Characterizing Consistency-Based Diagnoses

Sylvie Coste-Marquis[*+] and **Pierre Marquis**[*]
CRIL / Université d'Artois[*] & IUT de Lens[+]
rue de l'Université - S.P. 16 - F-62307 Lens Cedex
FRANCE
tel: (33) 3 21 79 32 77 - fax: (33) 3 21 79 32 72
e-mail: {coste, marquis}@cril.univ-artois.fr

Abstract. In this paper, we are mainly concerned with consistency-based diagnoses for systems represented in a propositional framework. Several characterizations of such diagnoses (mainly minimal diagnoses, kernel diagnoses, MT-diagnoses) have been proposed so far. Unfortunately, minimal diagnoses, and MT-diagnoses characterize only some consistency-based diagnoses and the set of kernel diagnoses for a system is often unsufficiently concise to be helpful; especially, the kernel diagnoses approach does not suggest any way to incorporate hierarchical relations between components within the diagnostic process. As an alternative to the kernel diagnoses characterization, we introduce the *theory minimal conflicts* for a system S. This characterization can be used to check whether a candidate diagnosis is a consistency-based (resp. a partial, a kernel) diagnosis for S in polynomial time. Additionally, it can be used to find out a consistency-based (resp. a partial, a kernel) diagnosis for S in polynomial time. Based on their theory prime implicates characterization, we show how theory minimal conflicts for S can be derived without requiring all the minimal conflicts of S to be generated.

1. Introduction

Diagnosing a system consists in determining the states of its components (*e.g.* normal or faulty or « don't care ») from what is known, mainly the system description and some observations. In the model-based approach, system description models the behavior of the system components and the connections; observations are typically inputs and outputs of the system. Many techniques for model-based diagnosis have been proposed so far. Among them, the *consistency-based* approaches (Reiter, 87) (De Kleer and Williams, 87) have the strong feature that the actual diagnosis of a system is a consistency-based one (*i.e.* a diagnosis consistent with what is known), provided that both the system description and the observations are certain. Unfortunately, since the system description and/or the observations are frequently incomplete, more than one consistency-based diagnosis can be derived in the general case. Indeed, in the worst case, exponentially many consistency-based diagnoses can be generated.

So as to focus on the most promising consistency-based diagnoses, some additional extra-logical knowledge can be taken into account as preference criteria. Among them are parsimony (preferred diagnoses are minimal diagnoses, *i.e.* those containing a minimal subset of faulty components w.r.t. set inclusion (Reiter, 87) (De Kleer and Williams, 87)), exoneration (preferred diagnoses are those in which all the components that behave as required by their models of correct behavior are assumed normal (Raiman, 90)), explainability (preferred diagnoses must explain the observations in the best way (El Ayeb *et al.*, 90) (Besnard and Cordier, 92)), or maximal probability (preferred diagnoses are those with highest *a priori* probability of failure).

Unfortunately, pointing out all preferred consistency-based diagnoses of a system is intractable for many preference criteria. The source of complexity is twofold: checking whether a candidate diagnosis is a (preferred) consistency-based diagnosis is NP-hard in the general case and the number of (preferred) diagnoses may easily be exponential in the number of components of the system. Especially, in many situations (e.g. when the system description is Horn), checking whether a candidate diagnosis is a (preferred) consistency-based diagnosis is tractable but computing all of them cannot always be achieved in practice. This raises up the problem of *characterizing* the (preferred) consistency-based diagnoses for a system without computing them extensively.

Many characterizations of such diagnoses have been proposed so far (e.g. minimal diagnoses (Reiter, 87), kernel diagnoses (De Kleer at al., 90, 92), MT-diagnoses (Friedrich, 93) and the notion of consequence of the observations (Darwiche, 95)). However, the problem of defining what a characterization should be has not been addressed till now. This paper contributes to fill this gap. Several properties that any characterization should satisfy are exhibited; the existing approaches are revisited along these lines and shown not fully adequate to the characterization purpose: minimal diagnoses and MT-diagnoses do not capture all the consistency-based diagnoses and kernel diagnoses are not concise enough in the general case since they do not take advantage of hierarchical relations between components within the diagnostic process.

As an alternative to the kernel diagnoses characterization, we introduce the notion of *theory minimal conflicts* for a system S. Following (Friedrich, 93), our purpose is to exploit knowledge about abstractions, especially component hierarchies, so as to derive concise descriptions of all the consistency-based diagnoses for S. Theory minimal conflicts are captured thanks to the notion of *theory prime implicate*, a generalization of the well-known concept of prime implicate. We show how theory minimal conflicts can be used as characterizations of the set of all consistency-based diagnoses for S. Interestingly, checking whether a candidate diagnosis is a consistency-based diagnosis, a partial diagnosis or a kernel diagnosis for a system can be achieved in time polynomial in the size of any of its theory minimal conflicts characterizations. More, the *function problems* consisting in pointing out a consistency-based diagnosis, a partial diagnosis or a kernel diagnosis for a system can be solved as well in polynomial time. All these decision and function problems are NP-hard in the general case.

The rest of this paper is organized as follows. After some formal preliminaries (Section 2), the problem of characterizing the consistency-based diagnoses for a system S is addressed. The existing approaches are shown not to fulfill all the properties that characterizations should possess (Section 3). Then, theory minimal conflicts are defined; we show that they characterize consistency-based diagnoses in a valuable way; especially, theory minimal conflicts characterizations can be exponentially more compact than kernel diagnoses ones, even if it is not always the case (Section 4); we also show how the theory minimal conflicts for S can be computed without requiring all the minimal conflicts of S to be generated. Finally, Section 5 concludes this paper. Proofs of propositions are sketched in an appendix.

2. Formal Preliminaries

The logical framework and the diagnostic framework in which our study takes place are successively presented.

2.1 The logical framework

Let PS be a finite set of propositional variables (or symbols). Let $PROP_{PS}$ denote the propositional language built in the usual way upon PS and the connectives. Elements Σ of $PROP_{PS}$ are called *formulas*. Var(Σ) is the set of all the propositional symbols of PS occurring in Σ. Any finite set of formulas (or *theory*) will be identified with the conjunctive formula whose conjuncts are the elements of the set. A *clause* is a (finite) disjunction of variables (*positive literals*) and negated variables (*negative literals*); it is viewed as a set of literals when it is convenient. A *term* is a finite conjunction of literals; it is viewed as the set of its literals when it is convenient. A formula is in CNF when it is under the form of a conjunction of clauses. Every propositional formula can be put in CNF in linear time, introducing new symbols. When Σ is a formula and l is a literal, we note Σ_l the formula obtained by replacing every occurrence of l in Σ by the boolean constant *true* and every occurrence of $\neg l$ in Σ by the boolean constant *false*. When x is a variable, the formula $(\neg x \wedge \Sigma_{\neg x}) \vee (x \wedge \Sigma_x)$ is referred to as the *Shannon expansion* of Σ around x.

