TR-688

Adaptive Model-Based Diagnos
is with Hierarchical Models

by

Y. Nakakuki, Y. Koseki & M. Tanaka (NEC)

September, 1991

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03)3456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

Adaptive Model-Based Diagnosis with Hierarchical Models

Yoichiro Nakakuki Yoshiyuki Koseki Midori Tanaka
C&C Systems Research Laboratories, NEC Corporation

4-1-1 Miyazaki, Miyamae-ku, Kawasaki 216, JAPAN
TEL: +81-44-856-2258
FAX: +81-44-856-2233
E-mail: nakakuki%btl.cl.nec.co.jp@uunet.UU.NET

Area: Enabling technology
AI Topic: Model-based diagnosis
Domain: Electronic device diagnosis

Abstract

This paper describes an efficient diagnostic mechanism which utilizes hierarchical models of a device. Model-based diagnosis is a general approach for diagnosing devices using the behavioral specification of a device. Although model-based systems are more robust than heuristic-based expert systems, they require more computation time. In order to develop an efficient model-based diagnostic system, it is important to utilize the information about hierarchical structure of a device. In general, a diagnosis using detailed level model is expensive. Therefore, in order to select an appropriate level to use, we must resolve the trade-off between the diagnosis cost and the diagnosis precision. This paper introduces the model diagnosability criterion to estimate how detailed diagnosis can be achieved with a model. This criterion enables to select an appropriate level to use. The techniques described in this paper is adaptive to several kinds of situations, such as, the required diagnosis precision, the given computation power, the observation cost, and the phase of a diagnosis.

1 Introduction

Model-based diagnosis is a general approach for diagnosing devices using the behavioral specification of devices [1, 3, 5, 9]. Although model-based systems are more robust than heuristic-based expert systems, they require more computation time. In general, the computational complexity of a model-based diagnosis grows rapidly with the complexity of a device model. In this paper we propose an efficient diagnostic mechanism which utilizes hierarchical models of a device.

In order to diagnose a device efficiently, several techniques have been investigated. One approach is to use probabilistic information. For instance, the minimum entropy technique (GDE [3], Sherlock [4]) and the focusing technique [2] are essential for practical use. Another approach is to utilize hierarchical models of a device. XDE [6] adopted a hierarchical model representation, which can represent the device at multiple levels of abstraction. In the early stage of a diagnosis, the system uses an abstract level model to eliminate portions of the device from consideration. In the later stage, it utilize a detailed level model. In general, a diagnostic computation at more detailed level is more complex and more expensive. Therefore, in order to select an appropriate level to use, we must resolve the trade-off between the diagnosis cost and the diagnosis precision.

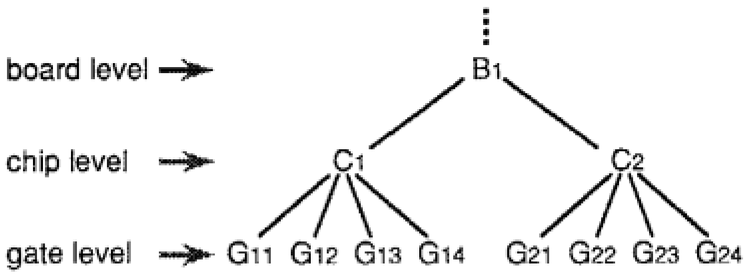
In the trade-off resolution process, a diagnostic system must consider the situation, especially the required diagnosis precision and the diagnosis cost. Consider that an electronic device which is composed of several boards, and that each board has several chips. In some situation, the diagnostic requirement may be to find a broken chip. However, in another situation, a repairman may want to know which board should be replaced. In the latter case, a diagnostic system does not necessary to pinpoint a broken chip. Hence, a diagnostic system should have adaptability to the required diagnosis precision. On the other hand, a diagnostic system should reduce the total diagnosis cost, i.e. the sum of the observation cost and the computation cost. The observation cost, however, depends on the situation. For example, it is costly to capture a digital signal at particular point of a device using a logic analyzer manually. However, an electron-beam tester can observe a signal at an arbitrary point in an LSI immediately. In the former case, the number of required observations greatly affects the total diagnosis cost. On the other hand, in the latter case, the total cost is mainly determined by the computation cost. Therefore, in order to reduce the total diagnosis cost, a diagnostic system should be adaptive to the observation cost. Moreover, the computation cost depends on the given computation power. Hence, a diagnostic system should also adapt to its computational environment.

