# Bayesian Model-based Diagnosis

Peter Lucas
Department of Computing Science
University of Aberdeen Aberdeen, AB24 3EU, Scotland, UK
E-mail: plucas@csd.abdn.ac.uk

**Abstract**

Model-based diagnosis concerns using a model of the structure and behaviour of a system or device in order to establish why the system or device is malfunctioning. Traditionally, little attention has been given to the problem of dealing with uncertainty in model-based diagnosis. Given the fact that determining a diagnosis for a problem almost always involves uncertainty, this situation is not entirely satisfactory. This paper builds upon and extends previous work in model-based diagnosis by supplementing the well-known model-based framework with mathematically sound ways for dealing with uncertainty. The resulting method integrates logical reasoning with probabilistic reasoning, and reasoning about the structure and behaviour of a system with reasoning by taking stochastic independence assumptions into account.

*Keywords*: model-based diagnosis, consistency-based diagnosis, Bayesian networks, probabilistic diagnosis, reasoning with uncertainty.

## 1 Introduction

There has been a great deal of research in the area of model-based diagnosis in the past two decades. Model-based diagnosis concerns using a model of the structure and behaviour of a system or device in order to establish why the system or device is malfunctioning. A number of different theories of diagnosis have been proposed, capable of dealing with different diagnostic problems. In fact, it has been recognised that there exists a wide range of different notions of diagnosis that can be taken as foundations for building diagnostic systems [12].

The two most popular theories of model-based diagnosis are abductive diagnosis and consistency-based diagnosis. In *abductive diagnosis*, typically a causal model of abnormal behaviour is used to explain observed abnormal findings in terms of a given causal theory [1, 11, 14, 15]. The theory possesses two different formalisations, one in terms of logic, as proposed by Console and Torasso [1], and by Poole [15]; the other one, proposed by Peng and Reggia, uses set theory [14]. The logical theory of abductive diagnosis is the more powerful one, which is a mere consequence of limitations imposed by Peng and Reggia on their set-theoretical formalisation, and not of limitations of set theory in general [11]. In contrast to abductive diagnosis, in *consistency-based diagnosis*, typically knowledge of the normal structure and behaviour of devices is used to determine what is wrong with a device or system [4, 7, 16]. The theory is in particular suitable when dealing with faults with which there exists little or no prior experience, such as done in troubleshooting of novel equipment and devices. The theory has the virtue of having a precise, formal underpinning in terms of logic [7, 16], supplemented by well-engineered computational environments [8, 6]. Although abductive diagnosis was originally meant for dealing with abnormal behaviour only, it was later extended to handle knowledge of normal behaviour as well [2]. Similarly, whereas the theory of consistency-based was originally focussed on diagnosis with a model of normal structure and behaviour, it was later extended to also incorporate abnormal behaviour [7]. However, the actual approaches followed in solving a diagnostic problem are still different: abductive reasoning versus consistency-based reasoning, yielding different solutions.

It has been argued that Bayesian networks offer a natural framework for dealing with uncertainty in abductive reasoning problems [13]. The reason for this is easy to understand: as Bayesian networks can be given a causal interpretation, there exists an obvious relationship between abductive reasoning in qualitative causal models and diagnostic reasoning in Bayesian networks. Such a relationship, however, does not exist with consistency-based diagnosis, as the models used in consistency-based diagnosis do not normally permit a causal interpretation. Handling uncertainty in consistency-based diagnosis has therefore been a difficult issue, for which no satisfactory solution has been found as yet. When researchers did explicitly consider the uncertainties involved in consistency-based diagnostic reasoning, they usually did so by rather restrictive methods.

In the present paper, we seek to develop methods for reasoning with uncertainty in consistency-based diagnosis that extends those described in the literature; probabilistic reasoning as offered by Bayesian networks and logical reasoning as done in consistency-based diagnosis are integrated. The aim is to combine the best of both worlds. As consistency-based diagnosis is not longer uniquely associated with models of normal structure and behaviour, but also with models of abnormal behaviour, we will assume that the latter type of knowledge may also be included.

The paper is organised as follows. In the next section, we briefly discuss previous work in dealing with uncertainty in consistency-based diagnosis. Next, the basic theory of consistency-based diagnosis is reviewed, and a number of properties required in subsequent sections are investigated. In the sections 4 and 5, a theory of Bayesian model-based diagnosis is developed by building upon previous work. The paper is rounded off by a discussion of what has been achieved.

## 2 Previous research

We start by summarising the major results of related previous research, and subsequently undertake to identify limitations and weaknesses of these results.

In his ground-breaking article, Reiter introduced for the first time a precise, formal description of model-based diagnosis of the consistency-based type [16]. This paper essentially discusses the logical structure of model-based diagnosis; the issue of how to deal with the uncertainties associated with the occurrence of faults in a device, however, was not touched upon. In a subsequent paper [6], De Kleer proposed to represent this uncertainty as a joint probability distribution $\Pr(C)$ on a set of components $C = \{C_1, \ldots, C_n\}$, where the adjustment of this probability distribution due to the observation of a particular finding $o$ would be computed by Bayes' rule:

$$\Pr'(C) = \Pr(C \mid o) = \frac{\Pr(o \mid C)\Pr(C)}{\Pr(o)}$$

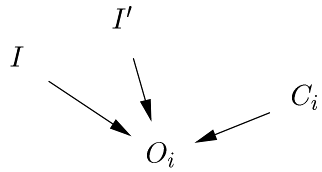


**Figure 1**: Typical structure of Bayesian-network fragment representing model-based diagnosis according to J. Pearl.

assuming $\Pr(o) > 0$. Bayes' rule reformulates the problem into determining the probabilities $\Pr(o|C)$ and $\Pr(C)$; the probability $\Pr(o)$ is a normalisation factor. Whereas the specification of the probability distributions $\Pr(o|C)$ and $\Pr(C)$ is exponential in their number of variables, computation of the other probability is hard in general.

In a more recent paper, Kohlas et al. propose another, yet related, approach [9]. Instead of adjusting a probability distribution when new evidence becomes available using Bayes' rule, they adjust it using knowledge of possible and impossible states of components as obtained by consistency-based reasoning. This may be viewed as determining the probabilistic diagnosis of faulty behaviour:

$$\Pr'(C) = \begin{cases} \frac{\Pr(C_1,\ldots,C_n)}{\Pr(D)} & \text{if } \{C_1,\ldots,C_n\} \in D \\ 0 & \text{otherwise} \end{cases}$$

where $D$ stands for a set of sets $C = \{C_1, \ldots, C_n\}$, with each set consisting of components that, when assumed to be behaving in a particular way, may be viewed as a diagnosis; the stochastic variables $C_i$, $1 \leq i \leq n$, indicate components that may or may not be faulty. The set $D$ contains all combinations of behaviours of components that, according to the theory of model-based diagnosis, are consistent with the observations; the other combinations of behaviours are inconsistent, and therefore left out. The probability distribution can thus be adjusted to reflect this new information.