An *interpretation* I (over $PROP_{PS}$) is a mapping from PS to {T, F}. The *semantics* of a formula Σ in an interpretation I, noted $[\Sigma](I)$, is an element of {T, F} defined in the usual way. We say that an interpretation I is a *model* of a formula Σ iff $[\Sigma](I) = T$. When a formula Σ has no model, it is *inconsistent* (otherwise, it is *consistent*). When every interpretation is a model of Σ, Σ is said valid. We will note

$\psi \models \xi$ when $[\xi](I) = T$, for every model I of ψ, and say that ξ is a *logical consequence* of ψ. When $\psi \models \xi$ and $\xi \models \psi$, ψ and ξ are said *equivalent*; this is noted $\psi \equiv \xi$. Every Shannon expansion of a formula is equivalent to it. $PROP_{PS} / \equiv$ denotes the quotient set of $PROP_{PS}$ induced by the equivalence relation $\equiv$. Let Φ be a formula. $\models_\Phi$ denotes the pre-ordering over $PROP_{PS}$ defined by $\psi \models_\Phi \xi$ iff $\Phi \models \psi \Rightarrow \xi$. Let $\equiv_\Phi$ denote the equivalence relation over $PROP_{PS}$ induced by $\models_\Phi$ and defined by $\psi \equiv_\Phi \xi$ iff $\psi \models_\Phi \xi$ and $\xi \models_\Phi \psi$. $PROP_{PS}/\equiv_\Phi$ denotes the quotient set of $PROP_{PS}$ induced by $\equiv_\Phi$. A formula Φ is said *tractable* iff there exists an algorithm for deciding whether $\Phi \models c$ (for any clause c) in time polynomial in the size of Φ plus the size of c.

Let Σ be a formula of $PROP_{PS}$. An *implicant* of Σ is a term π s.t. $\pi \models \Sigma$. An *implicate* of Σ is a clause π s.t. $\Sigma \models \pi$. A *prime implicant* of Σ is an implicant of Σ which is maximal w.r.t. $\models$. A *prime implicate* of Σ is an implicate of Σ which is minimal w.r.t. $\models$. $PI(\Sigma)$ denotes the set of prime implicates of Σ (up to logical equivalence). Let Σ, Φ be two formulas of $PROP_{PS}$. A Φ-*implicate* of Σ is a clause π s.t. $\Sigma \models_\Phi \pi$. A *theory prime implicate* of Σ w.r.t. Φ is a Φ-implicate of Σ which is minimal w.r.t. $\models_\Phi$. $TPI(\Sigma, \Phi)$ denotes the set of theory prime implicates of Σ w.r.t. Φ (up to Φ-equivalence).

2.2 The diagnosis framework

We adhere to full propositional logic as a representation language. Let us recall that every first-order formula over a finite domain can be reduced to a propositional formula (Henkin reduction). In this propositional setting, a system S is a 3-uple (SD, COMPS, OBS) where COMPS, the system components, is a finite set of constants $c_1, ..., c_n$ and SD and OBS are finite sets of formulas representing, respectively, the system description (models of the components and connections) and a set of observations (typically, the inputs and outputs of the system). Associated with each component c_i is a propositional symbol `ab_ci` (referred to as an *ab-symbol*) used to indicate that c_i is faulty. AB is the set of all the ab-symbols. An *ab-literal* is an ab-symbol or a negated one. An *ab-clause* (resp. *ab-term*, *ab-formula*) is a clause (resp. term, formula) built upon ab-literals only. Given a formula Σ of $PROP_{PS}$, the *ab-consequence* Σ_{ab} of Σ is the set (conjunctively considered) of all the (non-valid) ab-formulas ψ_{ab} that are logical consequences of Σ. Clearly enough, Σ_{ab} is a logical consequence of Σ and it is the most general ab-one, in the sense that every ab-formula ψ_{ab} that is a logical consequence of Σ also is a logical consequence of Σ_{ab}. Accordingly, Σ_{ab} can be computed by forgetting in Σ every symbol except the ab ones: $\Sigma_{ab} \equiv$ *forget*$(\Sigma, Var(\Sigma) \setminus AB)$, where *forget* is the operator introduced in (Lin and Reiter, 1994). $(SD \wedge OBS)_{ab}$ is also referred to as the consequence of OBS in (Darwiche, 1995) (Darwiche, 1997). Every clause (resp. minimal clause, resp. positive clause) from $(SD \wedge OBS)_{ab}$ is classically referred to as a *conflict* (resp. a *minimal conflict*, *positive conflict*) for S. The set of all positive conflicts for $(SD \wedge OBS)_{ab}$ is noted $(SD \wedge OBS)_{ab+}$. Because $(SD \wedge OBS)_{ab}$ can be put into CNF, it is equivalent to the set of all conflicts for S and it is also equivalent to the set of all minimal conflicts for S.

Given a system S, we will deal with the following notions:

- A *candidate diagnosis* Δ for S is an ab-term that is consistent.
- A *complete diagnosis* Δ for S is a candidate diagnosis for S in which every ab-symbol occurs (*i.e.* the status of every component of the system is represented in Δ). Every candidate diagnosis which is not complete is said *incomplete.*
- A *consistency-based diagnosis* Δ for S is a candidate diagnosis s.t. $(\Delta \wedge SD \wedge OBS)$ is consistent.
- A *minimal diagnosis* Δ for S is a complete consistency-based diagnosis for S in which a minimal subset (w.r.t. set inclusion) of positive ab-literals occurs. Δ is often represented as the set of such positive literals.
- A *partial diagnosis* Δ for S is a consistency-based diagnosis for S s.t. every candidate diagnosis which is a superset of Δ is a consistency-based diagnosis for S. In particular, complete consistency-based diagnoses for S are partial diagnoses for S.
- A *kernel diagnosis* Δ for S is a partial diagnosis for S s.t. no proper subset of it is a partial diagnosis for S.

Clearly enough, both minimal, partial and kernel diagnoses are consistency-based ones. The following characterizations of these notions have been pointed out in (De Kleer *et al.*, 92):

- A minimal diagnosis Δ for S is an ab-term s.t. the conjunction of all its positive literals is a prime implicant of $(SD \wedge OBS)_{ab+}$ and all the remaining ab-literals appear negatively in Δ [Theorem 3.13].
- A partial diagnosis Δ for S is an implicant of $(SD \wedge OBS)_{ab}$ [Theorem 4.5].
- A kernel diagnosis Δ for S is a prime implicant of $(SD \wedge OBS)_{ab}$ [Corollary 4.6].

Additionally, (De Kleer *et al.*, 92) shows that a complete candidate diagnosis Δ is a consistency-based diagnosis for S iff there exists a kernel diagnosis Δ_k for S s.t. $\Delta \models \Delta_k$ [Theorem 4.4]. All these results highlights the central role taken by $(SD \wedge OBS)_{ab}$ for characterizing consistency-based diagnoses (Darwiche, 95) (Darwiche, 97).

3. Characterizing Consistency-Based Diagnoses

They are many ways to characterize the consistency-based diagnoses for a system. At the extreme bounds of the spectrum are the intensional definition of consistency-based diagnoses (the one given in Section 2) and the extensional one (i.e. the set of consistency-based diagnoses for S itself). Intuitively, both cannot be considered as interesting characterizations. On the one hand, the former is of no help in the objective of deriving consistency-based diagnoses for S; from a computational point of view, it is « too far » from the set of consistency-based diagnoses. On the other hand, the latter is often too large to be explicitely computed.

