Solving Strong-Fault Diagnostic Models by Model Relaxation

Alexander Feldman[1] and **Gregory Provan**[2] and **Arjan van Gemund**[1]
[1]Delft University of Technology, Delft, The Netherlands
e-mail: {a.b.feldman,a.j.c.vangemund}@tudelft.nl
[2]University College Cork, College Road, Cork, Ireland, e-mail: g.provan@cs.ucc.ie

Abstract

In Model-Based Diagnosis (MBD), the problem of computing a diagnosis in a strong-fault model (SFM) is computationally much harder than in a weak-fault model (WFM). For example, in propositional Horn models, computing the first minimal diagnosis in a weak-fault model (WFM) is in P but is NP-hard for strong-fault models. As a result, SFM problems of practical significance have not been studied in great depth within the MBD community. In this paper we describe an algorithm that renders the problem of computing a diagnosis in several important SFM subclasses no harder than a similar computation in a WFM. We propose an approach for efficiently computing minimal diagnoses for these subclasses of SFM that extends existing conflict-based algorithms like GDE (Sherlock) and CDA*. Experiments on ISCAS85 combinational circuits show (1) inference speedups with CDA*of up to a factor of 8, and (2) an average of 28% reduction in the average conflict size, at the price of an extra low-polynomial-time consistency check for a candidate diagnosis.

1 Modeling for Diagnostic Inference

Model-based diagnosis (MBD), as formulated in terms of logic [Reiter, 1987], focuses on determining whether an assignment of failure status to a set of mode-variables is consistent with a system description and an observation (e.g., of sensor values). Hence, the diagnostic process consists of taking an observation OBS, and then inferring the failure-mode assignment (diagnosis) consistent with OBS.

Within MBD, two broad classes of model types have been specified: weak-fault models **WFM** [de Kleer *et al.*, 1992] and strong-fault models **SFM** [Struss and Dressler, 1989]. Traditionally, **WFM** has been considered to be computationally simple, and **SFM** computationally hard. Weak-fault models describe a system only in terms of its normal (non-faulty) behaviour, whereas strong-fault models include a definition of some aspects of abnormal behaviour. Strong-fault models can avoid violating physical rules (cf. [Struss and Dressler, 1989]), but at the cost of increased complexity: moving from a binary-valued model with n components (which is adequate for weak-fault models) to one with $m+1$ possible faulty values increases the maximum number of failure candidates from 2^n to $(m+1)^n$.

In terms of worst-case complexity, finding the first minimal diagnosis for a Horn model in **WFM** can be done in polynomial time, but finding the next minimal diagnosis is NP-complete [Friedrich *et al.*, 1990]. In contrast, inference in strong-fault models entails computing kernel diagnoses [de Kleer *et al.*, 1992], which is a Σ_2^P-hard task and is known to be computationally intensive in practice; for example, kernel diagnoses are given by the prime implicants of the minimal conflicts [de Kleer *et al.*, 1992]. Further, the average case complexity of reasoning in **WFM** versus **SFM** increases from poly-time in n (**WFM**) to exponential in n (**SFM**) [de Kleer *et al.*, 1992].

Given this intractability associated with inference using **SFM**, we show that, by closer examination of **SFM**, there is a spectrum of model types, and corresponding inference complexities. We identify two main categories of **SFM**, which we call literal-based SFM, **lSFM**, and function-based SFM, **fSFM**, and show that **lSFM** has the same properties (including inference complexity) as **WFM**, whereas **fSFM** has the properties traditionally assigned to **SFM**.

This paper is the first detailed analysis of **SFM**, to our knowledge, which exploits model structure for computational advantage. It demonstrates the spectrum of fault modeling choices available to the system designer, and the computational implications such choices impose on the resulting diagnostic inference.

We propose a **SFM** algorithm that: (1) decomposes a strong-fault model into strong and weak sub-models; (2) computes diagnoses first in the "relaxed" weak sub-model; and then (3) discards any diagnosis which is not also a diagnosis in the strong sub-model. We identify classes of **SFM** in which the **SFM** diagnosis verification (step 3 above) can be done efficiently. Using ISCAS85 benchmark circuits, we have empirically demonstrated that: (1) our algorithm reduces the diagnosis computation time in CDA*by up to a factor of 8; and (2) the average LTMS conflict size decreases (at the price of an extra consistency check, which has low-polynomial or better time-complexity for several classes of propositional strong-fault models).

2 Related Work

One of the key elements of our approach is decomposing a strong-fault model into strong and weak sub-models, and then

computing diagnoses first in the "relaxed" weak sub-model, and efficiently verifying if this diagnosis is also a diagnosis in the strong sub-model.

