

Theory Diagnoses: A Concise Characterization of Faulty Systems

Gerhard Friedrich

Technische Universität Wien, Institut für Informationssysteme

Christian Doppler Labor für Expertensysteme

Paniglgasse 16, A-1040 Wien, Austria

friedrich@vexpert.dbai.tuwien.ac.at

Abstract

In this article a new characterization of the set of diagnoses in the domain of consistency-based diagnosis is developed. The needs for such a characterization arise from the inability of the current characterizations (e.g. minimal diagnoses and kernel diagnoses) to deal with abstractions like component hierarchies. We propose a characterization based on theories which takes advantage of hierarchies and therefore allows a concise description of various sets of diagnoses (e.g. minimal diagnoses and kernel diagnoses).

1 Introduction

Consistency-based diagnosis systems ([3]) seek for sets of assumptions about the behavior of components which are consistent with the system description and the observations. The concept of kernel diagnoses was introduced in [3] in order to characterize the set of all consistency-based diagnoses. Kernel diagnoses are replacing the concept of minimal diagnoses which are not adequate in the case where supersets of minimal diagnoses can be inconsistent. If all supersets of minimal diagnoses are consistent or if supersets of minimal diagnoses are very unlikely, focusing on minimal diagnoses is still a good choice for practical reasons. Firstly, there sometimes exist exponentially more kernel diagnoses than minimal diagnoses. Secondly, let us assume that supersets of a minimal diagnosis are much more unlikely than the minimal diagnosis itself (which is the case in most practical applications). In this case further decisions like proposing the best next measurement are hardly influenced at all by considering supersets of minimal diagnoses. Therefore, minimal diagnoses or a subset of them are used as a starting point for the generation of non-minimal diagnoses if such diagnoses influence further decisions.

However, there is still the possibility of a combinatorial explosion although only minimal diagnoses are considered. Assume that there are several independent and faulty modules in a system, then every minimal diagnosis of the whole system is a combination of the minimal diagnoses of the modules.

It was argued by numerous researchers (e.g. [2, 7, 3, 9, 13]) that hierarchical diagnosis would avoid several

problems of a flat approach to diagnosis. However, exploiting hierarchies is still based on heuristic arguments rather than on firm grounds, e.g. selecting the best next measurement using hierarchies. Until now there has been no theory of diagnosis in the sense of [3] which allows to exploit such hierarchical structures for a characterization of diagnoses. Therefore, as soon as a composite component is broken up into its pieces, the descriptive power of hierarchies is lost.

Although kernel diagnoses provide a concise description of the set of all diagnoses, "this approach flattens out since there is no ability to enable hierarchical diagnosis, so achieving reduction of complexity" ([10], page 466). We therefore developed a new concept for a concise description of the set of all diagnoses where we are able to exploit additional knowledge, e.g. hierarchical relations. The main idea of our approach is that such knowledge expressed by a theory is exceptional. A characterization of diagnoses is done with respect to such a theory, i.e. a conjunction of "abnormal" literals (which we will call theory diagnosis) unified with a theory (e.g. describing abstractions based on component hierarchies) serves as a description of sets of diagnoses.

Although we focus in this paper on a concise description of the set of minimal diagnoses (because of application-oriented reasons) our approach naturally extends to a concise description of the set of all diagnoses which is discussed in the longer version of this paper.

The paper is organized as follows: In Section 2 we present an example showing that hierarchies flatten out if the traditional definition of consistency-based diagnosis is used. In Section 3 we propose an extension of the consistency-based approach where we use theories as means for a concise description of diagnoses. Hierarchical theories are formally introduced in Section 4. We show that hierarchical theories and theory-based diagnoses provide a computationally cheap generator for minimal diagnoses and how hierarchical theories can be adapted to preserve the descriptive power.

2 Hierarchies and Minimal Diagnoses

Consider the system and the observations depicted in Figure 1. The system consists of two buffer chains which are connected to an or-gate. We assume the usual behavior of these components. For the sake of simplicity we

don't consider fault modes. The inputs of the circuit are observed to be 1 and the output of $c1$ to be 0.