While several characterizations of consistency-based diagnoses have been pointed out to date, the problem of investigating what should be in such a characterization has been left unaddressed till now. Pointing out all the properties that a valuable characterization must exhibit is not an easy task. Our purpose is to put forward some necessary conditions, only, so as to prevent from considering as valuable characterizations the intensional definition and the extensional one. Thus, in our opinion, any valuable characterization $\chi(S)$ of the consistency-based diagnoses for a system S should be (at least):

• *information-preserving*: the set of all the consistency-based diagnoses for S can be derived from the characterization $\chi(S)$.

• *operational*: $\chi(S)$ must be computationally helpful. An algorithm for generating $\chi(S)$ from S must be available. Moreover, given a system S and a candidate diagnosis Δ, checking whether Δ is a consistency-based diagnosis for S (formally referred to as the CONSISTENCY-BASED DIAGNOSIS decision problem in the following) has to be achievable in time polynomial in $|\chi(S)|+|\Delta|$. Accordingly, the intensional definition of the consistency-based diagnoses for S is not operational in the general case.

• *concise*: $\chi(S)$ must be as compact as possible (especially, more compact than the set of all the consistency-based diagnoses for S). This prevents the extensional definition to be considered as a valuable characterization of S.

Clearly enough, these three properties are not independent one another. First, the operationality requirement constrains the conciseness requirement in the general case. Indeed, CONSISTENCY-BASED DIAGNOSIS is NP-complete[1]. Hence, characterizations χ s.t. $\chi(S)$ is operational and of size polynomial in the size of S for every S, are very unlikely: their existence would imply $NP \subseteq P/poly$. This is a consequence of [Theorem 5] from (Kautz and Selman, 1992) since a candidate diagnosis Δ is a consistency-based diagnosis for S iff the clause $\neg\Delta$ is *not* a logical consequence of $SD \wedge OBS$. Accordingly, there can be exponentially many minimal diagnoses, kernel diagnoses or MT-diagnoses for a system in the worst case. Such a possible exponential blowup is a severe limitation to which any operational characterization is liable. However, this complexity result neither questions the fact that a polynomial amount of space is sufficient for deriving operational characterizations of some sets of diagnoses, nor questions the fact that some operational characterizations are more concise than others.

Second, the information-preservation requirement may also have a strong influence on the conciseness one: characterizing some preferred consistency-based diagnoses can be either more easy or more difficult than characterizing the whole set of consistency-based diagnoses, depending on the preference criterion. Obviously, the fact that preferred consistency-based diagnoses are among the consis-

[1] This is a by-product of the general complexity results for diagnosis given in (Friedrich *et al.*, 90), especially [Theorem 4].

tency-based ones has nothing to do with the conciseness of their characterizations: consistency-based diagnoses for S are among the candidate diagnoses for S, which have both a concise and operational characterization (their intensional definition). At that time, the set of all minimal diagnoses for a system S can be characterized in a more compact way (e.g. by its extensional definition) than the set of all consistency-based ones (characterized by its extensional definition or the kernel diagnoses for S); especially, there cannot be more minimal diagnoses for S than kernel diagnoses for S and there can be exponentially more kernel diagnoses than minimal diagnoses. Unfortunately, parsimony is only a preference criterion,: while the actual diagnosis of a system is necessarily a consistency-based one (when both SD and OBS are certain), it can easily be the case that it is not a minimal one. Hence, only considering minimal diagnoses is not always a good idea (De Kleer *et al.*, 92).

Surprisingly, none of the characterizations of consistency-based diagnoses proposed so far fully satisfies the three properties above. Thus, it is well-known that the minimal diagnoses for a system S cannot be considered as a characterization of all its consistency-based diagnoses: every superset of a minimal diagnosis is not necessarily a consistency-based one (De Kleer *et al.*, 90, 92). Accordingly, the information-preservation requirement is not satisfied by the minimal diagnoses characterization.

The notion of kernel diagnosis (De Kleer *et al.*, 90, 92) has been introduced to overcome this problem. Actually, the set of all kernel diagnoses for a system S is a characterization of all consistency-based diagnoses for S; indeed, Δ is a consistency-based diagnosis for S iff there exists a kernel diagnosis Δ_k for S s.t. $\Delta \wedge \Delta_k$ is consistent. However, such a characterization is often insufficiently concise. In particular, kernel diagnoses do not offer ability to enable hierarchical reasoning, so as to achieve reduction of complexity (Provan and Poole, 91).

To cope with this problem, (Friedrich, 93) proposes minimal theory diagnoses (MT-diagnoses for short) as a way to describe all the minimal diagnoses of a system, while incorporating some additional knowledge, especially hierarchical relations between components. (Friedrich, 93) shows that there cannot be more MT-diagnoses for S than minimal diagnoses for S. Unfortunately, MT-diagnoses are oriented towards the characterization of minimal diagnoses (and not of all consistency-based ones). Furthermore, no algorithm for computing MT-diagnoses has been put forward. Hence, neither the information-preservation requirement nor the operationality one are fulfilled by the MT-diagnoses characterization.

Finally, the consequence of the observations, i.e. $(SD \wedge OBS)_{ab}$ (Darwiche, 95) (Darwiche, 97) clearly satisfies the information-preservation requirement. However, while minimal-cardinality diagnoses can be derived in time linear in the size of a decomposable negation normal form (NNF) of OBS, it is not clear whether or not such a form makes CONSISTENCY-BASED DIAGNOSIS tractable (especially when partial candidate diagnoses are considered). Moreover, the algorithm for deriving a decomposable NNF of OBS given in (Darwiche, 97) does not guarantee that the size of the derived formula always is polynomial in the size of S in the general case.

4. Theory Minimal Conflicts Characterizations

In the following, we suggest to take advantage of theory minimal conflicts for a system S as characterizations of its consistency-based diagnoses. Theory minimal conflicts are defined and characterized thanks to the notion of theory prime implicate. We show that component hierarchies can be used as theories in this framework. We prove that the theory minimal conflicts for S characterize all its consistency-based diagnoses and that such characterizations can be exponentially more concise than the kernel diagnoses ones in some situations. We show that the theory minimal conflicts for S can be used to check whether any candidate diagnosis is a consistency-based (resp. a partial, a kernel) diagnosis for S in polynomial time. In addition, we show how a consistency-based (resp. a partial, a kernel) diagnosis for S can be generated in time polynomial in the size of a theory minimal conflicts characterization. Finally, we show how theory minimal conflicts for S can be computed without requiring the minimal conflicts of S to be generated.

4.1 The theory minimal conflicts framework

Following (Friedrich, 93), our objective is to take advantage of knowledge about abstractions to derive characterizations that are *as concise as possible*; this knowledge, expressed by *theories*, is considered exceptional so as to avoid the approach to flatten out (Provan and Poole, 1991). Accordingly, our approach can be related to the the numerous works concerned with hierarchical diagnosis (*e.g.* (Davis, 84) (Hamscher, 91) (Mozetic, 91) (Struss, 92)).

In contrast to the minimal, kernel or MT-diagnoses approaches, we focus on characterizing the *complement* of the set of all consistency-based diagnoses for a system:

Definition 1 Let S = (SD, COMPS, OBS) be a system to be diagnosed. Let Φ be a tractable ab-formula belonging to $(SD \wedge OBS)_{ab}$. Φ is referred to as a *theory*. *A theory minimal conflict* for S w.r.t. Φ is a minimal conflict for S that is minimal w.r.t. $\models_{\Phi}$.
TMC(S, Φ) is the set of all the theory minimal conflicts for S w.r.t. Φ (up to Φ-equivalence). One representative per Φ-equivalence class is kept in the set, only.