Conventional systems [1, 6] adopted a simple technique to resolve the trade-off between the diagnosis cost and the diagnosis precision. They are not flexible, because they assume the observation costs are always costly, and they always descend (expand) the physical component hierarchy at most one level at a time. This paper presents a sophisticated way to select an appropriate model adaptively. First, we introduce a model diagnosability criterion to estimate how much information could be gained by using a model in a given situation. Then we propose an adaptive diagnosis mechanism with hierarchical models.

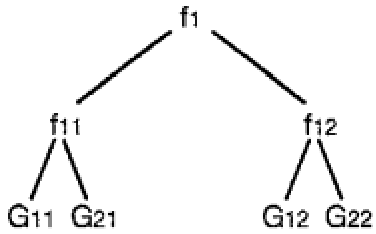
This paper is organized as follows. In the next section, we discuss about a model-based diagnosis with hierarchical models, especially the relationship between the required diagnosis precision and the entropy is considered. In section 3, we introduce a model diagnosability criterion to estimate how detailed diagnosis can be achieved with a model. An adaptive diagnosis mechanism is presented in section 4. The last section discusses the utility of the approach and the future work.

2 Diagnosis with Hierarchical Models

Most of the conventional hierarchical model-based approaches [1, 6], the structure of a device is represented as a physical hierarchy and a logical (functional) hierarchy. Usually, a required diagnosis precision is represented as a level in the physical hierarchy, for example, gate-level or chip-level, as shown in Fig. 2–1 (a).



(a) Physical hierarchy



(b) Logical hierarchy

Figure 2–1 Hierarchical structure

In general, a physical hierarchy and a logical hierarchy have different structures (e.g. Fig. 2–1 (b)). In this paper, to simplify the discussion, we assume that the two hierarchies have a same structure, and also assume that there is only single fault in a target device. However, the proposed techniques can easily be extended to the general case.

Here, we consider an example of a hierarchical model as shown in Fig. 2–2. A full adder is composed of 5 subcomponents, and 8 full adders construct a 8-bit ripple carry adder.

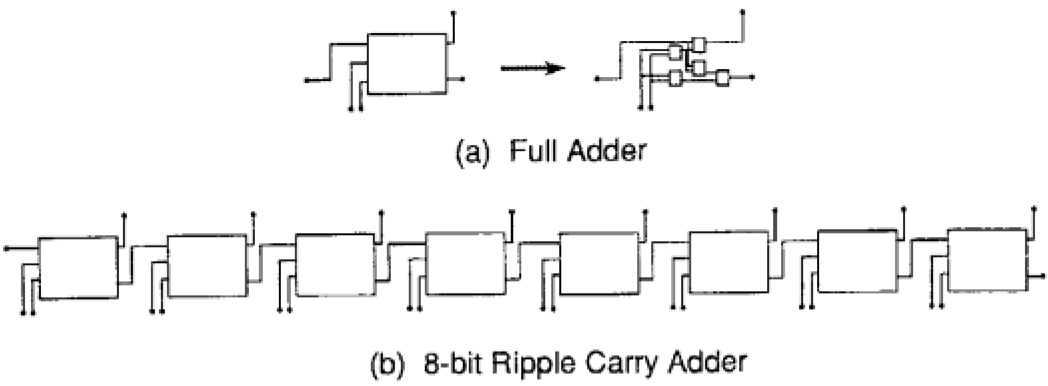


Figure 2–2 Example of Hierarchical Models

It is possible to generate several varieties of models for 8-bit ripple carry adder. Model X in Fig. 2–3 is the most abstract level model, and model Z is the most detailed one. Model Y is one of the intermediate models. In general, a diagnosis with more detailed level model is more expensive though it is more effective. Hence, it is important to select an appropriate level of model in the varieties. In the selection process, we must consider how much information can be gained by using each model. The information gain, however, depends on the required diagnosis precision. In the following section, we present a method to estimate the information gain.