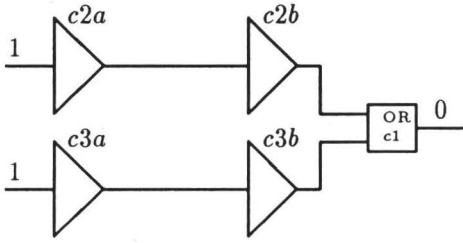


Figure 1: Two buffer chains and one or-gate

There is one minimal single fault diagnosis $ab(c1)$ and 4 minimal double fault diagnoses $ab(c2a) \wedge ab(c3a)$, $ab(c2a) \wedge ab(c3b)$, $ab(c2b) \wedge ab(c3a)$, and $ab(c2b) \wedge ab(c3b)$.¹ Note, that if there are n buffers in each chain then there are n^2 minimal double fault diagnoses.

Clearly, there is a much more comprehensive description than stating all pairs of possibly faulty buffers: $c1$ is abnormal or there is a double fault where one buffer of the upper and one buffer of the lower buffer chain are abnormal. In this case we have used an abstraction of the buffer chains where we don't care which buffer in a chain is faulty. There are many sources for such abstractions. E.g. assume that a buffer chain is a smallest replaceable unit. A further example is the case where multiple independent diagnosis problems are present in one system [5] (e.g. consider $c1$ is correct). Under this circumstances the diagnosis system should have the ability to generate abstractions in order to avoid a combinatorial explosion resulting from computing the cross product.

For these (and other) reasons many researchers have proposed the usage of hierarchies (e.g. [2, 7, 3, 9, 13]), i.e. the abstraction of components forming a composite component. Note, that they all use the usual definition of consistency-based diagnosis and therefore don't avoid the problem of flattening out the hierarchies. The abstraction of the buffer chains can be expressed very simply by the following equivalences forming a hierarchy of components:

$$ab(c2) \equiv ab(c2a) \vee ab(c2b). \quad ab(c3) \equiv ab(c3a) \vee ab(c3b).$$

and the addition of the components $c2$ and $c3$ to the set of components $COMPS$. We call the theory consisting of these two equivalences Σ_a .

The circuit extended by such composite components is depicted in Fig. 2. The equivalences express that if a composite component is faulty then at least one of its subcomponents is faulty. Conversely, if a subcomponent is faulty then the composite component is faulty. Note, that a faulty component (e.g. the composite component) need not show its abnormal behavior in a given situation. The fault is latent. But, if a subcomponent is at variance then also its supercomponent (see also [6] for a discussion on faulty behavior versus failure). In this example we use identities to express abstractions but our

¹We only mention the positive abnormal literals of minimal diagnoses, all other components are not abnormal by default. Furthermore, we assume that the reader is familiar with the basic definitions introduced in [3].

method allows arbitrary abstraction theories. The behavior of the composite components $c2$ and $c3$ is the same as for the buffers, i.e. output and input of a composite component are equal if it works correctly.

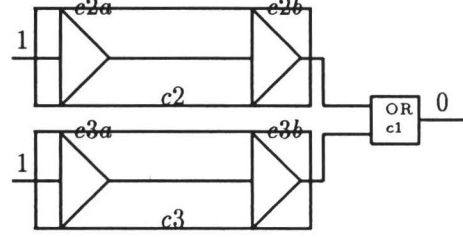


Figure 2: Two buffer chains and one or-gate with composite components

Extending the system $(SD, COMPS, OBS)$ by these axioms and components will lead to the following minimal conflicts and minimal diagnoses:

Minimal conflicts due to faulty behavior:

$$ab(c1) \vee ab(c2). \quad ab(c1) \vee ab(c3). \\ ab(c1) \vee ab(c2a) \vee ab(c2b). \quad ab(c1) \vee ab(c3a) \vee ab(c3b).$$

Minimal conflicts due to the hierarchical theory:

$$\neg ab(c2) \vee ab(c2a) \vee ab(c2b). \\ ab(c2) \vee \neg ab(c2a). \quad ab(c2) \vee \neg ab(c2b). \\ \neg ab(c3) \vee ab(c3a) \vee ab(c3b). \\ ab(c3) \vee \neg ab(c3a). \quad ab(c3) \vee \neg ab(c3b).$$

Minimal diagnoses:

$$ab(c1). \quad ab(c2) \wedge ab(c2x) \wedge ab(c3) \wedge ab(c3y). \quad x, y \in \{a, b\}$$

Note, that by using minimal diagnoses we do not have the ability to take advantage of hierarchies. Intuitively it should be possible to describe the set of minimal diagnoses and also the set of all diagnoses just by $ab(c1)$ and $ab(c2) \wedge ab(c3)$. Our basic idea is that the information about hierarchies in a system has a special status. Describing the set of minimal diagnoses exploiting this information leads to a new definition of diagnosis.