The two approaches mentioned above are very much in line with traditional model-based diagnosis. Somewhat different is the suggestion made by J. Pearl that model-based diagnosis can also be accommodated in the framework of Bayesian networks [13]. In the suggested representation, (faulty) components are represented as independent conditioning vertices, i.e. vertices without incoming arcs. Their state influences the component's output, which is modelled by vertices with incoming arcs from the appropriate component vertex, and possibly from input vertices; outgoing arcs correspond to the component's output connections. The typical structure of a Bayesian network that represents a model-based diagnostic problem as suggested by Pearl is shown in Figure 1. In the figure, $I$ and $I'$ represent inputs to the component $C_i$; $O_i$ stands for the associated output. Both the inputs $I$ and $I'$ and the output $O_i$ may have an outgoing arc connected to outputs of other components.

Of the three different ways to incorporate uncertainty in model-based diagnosis discussed above, none is really satisfactory; in all three, some unrealistic assumptions are made, as is in fact acknowledged by De Kleer in his paper [6]. In the first two approaches, it is assumed that the components are mutually independent. Since the event of failure of one component is unlikely to be completely independent of failure of all other components, this assumption is much too strong. Pearl's third approach assumes that the failure of components is unconditionally independent, but information concerning observations may make components dependent, which is known as *induced dependence* [13]. When in addition any information about failure of those components is entered into the network, by instantiating one or more of the component vertices, failure of the other components becomes less likely, a phenomenon known as *explaining away* [13]. Thus, Pearl's approach is able to cope with dependences among components to some extent. However, the structure of a Bayesian network only reflects the functional dependences as following from the specification of the structure and behaviour of a system. Possible additional stochastic dependences among components are ignored, which is unsatisfactory, because Bayesian networks are especially suited for that purpose.

Also note that the assumption of (marginal) independence tends to overestimate the likelihood of single-fault diagnoses, because the likelihood that a particular component $C_i$ is faulty is usually much smaller than it to be functioning normally. The prior probability

$$\Pr(C_1, \ldots, C_n) = \prod_{i=1}^{n} \Pr(C_i)$$

will therefore be largest for single faults. This also holds for the first two approaches when there are observations available that are inconsistent with the model. This consequence of the independence assumption is unfortunate, because one of the attractive features of the theory of model-based diagnosis is its capability of dealing with multiple, interacting faults. Something similar holds for Pearl's approach, as basically posterior probabilities $\Pr(C_i \mid O)$ are determined for every component $C_i \in C$ for given observations $O$. Obviously, these are not real multiple fault diagnoses. Furthermore, in the work of Kohlas et al. the probabilistic influence of evidence on the likelihood that particular components are faulty is dealt with in a limited way, viz. only to adjust the probability distribution with respect to possible system states.

The relaxation of these assumptions is the main subject of this paper. First, the foundations of consistency-based diagnosis will be briefly reviewed. The incorporation of uncertainty into the logic framework will then be dealt with in the subsequent sections.

## 3 Consistency-based diagnosis

The theory of *consistency-based diagnosis* was initially introduced by De Kleer et al. [4, 8], and formalised by Reiter [16] and De Kleer et al. [7]. The basic idea is that a specification of the structure and behaviour of a system or device is used for diagnosing problems encountered with the system or device. Knowledge about the structure and behaviour of a system is used as a basis for simulation. The thus obtained simulated behaviour is then compared to the behaviour as observed from the actual system. A discrepancy between predicted and actual behaviour is interpreted as indicating that the system must be faulty. When the discrepancy is resolved by assuming particular components of the system to be faulty, a diagnosis has been established. We shall briefly review the formalisation of this idea below.

In the theory of consistency-based diagnosis, the structure and behaviour of a *system* $\mathcal{S}$ is defined as a triple $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$, where

- SD denotes a finite set of formulae in first-order predicate logic, specifying normal structure and behaviour, but sometimes including abnormal behaviour as well, called the *system description*;

- COMPS denotes a finite set of nullary function symbols or constants in first-order logic, corresponding to the *components* of the system that can be faulty;
- OBS denotes a finite set of formulae in first-order predicate logic, representing *observations.*

It is, in principle, possible to specify normal as well as abnormal (faulty) behaviour within a system description SD, although the emphasis typically lies on the specification of normal behaviour. Formulae in SD having the form

$$\forall x((\mathrm{COMP}_j(x) \wedge \neg \mathrm{Ab}(x)) \rightarrow \mathrm{NormBehaviour}_j(x))$$

specify the *normal* behaviour of the components $c \in \mathrm{COMPS}$ for which $\mathrm{COMP}_j(c)$ would hold true. There may also be formulae specifying abnormal (faulty) behaviour; these formulae have the form

$$\forall x((\mathrm{COMP}_j(x) \wedge \mathrm{Ab}(x)) \rightarrow \mathrm{AbBehaviour}_j(x))$$

A literal $\mathrm{Ab}(c)$, where 'Ab' is short for abnormal, when taken true, is interpreted as the assumption that component $c$ is faulty; a literals $\neg\mathrm{Ab}(c)$, on the other hand, means that it is assumed that component $c$ is acting normally. An *Ab clause* is defined as a disjunction consisting of Ab literals. An *Ab conjunct* is defined as a conjunction of Ab literals. A set of literals is interpreted as a conjunction of those literals, and vice versa.

Adopting the definition from [7], a diagnosis in the theory of consistency-based diagnosis can be defined as follows.