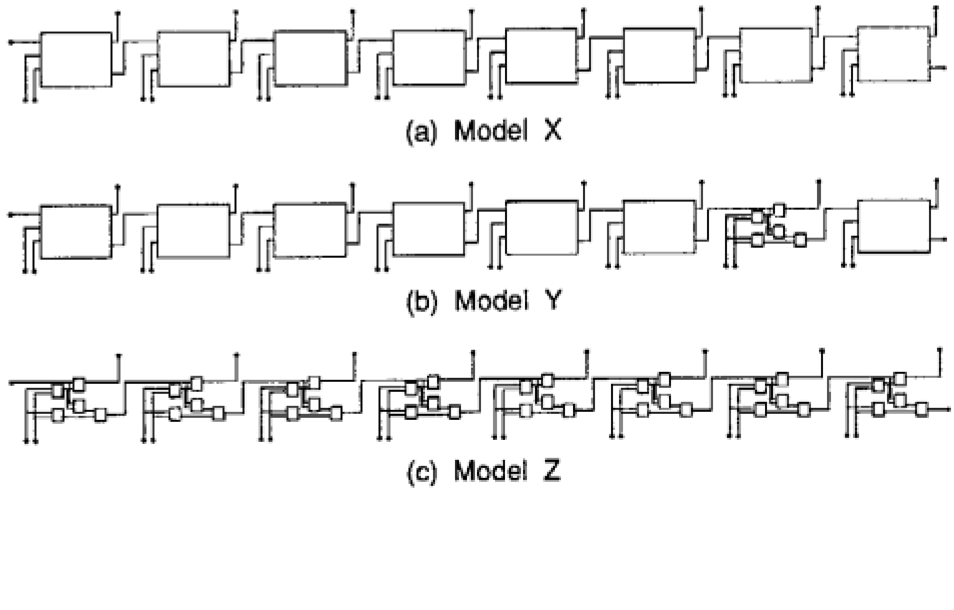
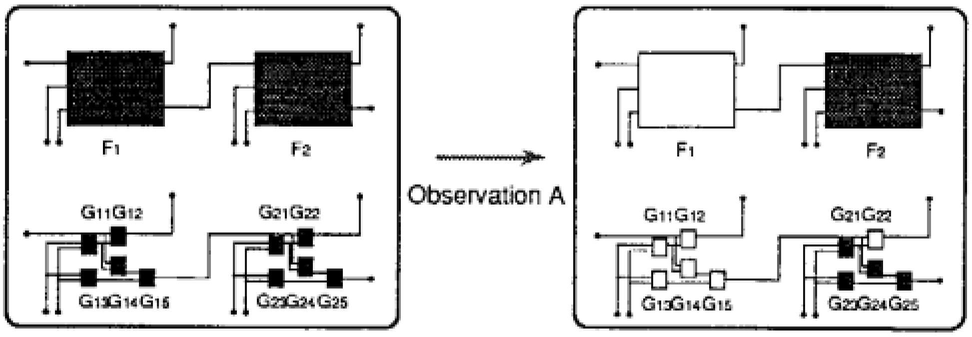


Figure 2–3 Varieties of Models for 8-bit Ripple Carry Adder

2.1 Diagnostic Precision and Entropy

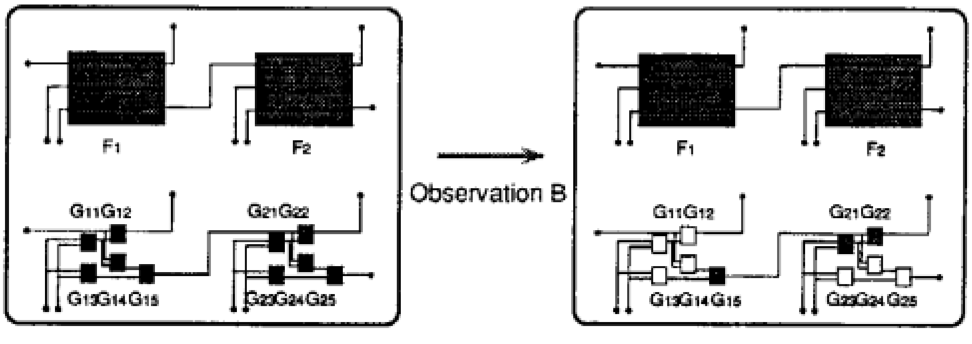
Several conventional systems [3, 4, 6, 7] make good use of the entropy of a set of suspected components in order to estimate the expected information needed to complete a diagnosis. However, in general, the expected information depends on the required diagnosis precision. The intuitions can be illustrated with the faulty 2-bit ripple carry adder shown in Fig. 2–4. Here we consider two varieties of models, i.e. a function-level model and a gate-level model, as shown in the figure. The hatched components are suspected, and the fault probability $P(C)$ for each suspected component C is also shown in the figure. The figure shows the changes of the diagnostic status after getting observation A or B.



