The Polybox Example using the Framework of Structured Hypothesis Tests

Report: LiTH-ISY-R-2335

Mattias Nyberg
Department of Electrical Engineering, Linköping University
S-581 83 Linköping, Sweden.
e-mail: matny@isy.liu.se

March 15, 2001

Abstract

The POLYBOX example is a standard example within the AI-field of model-based diagnosis research. Here, this example is discussed in the perspective of *structured hypothesis tests* (SHT). Even though the SHT framework was primarily developed for handling systems with noise, it has here been shown that it can perform very well in also noise-free systems. In the POLYBOX example, it manage to always give a complete and logically sound diagnosis statement, i.e. a complete and correct list of the possible fault modes. On the contrary, the established FDI framework (i.e. *structured residuals*) only manage to give a subset of the possible fault modes.

1 Introduction

The POLYBOX example is a standard example within the AI-field of model-based diagnosis research, e.g. see [1]. Here, this example is discussed in the perspective of *structured hypothesis tests* (SHT). A comparison with the established FDI framework *structured residuals* will at several places be done. Structural residuals is the most common framework for fault diagnosis within the FDI community. From now on, when we write write FDI framework, we really mean *structured residuals*.

The system consists of five components: three multipliers M1, M2, and M3, and two adders A1 and A2. The components are connected according to Figure 1. The observations consists of one sample of all inputs a, b, c, d, and e, and the outputs f and g, i.e. a vector of 6 elements.

When using SHT, the first step is to decide the set of *fault modes* that we want to consider. A fault mode is basically a certain type of fault. There are *single fault modes*, where only one component is faulty, or *multiple fault modes*, where more than one component is faulty. Each fault mode is associated with a model of the system, assumed to be valid when the corresponding fault mode

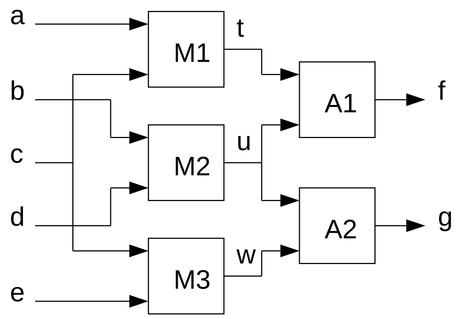

Figure 1: The Polybox example.

is present in the system. Below, first the single fault modes for the POLYBOX example are defined, and then also the multiple fault modes are defined.

2 Single Fault Modes

For each fault mode, we use an abbreviation. The complete list of single fault modes is:

NF	no-fault
A1	arbitrary fault in component A1
A2	arbitrary fault in component A2
M1	arbitrary fault in component M1
M2	arbitrary fault in component M2
M3	arbitrary fault in component M3

Note that also the fault-free case NF is a fault mode.

2.1 Models for Single Fault Modes

NF:
In the fault free case, we have the NF fault mode present, and the model describing the system is then

$$f = ac + bd \tag{1}$$
$$g = bd + ce \tag{2}$$

A1:
When the fault mode A1 is present, the system will behave according to

$$f = ac + bd + \mathbf{f_f} \tag{3}$$
$$g = bd + ce \tag{4}$$

The constant $\mathbf{f_f}$ is an unknown constant and we assume here that $\mathbf{f_f} \neq 0$. That is, with this fault modeling principle, we chose to use the *single fault exoneration assumption* (made in the FDI framework). However, as we will see below, in

Section 2.2, the single fault exoneration assumption can be avoided by using another fault modeling principle. Note here that we have made no assumptions about how the fault in component A1 occurs, only that the fault will cause the first equation of the fault free model to not be valid anymore.