3 Theory Diagnoses

In this section we define a new concept for the description of diagnoses which allows to take advantage of hierarchical theories resulting in a concise characterization. Although we focus on hierarchical theories in this paper (which are described in Section 4) the new concept for the description of diagnoses we propose is independent of hierarchical theories.

We assume that abstractions are expressed in terms of ab -clauses. Such clauses are conflicts if they are regarded as elements of SD . Consequently, the number of kernel or minimal diagnoses as well as the number of (positive) literals mentioned in these diagnoses can grow or shrink depending on which ab -clauses are used for defining a (hierarchical) theory and the set of conflicts implied by the system description and the observations. Note, that in the example of Section 2 the introduction of the hierarchies described by Σ_a does not reduce the number of minimal diagnoses but enlarges the number of positive literals of all but one of these diagnoses.

Since clauses like hierarchical theories play a special role, they have to be treated differently. Therefore, we define diagnosis with respect to an ab -theory Σ .

3.1 Characterizing Minimal Diagnoses

In this section we define a concept for the description of minimal diagnoses. We describe the set of all minimal diagnoses by a set of theory diagnoses and an *ab*-theory. $D(Cp, Cn)$ denotes $\bigwedge_{c \in Cp} ab(c) \wedge \bigwedge_{c \in Cn} \neg ab(c)$. SD , the system description, and OBS , a set of observations, are sets of first-order sentences; $COMPS$, the system components, is a finite set of constants.

Definition 1 (Theory Diagnosis for Minimal Diagnoses) A conjunction T of *ab*-literals is called a theory diagnosis for minimal diagnoses of $(SD, COMPS, OBS)$ and a set of *ab*-clauses $\Sigma \subseteq SD$ (called an *ab*-theory) iff for every minimal diagnosis $D(\Delta, COMPS - \Delta)$ of $(\{T\} \cup \Sigma, COMPS, \emptyset)$ follows that $D(\Delta, COMPS - \Delta)$ is a minimal diagnosis for $(SD, COMPS, OBS)$.

We call a theory diagnosis for minimal diagnoses a M(inimal)T(theory)-diagnosis. Note, that such MT-diagnoses are not necessarily irreducible but are a concept for describing minimal diagnoses of a system.

Given an *ab*-theory, an MT-diagnosis contains all the necessary information for describing a set of minimal diagnoses for the complete system $(SD, COMPS, OBS)$. The *ab*-theory and the MT-diagnosis serve as a generator for a set of minimal diagnoses. The *ab*-theory describes how an MT-diagnosis has to be extended in order to provide minimal diagnoses for $(SD, COMPS, OBS)$.

Example 1 Consider Fig. 2. $ab(c1)$ and $ab(c2) \wedge ab(c3)$ are MT-diagnoses for the given hierarchical theory Σ_a . Every minimal diagnosis for $\{ab(c1)\} \cup \Sigma_a$ and $\{ab(c2) \wedge ab(c3)\} \cup \Sigma_a$ is a minimal diagnosis for this circuit.

Let $MDIAGS(T)$ be the set of minimal diagnoses of $(\{T\} \cup \Sigma, COMPS, \emptyset)$ where T is an MT-diagnosis. We call this set the induced set of minimal diagnoses. In our example the set of all minimal diagnoses is equal to the union of the induced minimal diagnoses of $ab(c1)$ and $ab(c2) \wedge ab(c3)$, i.e. $MDIAGS(c1) \cup MDIAGS(ab(c2) \wedge ab(c3))$. As a consequence, we have achieved a reduction of the number of diagnoses necessary for the description of all minimal diagnoses.

Irreducible MT-diagnoses are defined as follows. We use the function $ATOMS(C)$ to denote the set of atoms of a logical sentence C .

Definition 2 (Irreducible MT-Diagnosis) An irreducible MT-diagnosis T is an MT-diagnosis with the property that the only MT-diagnosis T' which is a subconjunction of T , i.e. $ATOMS(T') \subseteq ATOMS(T)$, is the MT-diagnosis T itself.