Mozetič and Holzbaur [1994] have proposed a related approach: their diagnostic engine, IDA, computes diagnoses first with a *structural* model, which specifies the topological connections of components in the model. This model enables propagation of input values to components to output values, given that the component is OK, and is strictly less expressive than either weak- or strong-fault models. Given any diagnoses computed in a structural model, they are then verified by the weak fault-model, and finally the resulting minimal diagnoses are verified by the strong-fault model.

In contrast to Mozetič and Holzbaur, who relax models into an abstract structural relation, we relax models by partitioning the model into weak and strong sub-models, and then verifying the diagnoses of the weak sub-model by the strong sub-model (exploiting the efficient satisfiability of particular classes of strong sub-model, such as $\mathbf{lSFM}$). Furthermore, our work differs in that they make assumptions on the underlying MBD engines, incrementally updating the underlying conflict set, a step which is not necessary in our approach. In contrast to IDA, our relaxation scheme and algorithm are completely orthogonal to the diagnostic engine.

Our work also bears some relation to work on abstraction of propositional knowledge bases, e.g., [Selman and Kautz, 1996; del Val, 2000]. However, rather than relax a model to a Horn approximation, we relax a strong-fault diagnostic model to a more tractable sub-model, the weak-fault model.

This paper also characterises strong-fault models. In contrast to previous work discussing strong fault models, e.g., [Console and Torasso, 1991; Mozetič and Holzbaur, 1994; Struss and Dressler, 1989], we propose tractable sub-classes of strong-fault models that occur in practice, such as stuck-at circuit models.

Maier and Sachenbacher [2008] propose a general abstraction approach within the framework of constraint optimization. This approach is orthogonal to ours, in that it *adaptively* abstracts constraints to control the memory requirements of message-passing algorithms within a tree-decomposition framework, whereas we abstract the model *prior to* any inference, *independent of* any particular inference algorithm. It is possible to combine these two approaches, which is a topic of future work.

Strong-fault model relaxation is related to hierarchical abstraction [Chittaro and Ranon, 2004]. Due to the bad worst-case computational complexity of diagnosis, hierarchical diagnosis is often necessary, and strong-fault model relaxation can be applied at each level of a hierarchical abstraction to further improve inference efficiency.

3 Diagnostic Model Taxonomy

We represent an artifact as a propositional $\mathbf{Wff}$ over a set V of variables.

Definition 1 (Diagnostic System)**.** A diagnostic system DS is defined as the triple $\mathrm{DS} = \langle \mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS} \rangle$, where SD is a propositional theory over a set of variables V, $\mathrm{COMPS} \subseteq V$, $\mathrm{OBS} \subseteq V$, COMPS is the set of assumables, and OBS is the set of observables.

Throughout this paper we will assume that $\mathrm{SD} \not\models \bot$.

3.1 Diagnosis and Minimal Diagnosis

The traditional query in MBD computes terms of assumable variables which are explanations for the system description and an observation.

Definition 2 (Health Assignment)**.** Given a diagnostic system $\mathrm{DS} = \langle \mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS} \rangle$, an assignment HA to all variables in COMPS is defined as a health assignment.

A health assignment HA is a conjunction of propositional literals. $Lit^{-}(\mathrm{HA})$ is the set of negative literals in HA.

Definition 3 (Diagnosis)**.** Given a diagnostic system $\mathrm{DS} = \langle \mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS} \rangle$, an observation α over some variables in OBS, and a health assignment ω, ω is a diagnosis iff $\mathrm{SD} \wedge \alpha \wedge \omega \not\models \bot$.

Traditionally, other authors [de Kleer and Williams, 1987] arrive at minimal diagnoses by computing a minimal hitting set of the minimal conflicts (broadly, minimal health assignments incompatible with the system description and the observation), while this paper makes no use of conflicts, hence the equivalent direct definition above.

In the MBD literature, a range of types of "preferred" diagnosis has been proposed. This turns the MBD problem into an optimization problem. We assume that we have a preference relation $\preceq$ over diagnoses, which enables us to specify a minimal diagnosis with respect to $\preceq$.

Definition 4 (Minimal Diagnosis)**.** A diagnosis $\omega^{\preceq}$ is defined as minimal if no diagnosis $\tilde{\omega}^{\preceq}$ exists such that $Lit^{-}(\tilde{\omega}^{\preceq}) \subset Lit^{-}(\omega^{\preceq})$.