A2:

$$f = ac + bd \tag{5}$$
$$g = bd + ce + \mathbf{f_g} \tag{6}$$

M1:

$$f = ac + \mathbf{f_t} + bd \tag{7}$$
$$g = bd + ce \tag{8}$$

M2:

$$f = ac + bd + \mathbf{f_u} \tag{9}$$
$$g = bd + \mathbf{f_u} + ce \tag{10}$$

M3:

$$f = ac + bd \tag{11}$$
$$g = bd + ce + \mathbf{f_w} \tag{12}$$

Now to fit this into the framework of hypothesis testing, we view this as only one model $\mathcal{M}(\theta)$ with θ as a free parameter. This means that $\theta = [f_f\ f_g\ f_t\ f_u\ f_w]$, i.e. θ is a constant vector with 5 scalar elements. Depending on the specific fault, θ then takes different values. Further, the set of all possible θ-values is $\Theta = \mathcal{R}^5$.

2.2 Alternative Definitions of Fault Modes

In the above definitions of the fault modes, it is implicitly assumed that a single fault always affect the component, i.e. the single fault exoneration assumption. However, with another modeling strategy, it is possible to use SHT without the single fault exoneration assumption. Consider for example fault mode A1.

When the fault mode A1 is present, the system will behave according to

$$f = ac + bd + \mathbf{B_f} f_f \tag{13}$$
$$g = bd + ce \tag{14}$$

where B_f is now a binary variable and f_f an unknown arbitrary constant. When a fault in component $A1$ is present, $B_f = 1$, and otherwise $B_f = 0$. There are no constraint on the constant f_f. It is for example possible that $f_f = 0$, which means that the fault is "invisible". Thus, this is a way to relax the single fault exoneration assumption, if this is desirable.

In other words, when using the SHT framework, it is possible to chose whether one wants to make the exoneration assumption or not. As has been demonstrated here, this choice is done by using different fault modeling principles. However, in this report, the modeling principle described in Section 2.1 will be used and thus, the single fault exoneration assumption is used.

3 Multiple Fault Modes

Only multiple fault modes with two faulty components are considered. The multiple fault modes are abbreviated as for example A1&A2 for the fault mode with both the A1 and the A2 components faulty.

The models for the multiple fault modes are:

A1&A2:

$$f = ac + bd + \mathbf{f_f} \tag{15}$$
$$g = bd + ce + \mathbf{f_g} \tag{16}$$

A1&M1:

$$f = ac + \mathbf{f_t} + bd + \mathbf{f_f} \tag{17}$$
$$g = bd + ce \tag{18}$$

A1&M2:

$$f = ac + bd + \mathbf{f_u} + \mathbf{f_f} \tag{19}$$
$$g = bd + \mathbf{f_u} + ce \tag{20}$$

A1&M3:

$$f = ac + bd + \mathbf{f_f} \tag{21}$$
$$g = bd + ce + \mathbf{f_w} \tag{22}$$

A2&M1:

$$f = ac + bd + \mathbf{f_t} \tag{23}$$
$$g = bd + ce + \mathbf{f_g} \tag{24}$$

A2&M2:

$$f = ac + bd + \mathbf{f_u} \tag{25}$$
$$g = bd + \mathbf{f_u} + ce + \mathbf{f_g} \tag{26}$$

A2&M3:

$$f = ac + bd \tag{27}$$
$$g = bd + ce + \mathbf{f_w} + \mathbf{f_g} \tag{28}$$

M1&M2:

$$f = ac + \mathbf{f_t} + bd + \mathbf{f_u} \tag{29}$$
$$g = bd + \mathbf{f_u} + ce \tag{30}$$

M1&M3:

$$f = ac + \mathbf{f_t} + bd \tag{31}$$
$$g = bd + ce + \mathbf{f_w} \tag{32}$$

M2&M3:

$$f = ac + bd + \mathbf{f_u} \tag{33}$$
$$g = bd + \mathbf{f_u} + ce + \mathbf{f_w} \tag{34}$$

The set of all fault modes considered is:

$$\Omega = \{NF, A1, A2, M1, M2, M3, A1\&A2, A1\&M1, A1\&M2, A1\&M3, A2\&M1, A2\&M2, A2\&M3, M1\&M2, M1\&M3, M2\&M3\} \tag{35}$$

4 Basic Idea of Structured Hypothesis Tests

With the fault modes defined above, the diagnosis problem can be stated as follows:

> Given measured data, which different models, defined by the different fault modes, can explain the measured data?