The negation of a conflict for S is a candidate diagnosis that is *not* a consistency-based one, and the converse also holds. The key intuition behind our approach is that, in some situations, it is more easy to describe the candidate diagnoses that are not consistency-based than the consistency-based ones. Especially, there are sets of consistency-based diagnoses of size exponential in |COMPS| which can be represented by a set of conflicts of size linear in |COMPS| (see the example below).

In contrast to diagnoses, conflicts have the interesting feature that, in the general case (especially when fault models are allowed in SD), every superset of a conflict for S also is a conflict for S. Theory minimal conflicts for S w.r.t. Φ are the most abstract conflicts for the system. The existence of such an interpretation can prove helpful to guide the diagnostic process in the steps that are interleaved with the search one (*e.g.* taking some additional measurements, replacing a component ...).

Like conflicts and minimal conflicts, theory minimal conflicts are *implicates*. Such a characterization is of prime importance from a computational point of view (see Section 4.3):

Proposition 1 Let S = (SD, COMPS, OBS) be a system to be diagnosed. Let Φ be a tractable ab-formula belonging to $(SD \wedge OBS)_{ab}$. The theory minimal conflicts for S w.r.t. Φ are the theory prime implicates of $(SD \wedge OBS)_{ab}$ w.r.t. Φ (up to Φ-equivalence).

As evoked above, the available knowledge Φ about abstractions consists, for many systems, of component hierarchies:

Definition 2 (Friedrich, 93) A component hierarchy of S is a finite set of ab-formulas of the form:

$$\text{ab\_ci} \Leftrightarrow (\text{ab\_cj} \vee \ldots \vee \text{ab\_ck})$$

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.

Interestingly, such component hierarchies can be used as theories in our framework:

Proposition 2 Component hierarchies in CNF are tractable.

Let us illustrate these notions on a simple example from (Friedrich, 93). The system S consists of two buffer chains c_{2a}-c_{2b} and c_{3a}-c_{3b} connected to an OR-gate c_1 (see Figure 1 below). The normal behaviors of these components are assumed. The inputs of the circuit are observed to be 1 and the output of the OR-gate to be 0. The buffer chain c_{2a}-c_{2b} (resp. c_{3a}-c_{3b}) is abstracted into the component c_2 (resp. c_3). S is faulty if the OR-gate is faulty or if at least one buffer in each chain is faulty. Formally, let S = (SD, COMPS, OBS) where:

COMPS = $\{c_1, c_{2a}, c_{2b}, c_{3a}, c_{3b}, c_2, c_3\}$,
SD = {¬ab_c1 ⇒ (¬out_c1 ⇔ (¬in1_c1 ∧ ¬in2_c1)),
¬ab_c2a ⇒ (out_c2a ⇔ in1_c2a),
¬ab_c2b ⇒ (out_c2b ⇔ in1_c2b),
¬ab_c3a ⇒ (out_c3a ⇔ in1_c3a),
¬ab_c3b ⇒ (out_c3b ⇔ in1_c3b),
ab_c2 ⇔ (ab_c2a ∨ ab_c2b),
ab_c3 ⇔ (ab_c3a ∨ ab_c3b),
out_c2a ⇔ in_c2b, out_c3a ⇔ in_c3b ,
out_c2b ⇔ in1_c1, out_c3b ⇔ in2_c1},
OBS = {in_c2a, in_c3a, ¬out_c1}.

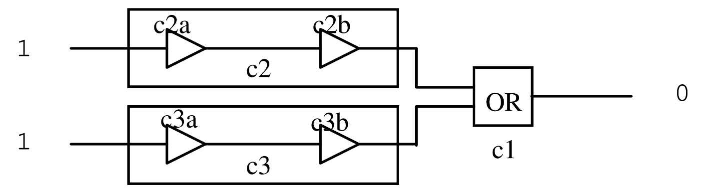

Figure 1: Two buffer chains and one OR-gate.

The system description and observations of S give rise to the following set of minimal conflicts $(SD \wedge OBS)_{ab}$: {ab_c1 ∨ ab_c2, ab_c1 ∨ ab_c3, ab_c1 ∨ ab_c2a ∨ ab_c2b, ab_c1 ∨ ab_c3a ∨ ab_c3b} ∪ Φ, where Φ = {¬ab_c2 ∨ ab_c2a ∨ ab_c2b, ab_c2 ∨ ¬ab_c2a, ab_c2 ∨ ¬ab_c2b, ¬ab_c3 ∨ ab_c3a ∨ ab_c3b, ab_c3 ∨ ¬ab_c3a, ab_c3 ∨ ¬ab_c3b} is a CNF of the formula (ab_c2 ⇔ (ab_c2a ∨ ab_c2b)) ∧ (ab_c3 ⇔ (ab_c3a ∨ ab_c3b)) of SD that expresses the component hierarchy of S. Roughly, Φ represents the fact that any buffer chain is faulty iff at least one component of it is faulty.

There are 10 minimal conflicts for S and only 2 theory minimal conflicts for S w.r.t. Φ (up to Φ-equivalence), namely ab_c1 ∨ ab_c2 and ab_c1 ∨ ab_c3. As we will see in the following, these two theory minimal conflicts for S characterize all the consistency-based diagnoses for S, especially the 20 complete ones and the 9 kernel ones. Moreover, the set of theory minimal conflicts for S can be interpreted within the diagnosis framework: S is faulty iff there is a faulty component in $\{c_1, c_2\}$ and a faulty component in$\{c_1, c_3\}$.

4.2 The theory minimal conflicts characterizations

While minimal and MT-diagnoses only characterize *some* consistency-based diagnoses for S, theory minimal conflicts for S characterize all its consistency-based diagnoses; hence, the information-preservation requirement is fully satisfied by our approach:

Proposition 3 Let S be a system to be diagnosed. Let Φ be a tractable ab-formula belonging to $(SD \wedge OBS)_{ab}$. Let Δ be a candidate diagnosis for S. Δ is a consistency-based diagnosis for S iff for all theory minimal conflicts X_{tmc} for S w.r.t. Φ, the formula $\Delta \wedge \Phi \wedge X_{tmc}$ is consistent.

Clearly enough, the number of theory minimal conflicts for a system w.r.t. a theory cannot exceed the number of its minimal conflicts; more, it can be exponentially smaller. More generally, the stronger Φ is (from a logical point of view) the smaller the corresponding set of theory minimal conflicts will be: if Φ and Φ' are tractable ab-formulas belonging to $(SD \wedge OBS)_{ab}$ s.t. $\Phi \models \Phi'$, then the cardinality of TMC(S, Φ) is always smaller or equal to the cardinality of TMC(S, Φ').

