Diagnosability Analysis based on Component Supported Analytical Redundancy Relations (May 2005)

L. TRAVE-MASSUYES, *Senior Member IEEE*, T. ESCOBET, X. OLIVE

***Abstract*— It is commonly accepted that the requirements for maintenance and diagnosis should be considered at the earliest stages of design. For this reason, methods for analysing the diagnosability of a system and determining which sensors are needed to achieve the desired degree of diagnosability, are highly valued. This paper clarifies the different diagnosability properties of a system and proposes a model based method for:**

1. Assessing the level of discriminability of a system, i.e. given a set of sensors, the number of faults which can be discriminated; and its degree of diagnosability, i.e. the discriminability level related to the total number of anticipated faults.

2. Characterizing and determining the minimal additional sensors that guarantee a specified degree of diagnosability.

The method takes advantage of the concept of component supported analytical redundancy relation, which considers recent results crossing over the FDI and DX communities. It uses a model of the system to analyze in an exhaustive manner the analytical redundancies associated to the availability of sensors and performs from that a full diagnosability assessment.

The method is applied to an industrial smart actuator that was used as benchmark in the DAMADICS European project.

Index Terms— Diagnosability, Sensor placement, Analytical redundancy, Model-based diagnosis, Structural analysis.

I. INTRODUCTION

IT is commonly accepted that diagnosis and maintenance requirements should be accounted for at the very early design stages of a system. For this purpose, methods for analysing properties such as diagnosability and characterising the instrumentation system in terms of the number of sensors and their placement are highly valuable. There is a significant amount of work dealing with diagnosability and sensor placement both in the AI Model Based Diagnosis (DX) community [5], [20], [8] and in the Control Model Based Diagnosis (FDI) community [2], [12], [10], [11].

This paper bridges DX and FDI results to propose a model-based method for:

- Assessing the level of discriminability of a system, i.e. given a set of sensors, which faults can be discriminated? and its degree of diagnosability, i.e. the discriminability level related to the number of faults,
- Characterizing and determining the Minimal Additional Sensor Sets (MASS) that guarantee a specified diagnosability degree.

The main ideas behind the method are to use a model of the behaviour of the system to analyse in an exhaustive manner the analytical redundancies introduced by hypothesized sensors and build an *Hypothetical Fault Signature Matrix* (HFS Matrix). The starting point of the method hypothesizes that all the variables are sensored and subsequent hypotheses proceed to the retraction of sensors. Our working assumption is that, for economic reasons, hardware redundancy in the instrumentation is not permitted, i.e. only one sensor per variable can be hypothesized. This assumption is acceptable, and even necessary, in many application domains.

The method follows a structural analysis approach familiar to the FDI community [3] extended to component supported Analytical Redundancy Relations (ARR) by tracing the ARRs'supports. ARRs'supports were introduced in [6] to bridge to the concept of conflict of the DX community. They are defined as the components, including possibly sensors, whose models are involved in the expression of the redundancy. The method builds on the work of [14] and [15] but handles the different operating regions (behavioral modes) of a system and explores all the model relation causal interpretations, hence resulting in the completeness of the produced set of hypothetical ARRs (H-ARRs) (under the assumption of no singularities leading to *instantiated ARRs* [7]). The supports of the obtained H-ARRs constitute the HFS matrix from which a complete diagnosability assessment can be performed.

The paper is organised as follows. Section II introduces the

Manuscript received May 24, 2005.

This work was supported in part by the European Community, Trial Application Project No: 27548 Tiger SHEBA, and the EC RTN contract RTN-1999-00392 DAMADICS, also by the Research Commission of the Generalitat of Catalunya Grant No. 2001SGR00236 and Spanish CICYT DPI2002-02147.

L. Travé-Massuyès is with LAAS-CNRS, 7 avenue du Colonel-Roche, 31077 Toulouse Cedex 4, France (telephone: +33 5 61 33 63 02, e-mail: louise@laas.fr).

T. Escobet is with Technical University of Catalonia (UPC), 10 Rambla de Sant Nebridi, 08222 Terrassa, Spain (telephone: +34 93 7 39 81 44, e-mail: teresa.escobet@upc.edu).

X. Olive is with LAAS-CNRS, 7 avenue du Colonel-Roche, 31077 Toulouse Cedex 4, France and with ACTIA, 25 Chemin de Pouvourville, 31432 Toulouse, France (telephone: +33 5 61 17 68 32, e-mail: olive@actia.fr).

basic concepts of analytical redundancy and presents the structural approach. Section III introduces a set of definitions to clarify the different diagnosability properties of a system, provides a recursive characterization of the sets of ARRs corresponding to successively included sensor sets, and proposes an operational method to generate ARRs and trace their support. Section IV approaches the diagnosability assessment problem. Section V illustrates the method on the well-know polybox example and applies the approach to an industrial actuator device. Finally, Section VI discusses related work and Section VII brings some conclusions and perspectives.

II. A Structural Approach for Analytical Redundancy

The (normal) behavior model of a system $\Sigma=(E,V)$ can be defined as a set of n relations E, which relate a set of m variables V. In a component-oriented model, these relations, called *primary relations*, are associated to the system's physical components, including the sensors. The set E is partitioned into *behavioral relations* E_{beh} which correspond to the internal components and *observation relations* E_{obs} which correspond to the sensors.

The structure of a model can be represented by a *Structural Matrix* which crosses model relations in rows and model variables in columns, or equivalently by the bipartite graph $G=(E\cup V, A)$ where A is a set of arcs such that $a(i,j)\in A$ iff variable $v_j\in V$ appears in relation $e_i\in E$.