The following theorem² is a simple consequence of the definition of an MT-diagnosis. It states that MT-diagnoses describe the set of all minimal diagnoses. Furthermore, this is also true if we only consider MT-diagnoses which contain only positive *ab*-literals. In this case we say, this theory diagnosis is positive.

Theorem 1 Let $\Sigma \subseteq SD$ be an *ab*-theory for a system $(SD, COMPS, OBS)$.

²The proof of this theorem and all other proofs can be found in the longer version of this paper.

If D is a minimal diagnosis for $(SD, COMPS, OBS)$ then there is a (positive) MT-diagnosis T for $(SD, COMPS, OBS)$ and Σ , and $ATOMS(T) \subseteq ATOMS(D)$.

Conversely, if T is a (positive) MT-diagnosis for $(SD, COMPS, OBS)$ and Σ then there is a minimal diagnosis D for $(SD, COMPS, OBS)$ and $ATOMS(T) \subseteq ATOMS(D)$.

Similar to the relation between minimal diagnoses and positive minimal conflicts, MT-diagnoses and positive conflicts can be related.

Theorem 2 T is an MT-diagnosis for a system $(SD, COMPS, OBS)$ and an *ab*-theory $\Sigma \subseteq SD$ iff $\{T\} \cup \Sigma$ implies each positive (minimal) conflict of $(SD, COMPS, OBS)$.

The following corollary is a simple consequence of the above theorem and relates irreducible MT-diagnoses and positive minimal conflicts.

Corollary 1 T is an irreducible MT-diagnosis for a system $(SD, COMPS, OBS)$ and an *ab*-theory $\Sigma \subseteq SD$ iff $\{T\} \cup \Sigma$ implies each positive (minimal) conflict of $(SD, COMPS, OBS)$ but for each subconjunction T' of T $\{T'\} \cup \Sigma$ does not imply each positive (minimal) conflict of $(SD, COMPS, OBS)$.

Example 2 Consider the circuit and the hierarchical theory presented in Section 2. Let the hierarchical theory Σ_a be the theory used for generating MT-diagnoses. For the given observations there are several MT-diagnoses. Let us consider $ab(c1)$ and $ab(c2) \wedge ab(c3)$. Since $\{ab(c1)\} \cup \Sigma_a$ as well as $\{ab(c2) \wedge ab(c3)\} \cup \Sigma_a$ implies all minimal positive conflicts, both $ab(c1)$ and $ab(c2) \wedge ab(c3)$ are MT-diagnoses. As they cannot be reduced without losing this property, they are irreducible MT-diagnoses. Note, that we have achieved a reduction of the number of positive literals compared to minimal diagnoses. However, in our example there also exist several other MT-diagnoses:

$$\begin{array}{ll} ab(c2a) \wedge ab(c3a). & ab(c2b) \wedge ab(c3a). \\ ab(c2a) \wedge ab(c3b). & ab(c2b) \wedge ab(c3b). \end{array}$$

Note, that these MT-diagnoses are more specific than the MT-diagnosis $ab(c2) \wedge ab(c3)$ since the hierarchical theory joined with such a more specific theory diagnosis implies $ab(c2) \wedge ab(c3)$.

The following corollary states that such specific MT-diagnoses can be pruned from the set of MT-diagnoses. It holds for positive MT-diagnoses.

Corollary 2 Let $\{T1, \dots, Tn\}$ be a set of different positive MT-diagnoses for $(SD, COMPS, OBS)$ and an *ab*-theory $\Sigma \subseteq SD$.

$\{T1\} \cup \Sigma \models T2 \vee \dots \vee Tn$ (the diagnoses $T2, \dots, Tn$ are at least as general as $T1$) iff every conjunction $D(\Delta, COMPS - \Delta)$ which is a minimal diagnosis for $(\{T1\} \cup \Sigma, COMPS, \emptyset)$ is also a minimal diagnosis for one of $(\{T2\} \cup \Sigma, COMPS, \emptyset)$ or $\dots$ or $(\{Tn\} \cup \Sigma, COMPS, \emptyset)$.

Stating it differently, the positive MT-diagnoses $T2, \dots, Tn$ are at least as general as the positive MT-diagnosis $T1$ iff the set of minimal diagnoses described

by $T2, \dots, Tn$ is a superset (not necessarily proper) of the set of minimal diagnoses described by $T1$. Therefore, $T1$ is redundant.