This definition allows us to capture the two best-known definitions of minimality, namely subset-minimality (using order $\subseteq$) and cardinality-minimality (using order $\leq$ which minimizes the number of negative literals).

The set of all diagnoses/minimal-cardinality (MC) diagnoses of a system description SD and an observation α is denoted as $\Omega(\mathrm{SD}, \alpha)$/$\Omega^{\leq}(\mathrm{SD}, \alpha)$, respectively. We are often interested in computing the size of $|\Omega^{\leq}(\mathrm{SD}, \alpha)|$. We refer to the latter problem as *counting* all MC diagnoses of SD and α.

3.2 Classification of Diagnostic Fault Models

We now introduce our classification of fault models. Given a health variable h_i and an arbitrary $\mathbf{Wff}$ F_i, normal behaviour for component i is denoted using the clause $h_i \Rightarrow F_i$, and abnormal behaviour by $\neg h_i \Rightarrow F_i$. Hence a weak-fault model, as depicted in row 2 of Table 1, is given by $\bigwedge_{i=1}^{n} (h_i \Rightarrow F_i)$; a strong-fault model consists of clauses denoting both normal and abnormal behaviour, i.e., $\bigwedge_{i=1}^{n} (h_i \Rightarrow F_{i,1}) \wedge (\neg h_i \Rightarrow F_{i,2})$.

We generalize the well-known class of *stuck-at* models using the strong-fault class $\mathbf{lSFM}$, in which each clause denoting abnormal behaviour is given by $(\neg h_i \Rightarrow l_i)$ for some literal l_i; for example, this captures component i (with output l_i being stuck-at-0) as given by $[(h_i = \text{stuck-at-0}) \Rightarrow (l_i = 0)]$.

The class $\mathbf{hSFM}$ of Horn strong-fault models has the consequent $\mathbf{Wff}$ F_i restricted to Horn clauses. The class $\mathbf{fSFM}$

Notation	Model Class	System Description	Restrictions
$\mathbf{M}$	abduction models	propositional $\mathbf{Wff}$	
$\mathbf{WFM}$	weak-fault models	$\bigwedge_{i=1}^{n}(h_i \Rightarrow F_i)$	none of h_i appears in F_j, $h_i \in$ COMPS ($1 \leq i,j \leq n$)
$\mathbf{SFM}$	strong-fault models	$\mathbf{SFM} = \mathbf{lSFM} \cup \mathbf{hSFM} \cup \mathbf{fSFM}$	
$\mathbf{lSFM}$	literal-based strong-fault models	$\bigwedge_{i=1}^{n}(h_i \Rightarrow F_i) \wedge (\neg h_i \Rightarrow l_i)$	none of h_i appears in F_j, $h_i \in$ COMPS, $l_i \notin$ COMPS, ($1 \leq i,j \leq n$)
$\mathbf{hSFM}$	Horn strong-fault models	$\bigwedge_{i=1}^{n}(h_i \Rightarrow F_i) \wedge (\neg h_i \Rightarrow H_i)$	none of h_i appears in F_j, $H_i = x_{i,1} \vee x_{i,2} \vee \ldots \vee x_{i,n_i}$ is a disjunction with at most one negative literal ($1 \leq i,j \leq n$)
$\mathbf{fSFM}$	functional strong-fault models	$\bigwedge_{i=1}^{n}(h_i \Rightarrow F_{i,1}) \wedge (\neg h_i \Rightarrow F_{i,2})$	none of h_i appears in $F_{j,1}$ and $F_{j,2}$ for $1 \leq i,j \leq n$
$\mathbf{nlSFM}$	negative literal strong-fault models	$\bigwedge_{i=1}^{n}(h_i \Rightarrow F_i) \wedge (\neg h_i \Rightarrow \neg F_i)$	none of h_i appears in F_j, $h_i \in$ COMPS ($1 \leq i,j \leq n$)

Table 1: Model classification

of functional strong-fault models allows the consequent $\mathbf{Wff}$ F_i to take on any form. Finally, the class $\mathbf{nlSFM}$ of negative-literal strong-fault models has the consequent $\mathbf{Wff}$ F_i restricted to defining the *negation* of the normal behaviour of component i given h_i.

4 SFM Algorithm

In this section we show that we can define an algorithm for specific classes of strong fault models which have complexity of the same class as weak fault models.

4.1 SFM Decomposability

We use the notion of model decomposability in this algorithm, which we introduce first.