This means, that the task of diagnosis is to investigate which of the different models that match the measured data. Because of model errors and measurement noise, a formalized decision procedure for this is naturally based on hypothesis testing. To construct the diagnosis system by combining a set of hypothesis tests is the basic idea of *structured hypothesis tests*. The hypothesis tests used are "binary", i.e. the task is to test one null hypothesis against one alternative hypothesis, see e.g. [2, 3].

Let F_p denote the present fault mode. To describe the k:th hypothesis test, introduce the set M_k. The null hypothesis and the alternative hypothesis can then be written

$H_k^0 : F_p \in M_k$ "some fault mode in M_k can explain meas. data"

$H_k^1 : F_p \in M_k^C$ "no fault mode in M_k can explain meas. data"

The convention used here and also commonly used in hypothesis testing literature, is that when H_k^0 is rejected, we *assume* that H_k^1 is true. This implies that the present fault mode can not belong to M_k, i.e. it must belong to M_k^C. Further, when H_k^0 is *not* rejected, we will not assume anything[1]. In this way, each hypothesis test contributes with a piece of information regarding which fault modes that can be matched to the data or not. This piece of information, i.e. the decision reached by each hypothesis test, is called a *subdiagnosis statement.*

How a set of different hypothesis tests are used to diagnose and isolate faults is illustrated by the following example.

Example 1

Assume that the diagnosis system contains the following set of three hypothesis tests:

$$\begin{aligned}
&H_1^0 : F_p \in M_1 = \{NF, F_1\} \quad H_1^1 : F_p \in M_1^C = \{F_2, F_3\} \\
&H_2^0 : F_p \in M_2 = \{NF, F_2\} \quad H_2^1 : F_p \in M_2^C = \{F_1, F_3\} \\
&H_3^0 : F_p \in M_3 = \{NF, F_3\} \quad H_3^1 : F_p \in M_3^C = \{F_1, F_2\}
\end{aligned}$$

[1] Sometimes it is possible to assume something also when H_k^0 is *not* rejected, see [4].

Then if only H_1^0 is rejected, we can draw the conclusion that $F_p \in M_1^C = \{F_2, F_3\}$, i.e. the present fault mode is either F_2 or F_3. If both H_1^0 and H_2^0 are rejected, we can draw the conclusion that $F_p \in M_1^C \cap M_2^C = \{F_2, F_3\} \cap \{F_1, F_3\} = \{F_3\}$, i.e. the present fault mode is F_3.

The decision reached by the whole diagnosis system is called the *diagnosis statement*. From the example above, it is clear that to find the diagnosis statement from the subdiagnosis statements, a simple intersection operation can be used.

When developing the actual hypothesis tests, we first need to choose the set of hypotheses to test. This choice can however not be done arbitrarily. One has to take into account special relations between the fault modes. This is shortly described next.

5 Relations Between the Fault Modes

By studying the models for the different fault modes we can realize that some fault modes can not be isolated from each other, since their corresponding models are in principle equivalent. This kind of relations between fault modes can be formalized by studying *observation sets* defined as follows:

Definition 1 (Observation Set) *The* observation set *of a fault mode F_i is denoted $\mathcal{O}_{F_i}$ and defined by*

$$\mathcal{O}_{F_i} = \{x \mid x \in \mathcal{M}(\theta) \ \wedge \ \theta \in \Theta_{Fi}\}$$

where the model $\mathcal{M}(\theta)$ is the set of possible observations x for a fixed θ, and Θ_{Fi} is the set of θ-values corresponding to fault mode Fi.