$$P(F_1) = P(F_2) = 0.5$$
$$P(G_{ij}) = 0.1\ (i = 1,2\ j = 1,2,3,4,5)$$

$$P(F_2) = 1.0$$
$$P(G_{2j}) = 0.25\ (j = 1,3,4,5)$$

(a)



$$P(F_1) = P(F_2) = 0.5$$
$$P(G_{ij}) = 0.1\ (i = 1,2\ j = 1,2,3,4,5)$$

$$P(F_1) = 0.33 \quad P(F_2) = 0.67$$
$$P(G_{15}) = P(G_{21}) = P(G_{22}) = 0.33$$

(b)

Figure 2–4 Changes of Diagnostic Status

First, we consider that the required precision is the function-level. Here, if we get observation A (Fig. 2–4(a)), then we can complete the diagnosis because F_1 turns out to be normal. On the other hand, if we get observation B (Fig. 2–4(b)), the components F_1 and F_2 are still suspected. Therefore, observation A seems to be more informative than observation B.

Next, we assume that the required precision is the gate-level. If we get observation A, then four components ($G_{21}, G_{23}, G_{24}, G_{25}$) are suspected with the same probability (1/4). On the other hand, in the case of observation B, three components (G_{15}, G_{21}, G_{22}) are suspected with the same probability (1/3). Therefore, in this case, observation B seems to be more informative.

In order to estimate the information gain appropriately, we introduce the entropy for each level in the physical hierarchy. For instance, in the above example, the entropy for function-level (E_F) and for gate-level (E_G) are defined as follows:

$$E_F = -\sum_i P(F_i) \log P(F_i)$$

$$E_G = -\sum_i \sum_j P(G_{ij}) \log P(G_{ij})$$

In Fig. 2–5, we illustrate the changes of each entropy of the above example. If observation A is given, the function-level entropy, E_F, gains 1(bit), while observation B gives only 0.08(bit). Hence, observation A gives more entropy gain than B. On the other hand, the gate-level entropy gain by observation A and B are 1.32(bit) and 1.74(bit), respectively.

These results show the entropy gain for required precision (i.e. the level in a physical hierarchy) agrees with our intuition.

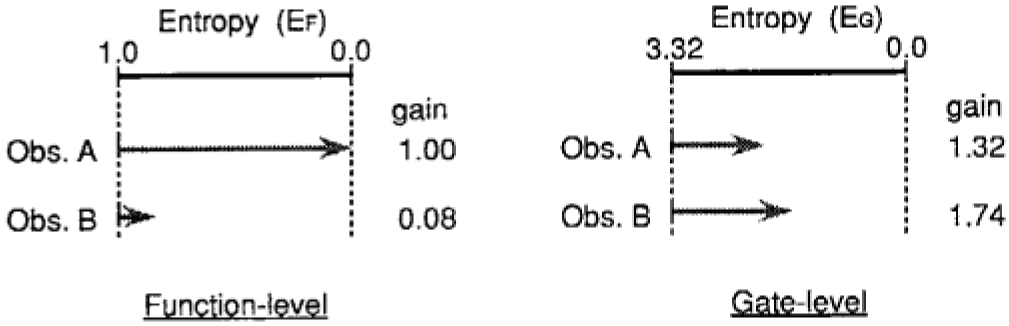


Figure 2–5 Changes of Entropy

3 Model Diagnosability Criterion

In this section we introduce a model diagnosability criterion to estimate how detailed diagnosis can be achieved with a model.

Consider three models for a 2-bit ripple carry adder as shown in Fig. 3–1. Here, we assume that the required diagnosis precision is the gate level, and that each of the 10 components in gate level has the same fault probability 0.1 (therefore, the fault probability for each function-level component, F_i, is 0.5).

