ELSEVIER

Artificial Intelligence 91 (1997) 103–130

Artificial Intelligence

A new approach to quantitative and credible diagnosis for multiple faults of components and sensors

T. Washio [a,*], M. Sakuma [b], M. Kitamura [b]

[a] *Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567, Japan*

[b] *Nuclear Engineering Department, Faculty of Engineering, Tohoku University, Sendai, Miyagi 980, Japan*

Received April 1996; revised September 1996

Abstract

Many practical applications of system diagnosis require the credible identification of multiple faults of nonlinear components and sensors in quantitative measures. However, the state of the art of diagnosis technique is considered to be still insufficient to meet these severe requirements. The approach of diagnosis using the traditional linear system identification theory can diagnose the disturbed parameters of a system in detail and evaluate the quantitative amplitude of the disturbance. However, it hardly provides the diagnosis of the multiple faults and the diagnosis of the components having high nonlinearity. On the other hand, some recent model-based diagnosis approaches can diagnose the multiple faults even for highly nonlinear components, though they do not provide the detailed diagnosis of elements indivisibly involved in components and the quantitative amplitudes of the faults.

The method proposed in this paper provides an efficient remedy to achieve all of the practical requirements, i.e., the credible, detailed and quantitative diagnosis of multiple faults of nonlinear components and sensors. Our study newly proposes the frameworks of optimal constraints and causal ordering of physical systems. Also, a systematic and strict theory to synthesize these frameworks together with the model-based diagnosis is provided to characterize an optimal consistency checking method in diagnosis and to evaluate quantitative amplitudes of faulty disturbances. First, the detection of faulty behaviors of an objective component is performed based on the quantitative consistency checking between observations and the optimal constraints, called as "minimal over-constraints", consisting of first principles in the components. Second, once if some inconsistencies are detected, a mathematical operation of model-based diagnosis derives the candidates of faulty elements and functions even under multiple fault conditions. Third, the anomalous quantities directly disturbed by the faulty elements are identified systematically based on causal ordering.

* Corresponding author. E-mail: washio@sanken.osaka-u.ac.jp.

PII S0004-3702(96)00060-4

Furthermore, the quantitative deviations of these quantities are evaluated by using the minimal over-constraints.

The performance of the proposed method is demonstrated through an example to diagnose an electric water heater. The ability of this diagnosis has been confirmed for the multiple faults in nonlinear and dynamic systems. © 1997 Elsevier Science B.V.

Keywords: Model-based diagnosis; Multiple faults; Minimal over-constrained subset; Causal ordering; Assumptive structural equation; First principle; Function; Process system

1. Introduction

The diagnosis of anomaly states is strongly needed in systems where high reliability is requested such as nuclear power plants and air planes. The diagnosis tasks in those applications require the credible, detailed and on-line identification of multiple faults of nonlinear components and sensors in quantitative measures.

The methodologies of diagnosis proposed so far can be categorized into (a) approaches based on the traditional linear system identification theory [3,9,24] and (b) approaches based on the recent theories and techniques developed in the artificial intelligence field. The latter can be further categorized into (b.1) synthesis of diagnostic rules based on knowledge of fault modes [10,16,22,32], generic diagnostic engine [20] and pattern ecognision [28], and (b.2) model-based consistency checking of causal constraints of normal systems [7,8,10,23].

Generally speaking, the methods (a) can diagnose the disturbed parameters of a system in detail and evaluate the quantitative amplitude of the disturbance under a given arrangement of sensors. However, it hardly provides the diagnosis of the multiple faults and the diagnosis of the components having high nonlinearity. On the other hand, most of the methods (b) can diagnose the multiple faults even for highly nonlinear components. Some methods of (b.1) can also provide detailed diagnosis on disturbed parameters and parts in components and sensors. However, the methods (b.1) have limitations of their applicability to unpredictable faults, since they use a priori knowledge of fault modes. In contrast, the methods (b.2) have an advantage to diagnose any unexpected faults without using knowledge of fault modes, though they do not provide the detailed diagnosis of elements indivisibly involved in components and the quantitative amplitudes of the faults. Also, many of the methods (b.2) assume a diagnostic environment that states at any point in a system can be probed. However, the arrangement of sensors in most of process systems are initially designed and fixed.

Accordingly, the state of the art does not provide an efficient remedy which addresses all of the following requirements.

(i) Diagnosis of highly nonlinear components,
(ii) diagnosis of elements indivisibly involved in components,
(iii) diagnosis of multiple element faults including sensor faults,
(iv) quantitative diagnosis of fault amplitudes,
(v) diagnosis under a given arrangement of sensors.

Some past works tried to address these issues. For example, the works based on the principle of the model-based diagnosis and the use of a nonlinear quantitative model of the objective system meet the requirements of (i), (iv) and (v) [21,31]. However, the main purpose of these works is the identification of the fault location in the granularity of the components in the objective system and not to address the issues of (ii). The other researches in the field of "sensor validation" to diagnose the integrity of sensors installed in the objective system have been reported in many literatures [3,9], since the reliability of sensors is not maintained in some operation conditions of nuclear power plants and air planes, where the sensors are exposed to the severe environments, e.g., high pressure and/or mechanical vibration. However, most of the methods require some assumptions such as integrity of some specific sensors and no faults in the components, and do not solve the issue of (iii).

This research proposes a generic method to overcome all of the issues previously stated under a premise that the objective component for diagnosis is represented by the constraints of first principles which may be nonlinear and dynamic. The approach presented here belongs to the aforementioned category of (b.2) in which only the knowledge of the model of a normal component is utilized to provide a highly credible diagnostic result. Nevertheless, the multiple faults of elements indivisibly involved in nonlinear and dynamic components can be diagnosed under a given arrangement of sensors. Our study newly proposes the frameworks of optimal constraints and causal ordering of physical systems. Also, a systematic and strict theory to synthesize these frameworks together with the model-based diagnosis is provided to characterize an optimal consistency checking method in diagnosis and to evaluate quantitative amplitudes of faulty disturbances. In the subsequent section, the overview of our method and an example problem for the demonstration throughout this paper are described. In the third section, the theory of each reasoning mechanism is explained, and its applicability is demonstrated through the example.

2. Overview of method and application

Fig. 1 shows the outline of the diagnosis method we propose. The knowledge required in the diagnostic reasoning is prepared in advance at the blocks of (A), (B) and (C) in off-line manners. First, a model consisting of first principles under the normal condition of the objective component including its sensors is given. Then in block (A), a certain type of optimum constraints is derived from the model to enable fault identification in high resolution [27,29,30]. In block (B), the knowledge of the correspondence between a set of first principles and a set of functions in the objective component is prepared. Block (C) is to derive the knowledge of dependency among quantities in the objective component by using an extended theory of *causal ordering* for physical systems [25–27,29,30]. This knowledge represents the orders of the determination of the values of quantities [11,12,19]. All of the diagnostic knowledge is derived based on the information of the normal component. The knowledge (A) and (C) are prepared systematically by off-line processing, while (B) depends on some expertise.

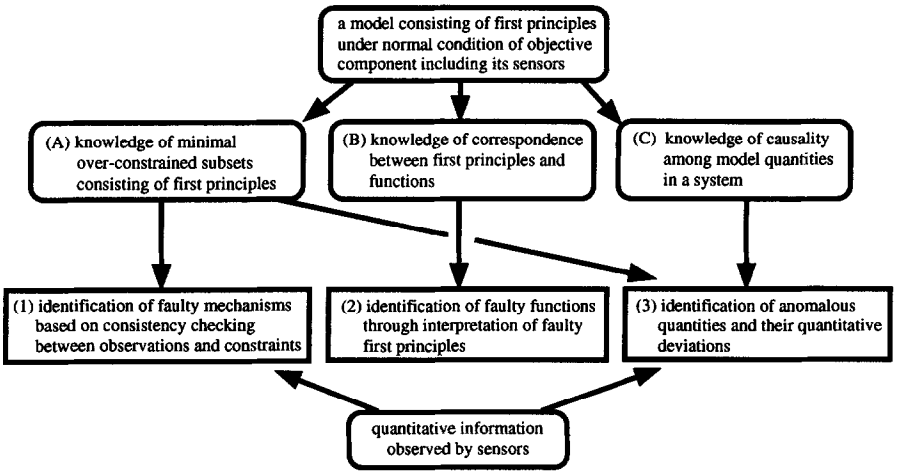


Fig. 1. Entire procedure of a proposed diagnosis method.

Once the diagnosis process is started, the identification of faults in the objective component proceeds in the order of the reasoning blocks (1), (2) and (3) in on-line manners. In reasoning block (1), the detection of faulty behaviors of an objective component is performed based on the quantitative consistency checking between the observed information and the knowledge (A). If some inconsistencies are detected, the model-based diagnosis [5,7,18] is applied. The constraints in the knowledge (A) are named as *minimal over-constrained subsets* [27,29,30]. They are defined to have the maximum resolution of the consistency checking to identify faulty elements in terms of first principles under the conditions that the arrangement of installed sensors is initially given and fixed during the operation of the component and that any quantitative expectations of dynamic component behaviors are not available without using the component description and the sensors' observations. Reasoning block (2) derives a set of suspicious functions through some operations on the resultant set of suspicious first principles and the knowledge (B). In short, suspicious first principles are interpreted into suspicious functions in the objective system by using the algorithm of the model-based diagnosis and the knowledge of the correspondence between the first principles and the functions. In the final reasoning block (3), the anomalous quantities directly disturbed by the faulty first principles are identified systematically based on the knowledge (C), i.e., the dependency information among the quantities. Furthermore, the quantitative deviations of these anomalous quantities are evaluated based on the knowledge (A). The on-line processing for the consistency checking and the deviation evaluation in reasoning blocks (1) and (3) does not require any combinatorial search, while the model-based diagnosis required in blocks (1) and (2) has the most computational complexity in the procedure to derive diagnoses from the inconsistency information.