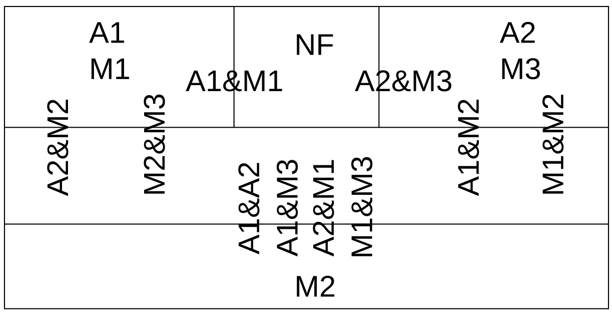

Figure 2: The observation sets and their relations in the polybox example.

The relations between the observation sets for the polybox example is illustrated in the Venn diagram in Figure 2. From the Venn diagram, we can see that for example the following relations hold:

$$\mathcal{O}_{A1} = \mathcal{O}_{M1} \tag{36}$$
$$\mathcal{O}_{NF} \subseteq \mathcal{O}_{A1\&M1} \tag{37}$$
$$\mathcal{O}_{A1} \subseteq \mathcal{O}_{M2\&M3} \tag{38}$$
$$\mathcal{O}_{A1} \cap \mathcal{O}_{A2} = \emptyset \tag{39}$$

and more...

For example the relation $\mathcal{O}_{A1} = \mathcal{O}_{M1}$ tells us that A1 can be obtained as a special case of M1, namely when $\mathbf{f_t} = \mathbf{f_f}$.

Next we study the closures of the observation sets. The motivation for this is that if continuous models are considered, then it is impossible to distinguish between the closure of a set and the set itself. Let the notation $A \subseteq^* B$ denote the relation $A \subseteq \bar{B}$, where $\bar{B}$ is the closure of the set B. Then, for example the following relations will hold:

$$\mathcal{O}_{NF} \subseteq^* \mathcal{O}_{A1} \tag{40}$$

$$\mathcal{O}_{NF} \subseteq^* \mathcal{O}_{A1\&A2} \tag{41}$$

$$\mathcal{O}_{M1\&M2} \subseteq^* \mathcal{O}_{M1\&M3} \tag{42}$$

$$\mathcal{O}_{A1\&M1} \subseteq^* \mathcal{O}_{M1} \tag{43}$$

$$\text{and more...}$$

For example the first relation $\mathcal{O}_{NF} \subseteq^* \mathcal{O}_{A1}$ tells us that the fault mode NF can (almost) be obtained as a special case of the fault mode A1, namely when $\mathbf{f_f}$ is very small.

By noting that $\subseteq$ implies $\subseteq^*$, and only concentrating on these latter relations, we get an equivalence relation defined by the following sets:

$$\{NF\}_{\subseteq^*} \tag{44a}$$

$$\{A1, M1, A1\&M1\}_{\subseteq^*} \tag{44b}$$

$$\{A2, M3, A2\&M3\}_{\subseteq^*} \tag{44c}$$

$$\{M2\}_{\subseteq^*} \tag{44d}$$

$$\{A1\&A2, A1\&M2, A1\&M3, A2\&M1, A2\&M2, M1\&M2, M1\&M3, M2\&M3\}_{\subseteq^*} \tag{44e}$$

All fault modes in such a set are related via the relation $\subseteq^*$. This means that by only studying the observed variables of the system, and considering continuous models and variables, it is not possible to distinguish (isolate) between these fault modes.

6 Hypothesis Tests

The hypothesis tests are only shortly described here. The hypothesis tests used in SHT are generally described in Sections 3.1 and 3.2 in my PhD thesis.