We cannot guarantee that there is always less theory minimal conflicts than kernel diagnoses for the same system; actually, this is not the case since a formula can possess more prime implicates than prime implicants. However, there are many situations in which the number of theory minimal conflicts sufficient to characterize the consistency-based diagnoses for a system S given a theory Φ is linear in the number of components of S, while the number of kernel diagnoses required to achieve the same goal would be exponential. In order to illustrate it, let us generalize the previous example and consider a circuit with p chains ($p \geq 2$) of buffers consisting of 2*k buffers each ($k \geq 1$) and the outputs of the buffer chains are connected to a single p-inputs OR-gate. Such a (parametric) system $S_{p,\ k}$ has 2*k*p + 1 « physical » components plus p « abstracted » ones. Assuming that the description of the correct behaviors of each component is available and that each buffer chain is abstracted into a unique component, giving rise to a component hierarchy Φ, we can easily prove that :

- There are $(2*k+1)^p$ kernel diagnoses for $S_{p,\ k}$, which is exponential in |COMPS| whenever k is fixed since p = (|COMPS|-1)/(2*k).
- There are only p theory minimal conflicts for $S_{p,\ k}$ w.r.t. Φ, which is always smaller than |COMPS|. Of course, Φ must be considered together with the set of theory minimal conflicts w.r.t. Φ to achieve the characterization purpose but this does not question the conciseness issue since $|\Phi| \leq |SD|$.

If no abstraction is performed (i.e. only the « physical » components of $S_{p,\ k}$ are considered), then there is still an exponential amount of kernel diagnoses for the system ($(2*k)^p + 1$ kernel diagnoses); thanks to [Theorem 3.13] of (De Kleer *et al.*, 1992), such diagnoses can be read as the minimal diagnoses for the system since every conflict of $S_{p,\ k}$ is positive when no hierarchy is considered.

Interestingly, Proposition 3 shows that the theory minimal conflicts for S w.r.t. Φ satisfy the requirement of making CONSISTENCY-BASED DIAGNOSIS tractable. Indeed, each elementary check ($\Delta \wedge \Phi \wedge X_{tmc}$ consistent) can be performed in time polynomial in $|\Delta| + |X_{tmc}| + |\Phi|$ when Φ is tractable since $\Delta \wedge \Phi \wedge X_{tmc}$ is inconsistent iff $\Phi \models \neg\Delta \vee \neg l$ for every literal l of X_{tmc}; therefore, checking whether a candidate diagnosis Δ is a consistency-based diagnosis for S can be done in time polynomial in $|\Delta| + |TMC(S, \Phi)| + |\Phi|$.

Additionally, *finding out* a consistency-based diagnosis for S from a theory minimal conflicts characterization can be achieved in polynomial time. More generally, a consistency-based diagnosis for S can be derived in time polynomial in the size of S whenever an oracle for CONSISTENCY-BASED DIAGNOSIS is available. The technique we use consists in building up such a diagnosis by considering successively the ab-symbols of S (see (Papadimitriou, 94) pp. 228 for details). Let COMPS = $\{c_1, ..., c_n\}$. Our algorithm first asks whether $\Delta = \{\}$ is a consistency-based for S (*i.e.* whether SD $\wedge$ OBS is consistent). If the answer is « no », no consistency-based diagnosis for S exists. Otherwise, our algorithm considers the two instances <S, {`ab_c1`}> and <S, {`¬ab_c1`}> of the CONSISTENCY-BASED DIAGNOSIS decision problem. Obviously, the answer of one of these two instances is « yes ». Our algorithm decides whether to keep the literal l = `ab_c1` or the literal l = `¬ab_c1` in the diagnosis depending on which of the two instances has a « yes » answer (if both have a « yes » answer, we keep any of the two). Then, c_1 is removed from COMPS and SD $\wedge$ OBS is simplified with the chosen literal l, *i.e.* it is replaced by $(SD \wedge OBS)_l$. With at most 2*n calls to the oracle, all with systems simpler than S, this algorithm generates a consistency-based diagnosis for S.

Additionally, theory minimal conflicts characterize partial and kernel diagnoses in an efficient way:

Proposition 4 Let S be a system to be diagnosed. Let Φ be a tractable ab-formula belonging to $(SD \wedge OBS)_{ab}$. Let Δ be a candidate diagnosis for S. Δ is a partial diagnosis for S iff $\Delta \models \Phi$ and for every theory minimal conflict X_{tmc} for S w.r.t. Φ, we have $\Delta \models X_{tmc}$.

Thanks to this proposition, checking whether a candidate diagnosis is a partial or a kernel diagnosis for S can be done in time polynomial in the size of any theory minimal conflicts characterization when Φ is in CNF. To be more precise, checking whether a candidate diagnosis is a partial diagnosis for S can be done in time linear in $|\Delta| + |TMC(S, \Phi)| + |\Phi|$ since $\Delta \models \Phi$ can be checked in $O(|\Delta| + |\Phi|)$ time when Φ is in CNF. Since complete consistency-based diagnoses are partial ones, Proposition 4 also shows how to check whether a complete diagnosis is a consistency-based one in linear time w.r.t. the size of the characterization. So as to decide whether a partial diagnosis Δ for S is a kernel diagnosis for

S, it is sufficient to check that, for every ab-literal of Δ, the candidate obtained by removing this literal from Δ is not a partial diagnosis for S. Accordingly, checking whether a candidate diagnosis for S is a kernel diagnosis for S can be done in $O((|\Delta|+1)*(|\Delta|+|TMC(S, \Phi)|+|\Phi|))$ time.

Interestingly, this strategy can be easily adapted for *deriving* a kernel diagnosis for S from a partial one in polynomial time (each time an ab-literal of Δ can be removed while giving rise to a partial diagnosis, it is removed from Δ and kept otherwise; at the end, a kernel diagnosis for S results). Since complete consistency-based diagnoses are partial ones, and since deriving a complete consistency-based diagnosis for S from the theory minimal conflicts for S can be done in polynomial time, finding out a kernel diagnosis for S from the theory minimal conflicts for it can also be done in polynomial time.

As mentioned above, the theory minimal conflicts approach presented in this paper is closely related to the MT-diagnoses one given in (Friedrich, 93); indeed, both take advantage of knowledge about abstractions into the diagnostic process. However, our approach is oriented towards the characterization of all the consistency-based diagnoses for a system, not the minimal ones only. Moreover, in our approach, the theory Φ is not restricted to a component hierarchy; especially, any tractable consequence of $(SD \wedge OBS)_{ab}$ can be used as a theory in our framework. This is not the case in the MT-diagnoses framework. Finally, in contrast to the MT-diagnoses approach, a strong feature of our approach is an algorithm for computing the characterization.

4.3 Computing theory minimal conflicts

Theory minimal conflicts are computed as the theory prime implicates of $(SD \wedge OBS)_{ab}$ w.r.t. Φ. In order to derive such theory prime implicates, the algorithm `TPI/BDD` given in (Marquis and Sadaoui, 96) can be easily modified. `TPI/BDD` is based on the following properties allowing us to compute the theory prime implicates of a formula Σ w.r.t. a theory Φ as the minimal implicates MinPI(Σ, Φ) of Σ w.r.t. Φ ; thanks to these properties, the minimal implicates of Σ w.r.t. Φ are computed recursively from the minimal implicates of its subformulas w.r.t. Φ:

- If $\Sigma \models \Phi$, then $TPI(\Sigma, \Phi) = MinPI(\Sigma, \Phi) = \min(PI(\Sigma), \models_{\Phi})$.
- If $(\neg x \wedge \Sigma_{\neg x}) \vee (x \wedge \Sigma_x)$ is the Shannon expansion of Σ around variable x, then
  $MinPI(\Sigma, \Phi) = \min(MinPI(\Sigma_{\neg x} \vee \Sigma_x, \Phi) \cup (\{\neg x\} \vee MinPI(\Sigma_x, \Phi)) \cup (\{x\} \vee MinPI(\Sigma_{\neg x}, \Phi)), \models_{\Phi})$.
- $MinPI(\Sigma_{\neg x} \vee \Sigma_x, \Phi) = \min(\{\pi_{\neg x} \vee \pi_x ; \pi_{\neg x} \in MinPI(\Sigma_{\neg x}, \Phi), \pi_x \in MinPI(\Sigma_x, \Phi)\}, \models_{\Phi})$.