The performance of the proposed method is exemplified through the diagnosis of an electric water heater depicted in Fig. 2. A resistant wire is electrically shielded from the surroundings, and its resistance has a nonlinear feedback effect from water temperature.

where $*$ indicates the measurements of sensors. Though each of them is quite simple stating only that the observed value of a sensor is equal to the actual value, the addition of these constraints enables the separation of the sensor failures from the component failures. The constraint-quantity matrix Q becomes as follows.

$$Q = \begin{pmatrix} I_p & I_g & I & R & V & F_h & H & M & T & I_p^* & I_g^* & V^* & M^* & T^* \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (13)$$

The rows correspond to Eqs. (1)–(12). Step P1 of the aforementioned procedure converts this matrix to the following Q' by removing the 10th–14th columns.

$$Q' = \begin{pmatrix} I_p & I_g & I & R & V & F_h & H & M & T \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (14)$$

As every row has some nonzero elements, step P2 is skipped. Step P3 in the loop derives a number of over-constrained subsets C^n. Some examples of C^n are

$$
\begin{aligned}
C^3 &= \{1,2,8,9\},\\
C^4 &= \{1,2,7,8,9\},\\
C^5 &= \{1,3,7,8,10,12\},\\
C^6 &= \{1,2,3,7,8,9,10,12\}.
\end{aligned}
\tag{15}
$$

In step P4, only the minimal over-constrained subsets are rested by removing every over-constrained subset which is a super set of any other over-constrained subset. In the example of Eq. (15), C^4 and C^6 are super sets of C^3 and C^5 respectively. Accordingly, C^4 and C^6 are removed. By continuing this process, finally the following 10 minimal over-constrained subsets are obtained.

$$
\begin{aligned}
M^3 &= \{1,2,8,9\},\\
M_1^5 &= \{1,3,7,8,10,12\},\\
M_2^5 &= \{2,3,7,9,10,12\},\\
M_1^7 &= \{1,4,5,6,8,10,11,12\},\\
M_2^7 &= \{2,4,5,6,9,10,11,12\},\\
M_3^7 &= \{3,4,5,6,7,10,11,12\},\\
M_1^8 &= \{1,3,4,5,6,7,8,10,11\},\\
M_2^8 &= \{1,3,4,5,6,7,8,11,12\},\\
M_3^8 &= \{2,3,4,5,6,7,9,10,11\},\\
M_4^8 &= \{2,3,4,5,6,7,9,11,12\}.
\end{aligned}
\tag{16}
$$

3.3. *Consistency checking*

Some definitions for the establishment of a systematic scheme of the consistency checking are presented first.

Definition 4 (*Deletion*). The deletion of a constraint c from SD is defined as an operation to remove the c while leaving the quantities involved in the c.

Definition 5 (*Self-contained subset* [12, 19]). A self-contained subset is a subset of SD in which the number of undetermined quantities is identical to that of the constraints while each constraint is mutually connected through the other quantities and constraints in the subset.

A self-contained subset determines the values of its quantities by itself. For example, if the deletion of Eq. (10) in the minimal over-constrained subset M_2^5 is performed, the resultant constraints in the M_2^5 become as follows.

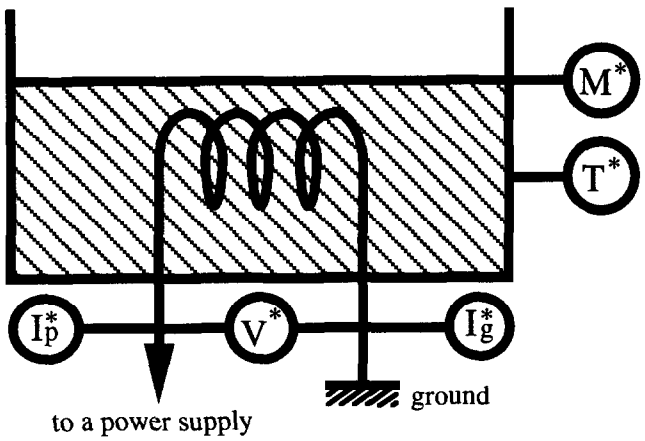


Fig. 2. An electric water heater.

The water is assumed to be always mixed well to avoid the spatial fluctuation of its temperature. The physical model of this process component is expressed in terms of first principles as follows.

$$I = I_{\mathrm{p}}, \tag{1}$$

$$I_{\mathrm{g}} = I, \tag{2}$$

$$V = IR, \tag{3}$$

$$F_{\mathrm{h}} = VI, \tag{4}$$

$$H = \int_{-\infty}^{t} F_{\mathrm{h}} \mathrm{d}t, \tag{5}$$

$$T = H/(cM), \tag{6}$$

$$R = r + k(T - t_c)^2, \tag{7}$$

- I_{p}: electric currents of a power supply,
- I: electric currents of a resistant wire,
- I_{g}: electric currents of the ground,
- R: resistance of the resistant wire,
- V: voltage of the resistant wire,
- F_{h}: heat generation rate of the resistant wire,
- H: contained heat of water,
- M: mass of water,
- T: temperature of water,
- c (= 4.2J/cal): specific heat coefficient of water,
- t_c (= 300K): standard temperature,
- r (= 100Ω): resistance of the wire at t_c,
- k (= 5Ω/K^2): temperature coefficient of the wire.

This example has the characteristics of nonlinearity, feedback loops and dynamic behaviors to demonstrate the generality of our proposing method.

3. Minimal over-constrained subsets and failure identification

3.1. Definition of minimal over-constrained subsets

The past methods of the model-based diagnosis have the following issues.

(i) Systematic criteria to select the constraints for the use are not well defined.

(ii) The limitation of the number and the location of sensors is not taken into account.

The first issue has been addressed by de Kleer et al. [4,6]. His approach is to identify faults at minimum probing steps. On the other hand, only the limited number of sensors are provided at specific points in components of the process systems such as nuclear power plants and air planes. Hence, the second issue must be also addressed in theoretical aspect. In this case, the most effective constraints for giving high resolution of diagnostic results under the limited information resource should be selected instead of ones having the best efficiency. For this purpose, Biswas and Yu proposed an approach of consistency checking to generate partial conflicts for each measurement quantity in process systems [2]. However, their approach assumes the linearity and steady state of the objective components. Our approach explained in this section addresses both issues while maintaining its applicability to the nonlinear and dynamic process components. The key idea of our approach is to derive a set of constraints giving the maximum resolution to identify faulty elements by using the information of sensor arrangement in the components.

First, Reiter's framework of the system definition is introduced for general discussion [18]. A system is a triple $(SD, COMPS, OBS)$ where the abbreviations stand for the system description, the system components and a set of observations, respectively. In our approach, each constraint c standing for a first principle in the complete model of a component belongs to SD, i.e., $c \in SD$, because they are used to derive the normal behaviors of the component. In addition, the constraints play another role in our framework: each constraint c provides a basic granule of anomaly, i.e., $c \in COMPS$. Accordingly, Eqs. (1)–(7) belong to both SD and $COMPS$ in our case. The constraints in SD are always over-constrained by the information in OBS obtained from a set of initially given sensor signals having some redundancy to monitor the state of the component. Especially, the over-constraints with one degree have the minimal sizes in the sense of the number of elements involved. They are expected to provide the maximum resolution in the consistency checking. Under this circumstance, the following definitions are proposed [27,29,30].

Definition 1. An *over-constrained subset of the nth order* (C^n) is a set of m constraints in SD involving n undetermined quantities where $m > n$ and each constraint is mutually connected through the other quantities and constraints in the set.

Definition 2. A *minimal over-constrained subset of the nth order* (M^n) is a set of $n+1$ constraints in *SD* involving n undetermined quantities and not involving any other over-constrained subsets where each constraint is mutually connected through the other quantities and constraints in the set.

The undetermined quantity is neither a directly observed quantity nor a quantity having a nominally fixed value in *SD* and *OBS*, and hence their values must be obtained by solving a simultaneous equation composed of the constraints. This categorization of undetermined and determined quantities explicitly introduces the information of sensor arrangement to the conflict generation in diagnosis. The following assumption must be introduced for the valid use of these definitions in the consistency checking.

Assumption 3. *The model constraints* $\{c \mid c \in SD\}$ *are mutually independent which provides each minimal over-constrained subset* M^n *to be well posed.*

The independency of model constraints in a nonlinear system is not always guaranteed, because the relations among quantities are state dependent. However, the model constraints of a process system are almost independent under normal operations in practical applications. Hence, the present over-constraint condition can be adopted widely to process systems.