**Definition 1** (*consistency-based diagnosis*). *Let* $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$ *be a system. Let*

$$\Delta_P = \{\mathrm{Ab}(c) \mid c \in \mathrm{COMPS}\}$$

*be the set of all positive* 'Ab' *literals, and*

$$\Delta_N = \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS}\}$$

*be the set of all negative* 'Ab' *literals. Furthermore, let* $\Delta \subseteq \Delta_P \cup \Delta_N$ *be a set, such that*

$$\Delta = \{\mathrm{Ab}(c) \mid c \in C\} \cup \{\neg\mathrm{Ab}(c) \mid \mathrm{COMPS} \backslash C\}$$

*for some* $C \subseteq \mathrm{COMPS}$. *Then,* $\Delta$ *is a* consistency-based diagnosis *of* $\mathcal{S}$ *if the following condition, called the* consistency condition, *holds:*

$$\mathrm{SD} \cup \Delta \cup \mathrm{OBS} \nvDash \bot \tag{1}$$

*i.e.* $\mathrm{SD} \cup \Delta \cup \mathrm{OBS}$ *is satisfiable.*

Here, $\nvDash$ stands for the negation of the logical entailment relation, and $\perp$ represents a contradiction. A diagnosis is thus defined as a maximally consistent Ab conjunct, indicating for each component $c \in \mathrm{COMPS}$ whether it is faulty or not.

The notion of conflict is dual to the notion of diagnosis; it is of central importance to the theory. We again follow [7] in defining it.

**Definition 2** (*conflict*). *Let* $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$ *be a system. If* $\phi$ *is a non-empty Ab clause, such that*

$$\mathrm{SD} \cup \mathrm{OBS} \vDash \phi$$

*then* $\phi$ *is called a* conflict.

The term 'conflict' comes from the earlier literature, in which an Ab conjunct $\psi$ for which $\mathrm{SD}\cup\psi\cup\mathrm{OBS}\vDash\bot$ would be called a conflict. Obviously, there exists in this case an Ab clause $\phi$ for which $\psi=\neg\phi$ and $\mathrm{SD}\cup\mathrm{OBS}\vDash\phi$. Hence, the relationship between these two definitions is straightforward.

We have the following important relationship between diagnoses and conflicts, using Definition 2, again based on [7].

**Theorem 1.** *Let* $\mathcal{S}=(\mathrm{SD},\mathrm{COMPS},\mathrm{OBS})$ *be a system with set of conflicts* $\Pi$, *then the maximally consistent Ab conjunct* $\Delta$ *is a diagnosis of* $\mathcal{S}$ *iff*

$$\Pi\cup\Delta\nvDash\bot$$

*Proof.* ($\Rightarrow$) Since $\Delta$ is a diagnosis, it holds that

$$\mathrm{SD}\cup\mathrm{OBS}\cup\Delta\nvDash\bot$$

It follows that $\mathrm{SD}\cup\mathrm{OBS}\nvDash\bot$, and any conflict $\phi\in\Pi$ will therefore be consistent with $\Delta$. ($\Leftarrow$) Suppose that $\Delta$ is not a diagnosis, i.e.

$$\mathrm{SD}\cup\mathrm{OBS}\cup\Delta\vDash\bot$$

and that $\Pi\cup\Delta\nvDash\bot$. Then, it follows that $\mathrm{SD}\cup\mathrm{OBS}\vDash\neg\Delta$. By definition, $\neg\Delta\in\Pi$, as $\neg\Delta$ is an Ab clause. But then, it must hold that $\Pi\cup\Delta\vDash\bot$, contradicting the assumption at the beginning. Hence, $\Delta$ must be a diagnosis. $\diamond$

This relationship between conflicts and diagnoses is not only of theoretical significance: conflicts are commonly used in diagnostic reasoning engines to act as a sort of intermediary result in computing diagnoses, as follows from the following corollary.

**Corollary.** *Let* $\mathcal{S}=(\mathrm{SD},\mathrm{COMPS},\mathrm{OBS})$ *be a system with set of conflicts* $\Pi$ *and set of diagnoses* $D$, *then for each* $\phi\in\Pi$ *and for each* $\Delta\in D$ *it holds that:* $\neg\phi\wedge\Delta\vDash\bot$.

*Proof.* Note that from Theorem 1 it follows that for each $\phi\in\Pi$:

$$\phi\wedge\Delta\nvDash\bot \tag{2}$$

Now, $\Delta$ contains an Ab literal for every component $c\in\mathrm{COMPS}$. From (2) it follows that the Ab clause $\phi$ contains at least one Ab literal with the same sign and concerning the same component as a literal in $\Delta$. $\diamond$

Hence diagnoses overlap, although usually not completely, with conflicts. This insight is the basis of a number of algorithms that construct diagnoses from sets of conflicts, such as the hitting-set algorithm by Reiter [16], the constructor algorithm in the General Diagnostic Engine (GDE) [8], and the component consequences algorithm by Darwiche [3].

We illustrate the concepts from this section with a classical example from the literature on consistency-based diagnosis (cf. [7]).

**Example 1.** Consider Figure 2, which depicts an electronic circuit with three multipliers,

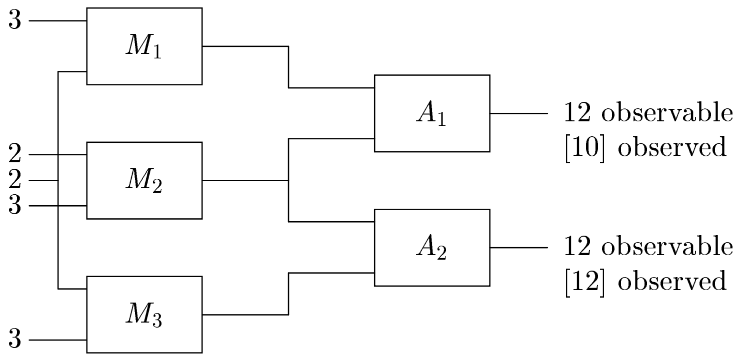


**Figure 2**: Multiplier-adder circuit.

referred to as $M_1$, $M_2$ and $M_3$, and two adders, denoted by $A_1$ and $A_2$, i.e. COMPS $= \{M_1, M_2, M_3, A_1, A_2\}$. The system description SD consists of formulae like

$$\forall x((\text{Multiplier}(x) \wedge \neg\text{Ab}(x)) \rightarrow (o(x) = i_1(x) \times i_2(x)))$$

for describing the behaviour of components, such as Multiplier($M_1$), and

$$o(M_1) = i_1(A_1)$$

(the output of $M_1$ is equal to the first input to $A_1$) for representing knowledge about the structure of the system. Now, the output of the system, $\{o(A_1) = 10, o(A_2) = 12\}$, differs from the one expected according to the simulation model, i.e. it gives rise to an inconsistency

$$\text{SD} \cup \text{OBS} \cup \{\neg\text{Ab}(c) \mid c \in \text{COMPS}\} \vDash \bot$$

as is also indicated in Figure 2. There exist two conflicts with minimal number of literals (the others are subsumed):

$$\begin{array}{r}\text{Ab}(A_1) \vee \text{Ab}(M_1) \vee \text{Ab}(M_2)\\ \text{Ab}(A_1) \vee \text{Ab}(M_1) \vee \text{Ab}(M_3) \vee \text{Ab}(A_2)\end{array}$$

One of the diagnoses is:

$$\text{Ab}(A_1) \wedge \neg\text{Ab}(A_2) \wedge \neg\text{Ab}(M_1) \wedge \neg\text{Ab}(M_2) \wedge \neg\text{Ab}(M_3)$$

i.e. when assuming component $A_1$ to be defective, the inconsistency is resolved. $\diamond$

Since conflicts are used as a basis for computing diagnoses, it is worthwhile to look for further relationships between the two concepts, in addition to the fundamental result of Theorem 1.