So as to avoid computing minimal implicates for several occurrences of the same subformula of Σ, a Binary Decision Diagram (BDD) of S is first computed. Let us recall that a BDD of a propositional formula Σ is a directed acyclic graph obtained by iterating the Shannon expansion of Σ around all the variables occurring in S (chosen according to a given order), then removing redundancy in the resulting Shannon tree by sharing common subtrees (see (Akers, 78) (Bryant, 92)).

Performing case analysis on every propositional symbol of $SD \wedge OBS$ except the ab ones enables deriving a representation of $(SD \wedge OBS)_{ab}$ (see (Darwiche, 97)) but such a representation is not a Shannon expansion of $(SD \wedge OBS)_{ab}$, that would remain to be computed. Interestingly, the ab-consequence $(SD \wedge OBS)_{ab}$ of $SD \wedge OBS$ must *not* be computed *before* calling `TPI/BDD` but its computation can be integrated within the first step of `TPI/BDD`, namely the generation of a Binary Decision Diagram (BDD) of $SD \wedge OBS$. So as to avoid the generation of $(SD \wedge OBS)_{ab}$ in a preliminary step, it is sufficient to iterate the Shannon expansion of $SD \wedge OBS$ around all of its ab-symbols. The restriction formulas ρ of $SD \wedge OBS$ labeling the leaves of the corresponding Shannon tree are then checked for consistency: if ρ is consistent (resp. inconsistent), ρ is replaced by the boolean constant *true* (resp. *false*). The resulting tree is a Shannon tree of $(SD \wedge OBS)_{ab}$. Then `TPI/BDD` is run normally and the theory prime implicates of $(SD \wedge OBS)_{ab}$ w.r.t. Φ are derived.

The correctness of this technique relies on the following proposition:

Proposition 5 Let $\vee_i (\Delta_i \wedge \rho_i)$ any formula equivalent to $SD \wedge OBS$ s.t. no ab-symbol occurs in any ρ_i and no symbol that is not an ab-symbol occurs in any Δ_i. Then $(SD \wedge OBS)_{ab} \equiv \vee_i \Delta_i$ s.t. ρ_i is consistent.

The key is that the formula $\vee_i\ (\Delta_i \wedge \rho_i)$ that can be read from any Shannon tree of SD $\wedge$ OBS around its ab-symbols (considered in any order) by identifying for every path of the tree the formula labeling its leaf with ρ_i and the term labeling the path by Δ_i, is equivalent to SD $\wedge$ OBS.

Clearly enough, every branch of the Shannon tree of $(SD \wedge OBS)_{ab}$ is associated with a consistency-based diagnosis for S (every term which can be read off, following the path from the root of the tree to any leaf of the tree, is a consistency-based diagnosis for S). However, the converse does not hold; hopefully, `TPI/BDD` does not require all the consistency-based diagnoses for S to be generated. Moreover, as shown in (Marquis and Sadaoui, 96), `TPI/BDD` does not require the prime implicates of $(SD \wedge OBS)_{ab}$ to be generated. Because the number of such prime implicates can easily be exponentially greater than the number of theory prime implicates, exponential time and space savings can be achieved by `TPI/BDD` w.r.t. prime implicates algorithms (see (Marquis and Sadaoui, 96) for details). Subsequently, because the prime implicates of $(SD \wedge OBS)_{ab}$ are the minimal conflicts of S, our algorithm for computing theory minimal conflicts for S does not require the minimal conflicts of S to be generated. This contrasts with many algorithms for deriving characterizations of the consistency-based diagnoses for S, especially kernel diagnoses (De Kleer *et al.*, 92), and this can save a lot of computing.

Let us illustrate our algorithm for computing theory minimal conflicts on our running example. The following table gives the BDD of $(SD \wedge OBS)_{ab}$ for the variable order `ab_c1 < ab_c2 ab_c3 < ab_c2a < ab_c2b < ab_c3a < ab_c3b` ; it also gives minimal conflicts and theory minimal conflicts computed at each node of the BDD. The first column gives the identifier `n` of the node, the second column indicates the variable `x` around which the formula Σ rooted at node `n` will be Shannon expanded. The third (resp. fourth) column gives the identifier of the left son (resp. right son) of node `n`; the formula rooted there is $\Sigma_{\neg x}$ (resp. Σ_x). The next column reports first the number of minimal conflicts of Σ, and these minimal conflicts. The rightmost one gives the number of theory minimal conflicts of Σ, and these theory minimal conflicts. Minimal conflicts are reported so as to illustrate the space savings that are achieved:

node	variable	left	right	minimal conflicts	theory minimal conflicts
1	leaf	-	-	(1) 0	(1) 0
2	leaf	-	-	(1) 1	(1) 1
3	ab_c3b	1	2	(1) ab_c3b	(1) ab_c3b
4	ab_c3b	2	1	(1) ¬ab_c3b	(1) ¬ab_c3b
5	ab_c3a	3	2	(1) ab_c3a ∨ ab_c3b	(1) ab_c3a ∨ ab_c3b
6	ab_c3a	4	1	(2) ¬ab_c3a, ¬ab_c3b	(2) ¬ab_c3a, ¬ab_c3b
7	ab_c2b	1	5	(2) ab_c2b, ab_c3a ∨ ab_c3b	(2) ab_c2b, ab_c3a ∨ ab_c3b
8	ab_c2b	1	6	(3) ab_c2b, ¬ab_c3a, ¬ab_c3b	(3) ab_c2b, ¬ab_c3b, ¬ab_c3a
9	ab_c2b	5	1	(2) ¬ab_c2b, ab_c3a ∨ ab_c3b	(2) ¬ab_c2b, ab_c3a ∨ ab_c3b
10	ab_c2b	6	1	(3) ¬ab_c2b, ¬ab_c3a, ¬ab_c3b	(3) ¬ab_c2b, ¬ab_c3a, ¬ab_c3b