Based on the relations defined by (44), it is possible to construct four different hypothesis tests:

$$H_0^0 : F_p \in \{NF\} \tag{45}$$

$$H_0^1 : F_p \notin \{NF\} \tag{46}$$

$$H_1^0 : F_p \in \{NF, A1, M1, A1\&M1\} \tag{47}$$

$$H_1^1 : F_p \notin \{NF, A1, M1, A1\&M1\} \tag{48}$$

$$H_2^0 : F_p \in \{NF, M2\} \tag{49}$$

$$H_2^1 : F_p \notin \{NF, M2\} \tag{50}$$

$$H_3^0 : F_p \in \{NF, A2, M3, A2\&M3\} \tag{51}$$

$$H_3^1 : F_p \notin \{NF, A2, M3, A2\&M3\} \tag{52}$$

Each hypothesis test consists of a null hypothesis and an alternative hypothesis. For example, in the first hypothesis test, the null hypothesis H_0^0 is $F_p \in \{NF\}$, meaning that the present fault mode F_p is in the set $\{NF\}$ (i.e. $F_p = NF$). The alternative hypothesis H_0^1 is that the present fault mode F_p is not in the set $\{NF\}$, i.e. $F_p \in (\Omega - \{NF\})$

The first of these hypothesis tests does not help us in the isolation task, but can be useful for detecting the faults. However, here we will assume that we use only the last three hypothesis tests, namely number 1,2, and 3.

If we want to construct the hypothesis tests, we have to define a test quantity (also called test statistic) and a threshold for each hypothesis test. However, for the discussion here, this is not necessary. It is in fact only important to have knowledge about how the different fault modes are defined, in terms of the models from Sections 2 and 3, and how they are related.

7 Incidence/Decision Structure

Related to the decision function of the hypothesis tests, we can set up the *Incidence/Decision Structure* (see Section 3.4 in my PhD thesis). This is closely related to the fault signature matrix used in the FDI community. However, we here use a distinction between the incidence structure, describing how faults ideally affect the test quantities, and the decision structure, describing how the diagnosis statement is derived using the individual hypothesis tests.

The main difference to the standard fault signature matrix, is that the incidence/decision structure includes X:s, which makes it possible to make correct conclusions in a noisy environment. Other reasons to include X:s are that faults can compensate each other (discussed in [5]), and that for example nonlinearities can make faults "invisible" in some operating conditions. In the paper [5], a fourth reason is discussed, namely that some faults are possibly not affecting the system at all (the single fault exoneration assumption). This reason is however not valid when using SHT, since this is taken care of by choosing different fault-modeling principles (as discussed in Section 2.2).

When studying the incidence/decision structures, we assume that we use test quantities that are "well-designed". That is, in the noise-free case we assume that the test quantity is zero if the measured data can be explained by the null hypothesis and non-zero otherwise. Then the incidence structure for each hypothesis test and thereby also the sets M_k are determined from the relations between the fault modes.

Assume we want to design a hypothesis test with the *desired* null-hypothesis $H_k^0 : F_p \in M_k = \{Fi_1, Fi_2, \ldots, Fi_n\}$. This may be possible, but depending on the relations between the fault modes, it is sometimes necessary to add some fault modes to the set M_k.

The following theorem tells us the relation between the fault mode relations and the incidence structure. It is here assumed that the desired null hypotheses only contain one fault modes. However the extension to more complex null hypotheses is trivial.

Theorem 1 ([6]) *Given a hypothesis test with a desired null-hypothesis* H_k^0 : $F_p \in M'_k$, *and a well-designed test quantity, the actual set* M_k *and the incidence structure are uniquely determined by the knowledge of the relations* $\mathcal{O}_{F1} \subseteq \mathcal{O}_{F2}$ *and* $\mathcal{O}_{F1} \cap \mathcal{O}_{F2} = \emptyset$ *as follows:*
The entry in the column corresponding to fault mode Fj *is*

$$
\begin{aligned}
&0 \text{ if } \mathcal{O}_{Fj} \subseteq \bigcup_{Fi \in M'_k} \mathcal{O}_{Fi} \\
&1 \text{ if } \mathcal{O}_{Fj} \cap \bigcup_{Fi \in M'_k} \mathcal{O}_{Fi} = \emptyset \\
&X \text{ if } \mathcal{O}_{Fj} \cap \bigcup_{Fi \in M'_k} \mathcal{O}_{Fi} \neq \emptyset \text{ and } \mathcal{O}_{Fj} \not\subseteq \bigcup_{Fi \in M'_k} \mathcal{O}_{Fi}
\end{aligned}
$$