Proposition 1. $\mathrm{SD} \in \mathbf{SFM}$ *is decomposable, i.e.,* $\mathrm{SD} = \mathrm{SD}_w \wedge \mathrm{SD}_s$, *where* $\mathrm{SD}_w, \mathrm{SD}_s$ *are the weak and strong subsets of clauses, respectively, such that the subsets have no clauses in common, i.e.,* $\mathrm{SD}_w \cap \mathrm{SD}_s = \emptyset$.

Proof. If we have a consistent SD, there can be no clause of the form $h_i \wedge \neg h_i \Rightarrow F_i$, where F_i is an arbitrary $\mathbf{Wff}$; i.e., every clause can have only one of h_i or $\neg h_i$ in it. Hence we must be able to partition the clauses into: SD_w which consists of clauses of the form $h_i \Rightarrow F_i$, for $i = 1, ..., m$, and SD_s which consists of clauses of the form $\neg h_i \Rightarrow F_i$, for $i = 1, ..., m$. As a consequence, we can decompose $\mathrm{SD} = \mathrm{SD}_w \cup \mathrm{SD}_s$, such that $\mathrm{SD}_w \cap \mathrm{SD}_s = \emptyset$. □

In order to make use of this decomposability property, we need to prove that the strong fault portion of the model SD_s will constrain the diagnoses that are generated by the weak part of the model SD_w, since we compute the intersection of the diagnosis sets of the weak $(\Omega(\mathrm{SD}_w, \mathrm{OBS}))$ and strong $(\Omega(\mathrm{SD}_s, \mathrm{OBS}))$ sub-models, i.e., $\Omega(\mathrm{SD}, \mathrm{OBS}) = \Omega(\mathrm{SD}_s, \mathrm{OBS}) \cap \Omega(\mathrm{SD}_w, \mathrm{OBS})$.

Lemma 1. *For a strong fault model* $\mathrm{SD} \in \mathbf{SFM}$ *which is decomposable such that* $\mathrm{SD} = \mathrm{SD}_s \wedge \mathrm{SD}_w$, *we must have* $\Omega(\mathrm{SD}, \alpha) = \Omega(\mathrm{SD}_w, \alpha) \setminus \overline{\Omega(\mathrm{SD}_s, \alpha)}$.

Proof. Since we are using a monotonic propositional logic, adding extra clauses to any formula F will reduce the number of logical models (diagnoses) of F. It is easy to show that $\Omega(\mathrm{SD}_w, \alpha) \supseteq \Omega(\mathrm{SD}_s, \alpha)$. The diagnoses of SD_s alone are given by $\Omega(\mathrm{SD}_s, \alpha)$, and diagnoses excluded by SD_s alone are given by $\overline{\Omega(\mathrm{SD}_s, \alpha)}$. Hence by adding SD_s to SD_w we obtain $\Omega(\mathrm{SD}, \alpha) = \Omega(\mathrm{SD}_w, \alpha) \setminus \overline{\Omega(\mathrm{SD}_s, \alpha)}$. □

4.2 SFM Algorithm

Algorithm 1 shows how SFM decomposability can be used to extend existing MBD engines which can benefit from reasoning in $\mathbf{WFM}$.

Algorithm 1 SFM Decomposition Algorithm

```
1:  function DIAGNOSE(DS, α) returns a set of diagnoses
        inputs: DS = ⟨SD, COMPS, OBS⟩, diag. system
                α, term, observation
        local variables: ω, term, diagnosis
                         z, term, internal variables
                         SD_w, SD_s, system descriptions
                         Ω, result, set of diagnoses
2:      SD_w ← NOMINALBEHAVIOR(SD)
3:      SD_s ← FAULTYBEHAVIOR(SD)
4:      while MOREDIAGNOSES?(SD_w, α) do
5:          ω ← NEXTDIAGNOSIS(SD_w, α)
6:          U ← COMPUTEINTERNALS(SD_w, α, ω)
7:          if SD_s ∧ α ∧ ω ∧ U ⊭ ⊥ then
8:              Ω ← Ω ∪ {ω}
9:          end if
10:     end while
11:     return Ω
12: end function
```

Note that Alg. 1 is independent of the diagnostic inference engine Ξ. In lines 2-3, we decompose the model; in line 5 we compute a diagnosis ω using the weak portion SD_w of the model using Ξ. Then, line 7 determines if ω is consistent with the strong portion of model, SD_s. Our SFM

algorithm relies on being able to assign values in function COMPUTEINTERNALS to all internal variables, the correctness of which we show in Lemma 3.

The complexity of Diagnose(DS, α) depends on the complexity of three steps:

1. Computing a diagnosis in a weak fault model using a diagnostic oracle Ξ;
2. Computing values of internal variables $U = V \setminus (\mathrm{OBS} \cup \mathrm{COMPS})$;
3. Computing a diagnosis in the strong portion of the fault model, SD_s.