Example 3 Consider the MT-diagnoses of our example:

$$\begin{array}{ll} ab(c2a) \wedge ab(c3a). & ab(c2b) \wedge ab(c3a). \\ ab(c2a) \wedge ab(c3b). & ab(c2b) \wedge ab(c3b). \end{array}$$

Each of these MT-diagnoses joined with the theory Σ_a implies the positive MT-diagnosis $ab(c2) \wedge ab(c3)$ and therefore all these MT-diagnoses are redundant.

Since the set of all MT-diagnoses is larger than needed for the description of all minimal diagnoses we are interested in a set of MT-diagnoses which is as small as possible.

Definition 3 (Irredundant set of MT-diagnoses) A set of MT-diagnoses MTS for $(SD, COMPS, OBS)$ and an ab -theory $\Sigma \subseteq SD$ is irredundant iff it is a smallest cardinality set with the property that every minimal diagnosis D for $(SD, COMPS, OBS)$ is an element of the induced minimal diagnoses of MTS , i.e. $D \in \bigcup_{T \in MTS} MDIAGS(T)$.

Stating it differently, if an element (i.e. an MT-diagnosis) of an irredundant set of MT-diagnoses is excluded then the remaining set of MT-diagnoses cannot describe the set of all minimal diagnoses of a system.

The following theorem states that it is possible to describe the set of all minimal diagnoses with a smaller (or at most equally large) set of MT-diagnoses.

Theorem 3 If $MINDIAGS$ is the set of minimal diagnoses for $(SD, COMPS, OBS)$ and $\Sigma \subseteq SD$ an ab -theory then there exists an irredundant set of MT-diagnoses $MTDIAGS$ for $(SD, COMPS, OBS)$ and Σ s.t. $|MTDIAGS| \leq |MINDIAGS|$.

The following theorem characterizes an irredundant set of positive MT-diagnoses:

Theorem 4 A set of positive MT-diagnoses $MTS = \{T1, \dots, Tn\}$ for $(SD, COMPS, OBS)$ and an ab -theory $\Sigma \subseteq SD$ is irredundant iff for all $Ti, Tj \in MTS$ and $Ti \neq Tj$ follows that $\{Ti\} \cup \Sigma \not\models Tj$ and every minimal diagnosis D for $(SD, COMPS, OBS)$ is an element of the induced minimal diagnoses of MTS .

Example 4 The set of MT-diagnoses $\{ab(c1), ab(c2) \wedge ab(c3)\}$ is irredundant. No MT-diagnosis of this set entails the other using the theory Σ_a .

Consequently, we need only two MT-diagnoses to describe the set of all minimal diagnoses of the example circuit. These two MT-diagnoses are still sufficient if the buffer chains are arbitrarily long.

4 Hierarchical Theories

In the previous section we developed the concept of MT-diagnoses, describing minimal diagnoses using an ab -theory. Until now we have not imposed any constraint on such an ab -theory. Note, that the set of all conflicts of a system may also be considered as an ab -theory for MT-diagnoses. In this case the empty set is the only irreducible MT-diagnosis. However, such an MT-diagnosis

and ab -theory would not be accepted as a diagnosis because from this description one cannot easily determine the minimal diagnoses. Even asking whether an additional minimal diagnosis exists is $\mathcal{NP}$ -complete. Therefore, we are interested in abstraction theories which allow the efficient generation of diagnoses as it is the case for kernel diagnoses.

In this section we introduce hierarchical theories which together with MT-diagnoses serve as a generator for all minimal diagnoses. Note, that the decision which minimal diagnoses should be constructed depends on the goals of the diagnosis process. Based on the properties of hierarchical theories we show that all minimal diagnoses can be efficiently computed, given an irredundant set of positive MT-diagnoses. Additionally, the test needed to show that a set of MT-diagnoses is irredundant can be done easily.

Definition 4 (Hierarchical Theory) A hierarchical theory is a set of ab -clauses of the form

$$ab(c_i) \equiv ab(c_j) \vee \dots \vee ab(c_k).$$

where all literals contained in such an identity are unique. The left hand side contains exactly one literal. The right hand side at least one literal.