3.2. Derivation of minimal over-constrained subsets

Although the efficiency of the derivation of all minimal over-constrained subsets is not the main issue for the off-line preparation of this knowledge in advance, a generic procedure maintaining the efficiency has been investigated [27,29,30]. First, a set of quantities *S* involved in the constraints in *SD* is given. Then, a constraint-quantity matrix of *SD* is defined as Q in the following manner.

Procedure 1.
If the ith constraint in *SD* involves the jth quantities in *S*,
then $Q(i,j) = 1$,
else $Q(i,j) = 0$,
where $Q(i,j)$ is the ij-element of the matrix Q.

The procedure depicted in Fig. 3 derives all minimal over-constrained subsets based on the constraint-quantity matrix Q. Step P1 derives another constraint-quantity matrix Q' representing the relations among undetermined quantities in the objective component by removing the columns of quantities determined by nominal values and sensor observations. This procedure introduces the knowledge of the arrangement of sensor and known parameters. Step P2 obtains minimal over-constrained subsets containing only a unique constraint. Such an example is the measurement of water amount in a container (*OBS*) to confirm if it is identical with the nominal value in *SD*. In the loop of the procedure, step P3 enumerates all over-constrained subsets where m constraints constrain n undetermined quantities under the condition of $m > n$. Step P4 chooses the minimal

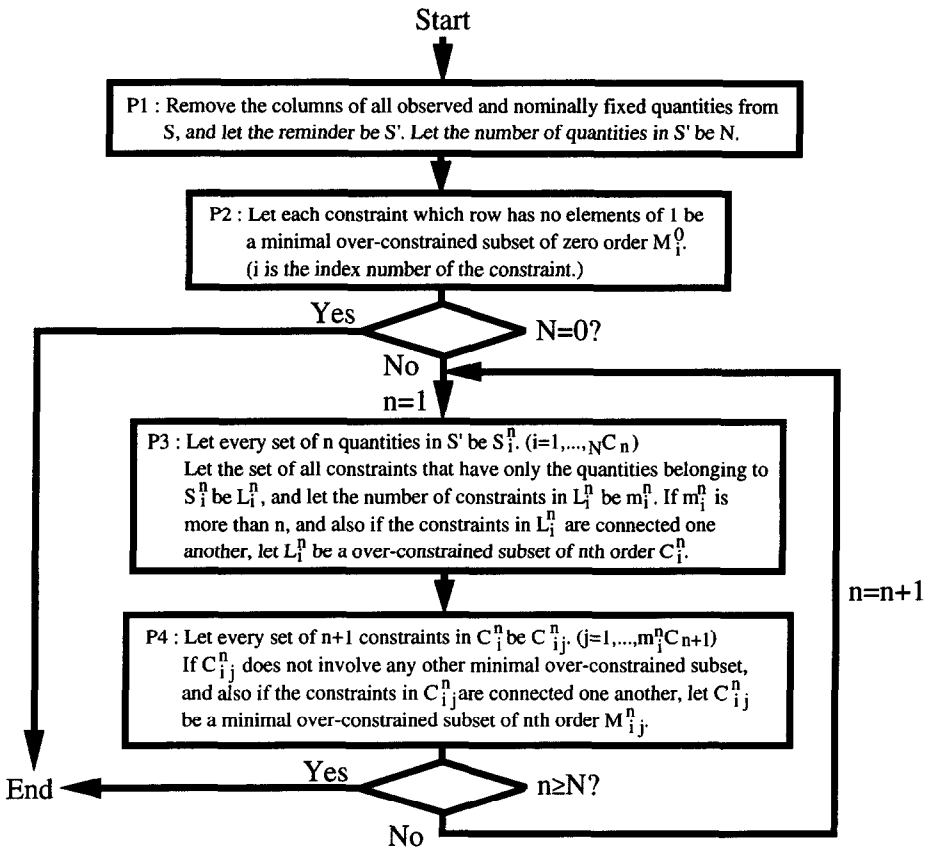


Fig. 3. A procedure to derive all minimal over-constrained subsets.

over-constrained subsets M^n having $n+1$ constraints among C^n obtained in P3. The complexity of this procedure is far less than the thorough search, since the enumeration number of subsets in *SD* with this procedure is almost proportional to 2^N where N is the number of the undetermined quantities in *SD*. Whereas, that of the thorough search is 2^K where K is the size of *SD* and larger than N.

We demonstrate these procedures through the aforementioned example of an electric water heater. Each sensor model is explicitly added to the component model to enable a uniform diagnosis of sensor failures and component failures.

$$I_p^* = I_p, \tag{8}$$

$$I_g^* = I_g, \tag{9}$$

$$V^* = V, \tag{10}$$

$$M^* = M, \tag{11}$$

$$T^* = T, \tag{12}$$

$$
\begin{aligned}
&I_g = I,\\
&V = IR,\\
&R = r + k(T - t_c)^2,\\
&I_g^* = I_g,\\
&V^*,\\
&T^* = T.
\end{aligned}
\tag{17}
$$

V^* and V remain in the model. The five constraints $\{I_g = I, V = IR, R = r+k(T-t_c)^2, I_g^* = I_g, T^* = T\}$ except V^* form a self-contained subset of five undetermined quantities $\{I, V, R, I_g, T\}$. In another case, if the deletion of Eq. (3) in the M_2^5 is performed, the resultant constraints in the M_2^5 become as follows.

$$
\begin{aligned}
&I_g = I,\\
&V, IR,\\
&R = r + k(T - t_c)^2,\\
&I_g^* = I_g,\\
&V^* = V,\\
&T^* = T.
\end{aligned}
\tag{18}
$$

Three self-contained subsets are obtained by this deletion. One is the set of two constraints $\{I_g = I, I_g^* = I_g\}$ containing two undetermined quantities $\{I, I_g\}$. Another is the set $\{R = r+k(T-t_c)^2, T^* = T\}$ containing two undetermined quantities $\{R, T\}$. Also, the set $\{V^* = V\}$ solely forms a self-contained subset of one undetermined quantity $\{V\}$.

The following theorems establish a systematic scheme of the consistency checking based on the above definitions and the minimal over-constrained subsets.

Theorem 6. *If the deletion of any one constraint c in a minimal over-constrained subset M is conducted, the M becomes one or more self-contained subset(s).*

Proof. By definition, the M becomes a subset of n constraints with n undetermined quantities by the deletion of a constraint.

(i) In case that the extra connections exist among quantities involved in the deleted c, the M remains to form a set of constraints where all quantities are connected through some constraints, and thus becomes a self-contained subset.

(ii) In case that the deleted c involves some unique connections among quantities, the M is partitioned into new Q subsets ($Q \geqslant 2$). Each new subset involves k_i undetermined quantities ($n = \sum_{i=1}^{Q} k_i$). As the M forms a set of constraints where all quantities are connected through some constraints by definition, each new subset also forms a set of connected constraints. Furthermore, as the M does not involve any other over-constrained subset by definition, each new subset

which is a part of the original M is not over-constrained. Therefore, each new subset involves k_i or less constraints. On the other hand, the total number of the undermined quantities in all new subsets, i.e., n, is identical to the total number of the remaining constraints. Accordingly, each new subset involves k_i constraints which is identical to the number of undetermined quantities in the subset, and hence each new subset is a self-contained subset. □

Theorem 7. *Any undetermined quantity x in a minimal over-constrained subset M appears in two or more constraints within the M.*

Proof. An assumption is introduced that an undetermined quantity x belongs to a unique constraint c in a minimal over-constrained subset M. In this case, the following smaller minimal over-constrained subset M' having n constraints and $n-1$ undetermined quantities can be always obtained by the removal of the c from the M.

$$M' = M - c \subset M.$$

This is contradictory to the definition of minimal over-constrained subsets. □

Theorem 8. *Two or more self-contained subsets which can independently determine the value of an undetermined quantity x always exist in a minimal over-constrained subset M.*

Proof. Due to Theorem 7, given an undetermined quantity x in M, a set of multiple constraints $C(x) = \{c_i \mid c_i \text{ involves } x, c_i \in M \text{ and } i = 1, \ldots, k(x)\}$, where $k(x) \geqslant 2$, always exists. On the other hand, the deletion of each $c_i \in C(x)$ always derives a set of self-contained subset(s) $S(c_i) = \{S_j \mid S_j \text{ is a self-contained subset derived by the deletion of } c_i \text{ from } M,\ j = 1, \ldots, m(c_i)\}$, where $m(c_i) \geqslant 1$, as stated in Theorem 6. Hence, the total number of the self-contained subsets in M which can determine the value of the x is $\sum_{i=1}^{k(x)} m(c_i)$, and it is always greater than or equal to 2. □

Any undetermined quantity x in a minimal over-constrained subset M can be chosen for the comparison among its values derived by the multiple self-contained subsets in M. Once the self-contained subsets for the derivation of the x have been set, the values of all undetermined quantities including the x in those subsets are sequentially determined by following the scheme of the causal ordering [11, 12, 19] while treating the directly observed and nominally fixed quantities as exogenous quantities. If the residuals among the values of the x exceed a certain threshold value, some constraints in M are considered to be faulty. This procedure is applied to every minimal over-constrained subset M in the *SD*. In the example of the electric water heater, one of the schemes of consistency checking is depicted in Fig. 4. This is the case to compare two values of V derived from the self-contained subset $\{I_g = I, V = IR, R = r + k(T - t_c)^2, I_g^* = I_g, T^* = T\}$ obtained by Eq. (17) and the set $\{V^* = V\}$ by Eq. (18).