Our algorithm is independent of the diagnostic oracle Ξ, so we state the worst-case complexity here, and do not focus on the average-case complexity of any particular inference algorithm.

The complexity of computing a diagnosis in a weak fault model $\mathrm{SD} \in \mathbf{WFM}$ (e.g., in the weak portion SD_w of the fault model) is as follows:

Lemma 2. *Given* $\mathrm{SD} \in \mathbf{WFM}$*, an observation* α*, the complexity of computing the first diagnosis* $\omega(\mathrm{SD}, \alpha)$ *is* Σ_1^p*-complete, and of computing subsequent diagnoses is* Σ_2^p*-complete.*

Proof. For the class of propositional Horn abduction problems, Bylander et al. [1991] show that computing the first diagnosis $\omega(\mathrm{SD}, \alpha)$ is polynomial in $|\mathrm{COMPS}| + |\mathrm{OBS}|)$, and of computing subsequent diagnoses is Σ_1^p-complete. Since we now deal with general propositional problems, the respective complexities will be up one level of the polynomial hierarchy, i.e., computing the first diagnosis $\omega(\mathrm{SD}, \alpha)$ is Σ_1^p-complete, and of computing subsequent diagnoses is Σ_2^p-complete. Further, these results concur with appropriate formulations in [Nordh and Zanuttini, 2008]. □

The complexity of computing the internal variables is as follows:

Lemma 3. *Given* $\mathrm{SD} \in \mathbf{WFM}$*, an observation* α *assigning values to all variables in* OBS*, and a diagnosis* ω*, we can assign values to all internal variables* $U = V \setminus (\mathrm{OBS} \cup \mathrm{COMPS})$ *in time* $O(|\mathrm{COMPS}| \cdot |V|)$.

Proof (Sketch). We first show the correctness of this procedure, and then show its complexity.

Correctness: Let $\mathrm{OBS} = \mathrm{IN} \cup \mathrm{OUT}$. For $\mathrm{SD} \in \mathbf{WFM}$ and α, diagnosis ω assigns all h_i for the clauses $h_i \Rightarrow F_i$. For each component i that has all inputs in IN, we can compute its outputs given h_i. Here we have to distinguish (1) h_i and (2) $\neg h_i$. In (1) the output of each gate can be computed directly from the inputs and the assumption that the gate is functioning correctly. In (2) the output of the gate is opposite to the one of a correctly functioning gate, otherwise the diagnostic inference cannot deduce $\neg h_i$. If the circuit is connected, which we assume is the case, we can recursively compute the outputs of all downstream components, using induction on COMPS and the connectivity assumption.

Complexity: In the worst case every component has $O(|V|)$ outputs, each of which can be determined in $O(1)$ time, hence a total of $O(|\mathrm{COMPS}| \cdot |V|)$ computations. □

For $\mathbf{nlSFM}$, we prove the following Lemma to be used later.

Lemma 4. *The strong part* SD_s *of* $\mathrm{SD} \in \mathbf{nlSFM}$ *is isomorphic to* $\mathrm{SD} \in \mathbf{WFM}$.

Proof. We show this isomorphism by rewriting SD_s so that it has the form of SD_w. SD_s has the form $\neg h_i \Rightarrow \neg F_i$, and SD_w has the form $h_i \Rightarrow F_i$. If we rename the literals in SD_s by replacing each literal $v \in V$ such that $\neg v_i$ is replaced with $\tilde{v_i}$, we obtain SD_s with the form $\tilde{h_i} \Rightarrow \tilde{F_i}$. The description for the strong part is now isomorphic to the description for the weak part, up to variable renaming. □

The complexity of computing a diagnosis in the strong portion SD_s of the fault model depends on the class of the model. We focus on the following strong fault model classes: $\mathbf{lSFM}$, $\mathbf{nlSFM}$, $\mathbf{hSFM}$, $\mathbf{fSFM}$.

Lemma 5. *Given the strong portion* SD_s *of* $\mathrm{SD} \in \mathbf{SFM}$*, an observation* α *assigning values to all variables in* OBS*, diagnosis* $\omega(\mathrm{SD}_w, \alpha)$ *for the weak portion* SD_s *of* $\mathrm{SD} \in \mathbf{SFM}$*, and values for internal variables* $U = V \setminus (\mathrm{OBS} \cup \mathrm{COMPS})$*, the worst-case complexity of the consistency check* $\mathrm{SD}_s \wedge \alpha \wedge \omega \wedge U \not\models \bot$ *is given by:*

- $O(|V|)$ *for* $\mathrm{SD} \in \mathbf{lSFM}$*;*
- $O(|V|)$ *for* $\mathrm{SD} \in \mathbf{hSFM}$*;*
- Σ_1^p*-complete for the first diagnosis for* $\mathrm{SD} \in \mathbf{nlSFM}$*;*
- Σ_2^p*-complete for* $\mathrm{SD} \in \mathbf{fSFM}$.