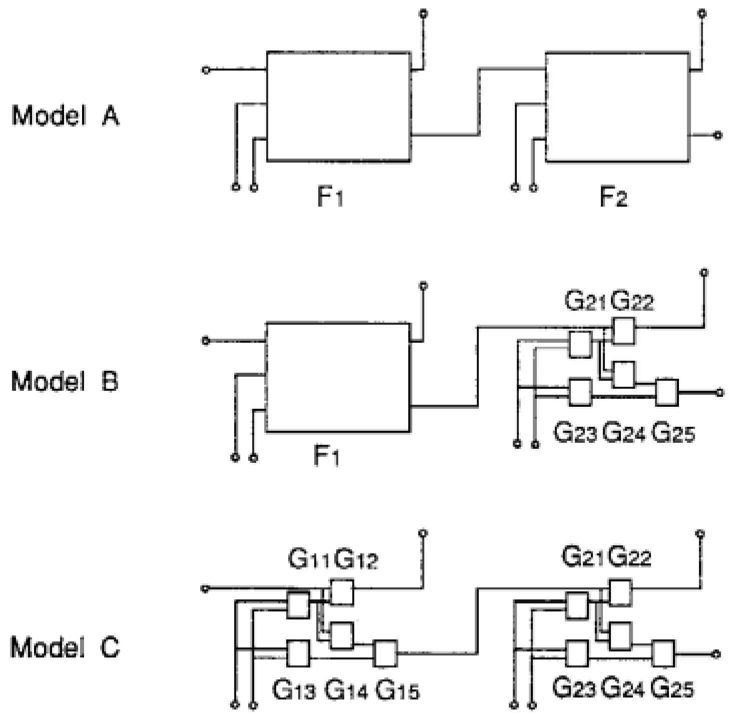


Figure 3–1 Models for a 2-bit Ripple Carry Adder

First, we consider a diagnosis with model A in the figure. In this case, the diagnosis can pinpoint which component in the function-level is faulty, if enough observations are given. However, it can not pinpoint which component in the gate level is faulty. For example, if the fault component is G_{11}, the system pinpoints the faulty component to be $G_{11}, G_{12}, G_{13}, G_{14}$, and G_{15} with probability 0.2 for each, and no more information can be gained. In other words, the gate-level entropy can not be reduced to less than $5 \cdot (-0.2 \log 0.2) = 2.32$ as far as using model A. That is, no matter how much observations are given, the system still can not gain 2.32-bit information (in average) to complete the diagnosis.

Next, we consider a diagnosis with model B. If the faulty component is G_{2i} (i = 1, 2, 3, 4, 5), then the system can conclude that the faulty component with probability 1.0, if enough observations are given. In this case, the gate-level entropy is reduced to 0. However, if the faulty component is G_{1i} (i =1, 2, 3, 4, 5), then the system can not decide which component in $\{G_1, G_2, \cdots, G_5\}$ is faulty, that is, the entropy can not be reduced to less than $5 \cdot (-0.2 \log 0.2) = 2.32$. Therefore, the expected lower bound for the entropy reduction is as follows (because $\sum_i P(G_{1i}) = \sum_i P(G_{2i}) = 0.5$):

$$0.5 \cdot 2.32 + 0.5 \cdot 0 = 1.16$$

Finally, if we use model C and enough observations are given, we can always pinpoint the faulty gate. Hence, the expected lower bound for the entropy reduction is 0.

In order to estimate how detailed diagnosis can be performed with a model, we define the model diagnosability $D(M)$ for a model M.

$$D(M) = \frac{\text{current entropy} - \text{expected lower bound for the entropy reduction with model M}}{\text{current entropy}}$$

The 'current entropy' expresses the expected information needed to complete a diagnosis. The numerator indicates how much information can be gained at most by using model M.

In the above example, current entropy is $10 \cdot (-0.1 \log 0.1) = 3.32$, at the initial stage of a diagnosis (Fig.3–2(a)). Therefore, the $D(M)$ for each model is calculated as follows:

$$\begin{aligned} D(modelA) &= \frac{3.32 - 2.32}{3.32} = 0.30 \\ D(modelB) &= \frac{3.32 - 1.16}{3.32} = 0.65 \\ D(modelC) &= \frac{3.32 - 0.00}{3.32} = 1.00 \end{aligned}$$

Fig. 3–2(b) illustrates this result. It is shown that a diagnosis with model A can gain at most 30% of necessary information. On the other hand, model C is powerful enough to gain all the necessary information.