**Proposition 1.** *Let $\mathcal{S} = (\text{SD}, \text{COMPS}, \text{OBS})$ be a system with a set of conflicts $\Pi$ and a set of diagnoses $D$, then $D \equiv \neg \bigvee\{\neg\phi \mid \phi \in \Pi\}$.*

*Proof.* For convenience, we shall interpret the disjunctive normal form of a formula as a set of elements; the set $D$ is also viewed as consisting of disjunctions of conjuncts (diagnoses). Let $\bar{\Pi} = \bigvee\{\neg\phi \mid \phi \in \Pi\}$, then we have that for each $\psi \in \bar{\Pi}$:

$$\text{SD} \cup \psi \cup \text{OBS} \vDash \bot$$

whereas for each $\Delta \in D$, it holds that:

$$\mathrm{SD} \cup \Delta \cup \mathrm{OBS} \nvDash \bot$$

and every possible maximally consistent combination of Ab literals is covered by $\bar{\Pi}$ and $D$ together. Hence,

$$\bar{\Pi} \vee D \equiv \top \tag{3}$$

Finally, from (3) it follows that $D \equiv \neg\bar{\Pi}$. $\diamond$

Hence, we may conclude that as soon as we have obtained the set of conflicts, finding the associated diagnoses is a trivial process.

The notion of cover is used in the following to investigate subsets of diagnoses.

**Definition 3** (*cover*). *Let $\phi$ and $\psi$ be Ab conjuncts, such that any interpretation $\mathcal{I}$ that satisfies $\phi$ also satisfies $\psi$, then $\psi$ is said to* cover *$\phi$. This is denoted by $\psi \preceq \phi$.*

Note that the cover relation defines a partial order on the set of Ab conjuncts.

In addition to the notion of diagnosis, De Kleer et al. define the notion of *partial diagnosis* [7]. A partial diagnosis combines information from different, but related diagnoses, in a single representation. For example, suppose that

$$\mathrm{Ab}(c_1) \wedge \mathrm{Ab}(c_2) \wedge \mathrm{Ab}(c_3)$$

and

$$\mathrm{Ab}(c_1) \wedge \mathrm{Ab}(c_2) \wedge \neg\mathrm{Ab}(c_3)$$

are two alternative diagnoses of a system $\mathcal{S}$. These two diagnoses are identical, with the exception of the last literal. Clearly, component $c_3$ may either be faulty or not faulty, as long as $c_1$ and $c_2$ are both faulty. The partial diagnosis $\mathrm{Ab}(c_1) \wedge \mathrm{Ab}(c_2)$ conveys exactly that information; note that both diagnoses above are covered by this partial diagnosis. As there may be several partial diagnoses related to each other in the sense discussed above, it seems desirable to single out special partial diagnoses having some minimality property. This is exactly what is achieved by the notion of *kernel diagnosis*, which is defined as a partial diagnosis that can only be covered by itself.

The following property of diagnoses will also be used in the following.

**Proposition 2.** *Let $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$ be a system, and let $\Delta$ and $\Delta'$ be two diagnoses of $\mathcal{S}$, such that $\Delta \not\equiv \Delta'$, then $\Delta \wedge \Delta' \equiv \bot$.*

*Proof.* The diagnoses $\Delta$ and $\Delta'$ are conjuncts, different in at least one literal. That literal, however, must concern the same component, hence the result. $\diamond$

Note that this property only holds for (full) diagnoses, and not generally for partial diagnoses.

# 4 Uncertainty in model-based diagnosis

Consistency-based diagnosis can be viewed as a form of assumption-based problem-solving [15]; resulting diagnoses may therefore be seen as assumptions that hold as long as no refuting evidence is available. Diagnoses are thus taken to be uncertain. However, without quantifying this uncertainty explicitly, there is no real choice here but to assume that all possible diagnoses are equally likely, even given that this is extremely rare. Some researchers have suggested using syntactic properties as a measure of likelihood, such as the number of positive Ab literals in a diagnostic solution, interpreting diagnoses with fewer positive literals as more likely [18]. However, these approaches are not really satisfactory, as they offer no systematic approach that will work in any problem domain. As discussed above, there have also been proposed several probabilistic approaches to this problem. Here, we take the work by Kohlas et al. [9], which was discussed above, as a starting point and undertake to generalise it.

### 4.1 Probabilistic independence structure

Uncertainty with respect to the normal or faulty behaviour of components is expressed by a joint probability distribution on the set of components COMPS:

$$\Pr(C_1, \ldots, C_n)$$

where $C_i$, for each $i$, $1 \leq i \leq n$, is a stochastic variable that when taking the value *true*, also denoted by $c_i$, indicates component $c_i$ to be faulty; when $C_i$ takes the value *false*, also denoted by $\neg c_i$, component $c_i$ is assumed to be normal. This yields a 1–1 correspondence between Ab formulae and Boolean expressions involving $c_i$ and its negation. It will usually be clear from the context whether by a particular expression a Boolean expression within probability theory is intended, or its corresponding logical formula within the theory of consistency-based diagnosis.