Proof. For $\mathrm{SD} \in \mathbf{lSFM}$, the strong part of the system description SD_s is a 2SAT Boolean function, since it has the form $\bigwedge_{i=1}^{n}(\neg h_i \Rightarrow l_i) \equiv \bigwedge_{i=1}^{n}(h_i \vee \neg l_i)$. As a consequence, SD_S can be solved in time linear in the number of literals (using unit propagation).

For $\mathrm{SD} \in \mathbf{hSFM}$, SD_s has the form $\neg h_i \Rightarrow H_i$. Given we have assignments to all the h_i, we need to check the consistency of a set of Horn formulae each of which is set to t or f. Since consistency checking of a set of Horn formulae is a linear-time operation, we can evaluate the consistency of SD_s in linear time.

For $\mathrm{SD} \in \mathbf{nlSFM}$, the strong part of the system description SD_s has the form $\bigwedge_{i=1}^{n}[\neg h_i \Rightarrow (\neg o_i \Leftrightarrow C_i)]$, where C_i is any CNF $\mathbf{Wff}$ that does not include o_i or h_i, for $i = 1, ..., n$. Using Lemma 4, we can show that this task has the same complexity as computing a weak-fault diagnosis.

For $\mathrm{SD} \in \mathbf{fSFM}$ (the case of general clauses), the complexity reverts to the worst-case result of Σ_2^p-completeness. □

5 Experimental Results

Our approach applies to any domain in which models can be defined as in Table 1. We empirically demonstrate our approach using the well-known ISCAS85 circuits [Brglez and Fujiwara, 1985]. In addition to the original ISCAS85 models, we have performed cone reductions as described by Siddiqi and Huang [2007] and de Kleer [2008].

Table 2 gives an overview of all models (V and C denote the total number of variables and number of clauses, respectively). For each circuit we have generated 100 non-masking

	original				reduced
id	|OBS|	|COMPS|	V	C	|COMPS|
c432	43	160	356	1 028	59
c499	73	202	445	1 428	58
c880	86	383	826	2 224	77
c1355	73	546	1 133	3 220	58
c1908	58	880	1 793	4 756	160
c2670	373	1 193	2 695	6 538	167
c3540	72	1 669	3 388	9 216	353
c5315	301	2 307	4 792	13 386	385
c6288	64	2 416	4 864	14 432	1 456
c7552	315	3 512	7 232	19 312	545

Table 2: **nlSFM** ISCAS85 models

double-faults and counted the number of minimal-cardinality diagnoses for each observation. The results are summarized in Table 3 for the **nlSFM** and **WFM** representations.

	original			reduced		
id	min	mean	max	min	mean	max
c432	88	165.8	370	18	55	114
c499	168	214.1	292	2	6.9	11
c880	783	1 032.5	1 944	15	28	99
c1355	1 200	1 623.3	1 996	3	6	18
c1908	1 782	2 321.9	5 614	37	84.4	625
c2670	2 747	3 621.7	7 275	9	16.8	90
c3540	1 364	1 642.2	2 650	128	169.1	312
c5315	3 312	6 202.1	17 423	10	21.6	121
c6288	6 246	8 526.1	15 795	2 385	3 040.4	4 970
c7552	16 617	23 641.2	46 003	22	61.3	657

Table 3: Number of MC diagnoses (**nlSFM**)

Next we have repeated the counting of MC diagnoses in stuck-at-0 and stuck-at-1 models (SD $\in$ **lSFM**). The results are shown in Table 4 (data for stuck-at-1 are very similar). With **lSFM**, there are observations α such that SD $\wedge\, \alpha \models \bot$. The number of these "inconsistent" observations are given in the θ columns of Table 4. Note that, in practice, θ increases with the cardinality of the MC diagnoses.