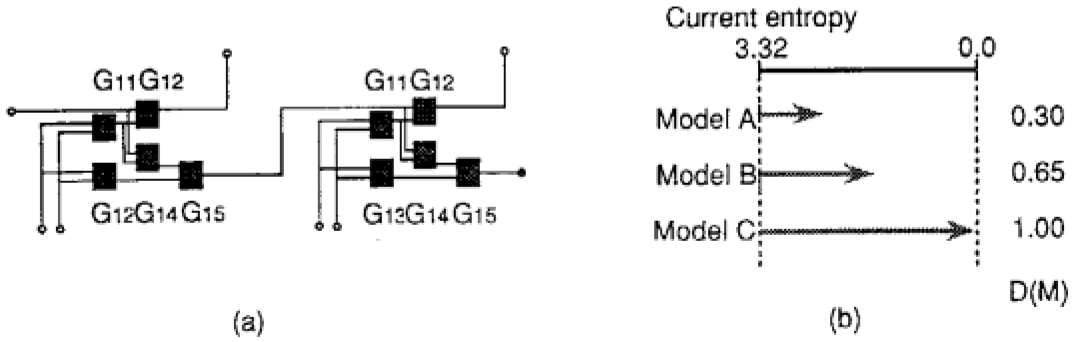


Figure 3–2 Model Diagnosability in the Initial Stage

Next we assume that the current diagnostic state has been changed by getting some observations (Fig. 3–3(a)), then current entropy is $5 \cdot (-0.2 \log 0.2) = 2.32$. Therefore the $D(M)$ for each model becomes as follows:

$$D(modelA) = \frac{2.32 - 2.32}{2.32} = 0.00$$
$$D(modelB) = \frac{2.32 - 0.00}{2.32} = 1.00$$
$$D(modelC) = \frac{2.32 - 0.00}{2.32} = 1.00$$

In this case, the result shows that no information can be gained by a diagnosis with model A (Fig. 3–3(b)). However, model B and C have an ability to gain the whole information needed to pinpoint a faulty gate.

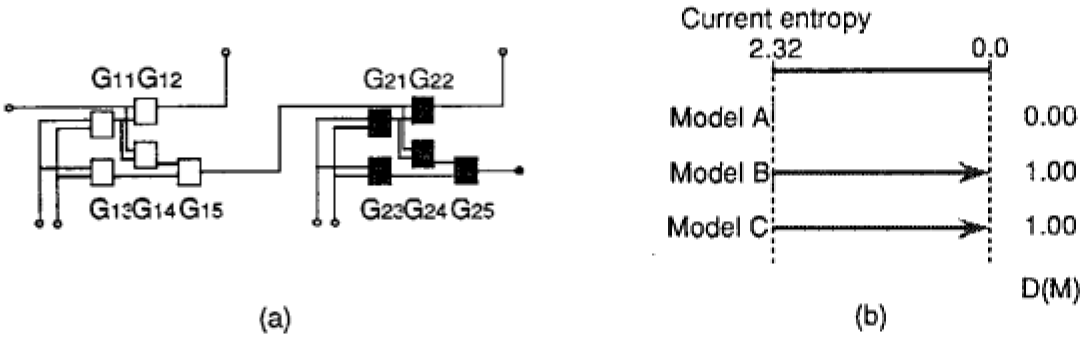


Figure 3–3 Model Diagnosability in the Later Stage

4 Adaptive Diagnosis Mechanism

In the previous section, we introduced the model diagnosability criterion. This section presents an adaptive diagnostic mechanism that selects an appropriate model at each stage of a diagnosis.

Here, let $D(M)$ be the diagnosability for model M, and let C be an average observation cost (required time). Because a diagnostic task is composed of several observation-computation cycle, $T(M)+C$ is an expected cost for a cycle, where $T(M)$ is the expected computation time for a diagnostic cycle with model M. Here, we assume that $T(M)$ can be estimated empirically or analytically, and that C is a (model independent) constant. In order to select an appropriate model, we evaluate each model by using the following criterion, $E(M)$.

$$E(M) = \frac{D(M)}{T(M)+C}$$

In each diagnostic cycle, the best model, i.e. a model with maximum $E(M)$, is selected.

This diagnostic mechanism has three kinds of adaptability. First, it adapts to the phase of a diagnosis. Second, it selects a model suitable for the given diagnosis precision. Finally, it adapts to the observation cost and computation cost. In the remainder of this section we illustrate some examples to show the advantages.

Example 1 Consider previously introduces three models of 8-bit ripple carry adder in Fig. 2–3. Here we assume that the required diagnosis precision is the gate-level, and that the expected computation time for model Y(Z) is about 1.5(5.0) times as large as $T(model\ X)$, respectively. That is:

$$\frac{T(model\ Y)}{T(model\ X)} \approx 1.5 \quad \text{and} \quad \frac{T(model\ Z)}{T(model\ X)} \approx 5.0$$

