDE can eventually identify good measurements to make next.

A second key advantage of the probabilistic framework is that it enables the diagnostic algorithm to focus on the most probable candidate diagnoses, not the minimal cardinality and not the minimal under subset as most other approaches do.

5 Using a Fixed ϵ-policy

We have implemented a new diagnostic algorithm called ϵGDE which accepts an arbitrary ϵ-policy and is logically complete (it identifies all conflicts and all variable value predictions efficiently). It is provided an ϵ-policy, fault probabilities, system model, and a set of input vectors and symptoms. Given this as input, ϵGDE, computes the average expected cost to diagnose the system. As this computes the cost over a large set of test-vectors it is an acid test of diagnostic approaches. Often clear advantages for certain classes of faults is balanced out by the additional cost of other faults.

The simplest, and trivial to implement, is a fixed policy in which the $epsilon_{ik}$ are fixed for all k. And where $\Sigma_{k=1}^{m} \epsilon_{ik} = 1$. Figure 2 graphs the cost of all single-faults for c432. This figure illustrates that no value of ϵ is of particular advantage for the single-faults of c432. Diagnostic cost is approximately 5.52 independent of ϵ. The small variations are result of the particular structure of the circuit and test-vectors used.

The somewhat surprising result that diagnostic cost is independent of a fixed ϵ reveals important properties of ϵ-policies. $p(x = v|\mathcal{D}) = \epsilon_{ik}$ is applied only if $\mathcal{D}$ do not predict any measurement. Initially, all inputs are known, and no outputs are given. When the first, symptomatic, measurement is made, no candidate diagnosis can predict this measurement. As a consequence, every candidate diagnosis posterior probability will be reduced by the same ϵ and the diagnostic cost of all single-faults is constant.

If the circuit contains more than one fault, the situation shifts dramatically. All single-faults will be ruled out by the observations, and diagnostic cost will be completely determined by candidates to which a ϵ policy applies to differentially. In general, if we are considering faults of size n, equation 3 will be applied at most n times to that candidate.

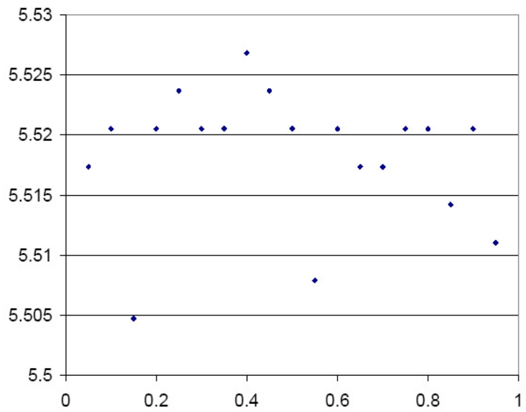

Figure 2: Average cost (single-faults) vs. ϵ_{i1} for circuit c432 All components fail with equal probability $p = 0.0001$. c432 has 160 gates and is a 27-channel interrupt controller from the ISCAS-85 test suite.

The consequences of ϵ policies on multiple-faults is dramatic. The data in Figure 3 shows the average expected costs for a suite of randomly generated double faults each with a fully populated input vector with one identified symptom. The diagnostic cost of finding the correct double diagnosis is substantially higher than that for single faults (13 vs 5). For this device, diagnostic cost is minimum if ϵ_{i1} is nearly 1, and nearly 0 for ϵ_{i0}. This is very different than the $\epsilon = .5$ estimate of GDE.

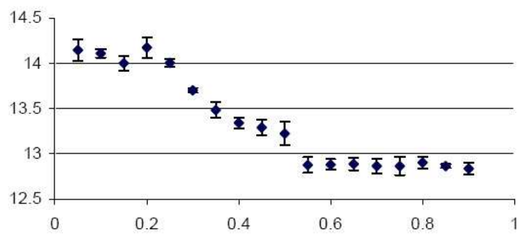

Figure 3: Average cost vs. ϵ for 1 for circuit c432 with .95 confidence interval.