For the demonstration of the consistency checking and the anomaly detection based on the minimal over-constrained subsets through the example, the following multiple failures are numerically simulated.

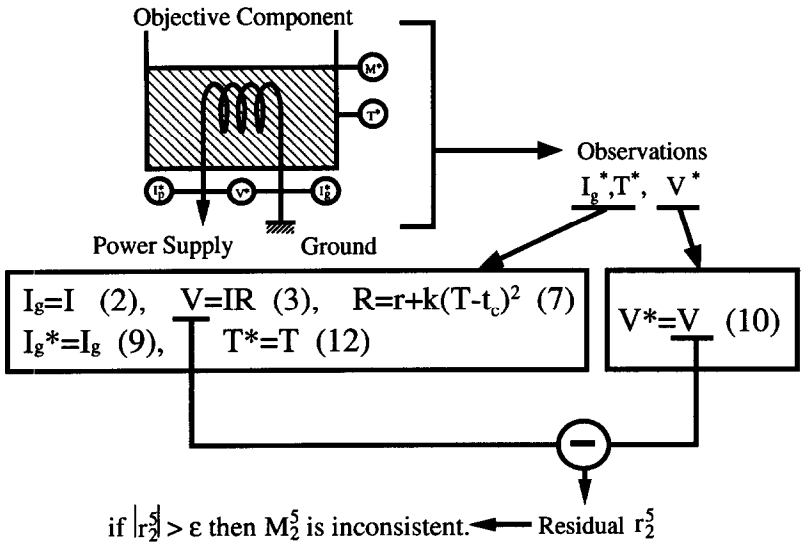


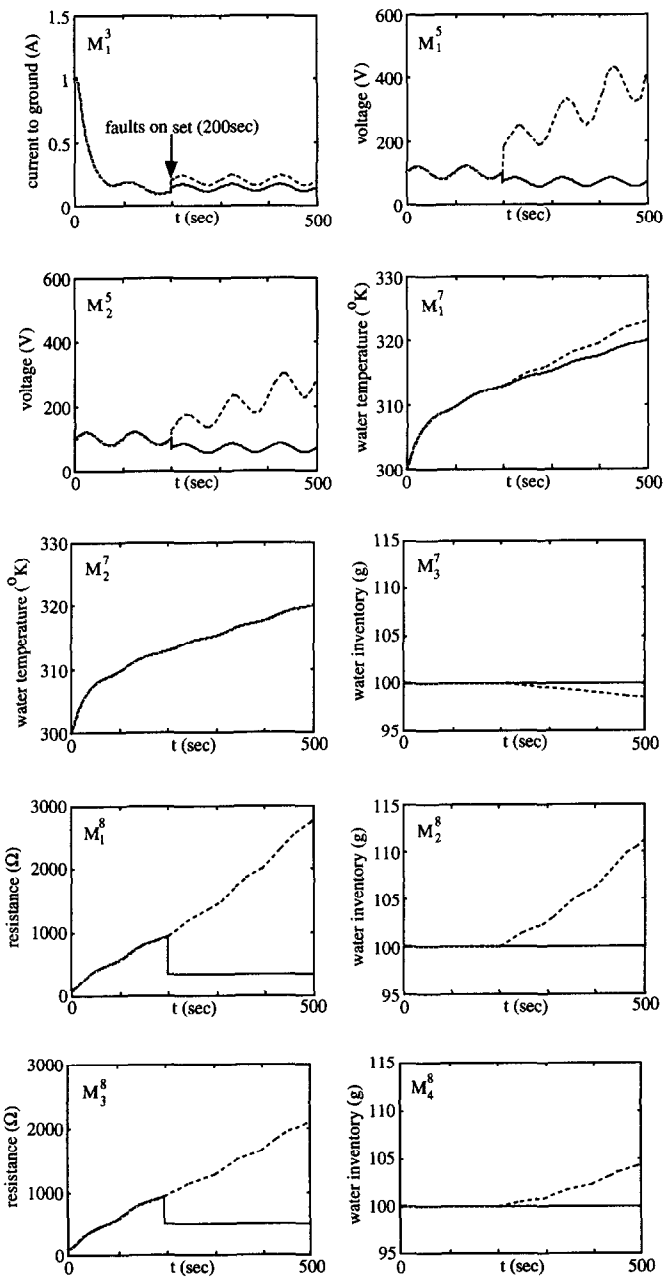
Fig. 4. A procedure to derive all minimal over-constrained subsets.

Case 1. The break of the electric shield and the deformation of the resistant wire happened by a mechanical shock at the time 200s. 30% of the electric current began to leak between the power supply and the resistant wire, and the resistance of the wire has been changed and fixed at the level of 500Ω.

Case 2. The electric shield of the resistant wire and the voltage sensor were broken by a mechanical shock at the time 200s. 30% of the electric current began to leak between the power supply and the resistant wire, and the indication of the voltage sensor has been changed and fixed at the level of 150V.

Ripples of 20% sine wave were added to the voltage of the power supply in order to evaluate the performance of the consistency checking in the dynamic behavior. Fig. 5 represents the result of the consistency checking for each minimal over-constrained subset M in the former case. The undetermined quantity x for the checking was arbitrarily chosen in each M. All subsets except M_2^7 became inconsistent at the time 200s. Fig. 6 shows the result of the latter where all subsets except M_4^8 became inconsistent at the time 200s.

Generally speaking, each minimal over-constrained subset is not very robust to the errors in the system model and the noise in the observation because of its low redundancy for consistency checking of an undetermined quantity x. However, various and efficient remedies in the field of numerical state estimation theory can be applied to this difficulty. For instance, the Kalman filter technique [14] provides a powerful measure to distinguish the physical inconsistency from the observation noise.



(solid lines: evaluations by a self contained subset,
dashed lines: evaluations by another self contained subset)

Fig. 5. A result of consistency checking in Case 1.

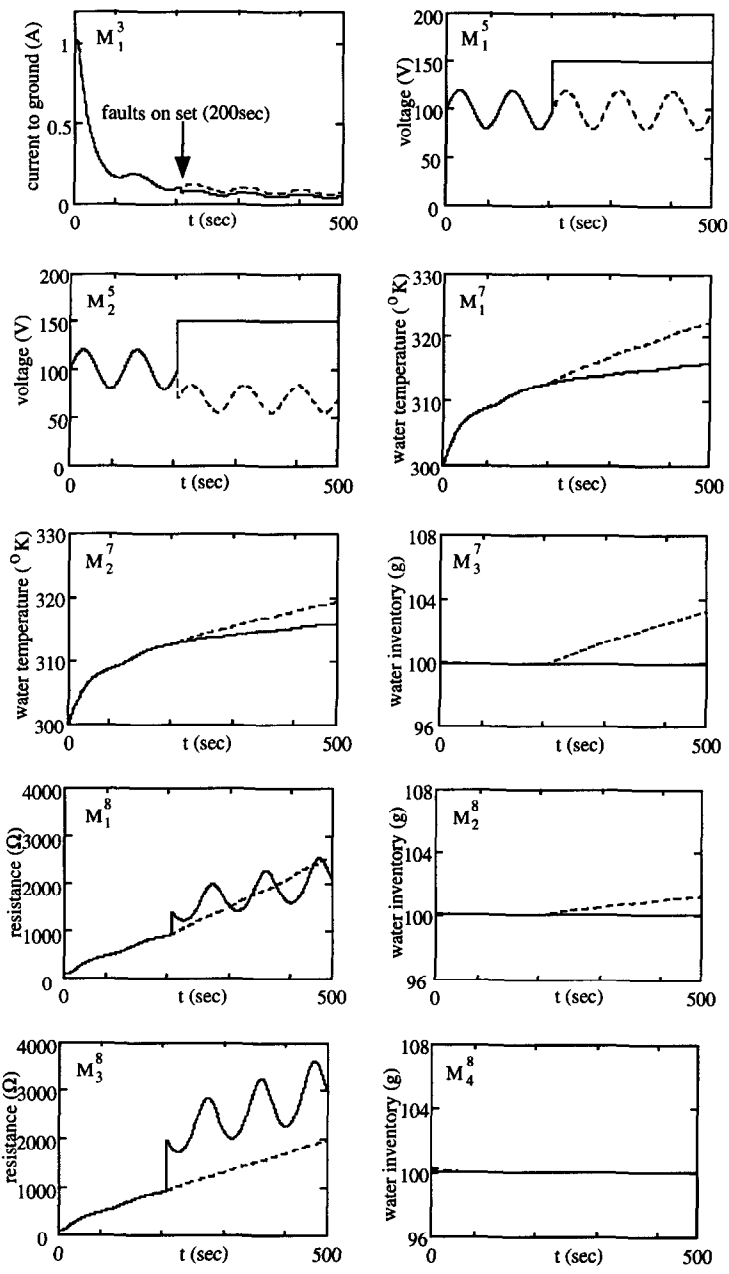


Fig. 6. A result of consistency checking in Case 2.

3.4. Identification of first principles disturbed by faults

First principles disturbed by faults are identified by applying model-based diagnosis to the result of the aforementioned consistency checking. This section explains some basics of the model-based diagnosis, and demonstrates their application to our framework.