Now first assume that the polybox-system operates in a noise-free environment. Then from the relations in the Venn-diagram in Figure 2 and Theorem 1, we get the following incidence structure:

	NF	A1	A2	M1	M2	M3	A1&A2	A1&M1	A1&M2	A1&M3	A2&M1	A2&M2	A2&M3	M1&M2	M1&M3	M2&M3
δ_1	0	0	1	0	1	1	1	0	1	1	1	X	X	1	1	X
δ_2	0	1	1	1	0	1	X	X	1	X	X	1	X	1	X	1
δ_3	0	1	0	1	1	0	1	X	X	1	1	1	0	X	1	1

The X:s in this table origins from the case when faults compensates each other.

In a noisy environment, we have, because of reasons explained in [6], have to replace all the 1:s with X:s and we thus obtain the decision structure

	NF	A1	A2	M1	M2	M3	A1&A2	A1&M1	A1&M2	A1&M3	A2&M1	A2&M2	A2&M3	M1&M2	M1&M3	M2&M3
δ_1	0	0	X	0	X	X	X	0	X	X	X	X	X	X	X	X
δ_2	0	X	X	X	0	X	X	X	X	X	X	X	X	X	X	X
δ_3	0	X	0	X	X	0	X	X	X	X	X	X	0	X	X	X

8 Diagnosis Examples

This section will discuss the different diagnosis statements obtained from SHT compared to the FDI framework (i.e. *structured residuals*). If we assume a noise-free environment, we can use the SHT framework and build our diagnosis statements based on the first incidence structure in Section 7 (containing 0:s, 1:s, and X:s). For different combinations of rejected and accepted hypotheses, the diagnosis statement becomes as shown in Table 1. For example, the first row in this table tells us that when no null hypotheses are rejected (i.e. all three are accepted), the diagnosis statement will be {NF, A1&M1, A2&M3}, meaning that all these three fault modes can explain what we observe. The cases marked with a "*" are those included in [5]. The diagnoses within parentheses are *non-minimal*, as defined in e.g. [5].

Now assume a noisy environment. Then the SHT framework can be used with the last decision structure from Section 7. The diagnosis statement becomes as shown in Table 2.

In the FDI framework, only 1:s are used. The diagnosis statement then becomes as shown in Table 3.

δ_1	δ_2	δ_3	diagnosis statement
0	0	0	NF, A1&M1, A2&M3
0	0	1	can not occur
0	1	0	can not occur
*0	1	1	A1, M1, (A1&M1), A2&M2, M2&M3
1	0	0	can not occur
*1	0	1	M2, A1&A2, A1&M3, A2&M1, M1&M3
1	1	0	A2, M3, A1&M2, (A2&M3), M1&M2
*1	1	1	A1&A2, A1&M2, A1&M3, A2&M1, A2&M2, M1&M2, M1&M3, M2&M3

Table 1: The diagnosis statement obtained with SHT using a decision structure with 0:s, 1:s, and X:s.