	original				reduced			
id	min	mean	max	θ	min	mean	max	θ
c432	19	66.2	192	0	1	32.1	103	0
c499	6	36.4	75	32	1	4.8	13	64
c880	114	217.9	510	0	1	5.4	21	10
c1355	286	503.6	688	0	1	4.8	11	54
c1908	410	716.5	1 823	0	1	51.5	522	7
c2670	46	952.2	1 950	0	1	7.1	30	30
c3540	42	333.3	768	0	1	29.1	112	57
c5315	372	1 591.6	4 615	0	1	13.1	50	30
c6288	704	1 083.1	1 755	0	188	374.3	784	0
c7552	3 200	6 168.9	15 655	0	2	33.9	241	22

Table 4: Number of MC diagnoses (**lSFM**, stuck-at-0)

The data in Table 3 and Table 4, show similar diagnostic density for **lSFM** and **nlSFM** models. For the non-reduced circuits, the ratio of the mean number of MC diagnoses in **lSFM** to **nlSFM** varies between 0.13 (for stuck-at-0, c6552) and 0.41 (for stuck-at-1, c6552). The mean ratio of 0.26 shows that more than a quarter of the **WFM** diagnoses are also diagnoses in **SFM**. This evidence, together with the large number of MC diagnoses in **WFM**, presume the existence of one or several *continuous* large **lSFM** subspaces.

Next we studied the effect of the fault modes on the average conflict size computed by a Logic-Based Truth Maintenance System (LTMS) [Forbus and de Kleer, 1993]. The conflict size is very important for algorithms like GDE [de Kleer and Williams, 1987] and CDA*[Williams and Ragno, 2007]. The LTMS conflict sizes for full models are shown in Table 5. We have measured similar reduction in the conflict size for reduced models. For each circuit and observation we computed the average conflict size from the "all nominal" candidate (SD $\wedge\, \alpha$ leads to a double fault, hence assuming each component is healthy leads to a conflict) and from all single-fault candidates. The **WFM** columns show these values averaged over all observations. The remaining three columns show the ratio of the average conflict sizes in **WFM** to the respective **SFM** subclasses.

id	**WFM**	stuck-at-0	stuck-at-1	**nlSFM**
c432	2.91	0.97	0.95	1.04
c499	3.42	0.94	0.92	0.98
c880	2.33	0.88	0.89	0.85
c1355	2.44	0.83	0.86	0.79
c1908	3.95	0.96	0.98	1.03
c2670	2.74	0.98	0.96	0.97
c3540	3.09	0.95	0.96	0.96
c5315	3.04	0.97	0.96	0.96
c6288	3.15	0.7	0.67	0.64
c7552	3.06	0.99	0.93	0.99

Table 5: Average LTMS conflict size

Interestingly, the savings in the average conflict size (between 30% and 36%) are biggest in c6288, which is considered the most difficult ISCAS85 circuit in MBD.

Finally, we studied the effect of Alg. 1 on CDA*, whose average-case complexity is governed by the following two NP-hard problems, which must be solved to compute minimal diagnoses, given a set of (non-minimal) conflicts: (1) compute the set of minimal conflicts by using, for example, directed resolution, and (2) compute the Minimal Hitting Sets (MHS) of all minimal conflicts [Reiter, 1987]. In this experiment we measured the time for computing the first 10 diagnoses in **lSFM** with (1) CDA*and (2) Alg. 1 (i.e., CDA*with **WFM** models and then discarding part of the **WFM** diagnoses after fast consistency check). Table 6 demonstrates speed-ups of up to 750%. Note that our CDA*implementation is pretty limited in solving only "easier" diagnostic problems, and the advantage of model relaxation would be more visible with problems leading to diagnoses of higher cardinalities (more conflicts).

Table 6 shows that Alg. 1 leads to a considerable diagnostic speedup (up to a factor of 8) and that the speedup increases

	original			reduced		
id	CDA* [s]	Alg. 1 [s]	speedup [%]	CDA* [s]	Alg. 1 [s]	speedup [%]
c432	3.2	1.2	267	0.1	0.1	100
c499	3.7	1.3	285	0.2	0.1	200
c880	12.2	10.1	121	1.4	0.2	700
c1355	25.8	17	152	3	0.6	500
c1908	49.1	15.3	321	5.2	4	130
c2670	191	43.1	443	15.9	10.4	153
c3540	514.7	148.6	346	28.1	18.8	149
c5315	712.9	94.8	752	81.2	72.3	112
c6288	2 240	412.8	543	1815.1	812.1	224
c7552	915.1	212.1	431	93.5	41.1	227

Table 6: Speed-up of CDA*due to model relaxation

with the model size. The speedup is better observed with the non-reduced models. We expect even better speedup for diagnosis instances leading to a higher number of conflicts (and respectively higher MC diagnosis cardinality).