The most representative theories of the model-based diagnosis are *minimal diagnoses* [7, 18] and *kernel diagnoses* [5]. The former derives the combinations of abnormal elements in *COMPS* from the information of inconsistency of constraints in *SD*. This theory has been extended to explicitly take the cancellation of inconsistency into account, where the influences of multiple faults mutually cancel not to cause any inconsistency in a set of constraints, and the theory of kernel diagnosis was established. The following definitions play central roles in these theories.

Definition 9 (*Conflict*). Let a clause C be a conjunction of literals $AB(c_i)$ and/or $\neg AB(c_j)$ where $c_i, c_j \in COMPS$ and $AB(c_i)$ stands for the abnormality of c_i. A conflict is the negation of a clause C, where C is not consistent with $SD \cup OBS$.

Definition 10 (*Minimal conflict*). A minimal conflict is a conflict which does not include any other conflict.

Once all minimal conflicts are given, the following standard procedure derives possible diagnoses [5, 7, 18].

Procedure 2.

(i) Multiply the minimal conflicts to give a disjunction of conjunctions.
(ii) Delete any conjunction containing a complementary pair of literals.
(iii) Delete any conjunction covered by some other conjunction.
(iv) The remaining conjunctions are the prime implicants of the original minimal conflicts.

These prime implicants are the possible interpretations of faulty states of the objective component. The following assumption made in the theory of minimal diagnoses [18] is adopted in our work as well as many other diagnosis methods.

Assumption 11 (Principle of parsimony). *A diagnosis is a conjecture that some minimal set of constraints are faulty.*

This principle derives the diagnoses which assume minimal numbers of faulty elements to explain the observed inconsistency.

Because each minimal over-constrained subset in *SD* is also the collection of constraints c belonging to *COMPS* in our approach, the result of the consistency checking on each minimal over-constrained subset directly entails a clause C which is a conjunction of literals $AB(c_i)$ and/or $\neg AB(c_j)$ where $c_i, c_j \in COMPS$ and C is not consistent

with $SD \cup OBS$. Thus, each result of the consistency checking yields a minimal conflict. In more detail, when a minimal over-constrained subset M^n is inconsistent, the proposition that "every constraint c_i is normal where $c_i \in M^n$" contradicts with the fact. Consequently, its minimal conflict becomes

$$\bigvee_{c \in M^n} AB(c). \tag{19}$$

For the consistent minimal over-constrained subset, some options provided by the diagnosis theories are applicable to derive its minimal conflicts. In our current work, the following assumption is introduced for process diagnosis.

Assumption 12. *The mutual cancellation of anomalous behaviors of multiple faults hardly occurs in process systems.*

This assumption makes our diagnosis basically equivalent to Raiman's approach [17]. As the proposition that "one or more constraints in M^n are abnormal" is against the consistent result of M^n, its negation

$$\neg AB(c) \quad \text{for every } c \in M^n, \tag{20}$$

becomes minimal conflicts.

In Case 1 of the example of the electric water heater, all subsets except M_2^7 became inconsistent at the time 200s as explained at the end of Section 3.3. The inconsistency of M^3 yields the following minimal conflict.

$$M^3\colon\ AB(1) \vee AB(2) \vee AB(8) \vee AB(9). \tag{21}$$

The minimal conflicts for the other inconsistent minimal over-constrained subsets are derived as well.

$$\begin{aligned}
M_1^5\colon\ & AB(1) \vee AB(3) \vee AB(7) \vee AB(8) \vee AB(10) \vee AB(12),\\
M_2^5\colon\ & AB(2) \vee AB(3) \vee AB(7) \vee AB(9) \vee AB(10) \vee AB(12),\\
M_1^7\colon\ & AB(1) \vee AB(4) \vee AB(5) \vee AB(6) \vee\\
& AB(8) \vee AB(10) \vee AB(11) \vee AB(12),\\
M_3^7\colon\ & AB(3) \vee AB(4) \vee AB(5) \vee AB(6) \vee\\
& AB(7) \vee AB(10) \vee AB(11) \vee AB(12),\\
M_1^8\colon\ & AB(1) \vee AB(3) \vee AB(4) \vee AB(5) \vee AB(6) \vee\\
& AB(7) \vee AB(8) \vee AB(10) \vee AB(11),\\
M_2^8\colon\ & AB(1) \vee AB(3) \vee AB(4) \vee AB(5) \vee AB(6) \vee\\
& AB(7) \vee AB(8) \vee AB(11) \vee AB(12),
\end{aligned} \tag{22}$$

$$M_3^8: \quad AB(2) \vee AB(3) \vee AB(4) \vee AB(5) \vee AB(6) \vee AB(7) \vee AB(9) \vee AB(10) \vee AB(11),$$

$$M_4^8: \quad AB(2) \vee AB(3) \vee AB(4) \vee AB(5) \vee AB(6) \vee AB(7) \vee AB(9) \vee AB(11) \vee AB(12).$$

The minimal conflicts of the consistent M_2^7 are derived as follows, since the proposition of "an equation in M_2^7 is abnormal" is against the consistency of M_2^7 under Assumption 12.

$$M_2^7: \quad \neg AB(2), \neg AB(4), \neg AB(5), \neg AB(6), \neg AB(9), \neg AB(10), \neg AB(11), \neg AB(12). \tag{23}$$

The aforementioned Procedure 2 derives the candidate diagnoses based on these minimal conflicts. In step (i), the following clauses are obtained for example.

$$AB(1) \wedge AB(2) \wedge \neg AB(2) \wedge AB(3) \wedge \neg AB(4) \wedge \neg AB(5) \wedge \neg AB(6) \wedge \neg AB(9) \wedge \neg AB(10) \wedge \neg AB(11) \wedge \neg AB(12), \tag{24}$$

$$AB(1) \wedge \neg AB(2) \wedge AB(3) \wedge \neg AB(4) \wedge \neg AB(5) \wedge \neg AB(6) \wedge AB(7) \wedge \neg AB(9) \wedge \neg AB(10) \wedge \neg AB(11) \wedge \neg AB(12), \tag{25}$$

$$AB(1) \wedge \neg AB(2) \wedge AB(3) \wedge \neg AB(4) \wedge \neg AB(5) \wedge \neg AB(6) \wedge \neg AB(9) \wedge \neg AB(10) \wedge \neg AB(11) \wedge \neg AB(12). \tag{26}$$

Formula (24) is removed at step (ii), as it involves a contradiction of $AB(2) \wedge \neg AB(2)$. Formula (25) is redundant and removed at step (iii), because it includes formula (26).

In this manner, formula (26) and the following solutions are determined.

$$AB(1) \wedge \neg AB(2) \wedge \neg AB(4) \wedge \neg AB(5) \wedge \neg AB(6) \wedge AB(7) \wedge \neg AB(9) \wedge \neg AB(10) \wedge \neg AB(11) \wedge \neg AB(12), \tag{27}$$

$$\neg AB(2) \wedge AB(3) \wedge \neg AB(4) \wedge \neg AB(5) \wedge \neg AB(6) \wedge AB(8) \wedge \neg AB(9) \wedge \neg AB(10) \wedge \neg AB(11) \wedge \neg AB(12), \tag{28}$$

$$\neg AB(2) \wedge \neg AB(4) \wedge \neg AB(5) \wedge \neg AB(6) \wedge AB(7) \wedge AB(8) \wedge \neg AB(9) \wedge \neg AB(10) \wedge \neg AB(11) \wedge \neg AB(12). \tag{29}$$

These solutions state the abnormality of the electric current balance of Eq. (1) and the wire resistance of Eq. (3), Eq. (1) and the wire resistance of Eq. (7), Eq. (3) and the electric current sensor of Eq. (8), and Eq. (7) and Eq. (8), respectively. Among these solutions, formula (26) and formula (27), saying the change of the wire resistance and the leakage of electric current between the power supply and the resistant wire, are the appropriate interpretations of Case 1 in the example.

The same approach derives the following two solutions in Case 2.

$$AB(1) \wedge \neg AB(2) \wedge \neg AB(3) \wedge \neg AB(4) \wedge \neg AB(5) \wedge \neg AB(6) \wedge \neg AB(7) \wedge \neg AB(9) \wedge AB(10) \wedge \neg AB(11) \wedge \neg AB(12), \tag{30}$$

$$\neg AB(2) \wedge \neg AB(3) \wedge \neg AB(4) \wedge \neg AB(5) \wedge \neg AB(6) \wedge \neg AB(7) \wedge AB(8) \wedge \neg AB(9) \wedge AB(10) \wedge \neg AB(11) \wedge \neg AB(12). \tag{31}$$

Formula (30) stands for the violation to the electric current balance between the power supply and the resistant wire (Eq. (1)) and the anomaly of the voltage sensor (Eq. (10)). This result is correct for the original faults in the simulation. An erroneous solution of formula (31) cannot be eliminated under the given *SD* and *OBS*.

If we apply the methods of minimal diagnoses and kernel diagnoses, the number of the candidate solutions may be significantly increased, because those theories take wider possibility of faults.