δ_1	δ_2	δ_3	diagnosis statement
0	0	0	Ω (i.e. all possible fault modes)
0	0	1	(can only occur when noise) A1, M1, M2, (A1&A2), (A1&M1), (A1&M2), (A1&M3), (A2&M1), (A2&M2), (M1&M2), (M1&M3), (M2&M3)
0	1	0	(can only occur when noise) A1, A2, M1, M3, (A1&A2), (A1&M1), (A1&M2), (A1&M3), (A2&M1), (A2&M2), (A2&M3), (M1&M2), (M1&M3), (M2&M3)
*0	1	1	A1, M1, (A1&A2), (A1&M1), (A1&M2), (A1&M3), (A2&M1), A2&M2, (M1&M2), (M1&M3), M2&M3
1	0	0	(can only occur when noise) A2, M2, M3, (A1&A2), (A1&M2), (A1&M3), (A2&M1), (A2&M2), (A2&M3), (M1&M2), (M1&M3), (M2&M3)
*1	0	1	M2, A1&A2, (A1&M2), A1&M3, A2&M1, (A2&M2), (M1&M2), M1&M3, (M2&M3)
1	1	0	A2, M3, (A1&A2), A1&M2, (A1&M3), (A2&M1), (A2&M2), (A2&M3), M1&M2, (M1&M3), (M2&M3)
*1	1	1	A1&A2, A1&M2, A1&M3, A2&M1, A2&M2, M1&M2, M1&M3, M2&M3

Table 2: The diagnosis statement obtained using SHT in a noisy environment, i.e. a decision structure with only 0:s and X:s.

δ_1	δ_2	δ_3	diagnosis statement
0	0	0	NF
0	0	1	(can only occur when noise)
0	1	0	(can only occur when noise)
*0	1	1	A1, M1, (A1&M1)
1	0	0	(can only occur when noise)
*1	0	1	M2
1	1	0	A2, M3, (A2&M3)
*1	1	1	A1&A2, A1&M2, A1&M3, A2&M1, A2&M2, M1&M2, M1&M3, M2&M3

Table 3: The diagnosis statement using the FDI framework, i.e. a decision structure with only 0:s and 1:s.

8.1 Comparison

We will now discuss the diagnosis statements given in Tables 1, 2, and 3; first the noise-free case, and then the noisy case. The main goal of this investigation is to find out which of the results that represents correct diagnoses.

8.1.1 Noise-free case

$[0, 0, 0]$:
For this case, i.e. when no null hypotheses are rejected, the diagnosis statement is different in all the three cases. Since we have a noise-free environment, the right diagnosis statement is {NF, A1&M1, A2&M3}. These are exactly the fault modes that can result in the response $[0, 0, 0]$. The NF is obvious, and the multiple fault modes because it is possible that faults compensate each other. In other words, the SHT framework manage to give the correct result. The FDI framework (with only 0:s and 1:s) has the no-compensation assumption and as a result, only NF is contained in the diagnosis statement. If a decision structure with only 0:s and X:s would be used, as in Table 2, the diagnosis statement would be the set of all fault modes, which is not correct.

$[0, 0, 1]$, $[0, 1, 0]$, and $[1, 0, 0]$:
These responses can not occur in a noise-free environment.

$[0, 1, 1]$, $[1, 0, 1]$, and $[1, 1, 0]$:
For all these three responses, the SHT framework manage to give the correct result. The FDI framework doesn't manage to give the complete list of possible fault modes for any of the three responses. If a decision structure with only 0:s and X:s would be used, the list of possible fault modes would be to large (see Table 2). However the minimal diagnoses are the same.

$[1, 1, 1]$:
The correct diagnosis statement is obtained from both the SHT and the FDI framework, and also when all 1:s are replaced with X:s.

8.1.2 Noisy case

When there is noise present, the SHT framework uses a decision structure with only 0:s and X:s.

$[0, 0, 0]$:
For this response, the diagnosis statement using the SHT framework, is that all possible fault modes are possible explanations for what we observe, i.e. the set Ω. The reason is that the faults could be small and then in a noisy environment, it is impossible to distinguish these from the NF fault mode. The FDI framework would still give only NF which in a noisy environment is the same as to implicitly assume that all faults are large (and as before, don't compensate each other).

$[0, 0, 1]$, $[0, 1, 0]$, $[0, 1, 1]$, $[1, 0, 0]$, $[1, 0, 1]$, $[1, 1, 0]$, and $[1, 1, 0]$:
In all these cases, the SHT framework give the correct diagnosis statement, when we assume that faults can be small. The FDI framework gives only a subset of the possible fault modes, because it implicitly assumes that all faults are large and don't compensate each other. Also interesting is that the FDI framework gives the incorrect result of an empty set for the cases $[0, 1, 1]$, $[1, 0, 1]$, and $[1, 1, 0]$.