11	ab_c2a	7	5	(2) ab_c2a ∨ ab_c2b, ab_c3b ∨ ab_c3a	(2) ab_c2a ∨ ab_c2b, ab_c3b ∨ ab_c3a
12	ab_c2a	8	6	(3) ab_c2a ∨ ab_c2b, ¬ab_c3b, ¬ab_c3a	(3) ab_c2a ∨ ab_c2b, ¬ab_c3b, ¬ab_c3a
13	ab_c2a	9	1	(3) ¬ab_c2a, ab_c3b ∨ ab_c3a, ¬ab_c2b	(3) ¬ab_c2a, ab_c3b ∨ ab_c3a, ¬ab_c2b
14	ab_c2a	10	1	(4) ¬ab_c2a, ¬ab_c3b, ¬ab_c3a, ¬ab_c2b	(4) ¬ab_c2a, ¬ab_c3b, ¬ab_c3a, ¬ab_c2b
15	ab_c3	12	11	(4) ¬ab_c3 ∨ ab_c3b ∨ ab_c3a, ab_c3 ∨ ¬ab_c3a, ab_c3 ∨ ¬ab_c3b, ab_c2b ∨ ab_c2a	(1) ab_c2b ∨ ab_c2a
16	ab_c3	14	13	(5) ¬ab_c3 ∨ ab_c3a ∨ ab_c3b, ab_c3 ∨ ¬ab_c3a, ab_c3 ∨ ¬ab_c3b, ¬ab_c2b, ¬ab_c2a	(2) ¬ab_c2b, ¬ab_c2a
17	ab_c3	1	11	(3) ab_c3, ab_c3a ∨ ab_c3b, ab_c2b∨ ab_c2a	(2) ab_c2b∨ ab_c2a, ab_c3a ∨ ab_c3b
18	ab_c2	16	15	(6) ab_c2 ∨ ¬ab_c2a, ab_c2 ∨ ¬ab_c2b, ¬ab_c2 ∨ ab_c2a ∨ ab_c2b, ¬ab_c3b ∨ ab_c3, ¬ab_c3a ∨ ab_c3, ab_c3b ∨ ab_c3a ∨ ¬ab_c3	(1) ab_c2 ∨ ¬ab_c2a
19	ab_c2	1	17	(4) ab_c2, ab_c2a∨ ab_c2b, ab_c3b ∨ ab_c3a, ab_c3	(2) ab_c2a ∨ ab_c2b, ab_c3b ∨ ab_c3a
20	ab_c1	19	18	(10) ab_c1 ∨ ab_c3, ab_c1 ∨ ab_c3a ∨ ab_c3b, ab_c1 ∨ ab_c2b ∨ ab_c2a, ab_c1 ∨ ab_c2, ab_c3 ∨ ¬ab_c3a, ab_c3 ∨ ¬ab_c3b, ¬ab_c3 ∨ ab_c3a ∨ ab_c3b, ab_c2b ∨ ab_c2a ∨ ¬ab_c2, ¬ab_c2b ∨ ab_c2, ¬ab_c2a ∨ ab_c2	(2) ab_c1 ∨ ab_c3b∨ ab_c3a, ab_c1 ∨ ab_c2a ∨ ab_c2b

Table 1: Minimal conflicts and theory minimal conflicts of the running example.

5. Conclusion

While several characterizations of consistency-based diagnoses for a system S have been proposed so far (e.g. minimal diagnoses, kernel diagnoses, MT-diagnoses), none of them captures all the consistency-based diagnoses for S and is concise enough in the general case. As an alternative to kernel diagnoses, we introduce theory minimal conflicts for S. Like in (Friedrich, 93), knowledge about abstractions (including component hierarchies) is incorporated into the diagnostic process. In contrast to minimal or MT-diagnoses, theory minimal conflicts characterize all the consistency-based diagnoses for S. Especially, the theory minimal conflicts for S can be used to check whether a candidate diagnosis is a consistency-based (resp. a partial, a kernel) diagnosis for S in polynomial time. In addition, pointing out a complete consistency-based (resp. a partial, a kernel) diagnosis for S can also be done in time polynomial in the size of the characterization. Finally, the theory minimal conflicts characterization of the consistency-based diagnoses for a system can prove exponentially more concise than the kernel diagnoses one in many situations.

References

(Akers, 78). S.B. Akers. Binary decision diagrams. *IEEE Transactions on Computers* C-27:509-516.

(Besnard and Cordier, 92). Ph. Besnard and M.-O. Cordier. Explanatory diagnoses and their computation by circumscription. Proc. *10th European Conference on Artificial Intelligence (ECAI'92),* Vienna, pp. 724-728, 1992.

(Bryant, 92). R.E. Bryant. Symbolic boolean manipulation with ordered binary-decision diagrams. *ACM Computing Surveys* 24(3) :293-318, 1992.

(Darwiche, 95). A. Darwiche. Model-based diagnosis using causal networks. Proc. *14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, Montreal , pp. 211-217, 1995.

(Darwiche, 97). A. Darwiche. New advances in structure-based diagnosis: a method for compiling devices. Proc. *8th International Workshop on Principles of Diagnosis (DX'97),* Le Mont St Michel, pp. 35-42, 1997.

(Davis, 84). R. Davis. Diagnostic reasoning based on structure and behaviour. *Artificial Intelligence* 24 :347-410, 1984.

(De Kleer and Williams, 87). J. de Kleer and B.C. Williams. Diagnosing multiple faults. *Artificial Intelligence* 32 :97-130, 1987.

(De Kleer *et al*., 90). J. de Kleer, A. Mackworth and R. Reiter. Characterizing diagnoses. Proc. *8th National Conference on Artificial Intelligence (AAAI'90),* Boston (MA), pp. 324-330, 1990.

(De Kleer *et al*., 92). J. de Kleer, A. Mackworth and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence* 56 :197-222, 1992.

(Del Val, 94). A. del Val. Tractable databases : how to make propositional resolution complete through compilation. Proc. *4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, Bonn, pp. 551-561, 1994.

(El Ayeb *et al*., 90). B. El Ayeb, P. Marquis and M. Rusinowitch. Deductive/abductive diagnoses : the DA-principles. Proc. *9th European Conference on Artificial Intelligence (ECAI'90),* Stockholm, pp. 47-52, 1990.

(Eshghi, 93). K. Eshghi. A tractable class of abduction problems. Proc. *13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, Chambéry , pp. 3-8, 1993.

(Friedrich, 93). G. Friedrich. Theory diagnoses : a concise characterization of faulty systems. Proc. *13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, Chambéry, pp. 1466-1471, 1993. See also *Annals of Mathematics and Artificial Intelligence* 11 :263-282, 1994.

(Friedrich *et al.*, 90). G. Friedrich, G. Gottlob and W. Nejdl. Physical impossibility instead of fault models. Proc. *8th National Conference on Artificial Intelligence (AAAI'90),* Boston (MA), pp. 331-336, 1990. See also an extended version in *Readings in Model-Based Diagnosis*, Morgan Kaufmann, 1992.

(Hamscher, 91). W.C. Hamscher. Modeling digital circuits for troubleshooting. *Artificial Intelligence* 51 :223 -271, 1991.

(Kautz and Selman, 92). H.A. Kautz and B. Selman. Forming concepts for fast inference. Proc. *10th National Conference on Artificial Intelligence (AAAI'92),* San Jose (CA), pp. 786-793, 1992.

(Lin and Reiter, 94). F. Lin and R. Reiter. Forget it ! Proc. *1994 AAAI Fall Symposium on Relevance*, New Orleans (LA), pp. 154-159, 1994.

(Marquis, 95). P. Marquis. Knowledge compilation using theory prime implicates. Proc. *14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, Montreal, pp. 837-843, 1995.

(Marquis and Sadaoui, 96). P. Marquis and S. Sadaoui. A new algorithm for computing theory prime implicates compilations. Proc. *13th National Conference on Artificial Intelligence (AAAI'96),* Portland (OR), pp. 504-509, 1996.

(Mozetic, 91). I. Mozetic. Hierarchical model-based diagnosis. *International Journal of Man-Machine Studies* 35(3):329-362, 1991.