4. Identification of faulty functions

The results of the aforementioned diagnosis are represented by the constraints of first principles in *SD* and *COMPS*. They do not show faulty functions we interpret in the objective component. However, the diagnoses in terms of detailed functions sometimes give more effective information for trouble shooting than that in form of first principles. For doing so, the knowledge of the correspondence between each constraint and a set of functions is needed. More strictly speaking, when a constraint c in *COMPS* corresponds to a function c' in a set of functions *COMPS'*, and it is represented as $c \Rightarrow c'$, the set of functions $FNC(c)$ corresponding to c is:

$$FNC(c) \equiv \{c' \mid \forall c' \in COMPS' \text{ where } c \Rightarrow c'\}. \tag{32}$$

The systematic method to derive this knowledge has not been established yet, and the derivation must rely on expertise. However some recent work on the systematic function-based modeling may relax this issue [15]. In this paper, we assume that the knowledge of $FNC(c)$ for each c is correctly given by experts or designers of the component. For a set $\{FNC(c) \mid \forall c \in COMPS\}$, the following procedure derives the conflicts in terms of functions under Assumption 12.

Procedure 3.

(i) If c has been identified as abnormal, then $\{\bigvee_{c' \in FNC(c)} AB(c')\}$ is a conflict, else $\{\neg AB(c)$ for every $c \in M^n\}$ are conflicts.

(ii) Apply Procedure 2 to the set of derived conflicts.

The conflicts given by step (i) are not always minimal conflicts, since some $FNC(c)$s may include the other $FNC(c)$s. Therefore, step (ii) must be applied to obtain the valid diagnoses.

For the example of the electric water heater, we initially give the following set of $FNC(c)$s.

$$
\begin{aligned}
FNC(1) &= \{\text{electric shield between power supply and resistant wire } (F_1),\\
&\qquad \text{electric shield of resistant wire } (F_2)\},\\
FNC(2) &= \{\text{electric shield between resistant wire and ground } (F_3),\\
&\qquad \text{electric shield of resistant wire } (F_2)\},\\
FNC(3) &= \{\text{electric conduction of resistant wire } (F_4)\},\\
FNC(4) &= \{\text{heat generation of resistant wire } (F_5),\\
&\qquad \text{heat conduction from resistant wire to water } (F_6)\},\\
FNC(5) &= \{\text{heat conduction from resistant wire to water } (F_6),\\
&\qquad \text{heat containment of water container } (F_7)\},\\
FNC(6) &= \{\text{water containment of water container } (F_8)\},\\
FNC(7) &= \{\text{electric conduction of resistant wire } (F_4),\\
&\qquad \text{heat conduction from resistant wire to water } (F_6)\},\\
FNC(8) &= \{\text{electric current sensing between power supply}\\
&\qquad \text{and resistant wire } (F_9)\},\\
FNC(9) &= \{\text{electric current sensing between resistant wire}\\
&\qquad \text{and ground } (F_{10})\},\\
FNC(10) &= \{\text{electric voltage sensing } (F_{11})\},\\
FNC(11) &= \{\text{water amount sensing } (F_{12})\},\\
FNC(12) &= \{\text{water temperature sensing } (F_{13})\}.
\end{aligned}
\tag{33}
$$

If we consider Case 1 in the example, Procedure 3 gives the following diagnoses for both solutions of formulae (26) and (27).

$$
\begin{aligned}
&AB(F_1) \wedge AB(F_4) \wedge \neg AB(F_2) \wedge \neg AB(F_3) \wedge \neg AB(F_5) \wedge \neg AB(F_6) \wedge\\
&\neg AB(F_7) \wedge \neg AB(F_8) \wedge \neg AB(F_{10}) \wedge \neg AB(F_{11}) \wedge \neg AB(F_{12}) \wedge \neg AB(F_{13}).
\end{aligned}
\tag{34}
$$

Also, the following is the solution for formulae (28) and (29).

$$
\begin{aligned}
&AB(F_4) \wedge AB(F_9) \wedge \neg AB(F_2) \wedge \neg AB(F_3) \wedge \neg AB(F_5) \wedge \neg AB(F_6) \wedge\\
&\neg AB(F_7) \wedge \neg AB(F_8) \wedge \neg AB(F_{10}) \wedge \neg AB(F_{11}) \wedge \neg AB(F_{12}) \wedge \neg AB(F_{13}).
\end{aligned}
\tag{35}
$$

As a result, the combination of "electric shield between power supply and resistant wire (F_1)" and "electric conduction of resistant wire (F_4)", and the combination of "electric current sensing between power supply and resistant wire (F_9)" and "electric conduction of resistant wire (F_4)" are the diagnoses.

In Case 2,

$$
\begin{aligned}
&AB(F_1) \wedge AB(F_{11}) \wedge \neg AB(F_2) \wedge \neg AB(F_3) \wedge \neg AB(F_4) \wedge \neg AB(F_5) \wedge\\
&\neg AB(F_6) \wedge \neg AB(F_7) \wedge \neg AB(F_8) \wedge \neg AB(F_{10}) \wedge \neg AB(F_{12}) \wedge \neg AB(F_{13})
\end{aligned}
\tag{36}
$$

is the solution for formula (30), and

$$AB(F_9) \wedge AB(F_{11}) \wedge \neg AB(F_2) \wedge \neg AB(F_3) \wedge \neg AB(F_4) \wedge \neg AB(F_5) \wedge \neg AB(F_6) \wedge \neg AB(F_7) \wedge \neg AB(F_8) \wedge \neg AB(F_{10}) \wedge \neg AB(F_{12}) \wedge \neg AB(F_{13}) \quad (37)$$

for formula (31). Accordingly, the combination of "electric shield between power supply and resistant wire (F_1)" and "electric voltage sensing (F_{11})", and the combination of "electric current sensing between power supply and resistant wire (F_9)" and "electric voltage sensing (F_{11})" are the diagnoses.

5. Identification of anomalous quantities and their quantitative deviations

5.1. Causal ordering and identification of anomalous quantities

Causal ordering [11, 12, 19] is required to identify anomalous quantities directly disturbed by faulty mechanisms. In the conventional framework, the determination orders of process quantities are derived based on the specification of exogenous quantities in the system and the time derivative quantities to change their integrals. However, the application of these criteria frequently misleads the result of the causal ordering. For instance, any of the inlet flow and the outlet flow can be exogenous in a water pipe, because they just mutually balance. Furthermore, in Faraday's law of induction, i.e.,

$$\frac{dB}{dt} = -rot(E) \quad \text{or} \quad B = -\int rot(E)dt, \quad (38)$$

where B is magnetic flux density, and E is electric field intensity, and the change of B directly determines the value of E. In other words, the time integral determines its time derivative within a fundamental physical law. Accordingly, the arbitrary specification of exogenous quantities and the unique assumption of the causality in time differential equations may derive inappropriate interpretations of causal structures for physical systems. This discussion we have made [25, 26] is also supported by Iwasaki and Simon [13].

Based on these discussions, the authors proposed an extended theory to reduce the ambiguity of the causal ordering for physical systems [25, 27, 29, 30]. The specific heat law (Eq. (6)) in our example defines the quantitative relation between H and T under the exogenously given heat capacity cM. Either of the values of H and T is physically determined in this law, but cM is not changed by H and T within this law. The authors named this type of restrictions on the direction of the disturbance propagation among quantities as *inherent causal structure*, where it is independent to the applications of the physical constraint [25, 29, 30]. The details of the generic method to determine the inherent causal structure of each equation can be seen in the authors' work [25, 26]. Once the inherent causal structure of each equation has been identified, its knowledge representation with the quantitative relation of the equation is given by the following manner [25, 26]. First, let X_ℓ be a set of exogenously given quantities in the equation, and let Y_ℓ be a set of the other quantities in the equation. Any element in Y_ℓ has a

possibility to be physically determined in the equation. Subsequently, the quantities in each set are located on either of the right-hand side and the left-hand side by the following rule.

$$\text{If } X_\ell \neq \{\emptyset\} \text{ then } G_\ell(Y_\ell) = F_\ell(X_\ell), \text{ else then } G_\ell(Y_\ell) = 0, \tag{39}$$

where $X_\ell \cap Y_\ell = \emptyset$, $Y_\ell \neq \emptyset$, F_ℓ is the right-hand side of equation, and G_ℓ is the left-hand side of equation. The symbol "=" does not merely represent the equality of the right-hand side and the left-hand side. It states the causality as the possible determination order of quantities from the right-hand side to the left-hand side. This knowledge representation of an equation is called as an "assumptive structural equation". If an assumptive structural equation has only one quantity on its left-hand side, the value of the quantity is uniquely determined by the other quantities on the right-hand side. Thus, the following definitions can be made.

Definition 13. A *determining equation* is an equation having a unique quantity on the left-hand side.

Definition 14. A *determined quantity* is the unique quantity on the left-hand side of a determining equation.

When the model of the objective system is a set L of the assumptive structural equations, let a set of all quantities in L be S. The unambiguous determination orders of the quantities in L can be derived by the systematic procedure depicted in Fig. 7. Its resultant revised equations stand for the determination orders of the quantities from their right-hand sides to the left-hand sides.