$[1, 1, 1]$:
The correct diagnosis statement is obtained from both the SHT and the FDI

framework.

9 Conclusions

In this report, I have compared the performance of the SHT and the FDI framework using the POLYBOX example. However, it should be noted that in my point of view, the FDI framework is generally not a well defined framework. Only at a few places in the literature (e.g. [5]) there is a formal definition of it, and it is not clear if all people agree on the same definition. However, by adopting the definition in [5] of the FDI framework, which is in principle the same definition as in my PhD thesis, it has been possible to make this comparison with the SHT framework.

Even though the SHT framework was primarily developed for handling systems with noise, it has here been shown that it can perform very well in also noise-free systems. In the POLYBOX example, it managed to always give a complete and logically sound diagnosis statement, i.e. a complete and correct list of the possible fault modes. On the contrary, the FDI framework only managed to give a subset of the possible fault modes.

As described in [5], the problems with the FDI framework are the single-fault exoneration and the no-compensation assumptions. In SHT we have a choice if we should make the single-fault exoneration assumption or not. This is because we have the powerful tool of fault modeling. Another reason to work with fault models is the increased possibility to isolate different faults. For example, by knowing that two different faults are acting in a different way, we can distinguish between the two even though they are acting on the same component.

The no-compensation assumption is not made in the SHT framework. In a noise free environment, it can be argued that this is not so important since the faults that can compensate are "singular", and thus the "probability" for compensation is in principle zero. However in a noisy environment it can happen that two faults approximately compensates each other, and because of the noise, this can not be distinguished from the no-fault case.

In the POLYBOX example, the system is static and the faults very simple. However, as has been shown in for example my works with the automotive engine examples, the SHT framework also handles dynamic systems with also more complex fault models.

Something that may not have been obvious in this report is why it is an advantage to view the diagnosis system as a set of hypothesis tests. The main reason for this is because when it comes to the design and analysis of the individual hypothesis tests, the established area of hypothesis testing provides us with a theoretical framework for making decision in a noisy environment. Also, a lot of theory exists which can be used for design and evaluation of the hypothesis tests. An example of the latter can be seen in [7] or my PhD thesis (Section 4.6 and Chapter 6).

Finally it should be mentioned that I also made a comparison between the SHT and the FDI frameworks in my PhD thesis (see Section 3.5). There I came to the conclusion that, in a noisy environment, the 1:s used in the fault-signature matrix should not be used since it implies the assumption that even small faults always result in that the test quantity is above the threshold, something that of course not is true. Normally, the 1:s should therefore be replaced by X:s.

It is interesting to see that the paper [5], came to the same conclusion but with another main argument, namely the exoneration and the no-compensation assumption.

References

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

[2] E. L Lehmann. *Testing Statistical Hypotheses*. Springer Verlag, second edition, 1986.

[3] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury Press, 1990.

[4] Mattias Nyberg. *Model Based Fault Diagnosis: Methods, Theory, and Automotive Engine Applications*. PhD thesis, Linköping University. URL: http://www.fs.isy.liu.se/Publications/, May 1999.

[5] M.O. Cordier, P. Dague, M. Dumas, F. Levy, J. Montmain, M. Staroswiecki, and L. Trave-Massuyes. AI and automatic control approaches of model-based diagnosis: Links and underlying hypothesis. SAFEPROCESS, pages 274–279, Budapest, Hungary, 2000. IFAC.

[6] M. Nyberg. A general framework for model based diagnosis based on statistical hypothesis testing. 12:th International Workshop on Principles of Diagnosis, pages 135–142, Sansicario, Via Lattea, Italy, 2001.

[7] M. Nyberg. Automatic design of diagnosis systems with application to an automotive engine. *Control Engineering Practice*, 7(8):993–1005, 1999.