In contrast to [6] and [9], it is not assumed that the variables $C_i$ are mutually independent, because information of whether a component is faulty or not usually influences our knowledge about other components. We adopt, therefore, the very general idea that the joint probability distribution Pr can be factorised according to our knowledge of a given problem domain, such as knowledge of causal relationships between (mal)function of components and of dependences and (un)conditional independences among components:

$$\Pr(C_1, \ldots, C_n) = \prod_{i=1}^{n} \Pr(C_i \mid \pi(C_i))$$

where $\pi(C_i)$ are the variables on which the variable $C_i$ is conditioned according to the factorisation, possibly taking conditional independence information into account. Components that are behaving normally, however, are assumed to be mutually independent, i.e.

$$\Pr(\neg c_i \mid \hat{\pi}_N(C_i)) = \Pr(\neg c_i)$$

where the expression $\hat{\pi}_N(C_i)$ stands for $\neg c_{j_1}, \ldots, \neg c_{j_{m_i}}$, and where $\Pr(\neg c_i \,|\, \hat{\pi}_N(C_i))$ is obtained from the factorisation. Thus:

$$\Pr(\neg c_1, \ldots, \neg c_n) = \prod_{i=1}^{n} \Pr(\neg c_i)$$

Hence, knowing that particular components are behaving normally does not influence our knowledge of other components. This seems the only general independence assumption that can be made without introducing unrealistic limitations.

A factorisation of a probability distribution can also be depicted as a directed acyclic graph $G = (C, A)$, where in this case the set of vertices $C$ corresponds to the set of components COMPS; the set of arcs $A \subseteq C \times C$ reflects the stochastic dependences and independences among the variables, and follows the structure of the factorisation. One could also start with defining a directed acyclic graph $G$ to model the independence structure of the problem, and next define a joint probability distribution Pr on the variables corresponding to the vertices $C$. The result will be a Bayesian network $\mathcal{B}_{\mathcal{S}} = (G, \Pr)$ of the system $\mathcal{S}$ [10]. In the following, this is the approach that will be adopted.

### 4.2 Handling evidence

Now, let us assume that some observations are obtained for a system $\mathcal{S}$. These observations could in principle influence our knowledge of the likelihood of malfunction of certain components of the system. However, the extent of influence cannot be established by instantiating variables in the Bayesian network $\mathcal{B}_{\mathcal{S}}$, as is usually done in Bayesian-network applications, as observation variables have not been included. Instead, we use information that is obtained from diagnostic problem-solving with the system $\mathcal{S}$ that is associated with the Bayesian network $\mathcal{B}_{\mathcal{S}}$.

As before, let $\Pi$ be the set of all conflicts that have been obtained for the system $\mathcal{S}$. For each element $\phi \in \Pi$ it holds that:

$$\mathrm{SD} \cup \mathrm{OBS} \cup \{\neg\phi\} \vDash \perp$$

In other words, the system cannot be in state $\neg\phi$, because that gives rise to a contradiction. Hence, we know that

$$\Pr'(\neg\phi) = 0$$

must hold, where $\Pr'$ is the probability distribution obtained by conditioning on information obtained from performing consistency-based diagnosis. Now, let $\bar{\Pi}$ be defined as follows:

$$\bar{\Pi} = \bigvee\{\neg\phi \mid \phi \in \Pi\}$$

From Proposition 1 we know that the set of diagnoses $D \equiv \neg\bar{\Pi}$. Let us assume that $D$ is in disjunctive normal form, where a diagnosis $\{C_1, \ldots, C_n\} \in D$ is again an Ab conjunct, then it holds that:

$$\Pr'(C_1, \ldots, C_n) = \begin{cases} \frac{\Pr(C_1, \ldots, C_n)}{\Pr(D)} & \text{if } \{C_1, \ldots, C_n\} \in D \\ 0 & \text{otherwise} \end{cases}$$

where

$$\Pr(D) = \sum_{\Delta \in D} \Pr(\Delta)$$

because according to Proposition 2 it holds, for each $\Delta, \Delta' \in D$, with $\Delta \not\equiv \Delta'$, that $\Delta \wedge \Delta' \equiv \perp$. The probabilities $\Pr(\Delta)$ can be computed from the associated Bayesian network $\mathcal{B}_{\mathcal{S}}$, with a computational complexity depending on the density of the topology of the network. Alternatively, the probability $\Pr(D)$ may be computed directly from the conflict set $\bar{\Pi}$ as

$$\Pr(D) = 1 - \Pr(\bar{\Pi})$$

or, when focussing on individual conflicts:

$$\Pr(D) = 1 - \sum_{\phi \in \bar{\Pi}'} \Pr(\phi)$$

with

$$\bar{\Pi}' = \{\phi \mid \phi = \{\mathrm{Ab}(c) \mid c \in C\} \cup \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS} \backslash C\}, C \subseteq \mathrm{COMPS}, \phi \in \bar{\Pi}\}$$

i.e. only maximally consistent conflicts should be taken into account to prevent counting probabilities more than once.

Based on the results above, it is now also possible to determine the probability of partial diagnoses. Partial diagnoses are no longer merely simplified representations of alternative diagnoses, as we have to take their likelihood into account as well.

**Proposition 3.** *Let $\Psi$ be a partial diagnosis of the system $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$, then for any diagnosis $\Delta$ of $\mathcal{S}$ with $\Psi \preceq \Delta$, it holds that $\Pr(\Delta) \leq \Pr(\Psi)$.*

*Proof.* Using marginalisation, it holds that

$$\Pr(\Psi) = \sum_{\Psi \preceq \Delta} \Pr(\Delta)$$

where $\Delta$s are diagnoses, from which the premise follows immediately. ◇

The following corollary comprises the result we are really after.

**Corollary.** *If $\Psi$ is a kernel diagnosis of the system $\mathcal{S}$, then for any partial diagnosis $\Psi'$ of $\mathcal{S}$ with $\Psi \preceq \Psi'$ it holds that $\Pr(\Psi') \leq \Pr(\Psi)$.*

Hence, kernel diagnoses are the most likely diagnoses. As a consequence, when it is necessary to restrict the number of diagnoses to be considered, it might be worthwhile to determine kernel diagnoses only, and to focus on the repair of the components included in the kernel diagnoses.

The results of this section are related to the results obtained by Kohlas et al. [9], but here they are achieved without making, possibly unrealistic, independence assumptions about the problem domain. Yet, the method is computationally feasible under realistic conditions. The heaviest part of the computation will usually be done by the logical reasoning engine, which may be implemented by an ATMS [5], by using the hitting-set algorithm [16], or by the component consequences algorithm [3]. Furthermore, it was shown that the methods could be extended in a straightforward way to deal with partial and kernel diagnoses, yielding some practically useful results.

# 5 Incorporating observations: the Bayesian approach

Up until now, we have handled evidence in model-based diagnosis by renormalisation of a given joint probability distribution $\Pr(C)$ based on information concerning conflict sets. These were computed using traditional logical techniques from consistency-based diagnosis. Although only realistic assumptions were made, this method is still rather crude. Whereas this type of model-based diagnosis does indeed enable the usage of knowledge of abnormal behaviour and of abnormality observations for reducing the number of alternative diagnoses

[6], it is not possible to employ such evidence effectively to influence the likelihood of the individual diagnoses. It is this aspect that will be studied in this section.