The model of the electric water heater can be represented by the following assumptive structural equations [25,26].

$$I - I_p = 0, \tag{1'}$$

$$I_g - I = 0, \tag{2'}$$

$$V/I = R, \tag{3'}$$

$$F_h = VI, \tag{4'}$$

$$H = \int_{-\infty}^{t} F_h dt, \tag{5'}$$

$$H/T = cM, \tag{6'}$$

$$R = r + k(T - t_c)^2, \tag{7'}$$

$$I_p^* = I_p, \tag{8'}$$

$$I_g^* = I_g, \tag{9'}$$

$$V^* = V, \tag{10'}$$

$$M^* = M, \tag{11'}$$

$$T^* = T. \tag{12'}$$

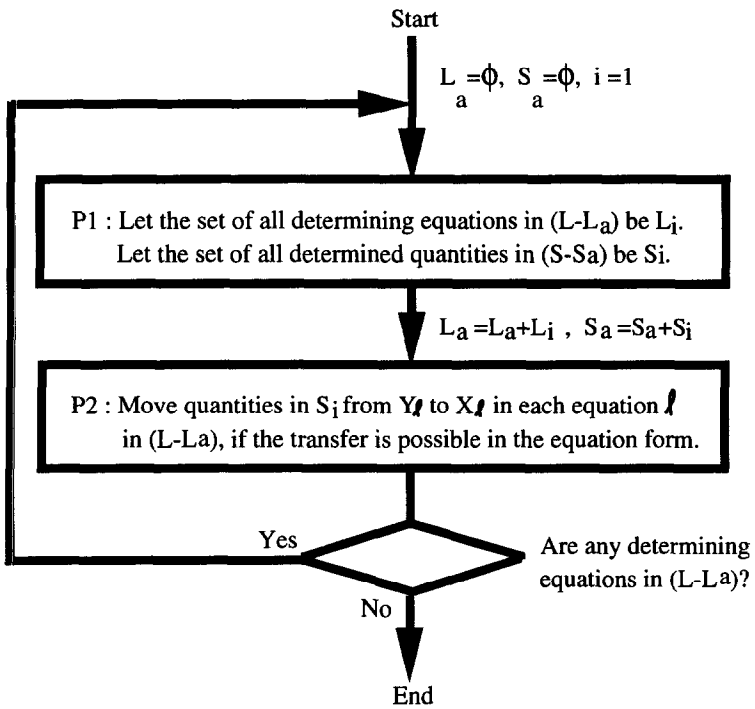


Fig. 7. A procedure of extended causal ordering.

All quantities in Eqs. (1′) and (2′) are located on their left-hand sides, because they are balance equations. Eq. (3′) (Ohm's law) defines the relation between V and I under an exogenously given resistance R. The heat generation rate F_h in Eq. (4′) (Joule's law) is unidirectionally determined by V and I, because this law represents an irreversible process in a thermodynamic phenomenon. Eq. (5′) stands for a standard time evolution. The structure of Eq. (6′) has been aforementioned. Eq. (7′) represents another irreversible process from T to R. The rests are for sensors, and their structures are trivial. The causal ordering procedure of Fig. 7 is applied to this model. In step P1, Eqs. (4′), (5′) and (7′)–(12′) are identified as determining equations. In step P2, a determined quantity H of Eq. (5′) is moved from the left-hand side to the right-hand side in Eq. (6′). Thus,

$$T = H/(cM). \tag{6''}$$

As no other determined quantities appear in any left-hand sides, the procedure goes back to step P1. Then a new determined quantity T in Eq. (6″) is identified. However, the loop is halted in step P2, because no T exists in any left-hand sides. The resultant equations of Eqs. (1′)–(5′), (6″), and (7′)–(12′) indicate the determination orders of the quantities. The orders are depicted in the form of a causal network in Fig. 8. The quantities remaining on each left-hand side of Eqs. (1′), (2′) and (3′) influence bidirectionally. Based on this result of the causal ordering, the quantities directly disturbed by the faults

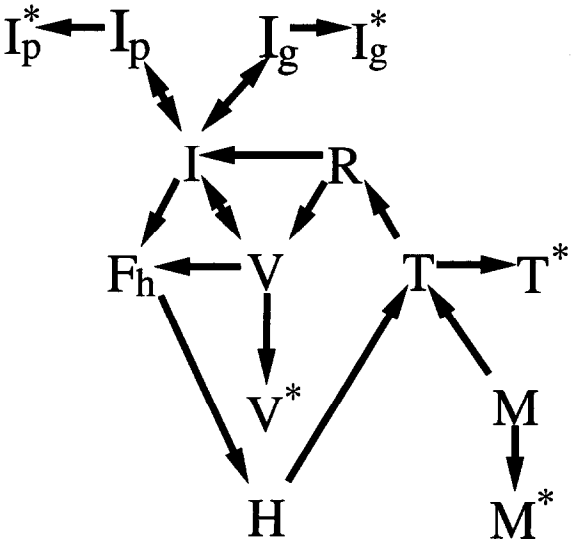


Fig. 8. The causal network of an electric water heater.

are identified. In the solution of formula (27) of Case 1, the directly disturbed quantities by the fault of Eq. (1) are known to be any of I_p and I based on the final structure of Eq. (1′). Moreover, the directly disturbed quantity by the fault of Eq. (7) is R. In case of formula (30) resulted in Case 2, the anomalous quantities directly disturbed by the fault of Eq. (1) are any of I_p and I as well. In practice, any of I_p and I can be changed by the break of the electric shield between the power supply and the resistant wire. Also, the quantity directly disturbed by Eq. (10) is identified as V^*.

Many physical systems partially involve the bidirectional causality as shown in this example, and the derivation of the exact causal structure of large systems is highly difficult within our physical intuition. Accordingly, this systematic causal ordering method provides an efficient remedy to identify anomalous quantities directly disturbed by faulty constraints.

5.2. *Evaluation of quantitative deviations*

The following theorem ensures the ability to evaluate the quantitative deviation of any anomalous quantity directly disturbed by any multiple faults identified in Section 3.4.

Theorem 15. *For any abnormal constraint $AB(c)$ belonging to a diagnosis D, at least one inconsistent minimal over-constrained subset exists which involves $AB(c)$ and does not involve the other abnormal constraints in the D.*

Proof. An assumption is introduced that any inconsistent minimal over-constrained subset involving the $AB(c)$ involves some other abnormal constraints in a D. In step (i) of the standard Procedure 2 described in Section 3.4, the following smaller diagnosis D'

can be always obtained by selecting an abnormal constraint except the $AB(c)$ from every minimal conflict corresponding to each inconsistent minimal over-constrained subsets.

$$D' = D - AB(c) \subset D.$$

This is contradictory to the requirement in step (iii) of Procedure 2 that D does not involve any other diagnoses. □

Theorem 16. *When a quantity x is contained in an $AB(c)$, any minimal over-constrained subset M involving the c includes a self-contained subset which determines the value of x without including the c. In the mean time, the minimal over-constrained subset M involves another self-contained subset which determines the value of x by using the c.*

Proof. Due to the aforementioned Theorem 8, the former self-contained subset is derived by the deletion of the c in the M. The latter is obtained by the deletion of a constraint connected with the c through x in the M. □

As a consequence of Theorem 15, for every $AB(c)$ which directly disturbs an anomalous quantity x, one minimal over-constrained subset always exists in which the $AB(c)$ is the unique abnormal equation. Accordingly, the actual anomalous value and the normal value of each anomalous quantity x can be always evaluated by the former and latter self-contained subsets in Theorem 16. The value of the x in these subsets is determined similarly to Section 3.3.

In case of the diagnosis formula (27) of Case 1, the directly disturbed quantities are any of I_p and I by the fault of Eq. (1) and R by the fault of Eq. (7). Some minimal over-constrained subsets contain a unique faulty constraint for both Eq. (1) and Eq. (7). We choose M^3 for Eq. (1) and M_2^5 for Eq. (7) having the smallest cardinal numbers, where they are convenient to save computational load. For I_p, its normal value is evaluated through Eqs. (1), (2) and (9) which are obtained by the deletion of Eq. (8). Its actual anomalous value is evaluated by an Eq. (8) which is made by the deletion of Eq. (1). Similarly, the normal value of I is derived from Eqs. (1) and (8), and its actual anomalous value from Eqs. (2) and (9). Furthermore, the normal value of R is derived from Eqs. (7) and (12) in M_2^5, and its actual one is from Eqs. (2), (3), (9) and (10). The quantitative deviations between those normal values and actual values are depicted in (a), (b) and (c) of Fig. 9. The actual value of I_p is greater than its normal value, and the actual value of I is smaller than it normal value. These are because of the leakage of the electric current between the power supply and the resistant wire. The value of the resistance R seems to be fixed at the level of 500Ω. These results are consistent with the actual conditions given in Case 1.