6 Conclusions

We have provided a classification of strong-fault diagnostics models, and have shown some computational properties of key sub-classes of such models. In particular, by decomposing a strong-fault model into disjoint strong and weak sub-models, we have shown a relaxation algorithm which improves the efficiency of conflict-based inference algorithms by computing first the diagnoses for the weak sub-model, $\Omega(\mathrm{SD}_w, \alpha)$, and then removing any diagnoses in $\Omega(\mathrm{SD}_w, \alpha)$ inconsistent with the strong sub-model, which can be performed in low-polynomial or better time for several classes of propositional strong-fault models.

Our method is complementary to algorithms (such as conflict-based algorithms), which, on average, work faster with $\mathbf{WFM}$ models. Using ISCAS85 circuits, we have empirically demonstrated that: (1) a large portion of the $\mathbf{WFM}$ models are also diagnoses in the corresponding $\mathbf{SFM}$ models; (2) the average LTMS conflict size decreases; and (3) the diagnosis computation time in CDA*decreases by up to a factor of 8. Our method gives best results with the "difficult" c6288 circuit, and we conjecture that its speedups will increase with the complexity of diagnostic inference.

In the future we plan to extend our methods beyond propositional logic. One straightforward extension of our strong-fault model classification is to many-valued logic, where components may have more than one nominal state; similar relaxation (or strengthening) techniques are also applicable to temporal logic.

References

[Brglez and Fujiwara, 1985] Franc Brglez and Hideo Fujiwara. A neutral netlist of 10 combinational benchmark circuits and a target translator in Fortran. In *Proc. ISCAS'85*, pages 695–698, 1985.

[Bylander *et al.*, 1991] Tom Bylander, Dean Allemang, Michael Tanner, and John Josephson. The computational complexity of abduction. *Artificial Intelligence*, 49:25–60, 1991.

[Chittaro and Ranon, 2004] Luca Chittaro and Roberto Ranon. Hierarchical model-based diagnosis based on structural abstraction. *Artificial Intelligence*, 155(1-2):147–182, 2004.

[Console and Torasso, 1991] Luca Console and Pietro Torasso. A spectrum of logical definitions of model-based diagnosis 1. *Computational Intelligence*, 7(3):133–141, 1991.

[de Kleer and Williams, 1987] Johan de Kleer and Brian Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32(1):97–130, 1987.

[de Kleer *et al.*, 1992] Johan de Kleer, Alan Mackworth, and Raymond Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56(2-3):197–222, 1992.

[de Kleer, 2008] Johan de Kleer. An improved approach for generating Max-Fault Min-Cardinality diagnoses. In *Proc. DX'08*, pages 247–252, 2008.

[del Val, 2000] Alvaro del Val. The complexity of restricted consequence finding and abduction. In *Proc. AAAI'00*, pages 337–342, 2000.

[Forbus and de Kleer, 1993] Kenneth Forbus and Johan de Kleer. *Building Problem Solvers*. MIT Press, 1993.

[Friedrich *et al.*, 1990] Gerhard Friedrich, Georg Gottlob, and Wolfgang Nejdl. Physical impossibility instead of fault models. In *Proc. AAAI*, 1990.

[Maier and Sachenbacher, 2008] Paul Maier and Martin Sachenbacher. Constraint optimization and abstraction for embedded intelligent systems. In *Proc. CPAIOR'08*, pages 338–342, 2008.

[Mozetič and Holzbaur, 1994] Igor Mozetič and Christian Holzbaur. Controlling the complexity in model-based diagnosis. *Annals of Mathematics and Artificial Intelligence*, 11(1):297–314, 1994.

[Nordh and Zanuttini, 2008] Gustav Nordh and Bruno Zanuttini. What makes propositional abduction tractable. *Artificial Intelligence*, 172(10):1245–1284, 2008.

[Reiter, 1987] Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, 1987.

[Selman and Kautz, 1996] Bart Selman and Henry Kautz. Knowledge compilation and theory approximation. *Journal of the ACM*, 43(2):193–224, 1996.

[Siddiqi and Huang, 2007] Sajjad Siddiqi and Jinbo Huang. Hierarchical diagnosis of multiple faults. In *Proc. IJCAI'07*, pages 581–586, 2007.

[Struss and Dressler, 1989] Peter Struss and Oskar Dressler. Physical negation: Integrating fault models into the general diagnosis engine. In *Proc. IJCAI'89*, pages 1318–1323, 1989.

[Williams and Ragno, 2007] Brian Williams and Robert Ragno. Conflict-directed A* and its role in model-based embedded systems. *Journal of Discrete Applied Mathematics*, 155(12):1562–1595, 2007.