Figure 4 shows the results for the same task on circuit c499 which has 202 components. For this task, there is a sharp notch around .5 which corresponds to GDE's estimate.

6 Towards a Dynamic ϵ-policy

Figures 3 and 4 suggest adaptive ϵ-policies can improve diagnostic costs. We would like to devise a dynamic ϵ-policy appropriate to each diagnostic task. It is also important that this policy be easy to compute, otherwise it competes with the alternative of expensive multi-step lookahead.

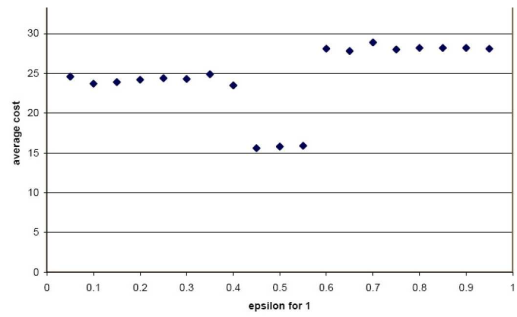

Figure 4: Average cost vs. ϵ for 1 for circuit c499. c499 has 202 gates and is a 32-bit single-error-correcting-circuit from the ISCAS-85 test suite.

Consider the oversimplistic example of a single inverter (say A in Figure 1). Assume that $p(\neg AB(A)) = \alpha$ and we measured first $a = 0$ and then $b = 1$. The priors are $p(AB(A)) = (1 - \alpha)$, $p(\neg AB(A)) = \alpha$, and clearly $p(AB(A)|a = 0) = p(AB(A)) = (1 - \alpha)$ and $p(\neg AB(A)|a = 0) = p(\neg AB(A)) = \alpha$. Applying equation 2 with $\epsilon = .5$ to both possible diagnoses gives:

$$p(\neg AB(A)|b = 1, a = 0) = \frac{\alpha}{p(b = 1)},$$

$$p(\neg AB(A)|b = 1, a = 0) = \frac{\frac{1}{2}(1 - \alpha)}{p(b = 1)}.$$

As the sum of the probabilities of all diagnoses must always be 1, we obtain:

$$p(b = 1) = \frac{2\alpha}{1 + \alpha}.$$

Hence, given the evidence,

$$p(\neg AB(A)) = \frac{2\alpha}{\alpha + 1}.$$

Definition 4 *[Raiman* et al.*, 1991] A component behaves non-intermittently if its outputs are a function of its inputs.*

Table 2: The 4 possible binary functions of one input and one output. i is the binary input, and each column f_i list the outputs for the corresponding input.

i	f_0	f_1	f_2	f_3
0	0	0	1	1
1	0	1	1	0

Table 2 describes all possible binary functions of one input and one output. f_3 describes the correct behavior of an inverter, f_0 is the fault "stuck-at-0," f_2 is the fault "stuck-at-1," and f_1 an unexpected short of input to output for the inverter. Assuming that each fault mode $f_{i\neq 3}$ has equal probability ($\frac{\alpha}{3}$), the correct posterior should be $\frac{3\alpha}{2\alpha+1}$. This corresponds to $\epsilon = \frac{1}{3}$. The difference is:

$$\frac{\alpha(\alpha - 1)}{(2\alpha + 1)(\alpha + 1)}.$$

Suppose that we had measured $b = 0$ instead. Table 2 shows that only one of the three possible faulty behaviors are eliminated: f_2 is eliminated, but f_0 and f_1 remain. Therefore, $p(x = v|\mathcal{D}) = \frac{2}{3}$.