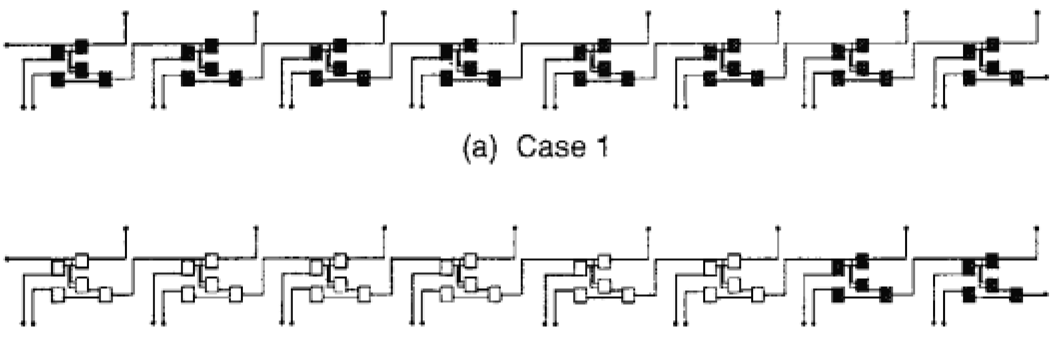
Moreover, in order to simplify the discussion, we assume that $C \ll T(M)$ for each model M. Therefore, $E(M)$ can be approximated as:

$$E(M) \approx \frac{D(M)}{T(M)}$$

[Case 1] As shown in Fig. 4–1(a), if all gates are suspected with the same probability(1/20), model X is selected by the proposed criterion because there is a possibility that a diagnosis with model X can gain about 56% of the necessary information with low cost. $E(M)$ for each model is shown in Fig. 4–1(c).

[Case 2] On the other hand, Fig. 4–1(b) shows another stage of a diagnosis. In this case, the suspected components have been narrowed down by some observations. Then, model Y is selected because it has the best $D(M)$ per cost in the three models (Fig. 4–1(c)). That is, a diagnosis with model X can gain at most only 30% of necessary information, and model Z is costly.

The example shows the ability to select a model suitable for each diagnostic stage.



(a) Case 1

(b) Case 2

Case 1

Model	D(M)	Cost	E(M)=D(M)/Cost
Model X	0.56	1.0	0.56
Model Y	0.62	1.5	0.41
Model Z	1.00	5.0	0.20

Case 2

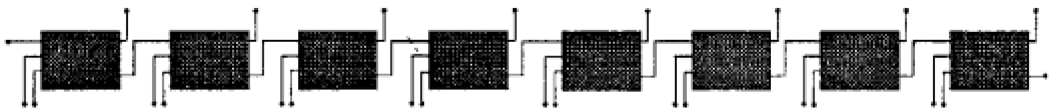
Model	D(M)	Cost	E(M)=D(M)/Cost
Model X	0.30	1.0	0.30
Model Y	0.65	1.5	0.43
Model Z	1.00	5.0	0.20

(c) E(M) for Each Model

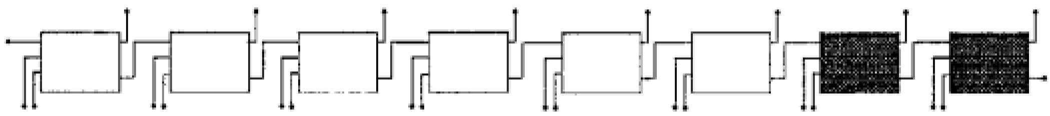
Figure 4–1 Example 1

Example 2 Here, we assume that the required diagnosis precision is the function-level, in the above example. In the early stage (Case 1' in Fig. 4–2), 8 components (full adders) are suspected. As shown in Fig. 4–2(c), $D(M)$ for each model is 1.0, which is the maximal value. Therefore, model X is selected because it has the least cost among the three.

Next we consider a diagnosis in the later stage (Case 2' in Fig. 4–2). Here, model X is precise enough to determine which full adder is broken. On the other hand, model Y and Z are so precise that they cost too much. Therefore, model X is also selected in this case.



(a) Case 1'



(b) Case 2'

Case 1'

Model	D(M)	Cost	E(M)=D(M)/Cost
Model X	1.00	1.0	1.00
Model Y	1.00	1.5	0.67
Model Z	1.00	5.0	0.20

Case 2'

Model	D(M)	Cost	E(M)=D(M)/Cost
Model X	1.00	1.0	1.00
Model Y	1.00	1.5	0.67
Model Z	1.00	5.0	0.20

(c) E(M) for Each Model

Figure 4–2 Example 2

Example 3 Consider that the observation cost is much larger than the computation cost, i.e. $T(M) \ll C$ for each model M. Then, the diagnosis cost is hardly affected by the computation cost, $T(M)$. For example, here, we assume that $C = 100.0$, in Example 1. The diagnosis cost, that is the sum of the observation cost and the computation cost, for model X, Y and Z are 101.0, 101.5, 105.0, respectively. Therefore, in both case, that is initial stage (Case 1") and later stage (Case 2"), model Z is selected because of its high diagnosability (Fig. 4–3).