There are no two equivalences having a same literal on the left hand side, and there are no two equivalences having a same literal on the right hand side.

Note, that such an equivalence corresponds to a set of clauses of the form:

$$\begin{array}{l} \neg ab(c_i) \vee ab(c_j) \vee \dots \vee ab(c_k). \\ ab(c_i) \vee \neg ab(c_j). \quad \dots \quad ab(c_i) \vee \neg ab(c_k). \end{array}$$

Such a hierarchical theory defines a set of directed trees where $ab(c_i)$ is the direct predecessor of each $ab(c_l) \in \{ab(c_j), \dots, ab(c_k)\}$. Each identity uniquely defines all successors of an $ab(c_i)$. The nodes are labeled by these literals.

4.1 Testing for Irredundancy

In order to show that a set of positive MT-diagnoses is irredundant, we have to perform tests of the form $\{T1\} \cup \Sigma \models T2$ where $T1, T2$ are positive MT-diagnoses and Σ is a hierarchical theory. For this test we have to compute the set of positive ab -literals entailed by $\{T1\} \cup \Sigma$.

Corollary 3 Let $T1, T2$ be positive MT-diagnoses and Σ a hierarchical theory. For testing if $\{T1\} \cup \Sigma \models T2$ only the Horn-clauses of Σ have to be considered.

As a consequence, the test whether $\{T1\} \cup \Sigma \models T2$ can be carried out in linear time.

4.2 Generating Minimal Diagnoses

Since hierarchical theories serve as a generator for minimal diagnoses we have to show how minimal diagnoses can be computed, given a conjunction of ab -literals (e.g. an MT-diagnosis). Since an irredundant set of MT-diagnoses can be easily computed using only positive MT-diagnoses we focus on the generation of minimal diagnoses given a positive MT-diagnosis.

For the construction we use the following properties of hierarchical theories: Each positive ab -literal implies

that all predecessor literals are positive. Stating it differently, if a component is assumed to be abnormal then every super-component is also assumed to be abnormal. Therefore, given a positive MT-diagnosis, every predecessor of a literal in this diagnosis can be pruned since it is implied by the hierarchical theory unified with another literal of the diagnosis. We call such a positive MT-diagnosis a pruned positive MT-diagnosis.

Let $pred(ab(ci))$ be the set of predecessors of a literal $ab(ci)$. Let $path(ab(ci), ab(cj))$ be the set of literals between $ab(ci)$ and $ab(cj)$ including these literals. The components mentioned by a set of ab-literals (or a logical sentence) S are denoted by $comp(S)$. The following corollary describes how minimal diagnoses can be computed.

Corollary 4 Let T be a pruned positive MT-diagnosis.

$$\Delta = \bigcup_{ab(ci) \in ATOMS(T)} pred(ab(ci)) \cup path(ab(ci), ab(cj))$$

where $ab(cj)$ is a leaf node connected to $ab(ci)$ (i.e. $ab(cj)$ is a successor of $ab(ci)$ and $ab(cj)$ has no successor) iff $D(comp(\Delta), COMPS - comp(\Delta))$ is a minimal diagnosis for $(\{T\} \cup \Sigma, COMPS, \emptyset)$ and therefore it is a minimal diagnosis for $(SD, COMPS, OBS)$.

Stating it differently, minimal diagnoses can be simply generated from a pruned positive MT-diagnosis T and a hierarchical theory in the following way: For each literal l in T all predecessors have to be included plus the literals along the path from l to a leaf connected to l . This defines all components to be assumed as abnormal.

Example 5 Consider the MT-diagnosis $ab(c2) \wedge ab(c3)$ of our example and the hierarchical theory Σ_a . From this MT-diagnosis and Σ_a the following minimal diagnoses can be computed according to the above corollary:

$$ab(c2) \wedge ab(c2x) \wedge ab(c3) \wedge ab(c3y). \quad x, y \in \{a, b\}$$

4.3 Translating Non-positive MT-Diagnoses

In the previous sections we showed that positive MT-diagnoses have extreme good properties with respect to the irredundancy test and the generation of minimal diagnoses. However, one can argue that positive MT-diagnoses tend to flatten out if components are exonerated³.

Example 6 Consider a slight extension of our example, where the buffer chains consist of n buffers and let us assume that some buffers, e.g. the first k ($k < n$) buffers of each chain are known to be not faulty.