### 5.1 An extended Bayesian-network representation

In order to deal with the unrestricted probabilistic influence of observed findings on a probability distribution, it is necessary to include observations into Bayesian-network models. We will do so accordingly.

**Definition 4** (*Bayesian observation model*). *A* Bayesian observation model $M_{\mathcal{S}}$ *of system* $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$ *is a tuple* $M_{\mathcal{S}} = (C, O, A, E, \Pr)$*, such that*

- $C$ *represents a set of stochastic (component) variables, corresponding to a system's components;*
- $O$ *represents a set of stochastic (observation) variables, corresponding to observations;*
- $A \subseteq C \times (C \cup O)$ *is a set of (directed) arcs, portraying stochastic dependences among the variables;*
- $E \subseteq O \times \mathcal{B}(C \cup O)$ *is a set of Boolean expressions concerning the components and observations, indexed by particular observations;*
- $\Pr$ *is a joint probability distribution defined on* $C \cup O$ *that reflects all stochastic (in)dependence information represented in* $A$ *and* $E$*.*

Note that although we speak of 'observation variables', this does not mean that they need actually be observed. A more proper name would have been 'observable variables'. However, as the role of a variable in probability theory depends on whether or not it is conditioning other variables—if it is conditioning it is assumed to be observed, if not, it is assumed to be observable—, which could easily happen by transforming probabilistic formulae, it seems justified to ignore such subtleties. However, it should be recognised that the role of observation variables may change in this process.

In the following, the structure of an observation model $M_{\mathcal{S}}$ will be depicted as a directed acyclic graph, where elements of $A$ will be represented as solid arcs; $E$ is used to augment the probability distribution of observation variables with extra information. Note that the set of arcs $A$ contains both knowledge about dependences among components, on the one hand, and among components and observations, on the other hand. The set $E$ includes knowledge of dependences among components and observation variables, expressed as conditional probabilities. It is assumed that there are no immediate arcs in $A$ between any two observation variables. However, the set $E$ may incorporate more complicated interactions, including those among observation variables.

Now, let us assume that the joint probability distribution

$$\Pr(O, C_1, \ldots, C_n) = \Pr(C_1, \ldots, C_n) \Pr(O \mid C_1, \ldots, C_n) \tag{4}$$

is decomposed in such way that

$$\Pr(C_1, \ldots, C_n) = \prod_{i=1}^{n} \Pr(C_i \mid \pi_A(C_i)) \tag{5}$$

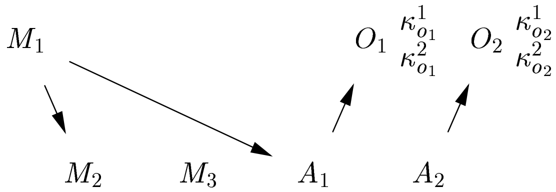


**Figure 3**: Bayesian observation model.

according to the set of arcs $A$. The second component of Pr, i.e. the conditional probability distribution $\Pr(O \mid C_1, \ldots, C_n)$, is defined by the set $E$ together with the set of arcs $A$ by a decomposition into factors using the chain rule of probability theory; factors have the following form:

$$\Pr(O_i \mid c_1^{e_1}, \ldots, c_n^{e_n}, O') = \begin{cases} \Pr(O_i \mid \kappa) & \text{if } \kappa \subseteq \{c_1^{e_1}, \ldots, c_n^{e_n}\} \cup O', \\ & \quad \text{and } (O_i, \kappa) \in E \\ \Pr(O_i \mid \hat{\pi}_A(O_i)) & \text{otherwise} \end{cases} \tag{6}$$

for each $O_i \in O$, where $\{O_i\} \cup O' \subseteq O$; furthermore, $c_i^{e_i}$ is either equal to $c_i$ or to $\neg c_i$. We assume that the factorisation is determined, firstly, by the factors $\Pr(O_i \mid \kappa)$, with $(O_i, \kappa) \in E$, and only in the second place by the factors $\Pr(O_i \mid \hat{\pi}_A(O_i))$.

The principal idea is that the set $E$ contains all additional dependences, when the independences portrayed by $A$ are not fulfilled for all values (instances) of the variables. This has the advantage that the resulting Bayesian-network model can still be sparse, even though at the variable level some independences, including those among observation variables, are not fulfilled. The elements $(O, \kappa) \in E$ will be denoted in the following by $\kappa_o$.

The set $E$ together with $A$ and their associated joint probability distribution Pr is now used to model the behaviour of a given system $\mathcal{S}$, i.e. to predict observations given other observed output and state assumptions regarding the components. We assume that the following correspondence exists between the two formalisations. Let $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$ be a system, then

$$\mathrm{SD} \cup O \cup \psi \vDash o^e \Leftrightarrow \Pr(o^e \mid \psi, O) = 1$$

with $\{o^e\} \cup O \subseteq \mathrm{OBS}$, and for each $O' \subseteq \mathrm{OBS}$, $O' \subset O$: $\mathrm{SD} \cup O' \cup \psi \nvDash o^e$, i.e. $O \subseteq \mathrm{OBS}$ is $\subseteq$-minimal. In addition, it is assumed that $\psi \subseteq \Delta_P \cup \Delta_N$ is $\subseteq$-minimal. Finally, $o^e$ is either $o$ or $\neg o$. These conditions ensure that the specification of $E$ will be as small as possible, still capturing the essential behaviour of $\mathcal{S}$.

It follows that the factors (6) may not be uniquely defined, as there may be more than one $(O, \psi)$ combination that enables us to derive $o^e$. This indicates that there may exist more than one partial behaviour explaining the overall behaviour of the system, and we could have chosen any of these, as their probabilities will all be equal to 1.