For simplicity consider only the first three inverters and the double fault diagnosis of $\mathcal{D}(\{B, C\}, \{A\})$. The prior is $(1 - \alpha)\alpha^2$. After measuring $d = 0$, GDE reduces its probability by $\frac{1}{2}$. Given Table 2 we can compute the posterior probability exactly as follows. Inverter C is faulted with output 0 and thus it can only be behaving according to functions f_0 or f_1. The input to the faulty inverter B is 1, but that alone does not provide any evidence for changing its probability. For the f_1 mode of inverter C to produce $d = 0$, c must be 0. This is inconsistent with modes f_1 or f_2 of B. Therefore, there are only 4 consistent combinations of modes for B and C: $\{\langle f_0, f_0\rangle, \langle f_0, f_1\rangle, \langle f_1, f_0\rangle, \langle f_2, f_0\rangle\}$. As only 4 out of the 9 combinations survive, the posterior probability is reduced by $\frac{4}{9}$. Measuring $c = 0$ eliminates only $\langle f_2, f_0\rangle$ to a final reduction of $\frac{3}{9}$. Tables 3 and 4 summarize these calculations.

Table 3: GDE vs. correct probability changes for $\mathcal{D}(\{B, C\}, \{A\})$

OBS	$\epsilon = 1/2$	correct
$d = 0$	$\frac{1}{2}$	$\frac{4}{9}$
$c = 0$	$\frac{1}{4}$	$\frac{3}{9}$

Table 4: GDE vs. correct probability changes for $\mathcal{D}(\{A, B, C\}, \{\})$

OBS	$\epsilon = 1/2$	correct
$d = 0$	$\frac{1}{2}$	$\frac{2}{3}$
$c = 0$	$\frac{1}{4}$	$\frac{4}{9}$
$b = 0$	$\frac{1}{8}$	$\frac{8}{27}$

Consider a simple 2-input *and* gate. Table 5 lists all possible behaviors for a 2-input/1-output gate. The correct behavior for the *and* gate is given by f_2. All remaining behaviors correspond to fault modes.

This analysis and the results of Figures 3 and 4 suggest that an adaptive ϵ-policy might be exploited to further improve the diagnostic cost of isolating faulty components.

7 Extension to Fault Modes

The probabilistic framework outlined in Sections 2 and 3 can be directly expanded to include component fault modes [de Kleer and Williams, 1989]. The $AB/\neg AB$ framework can viewed as assigning each component a "G", good, mode or an "U", unknown, faulty mode. The good model for an *and* gate is given by column f_2 of Table 5, and faulty behaviors correspond to all the remaining columns. Two common fault models for a *and* gate "SA1" (output stuck at 1) and "SA0" (output stuck at 0). This model for the gate has 4 modes: "G" (f_2) "SA1"(f_{15}), "SA0" (f_0), and "U" (which corresponds to

Table 5: Possible functions of two inputs and one output.

i_0	i_1	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	1	1	1
1	1	0	0	1	1	0	0	1	1
1	0	0	1	0	1	0	1	0	1

i_0	i_1	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}
0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1
1	1	0	0	1	1	0	0	1	1
1	0	0	1	0	1	0	1	0	1

the remaining 13 columns). All the analyses of this paper directly extend to multiple fault modes, but $p(U)$ will always be reduced because some of the fault modes are explicitly modeled with their own probabilities. As the extension to fault modes is direct, we do not formalize them in this paper.

8 Dynamic Epsilon Policies

Using the framework of the previous sections and two additional assumptions, we now define an ϵ-policy which is computes ϵ exactly for each individual variable and candidate diagnosis. Intuitively, we apply the diagnostic framework laid out in Section 2 recursively for each possible candidate. We assume that each component model specifies an output value when all its inputs are known. This assumption holds for digital circuits, but may not apply for some qualitative modeling paradigms where adding a qualitative "$+$" and "$-$" yields no result. In addition, we assume that we are provided the prior probabilities for each possible function of a component. For example, in the case of an *and* gate we are given the prior probability of each possible function f_i of Table 5. Let $p(f_i(c))$ be the prior probability that component c behaves according to function f_i. We know that:

$$\sum_{f_i} p(f_i(c)) = 1,$$

and,

$$\sum_{f_i \in F(c)} p(f_i(c)) = p(c),$$

where $F(c)$ is the set of all faulty functions f_i and $p(c)$ is the prior probability that component c is faulted as defined in Equation 1. Consider a diagnosis $\mathcal{D} = \mathcal{D}(B, G)$ which fails to predict some $x = v$. Restating Equation 3:

$$p(x = v|\mathcal{D}(B, G)) = \epsilon_{ik} p(\mathcal{D}(B, G)).$$

Definition 5 *A micro candidate diagnosis M for candidate diagnosis $\mathcal{D} = \mathcal{D}(B, G)$ is a conjunction:*

$$\bigwedge_{c \in B} f_i'(c),$$

where f_i' is formula describing the behavior of f_i as a propositional formula, and M is consistent with $\mathcal{D} \cup SD \cup OBS$.

$p(M)$ follows straightforwardly from Bayes Rule:

$$p(M) = \frac{\prod_{c \in B} p(f_i(c))}{\sum p(M)}.$$

Thus,

$$p(x = v|\mathcal{D}) = \sum_{M s.t. M \cup \mathcal{D} \cup SD \cup OBS \vdash x=v} p(M)p(\mathcal{D}),$$

or,

$$\epsilon_{ik} = \sum_{M s.t. M \cup \mathcal{D} \cup SD \cup OBS \vdash x=v} p(M),$$

where the M are the micro candidate diagnoses for $\mathcal{D}$.

If all the faulted $p(f_i(c))$ are equal for each component c, this new framework reduces to that of the previous section. This approach is most powerful when the individual $p(f_i(c))$ vary significantly. In these cases, the resulting diagnostic efficiency improvement can be significant.

Using GDE with fault modes, identical results would be obtained if explicit fault modes were introduced for each possible faulty function of each component. Unfortunately, this approach is computationally intractable. For a digital circuit where the sum of the number of input terminals was l, the complexity would be 2^{2^l}.

9 Conclusions

This paper presents three advances. First, it presents a generalization of the information gain equation used in evaluating possible measurements. For single-faults, the $\epsilon = \frac{1}{m}$ is nearly optimal. But analysis of double-faults opens the possibility of dramatically increased diagnostic cost with a more adaptive ϵ policy.

10 Acknowledgments

Conversations with Olivier Raiman and Brian Williams helped clarify many of these concepts.

References

[Brglez and Fujiwara, 1985] F. Brglez and H. Fujiwara. A neutral netlist of 10 combinational benchmark circuits and a target translator in fortran. In *Proc. IEEE Int. Symposium on Circuits and Systems*, pages 695–698, June 1985.

[Brusoni *et al.*, 1998] Vittorio Brusoni, Luca Console, Paolo Terenziani, and Daniele Theseider Dupre. A spectrum of definitions for temporal model-based diagnosis. *Artificial Intelligence*, 102(1):39–79, 1998.

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

[Köb and Wotawa, 2004] Daniel Köb and Franz Wotawa. Introducing alias information into model-based debugging. In *16th European Conference on Artificial Intelligence (ECAI)*, Valencia, Spain, August 2004.

[Raiman *et al.*, 1991] O. Raiman, J. de Kleer, V. Saraswat, and M. H. Shirley. Characterizing non-intermittent faults. In *Proc. 9th National Conf. on Artificial Intelligence*, pages 849–854, Anaheim, CA, July 1991.

[Steinbauer and Wotawa, 2005] Gerald Steinbauer and Franz Wotawa. Detecting and locating faults in the control software of autonomous mobile robots. In *Proceedings of the* 19^{th} *International Joint Conference on AI (IJCAI-05)*, pages 1742–1743, Edinburgh, UK, 2005.

[Struss and Price, 2004] Peter Struss and Chris Price. Model-based systems in the automotive industry. *AI Magazine*, 24(4):17–34, 2004.

[Williams and Nayak, 1996] B. C. Williams and P. P. Nayak. A model-based approach to reactive self-configuring systems. In *Proc. 14th National Conf. on Artificial Intelligence*, pages 971–978, 1996.