This example shows that the proposed mechanism has an ability to adapt to the observation cost. The mechanism is also adaptive to the computation cost, which depends on the computation power. For example, a 1-MIPS computer requires great deal of computation time in comparison with a 100-MIPS machine. Because the proposed mechanism considers such a factor, it can select an appropriate diagnostic strategy.

Case 1"

Model	D(M)	Cost	E(M)=D(M)/Cost
Model X	0.56	101.0	0.0055
Model Y	0.62	101.5	0.0061
Model Z	1.00	105.0	0.0095

Case 2"

Model	D(M)	Cost	E(M)=D(M)/Cost
Model X	0.30	101.0	0.0030
Model Y	0.65	101.5	0.0064
Model Z	1.00	105.0	0.0095

Figure 4–3 Example 3

5 Discussion

The technique described in this paper is adaptive to several kinds of situation, such as the required precision, the given computation power, the observation cost, and the phase of a diagnosis. Although we assumed several restrictions to the diagnosis problem, the proposed mechanism is general and can naturally be extended to more general cases. First, the diagnosis precision can be specified more flexibly, instead of specifying some level in a physical hierarchy. For example, chip-level precision may be required for certain part of the target device, and board-level precision may be required for the others. Such a situation often occurs depending on the spare parts availability. Second, the physical hierarchy and the logical hierarchy are not necessary to be identical. Third, the observation cost may depend on a model, while we assumed that it is model independent. For instance, an output signal of a board may be easily observed, however it may require more cost to observe an output signal of a chip.

In order to apply the proposed techniques to a practical system, we must consider following problems. First, the computation for model selection must not be expensive, therefore, we should not evaluate all of the possible models. Hence, we must focus on several promising models heuristically. Second, in some domain, it may be preferable to improve the model diagnosability criterion because it does not estimate the number of required observations. For example, even if $D(m) = 0.9$ for a certain model m, a diagnosis with the model may require dozens of observations to gain the 90% of the necessary information. This means the criterion does not always estimate the diagnosability exactly. Finally, it is important to estimate the computation cost and the fault probability for each component appropriately. Inductive learning techniques[8] or analytical methods should be included.

Acknowledgment

This research has been carried out as a part of the Fifth Generation Computer Project. The authors would like to thank Katsumi Nitta of the Institute for New Generation Computer Technology for his support. They also express their appreciation to Masahiro Yamamoto and Takeshi Yoshimura for their encouragement in this work.

References

[1] Davis, R., "Diagnostic reasoning based on structure and behavior," *Artificial Intelligence,* Vol. 24, pp. 347–410, 1984.

[2] de Kleer, J., "Focusing on probable diagnosis," *Proc. AAAI-91,* Vol. 2, pp. 842–848, 1991.

[3] de Kleer, J. and Williams, B. C., "Diagnosing multiple faults," *Artificial Intelligence,* Vol. 32, pp. 97–130, 1987.

[4] de Kleer, J. and Williams, B. C., "Diagnosis with behavioral modes," *Proc. IJCAI-89,* Vol. 2, pp. 1324–1330, 1989.

[5] Genesereth, M. R., "The use of design descriptions in automated diagnosis," *Artificial Intelligence,* Vol. 24, pp. 411–436, 1984.

[6] Hamscher, W., "XDE: Diagnosing devices with hierarchic structure and known component failure modes," *Proc. CAIA-90,* Vol. 1, pp. 48–54, 1990.

[7] Koseki, Y., Nakakuki, Y., and Tanaka, M., "An adaptive model-Based diagnostic system," *Proc. PRICAI'90,* Vol. 1, pp. 104–109, 1990.

[8] Nakakuki, Y., Koseki, Y., and Tanaka, M., "Inductive learning in probabilistic domain," *Proc. AAAI-90,* Vol. 2, pp. 809–814, 1990.

[9] Reiter, R., "A theory of diagnosis from first principles," *Artificial Intelligence,* Vol. 32, pp. 57–95, 1987.