In case of the diagnosis of formula (30) in Case 2, the anomalous quantities are I_p and I disturbed by Eq. (1) and V^* by Eq. (10). For I_p and I, the procedure to evaluate their quantitative deviations is identical, and the same results are obtained. For the quantity V^*, its actual anomalous value is obtained by its direct measurement. The normal value is derived by the M_2^5 which involves Eq. (10) but not Eq. (1). Fig. 9(d) shows the quantitative deviations of this anomalous quantity, where the actual value

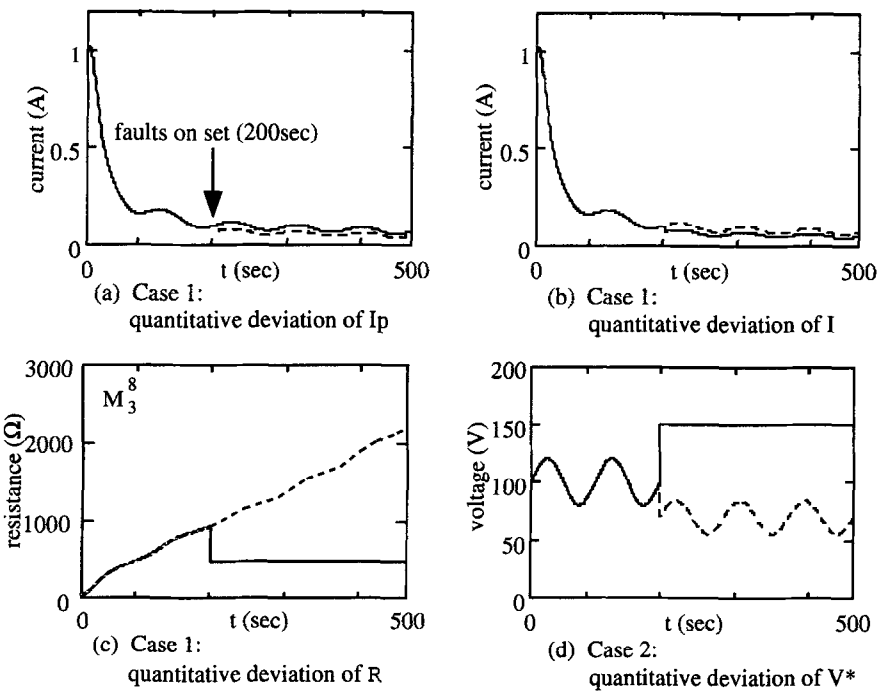


Fig. 9. The deviations of anomalous quantities.

of V^* is fixed at the level of 150V. The results are quantitatively consistent with the multiple failures introduced in Case 2.

6. Related work

In the ATMS-based methodology [7], conflicts are generated incrementally as new me s are made. A heuristic probing of the obvious and semi-obvious conflicts usi ty information [1] and one step look ahead random probing [6] indicate goc cy to identify faults, when the objective system is large and has many possible probing points. On the other hand, the preparation of all minimal over-constraints beforehand in our approach usually does not face the difficulty of the combinatorial explosion, since the size of *COMPS* and the number of given sensing points in a process component are quite limited. The definition of minimal over-constraints does not depend on any causality information.

The idea to prepare all schemes for consistency checking in advance has also been presented by Biswas and Yu [2]. They proposed "partial conflicts" to derive a conflict for each observation. The elements of *COMPS* in their work are parameters attributed to each

process mechanism. Their framework essentially requires the linearization of process models and the steady state assumption of the process, and hence is not applicable to highly nonlinear and dynamic systems. On the contrary, the basic element in our approach is a constraint among the parameters and state variables. The nonlinear and dynamic features of the system do not limit its application.

7. Conclusion

The operations used in this method are well defined and well combined to synthesize an efficient and credible procedure for diagnosis. This proposed method can diagnose multiple faults of elements and sensors occurred in a component. Non-linear and dynamic processes in which the quantities are intimately connected to one another can be diagnosed in high resolution. As the computational load required in the on-line processing is quite limited, the real-time and quantitative diagnosis can be performed without loosing the maximum performance of this method. The high possibility of this method to meet the severe requirements in practical applications has been demonstrated.

Acknowledgements

The authors wish to express their thanks to Dr. Hiroshi Motoda in Osaka University and Prof. Toyoaki Nishida in Nara Institute of Science and Technology for their useful comments. The authors extend their gratitude to Dr. Shuichi Koike and Dr. Hideaki Takahashi in Mitsubishi Research Institute, Inc. for their extensive support.

References

[1] R.R. Bakker and M. Bourseau, Pragmatic reasoning in model-based diagnosis, in: *Proceeding 10th European Conference on Artificial Intelligence*, Vienna (1992) 734–738.

[2] G. Biswas and X. Yu, A formal modeling scheme for continuous systems: focus on diagnosis, in: *Proceeding IJCAI-93*, Chambery (1993) 1474–1479.

[3] R.N. Clark and B. Campbell, Instrument fault detection in a pressurized water reactor pressurizer, *Nuclear Technol.* **56** (1982) 23–32.

[4] J. de Kleer, Focusing on probable diagnosis, in: *Readings in Model-Based Diagnosis* (Morgan Kaufmann, Los Altos, CA, 1992) 131–137.

[5] J. de Kleer, A.K. Mackworth and R. Reiter, Characterizing diagnosis and systems, *Artif. Intell.* **56** (1992) 197–222.

[6] J. de Kleer, O. Raiman and M.H. Shirley, One step lookahead is pretty good, in: *Readings in Model-Based Diagnosis* (Morgan Kaufmann, Los Altos, CA, 1992) 138–142.

[7] J. de Kleer and B.C. Williams, Diagnosing multiple faults, *Artif. Intell.* **32** (1987) 97–130.

[8] T.P. Hamilton, HELIX: a helicopter diagnostic system based on qualitative physics, *Artif. Intell. Eng.* **3** (3) (1988) 141–150.

[9] K.E. Holbert and B.R. Upadhyaya, An integrated signal validation system for nuclear power plants, *Nuclear Technol.* **3** (1990) 411–427.

[10] Y. Ishida and L. Eshelman, AQUA: integrating model-based diagnosis and symdrome-based diagnosis, Tech. Rept. CMU-CS-87-111, Carnegie-Mellon University, Pittsburg, PA (1987).

[11] Y. Iwasaki, Causal ordering in a mixed structure, in: *Proceedings AAAI-88*, St. Paul, MN (1988) 313–318.
[12] Y. Iwasaki and H.A. Simon, Causality in device behavior, *Artif. Intell.* **29** (1986) 3–32.
[13] Y. Iwasaki and H.A. Simon, Retrospective on "Causality in device behavior", *Artif. Intell.* **59** (1993) 141–146.
[14] R.E. Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME Ser. D; J. Basic Eng.* **82** (1960) 33–45.
[15] M. Lind, *Human Detection and Diagnosis of System Failures: The Use of Flow Models for Automated Plant Diagnosis* (Plenum Press, New York, 1981).
[16] D.A. Pearce, The induction of fault diagnosis systems from qualitative models, in: *Proceedings AAAI-88*, St. Paul, MN (1988) 353–357.
[17] O. Raiman, A circumscribed diagnosis engine, in: *Proceedings International Workshop on Expert Systems in Engineering*, Lecture Notes in Artificial Intelligence **462** (Springer, Berlin, 1990) 90–101.
[18] R. Reiter, A theory of diagnosis from first principles, *Artif. Intell.* **32** (1987) 57–95.
[19] H.A. Simon, *Models of Discovery* (Reidel, Dordrecht, 1977).
[20] P. Struss and O. Dressler, "Physical negation"—integrating fault models into the general diagnostic engine, in: *Proceeding IJCAI-89*, Detroit, MI (1989) 1318–1323.
[21] M. Takahashi, M. Kitamura and K. Sugiyama, Application of non-monotonic logic to failure diagnosis of nuclear power plant, in: *Proceedings 7th Power Plant Dynamics, Control & Testing Symposium*, Knoxville, TN (1989) 57.01–57.19.
[22] M. Takahashi, M. Kitamura and K. Sugiyama, Representation of generalized failure mechanism knowledge for diagnosis of nuclear power plant, *J. Atomic Energy Soc. Japan* **34** (1992) 678–692 (in Japanese).
[23] P. Torasso and L. Console, *Diagnostic Problem Solving* (North Oxford Academic, Oxford, 1989).
[24] J.L. Tylee, A generalized likelihood ratio approach to detecting and identifying failures in pressurizer instrumentation, *J. Nuclear Technol.* **56** (1982) 485–492.
[25] T. Washio, Causal ordering methods based on physical laws of plant systems, Tech. Rept. MITNRL-033, Nuclear Reactor Laboratory, Massachusetts Institute of Technology, Cambridge, MA (1989).
[26] T. Washio, Derivation of exogenously driven causality based on physical laws, *J. Japan. Soc. Artif. Intell.* **5** (1990) 482–491 (in Japanese).
[27] T. Washio and M. Kitamura, A new approach for plant component diagnosis based on credible and transparent physical knowledge, in: *Proceedings 8th Power Plant Dynamics, Control & Testing Symposium*, Knoxville, TN (1992) 15.01–15.06.
[28] T. Washio, M. Kitamura and K. Sugiyama, Development of failure diagnosis method based on transient information of nuclear power plant, *J. Nuclear Sci. Technol.* **24** (1987) 452–461.
[29] T. Washio, M. Sakuma and M. Kitamura, A diagnosis method for multiple process failures, in: *Working Papers DX-93: Fourth International Workshop on Principles of Diagnosis*, Aberystwyth, Wales (1993) 327–340.
[30] T. Washio, M. Sakuma and M. Kitamura, A diagnosis method for multiple process failures in a nonlinear and dynamic process, in: *Proceedings Tenth Biennial Conference of the Canadian Society for Computational Studies of Intelligence*, Banff, Alta. (1994) 139–146.
[31] N. Yamada and H. Motoda, A diagnosis method of dynamic system using the knowledge on system description, in: *Proceedings IJCAI-83*, Tokyo (1983) 255–229.
[32] T. Yamaguchi et al., Knowledge compiler II based on domain model and failure model, *J. Japan. Soc. Artif. Intell.* **7** (1990) 663–674 (in Japanese).