The corresponding hierarchical theory Σ_a is:

$$ab(c2) \equiv ab(c2a) \vee \dots \vee ab(c2n).$$

$$ab(c3) \equiv ab(c3a) \vee \dots \vee ab(c3n).$$

In this case the MT-diagnosis $ab(c2) \wedge ab(c3)$ w.r.t. Σ_a is not longer an MT-diagnosis. One way to deal with this problem is to consider Σ_a as fixed and to break up this MT-diagnosis into a set of positive MT-diagnoses using the subcomponents of $c2$ and $c3$. Another possibility is to use non-positive MT-diagnoses. The conjunction

³Note, that this is not possible if the unknown fault mode is included in the behavior description of every component.

$ab(c2) \wedge ab(c3) \wedge \neg ab(c2a) \wedge \dots \wedge \neg ab(c2k) \wedge \neg ab(c3a) \wedge \dots \wedge \neg ab(c3k)$ is an MT-diagnosis w.r.t. Σ_a .

But since we model the abstraction hierarchy explicitly, we are able to reconstruct this hierarchy according to our needs. Therefore, there exists another option where the hierarchical theory is changed. We introduce two additional abstract components $c2'$ and $c3'$ and change the subcomponents of $c2$ and $c3$ accordingly. This leads to the following new abstraction theory Σ_{new} :

$$ab(c2') \equiv ab(c2a) \vee \dots \vee ab(c2k).$$

$$ab(c2) \equiv ab(c2(k+1)) \vee \dots \vee ab(c2n).$$

$$ab(c3') \equiv ab(c3a) \vee \dots \vee ab(c3k).$$

$$ab(c3) \equiv ab(c3(k+1)) \vee \dots \vee ab(c3n).$$

The corresponding MT-diagnosis is still $ab(c2) \wedge ab(c3)$. Again, we avoided the combinatorial explosion leading to an $O(n^2)$ description.

This example can be generalized in order to convert a set of MT-diagnoses MTS (including non-positive ones) and a hierarchical theory Σ into a set of positive MT-diagnoses MTS_{new} and Σ_{new} where the set of minimal diagnoses described stays the same. The complete algorithm is given in the extended version of this paper.

By this method we are able to change a set of MT-diagnoses (containing negative literals) and its corresponding hierarchical theory in a set of positive MT-diagnoses and a new hierarchical theory. Thus preserving the good properties of positive MT-diagnoses. Note, that the cardinality of an irredundant set of positive MT-diagnoses is always less or equal to the number of minimal diagnoses. In this method the given structure of the device (usually corresponding to the structure of a designer's mental view of the device) is preserved.

5 Comparison

The use of abstractions of devices for a more efficient diagnosis process was proposed by many researchers, e.g. [2, 7, 3, 9, 13]. Broadly speaking, their aim is to exploit abstractions like component hierarchies or behavior abstractions to make candidate generation and verification more efficient. However, in all these approaches no concise description of at least the minimal diagnoses is achieved. The works of [4], [14], and [12] are exceptional because they aim at characterizing sets of diagnoses.

The approach of [4] deals mainly with the integration of working hypotheses which are made by a diagnostic reasoner during the process of diagnosis. However, the combinatorial explosion of characterizing the minimal diagnoses of our example is not avoided. The description of diagnoses in [14] uses a two-level and/or graph for the description of diagnoses and can therefore be seen as a special case of our approach.

A further interesting approach is proposed in [12]. They introduce the concepts of abstract and concrete diagnosis. Abstract diagnoses are defined as ab-theories and no further syntactical constraints are required. Consequently, the set of all minimal conflicts is the unique maximally abstract diagnosis. However, generating minimal diagnoses from this description is $\mathcal{NP}$ -hard. The same property holds for abductive diagnoses as defined

by [1, 8] if these diagnoses have to be extended to consistency based (e.g. minimal) diagnoses in order to assign probabilities. Since in our approach the underlying abstraction theory is explicit, we can restrict this theory (as it is the case for hierarchical theories) in order to provide efficient generators for minimal diagnoses. Concrete diagnoses force a mode assignment to all components and therefore they also flatten out.

In the most recent work by [11] combinatorial explosion is avoided by identifying critical conflicts. In opposite to our approach not all minimal diagnoses are characterized which can lead to suboptimal probes as it is the case for our example depending on the relation between the failure probability of buffers and or-gate.