(Papadimitriou, 94). Ch. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

(Provan and Poole, 91). G.M. Provan and D. Poole. The utility of consistency-based diagnostic techniques. Proc. *2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, Cambridge (MA), pp. 461-472, 1991.

(Raiman, 90). O. Raiman. A circumscribed diagnosis engine. Proc. *International Workshop on Expert Systems in Engineering*, Vienna, Lecture Notes in Artificial Intelligence 462, Springer-Verlag, pp. 90-101, 1990.

(Reiter, 87). R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence* 32:57-97, 1987.

(Struss, 92). P. Struss. What's in SD ? Towards a theory of modeling for diagnosis, In *Readings in Model-Based Diagnosis*, Morgan Kaufmann, pp. 419-449, 1992.

Appendix

The proofs of the propositions are given in the following:

Proposition 1 By definition, a theory minimal conflict for S w.r.t. Φ is a minimal conflict for S that is minimal w.r.t. $\models_\Phi$. Since a minimal conflict for S is a prime implicate of $(SD \wedge OBS)_{ab}$, a theory minimal conflict for S w.r.t. Φ is a prime implicate of $(SD \wedge OBS)_{ab}$ that is minimal w.r.t. $\models_\Phi$. Whenever Φ belongs to $(SD \wedge OBS)_{ab}$, the concept of implicate of $(SD \wedge OBS)_{ab}$ and the concept of Φ-implicate of $(SD \wedge OBS)_{ab}$ coincide. Because $\models \subseteq \models_\Phi$ for every formula Φ, the implicates of $(SD \wedge OBS)_{ab}$ that are minimal w.r.t. $\models_\Phi$ are among the prime implicates of $(SD \wedge OBS)_{ab}$ (up to Φ-equivalence). Hence, the theory minimal conflicts for S w.r.t. Φ are the theory prime implicates of $(SD \wedge OBS)_{ab}$ w.r.t. Φ (up to Φ-equivalence).

Proposition 2 Because there are no *tied chains* in hierarchical theories Φ in CNF (see (Eshghi, 93) for a definition), they are *unit-refutation complete* (Del Val, 94), hence tractable: for every clause c, $\Phi \models c$ holds iff there exists a unit-refutation from $\Phi \cup \{\neg c\}$, which can be checked in $O(|\Phi|+|c|)$ time.

Proposition 3 Let us first recall that the theory prime implicates of a formula Σ (w.r.t. a theory Φ) cover all its Φ-implicates: for any clause c, we have $\Sigma \models_\Phi c$ iff there exists a theory prime implicate π of Σ w.r.t. Φ s.t. $\pi \models_\Phi c$. Additionally, whenever Φ is a logical consequence of Σ, we have $\Sigma \models c$ iff

$\Sigma \models_\Phi c$. Now, a candidate diagnosis Δ is a consistency-based diagnosis for S iff $\Delta \wedge (SD \wedge OBS)_{ab}$ is consistent iff $\neg\Delta$ is *not* a logical consequence of $(SD \wedge OBS)_{ab}$ iff there does *not* exist a theory prime implicate X_{tmc} of $(SD \wedge OBS)_{ab}$ w.r.t. Φ s.t. $X_{tmc} \models_\Phi \neg\Delta$ holds (since Φ is a logical consequence of $(SD \wedge OBS)_{ab}$) iff there does not exist a theory prime implicate X_{tmc} of $(SD \wedge OBS)_{ab}$ w.r.t. Φ s.t. $\Delta \wedge \Phi \wedge X_{tmc}$ is inconsistent.

Proposition 4 Δ is a partial diagnosis for S iff $\Delta \models (SD \wedge OBS)_{ab}$ iff $\Delta \models (SD \wedge OBS)_{ab} \wedge \Phi$ (since Φ is a logical consequence of $(SD \wedge OBS)_{ab}$) iff $\Delta \models (\neg\Phi \vee (SD \wedge OBS)_{ab}) \wedge \Phi$ iff $\Delta \models (\neg\Phi \vee (SD \wedge OBS)_{ab})$ and $\Delta \models \Phi$ iff $\Delta \models_\Phi (SD \wedge OBS)_{ab}$ and $\Delta \models \Phi$ iff $\Delta \models \Phi$ and for every theory prime implicate π of $(SD \wedge OBS)_{ab}$, we have $\Delta \models_\Phi \pi$ (indeed, $(SD \wedge OBS)_{ab}$ is Φ-equivalent to the set of its theory prime implicates w.r.t. Φ, see (Marquis, 95) for details)) iff $\Delta \models \Phi$ and for every theory prime implicate π of $(SD \wedge OBS)_{ab}$, we have $\Delta \models \pi$ (since $\Delta \models \Phi$, we have $\Delta \models_\Phi \pi$ iff $\Delta \models \pi$) iff $\Delta \models \Phi$ and for every theory minimal conflict X_{tmc} for S w.r.t. Φ, we have $\Delta \models X_{tmc}$.

Proposition 5 Let us recall that $(SD \wedge OBS)_{ab} \equiv$ *forget*$(SD \wedge OBS, Var(SD \wedge OBS) \setminus AB)$ and that *forget* is inductively defined by :

- *forget*$(SD \wedge OBS, \{x\}) = (SD \wedge OBS)_{\neg x} \vee (SD \wedge OBS)_{x}$,
- *forget*$(SD \wedge OBS, X \cup \{x\}) =$ *forget*(*forget*$(SD \wedge OBS, \{x\}), X)$

(see (Lin and Reiter, 94) for details).

We have *forget*$(SD \wedge OBS, Var(SD \wedge OBS) \setminus AB) \equiv$ *forget*$(\vee_i (\Delta_i \wedge \rho_i), Var(SD \wedge OBS) \setminus AB)$. Let x be a variable from $Var(SD \wedge OBS) \setminus AB$. We have *forget*$(\vee_i (\Delta_i \wedge \rho_i), \{x\}) \equiv \vee_i (\Delta_i \wedge \rho_i)_{\neg x} \vee \vee_i (\Delta_i \wedge \rho_i)_x \equiv \vee_i ((\Delta_i \wedge \rho_i)_{\neg x} \vee (\Delta_i \wedge \rho_i)_x)$. Since Δ_i does not contain any occurrence of x for any x from $Var(SD \wedge OBS) \setminus AB$, we have $\vee_i ((\Delta_i \wedge \rho_i)_{\neg x} \vee (\Delta_i \wedge \rho_i)_x) \equiv \vee_i (\Delta_i \wedge (\rho_{i\,\neg x} \vee \rho_{i\,x})) \equiv \vee_i (\Delta_i \wedge$ *forget*$(\rho_i, \{x\}))$. By induction on the number of variables in $Var(SD \wedge OBS) \setminus AB$, we can easily show that *forget*$(SD \wedge OBS, Var(SD \wedge OBS) \setminus AB) \equiv \vee_i (\Delta_i \wedge$ *forget*$(\rho_i, Var(SD \wedge OBS) \setminus AB))$. Since $Var(\rho_i) \subseteq Var(SD \wedge OBS) \setminus AB)$, *forget*$(\rho_i, Var(SD \wedge OBS) \setminus AB))$ is equivalent to *true* if ρ_i is consistent and is equivalent to *false* otherwise. Hence, *forget*$(SD \wedge OBS, Var(SD \wedge OBS) \setminus AB) \equiv \vee_i \Delta_i$ s.t. ρ_i is consistent.