**Example 2.** Reconsider the multiplier-adder circuit $\mathcal{S}$ from Example 1. In Figure 3, a Bayesian observation model $M_{\mathcal{S}}$ of that system is shown. For ease of exposition, we assume that the input to the circuit is fixed; it is therefore not necessary to represent the input explicitly. We have $C = \{M_1, M_2, M_3, A_1, A_2\}$ and $O = \{O_1, O_2\}$. Furthermore, the following probability distribution is defined for this model:

$$
\begin{array}{ll}
\Pr(o_1 \mid a_1) = 1 & \Pr(o_2 \mid a_2) = 1 \\
\Pr(o_1 \mid \neg a_1) = 0.04 & \Pr(o_2 \mid \neg a_2) = 0.05 \\
\Pr(m_3) = 0.2 & \Pr(a_1 \mid m_1) = 0.02 \\
\Pr(m_1) = 0.02 & \Pr(a_1 \mid \neg m_1) = 0.01 \\
\Pr(m_2 \mid m_1) = 0.01 & \Pr(a_2) = 0.03 \\
\Pr(m_2 \mid \neg m_1) = 0.005 &
\end{array}
$$

In addition, $E = \{\kappa_{o_1}^1, \kappa_{o_1}^2, \kappa_{o_2}^1, \kappa_{o_2}^2\}$, where

$$
\begin{array}{lcl}
\kappa_{o_1}^1 & = & \neg a_1 \wedge \neg m_1 \wedge \neg m_2 \\
\kappa_{o_1}^2 & = & \neg a_1 \wedge \neg a_2 \wedge \neg m_1 \wedge \neg m_3 \wedge \neg o_2 \\
\kappa_{o_2}^1 & = & \neg a_2 \wedge \neg m_2 \wedge \neg m_3 \\
\kappa_{o_2}^2 & = & \neg a_1 \wedge \neg a_2 \wedge \neg m_1 \wedge \neg m_3 \wedge \neg o_1
\end{array}
$$

i.e. only normality assumptions are represented in $E$. Finally, the following probabilities are defined for the elements in $E$:

$$
\begin{array}{rcl}
\Pr(\neg o_1 \mid \neg a_1, \neg m_1, \neg m_2) & = & 1 \\
\Pr(\neg o_1 \mid \neg a_1, \neg a_2, \neg m_1, \neg m_3, \neg o_2) & = & 1 \\
\Pr(\neg o_2 \mid \neg a_2, \neg m_2, \neg m_3) & = & 1 \\
\Pr(\neg o_2 \mid \neg a_1, \neg a_2, \neg m_1, \neg m_3, \neg o_1) & = & 1
\end{array}
$$

Note that e.g. $\Pr(o_1 \mid \neg a_1, \neg m_1, \neg m_2) = 0$ replaces the probability $\Pr(o_1 \mid \neg a_1) = 0.04$ in the Bayesian network specified above, when used to define the joint probability distribution (4) using equation (6). As mentioned above, both

$$\Pr(\neg o_1 \mid \neg a_1, \neg m_1, \neg m_2)$$

and

$$\Pr(\neg o_1 \mid \neg a_1, \neg a_2, \neg m_1, \neg m_3, \neg o_2)$$

may be used in computing

$$\Pr(\neg o_1, \neg o_2 \mid \neg a_1, \neg a_2, \neg m_1, \neg m_2, \neg m_3)$$

as both are possible factors as defined by equation (6). $\diamond$

## 5.2 Probabilistic reasoning

As we have redefined model-based systems in terms of probability distributions, there is now a close correspondence between the two. One may thus expect that the probability distribution that represents a system will have been defined in such a way that it respects the logical behaviour of the corresponding system $\mathcal{S}$, i.e.

$$\Pr(C \mid O) = 0$$

iff $\mathrm{SD} \cup O \cup C \vDash \perp$, i.e. when $C$ is a conflict, where the negative form of conflicts $\bar{\Pi}$ is used. We now have the following proposition, in which use is made of the relation $E$.

**Proposition 4.** *Let* $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$ *be a system, and let* $\Pr(C), \Pr(O) > 0$, *then*

*$C$ is a conflict iff* $\Pr(C \mid O) = 0$.

*Proof.* ($\Rightarrow$) Note that $\Pr(O \mid C) = \Pr(O_i \mid C, O') \cdot \Pr(O' \mid C)$, with $\{O_i\} \cup O' = O$, and $O_i \notin O'$. If $\mathrm{SD} \cup O \cup C \vDash \perp$ then for some $O_i \in O$, $\Pr(O_i \mid C, O') = 0$ according to the Bayesian observation model $M_{\mathcal{S}}$. The remainder of the proof follows trivially from Bayes' rule.

($\Leftarrow$) If $\Pr(C|O) = 0$ then $\Pr(O|C) = 0$, meaning that $C$ must be a conflict of the corresponding system $\mathcal{S}$. $\diamond$

Hence, there exists a 1–1 correspondence between the notion of conflict in consistency-based diagnosis and in Bayesian model-based diagnosis. As in consistency-based diagnosis, the set of conflicts can be used as a basis for determining diagnoses. The known algorithms for consistency-based diagnosis are quite suitable for that purpose.

There is, however, more knowledge encoded in the Bayesian observation model than simply the type of knowledge that can be used to compute conflicts. This knowledge can be used to compute the likelihood of the diagnoses:

$$\Pr(C_1, \ldots, C_n \mid O) = \Pr(O \mid C_1, \ldots, C_n) \prod_{i=1}^{n} \Pr(C_i \mid \pi_A(C_i)) / \Pr(O)$$

where $\Pr(O)$ is a normalisation factor, obtained by computing $\Pr(C_1, \ldots, C_n \mid O)$ for every diagnosis. Note that according to Proposition (4) we have that $\Pr(C \mid O) = 0$, for every conflict $C$, so these probabilities need not be computed. Sometimes, the observation variables are independent, i.e. $\Pr(O) = \prod_{j=1}^{m} \Pr(O_j)$ holds.

Furthermore, $\Pr(O \mid C_1, \ldots, C_n)$ may be decomposed as discussed in the previous section for the Bayesian observation model; sometimes the simpler condition

$$\Pr(O \mid C_1, \ldots, C_n) = \prod_{j=1}^{m} \Pr(O_j \mid C_1, \ldots, C_n)$$

holds for the Bayesian observation model; this certainly holds when it is possible to restrict to the part of the Bayesian observation model concerning $A$.

These two approaches—the qualitative, logical one, and the numerical, probabilistic one—can be combined, yielding the following two-step procedure:

1. A logical consistency-based algorithm computes all potential diagnoses;

2. The likelihood of every diagnosis is computed using the Bayesian observation model.

The set $E$ may be used as a tool to reduce the number of diagnoses to be considered; the more knowledge it contains, the smaller will be the number of diagnoses to be considered [7]. We continue with the example in the previous section, to illustrate the method.