6 Conclusions

Although the concepts of minimal diagnoses and kernel diagnoses provide a concise description of the set of all diagnoses, we pointed out that within this formalism there is no ability to exploit knowledge about abstractions in order to achieve a concise *characterization* of all diagnoses. Consequently, hierarchies of components (forming composite components) flatten out. We proposed a new approach for the characterization of diagnoses exploiting the descriptive power of theories, e.g. hierarchical theories. We call these diagnoses theory diagnoses because they are defined with respect to a theory. By such a descriptive characterization we can separate very clearly problem description from problem solving following the approach taken in model-based diagnosis ([3]).

In this paper we focused on theory diagnoses for the description of minimal diagnoses (called MT-diagnoses). However, our approach can be easily adapted to theory diagnoses for the description of all diagnoses, which is discussed in the longer version of this paper. We showed that MT-diagnoses always provide a concise description of the set of all minimal diagnoses by reducing the number of diagnoses and literals needed for a complete characterization. Furthermore, we stated the relation between conflicts and MT-diagnoses and gave a condition whether or not a set of positive MT-diagnoses is irredundant. We proceeded by discussing hierarchical theories in detail, showing how minimal diagnoses can be easily generated given a set of positive MT-diagnoses. Additionally, we stated that the test whether a set of MT-diagnoses is irredundant or not as well as detecting which MT-diagnosis is redundant can be carried out efficiently. Finally, we pointed out how the hierarchical theory can be automatically rewritten in order to preserve these good properties of positive MT-diagnoses on one hand and to avoid hierarchies to flatten out in case of exoneration on the other hand.

Based on a clear definition of diagnosis where it is possible to exploit (hierarchical) theories, several further research issues can be investigated such as the dynamic generation of abstractions.

Acknowledgments

We would like to thank Philippe Dague, Walter Hamscher, and anonymous referees for valuable comments on this work.

References

- [1] L. Console, D. T. Dupré, and P. Torasso. A theory of diagnosis for incomplete causal models. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1311–1317, Detroit, Aug. 1989. Morgan Kaufmann Publishers, Inc.
- [2] R. Davis. Diagnostic reasoning based on structure and behaviour. *Artificial Intelligence*, 24:347–410, 1984.
- [3] J. de Kleer, A. K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56:197–222, 1992.
- [4] O. Dressler and P. Struss. Back to defaults: Characterizing and computing diagnoses as coherent assumption sets. In *Third International Workshop on Principles of Diagnosis*, Rosario (WA), Oct. 1992.
- [5] H. Freitag and G. Friedrich. Focusing on independent diagnosis problems. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, MA, Oct. 1992.
- [6] G. Friedrich and F. Lackinger. Diagnosing temporal misbehavior. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1116–1122, Sydney, Australia, Aug. 1991. Morgan Kaufmann Publishers, Inc.
- [7] W. C. Hamscher. Modeling digital circuits for troubleshooting. *Artificial Intelligence*, 51(1-3):223–271, Oct. 1991.
- [8] K. Konolige. Abduction versus closure in causal theories. *Artificial Intelligence*, 53:255–272, 1992.
- [9] I. Mozetič. Hierarchical model-based diagnosis. *International Journal of Man-Machine Studies*, 35:329–362, 1991.
- [10] G. M. Provan and D. Poole. The utility of consistency-based diagnostic techniques. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, pages 461–472, Cambridge, MA, Apr. 1991. Morgan Kaufmann Publishers, Inc.
- [11] O. Raiman and J. de Kleer. Critical reasoning. In *Third International Workshop on Principles of Diagnosis*, Rosario (WA), Oct. 1992.
- [12] V. A. Saraswat, J. de Kleer, and O. Raiman. Contributions to a theory of diagnosis. In *First International Workshop on Principles of Diagnosis*, pages 33–38, Stanford, July 1990.
- [13] P. Struss. What's in SD? towards a theory of modeling for diagnosis. In W. Hamscher, L. Console, and J. de Kleer, editors, *Readings in Model-Based Diagnosis*, pages 419–449. Morgan Kaufmann Publishers, Inc., 1992.
- [14] T. D. Wu. Efficient diagnosis of multiple disorders based on a symptom clustering approach. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, Boston, Aug. 1990.