**Example 3.** In Example 1, the potential diagnoses were already established. The posterior probability of such diagnoses can now be computed as follows, using the Bayesian observation model $M_{\mathcal{S}}$:

$$\begin{aligned}
&\Pr(a_1, \neg a_2, \neg m_1, \neg m_2, \neg m_3 \mid o_1, \neg o_2) = \\
&\quad = \Pr(o_1 \mid a_1, \neg a_2, \neg m_1, \neg m_2, \neg m_3) \cdot \\
&\quad\quad \Pr(\neg o_2 \mid a_1, \neg a_2, \neg m_1, \neg m_2, \neg m_3) \cdot \\
&\quad\quad \Pr(\neg m_1) \cdot \Pr(\neg m_2 \mid \neg m_1) \cdot \Pr(\neg m_3) \cdot
\end{aligned}$$

$$
\begin{aligned}
&\Pr(a_1 \mid \neg m_1) \cdot \Pr(\neg a_2) / \Pr(o_1, \neg o_2) \\
&= 1 \cdot 1 \cdot 0.98 \cdot 0.995 \cdot 0.80 \cdot 0.01 \cdot 0.97/0.046 \\
&\approx 0.16
\end{aligned}
$$

where, according to the definition in the previous section, it holds that:

$$\Pr(o_1 \mid a_1, \neg a_2, \neg m_1, \neg m_2, \neg m_3) = \Pr(o_1 \mid a_1)$$

and

$$\Pr(\neg o_2 \mid a_1, \neg a_2, \neg m_1, \neg m_2, \neg m_3) = \Pr(\neg o_2 \mid \neg a_2, \neg m_2, \neg m_3)$$

Furthermore, $\Pr(o_1, \neg o_2) = \Pr(o_1) \Pr(\neg o_2)$ was computed straight from the Bayesian observation model, using a standard Bayesian-network inference algorithm. $\diamond$

Finally, note that the results of the previous section are a special case of the results of the present section. When the observation variables $O$ are independent of the component variables $C$, i.e. the set of arcs $A$ does not include elements $(C_i, O_j)$, whereas the set $E$ only represents normality assumptions, the results of Section 4.2 coincide with the achievements of this section.

# 6 Discussion

Methods from the field of model-based diagnosis are especially good at incorporating knowledge of the structure and behaviour of systems for the purpose of diagnosis, but are weak at the representation of the uncertainties involved. Bayesian networks are good at representing the stochastic (in)dependences and uncertainties involved in processes, but are not really suitable for the representation of their associated structure and behaviour. Bayesian model-based diagnosis, on the other hand, integrates the methods from traditional logic-based consistency-based diagnosis and Bayesian networks, incorporating results from these research fields as special cases. Whereas De Kleer [6] already suggested the use of Bayes' rule, his approach implies using very strong independence assumptions, which is unnecessarily restrictive. De work of Kohlas et al. [9] does not fully recognise the importance of observations to rank resulting diagnoses, and also ignores the possibilities offered by Bayesian networks for computing diagnoses efficiently, without having to make abundant independence assumptions. Finally, J. Pearl [13] does not fully appreciate the power of model-based reasoning techniques, on the one hand, and does not give proper attention to the modelling of interactions between components, on the other hand. The present work does not suffer from such drawbacks.

If the likelihood of every diagnosis is determined, the research described in this paper may be seen as a model-based approach to the MAP (maximum a posteriori) assignment problem for Bayesian networks. In the MAP problem the instantiation $I$ yielding the largest a posteriori probability $\Pr(I \mid E)$ for evidence $E$ is determined; this problem is known to be NP-hard [17]. In our case, we try to reduce the computational burden by eliminating potential candidates $I$ using consistency-based reasoning. Of course, whether this approach is effective in practice is determined by the actual Bayesian observation model chosen for a problem. If the network topology is dense, the computational burden will be large, but if it is sparse, and the set $E$ includes sufficiently strong constraints, computation of the most likely diagnosis will be feasible.

**Aknowledgement.** I would like to thank the reviewers and the organisers of the CANEW2000 workshop for their helpful comments.

## References

[1] L. Console, D. Theseider Dupré and P. Torasso (1989). A theory of diagnosis for incomplete causal models. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pp. 1311–1317.

[2] L. Console and P. Torasso (1990). Integrating models of correct behaviour into abductive diagnosis. In *Proceedings of ECAI'90*, pp. 160–166.

[3] A. Darwiche (1998). Model-based diagnosis using structured system descriptions. *Journal of Artificial Intelligence Research*, **8**, 165–222.

[4] J. de Kleer (1976). Local methods for localizing faults in electronic circuits. *MIT AI Memo* 394. Massachusetts Institute of Technology, Cambridge, MA.

[5] J. de Kleer (1986). An assumption-based TMS. *Artificial Intelligence*, **28**, 127–162.

[6] J. de Kleer (1990). Using crude probability estimates to guide diagnosis. *Artificial Intelligence*, **45**, 381–392.

[7] J. de Kleer, A.K. Mackworth and R. Reiter (1992). Characterizing diagnoses and systems. *Artificial Intelligence*, **52**, 197–222.

[8] J. de Kleer and B.C. Williams (1987). Diagnosing multiple faults. *Artificial Intelligence*, **32**, 97–130.

[9] J. Kohlas, B. Anrig, R. Haenni and P.A. Monney (1998). Model-based diagnosis and probabilistic assumption-based reasoning. *Artificial Intelligence*, **104**, 71–106.

[10] S.L. Lauritzen, D.J. Spiegelhalter (1987). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society (Series B)*, **50**, 157–224.

[11] P.J.F. Lucas (1997). Symbolic diagnosis and its formalisation. *The Knowledge Engineering Review*, **12**(2), 109–146.

[12] P.J.F. Lucas (1998). Analysis of notions of diagnosis. *Artificial Intelligence*, **105**(1-2), 293–341.

[13] J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufman, San Mateo, California, 1988.

[14] Y. Peng and J.A. Reggia (1990). *Abductive Inference Models for Diagnostic Problem Solving.* Springer-Verlag, New York.

[15] D. Poole (1990). A methodology for using a default and abductive reasoning system. *International Journal of Intelligent Systems*, **5**(5), 521–548.

[16] R. Reiter (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, **32**, 57–95.

[17] S.E. Shimoni (1994). Finding MAPs for belief networks is NP-hard. *Artificial Intelligence*, **68**, 399–410.

[18] S. Tuhrim, J.A. Reggia and S. Goodall (1991). An experimental study of criteria for hypothesis plausibility. *Journal of Experimental and Theoretical Artificial Intelligence*, **3**, 129–144.