The set of variables V can be partitioned as $V=X\cup O$, where O is the set of observed (measured) variables and X is the set of unknown variables[1]. Then, the structural approach of [3] is based on determining a complete matching $\mathcal{M}$ between E and X in the bipartite graph G. A matching on a graph G is a set of edges of G such that no two of them share a vertex in common. A complete matching between E and X in a bipartite graph $G=(E\cup X\cup O, A)$, or equivalently in $G=(E\cup X, A')$ where A' is a subset of A, is one that saturates all of the vertices in E or X. It corresponds to a selection of line-independent entries, i.e. which are not in the same row or column, in the structural submatrix crossing E and X. If the relation e_i is associated to the variable x_j by $\mathcal{M}$, then e_i can be interpreted as a mechanism for solving for x_j. The *Resolution Process Graph* (RPG) is defined as the oriented graph obtained from G by orienting the edges of A from x_j towards e_i if $a(i,j)\notin\mathcal{M}$ and from e_i towards x_j if $a(i,j)\in\mathcal{M}$. It provides the orientation of calculability (or causal interpretation) associated to $\mathcal{M}$. The determination of $\mathcal{M}$ must account for the possibly restricted causal interpretation of some relations, e.g. a given relation may not be invertible and hence can only be used in a predefined direction. In practice, this is performed by orienting the corresponding edges *a priori*.

[1] If v is a physical quantity of the system, and has an associated sensor providing a reading v_{obs}, then $v\in X$ and $v_{obs}\in O$.

Reference [3] shows that this graph can be used to derive the *Analytical Redundant Relations* (ARR). Given our working assumption that instrumentation is not redundant, i.e. there is only one sensor for the same variable or quantity, ARRs exist if and only if the number of relations *card(E)* is strictly greater than the number of unknown variables *card(X)*. In this case the complete matching is of X into E and ARRs correspond to the relations which are not involved in the complete matching, and consequently are not needed to determine the values of the unknow variables. These "extra-relations" appear as sink nodes of the RPG. ARR of the form $r = 0$ where r is called the *residual* of the ARR, are obtained from the extra-relations by replacing the unknown variables by their formal expression in terms of observable variables, tracing back the analytical paths defined by the RPG.

If *card(E)=card(X)*, then there are no ARRs and the system is said to be *non monitorable* [3][2]. The above is illustrated in Figure 1 with a toy example of five relations e_1 to e_5 and four unknown variables x_1 to x_4.

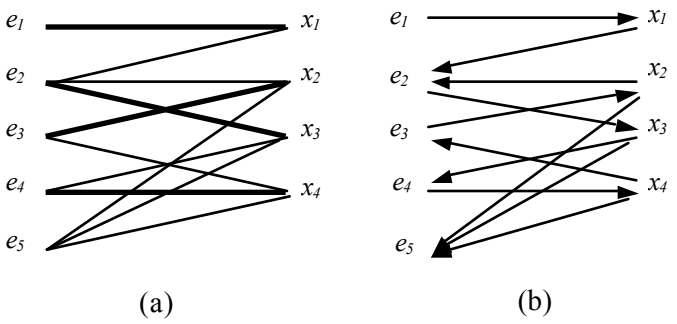

Fig. 1. (a) Bipartite graph $G=(E\cup X, A')$ and complete matching of X into E indicated by bold arcs; (b) RPG indicating that e_5 is a redundant relation.

To summarize, an ARR arises from a causal interpretation of the underlying model primary relations. It only contains observed variables and can hence be evaluated from the observations. If r is evaluated to 0, then the ARR is satisfied, and conversely. The binary evaluation (0 or non 0) of every ARR provides a means to characterize the system at a given time. This pattern is also referred to as the *observed signature* whereas the expected binary pattern for a given fault is called the *fault signature*.

The *support* of an ARR [7] refers to the underlying components whose corresponding primary relations are involved in the ARR and is noted *Supp(ARR)*. In the general case, *Supp(ARR)* is composed of a set of components and a set of sensors and can be spited in a component-support *Comp-Supp(ARR)* and a sensor-support *Sens-Supp(ARR)*[3]. An ARR together with its support *Supp(ARR)* is called a *Supported ARR*. When not ambiguous, ARR is used for short.

[2] The behavioral system model is assumed to be just determinated. The redundancy within the system model originates from the available observations (sensors) only.

[3] Sensors must be distinguished from other components because the diagnosability method stands on hypothesizing sensors, which must hence have a different status.

In the FDI terminology, the fault signature (FS) matrix crosses ARRs in rows and (sets of) faults in columns [6]. Every row provides the support of an ARR and every column the fault signature of a fault. Let us assume that F_j denotes a fault on component C_j, then the interpretation of some entry γ_{ij} being 0 is that component C_j does not belong to $Supp(ARR_i)$, i.e. the occurrence of the fault F_j does not affect ARR_i, meaning that ARR_i is satisfied in the presence of that fault. γ_{ij}=1 means that C_j belongs to $Supp(\mathrm{ARR_i})$, i.e. ARR_i *is expected* to be affected by fault F_j, but it is *not guaranteed* that it will really be (the fault might be non detectable by this ARR) [13].

The *ARR-based exoneration assumption* is generally adopted in the FDI approach, meaning that a fault F_j is assumed to always affect the ARRs whose support include C_j. In this case, γ_{ij} = 1 is interpreted as ARR_i is violated in the presence of fault F_j.

Proposition 1 The ARR-based exoneration assumption implies that F_j is strongly detectable in the sense of the definition provided by [13], i.e. the residual maintains a non zero value after its transition time[4].

Proof of Proposition 1. The ARR-based exoneration assumption implies that under the occurrence of a fault F_j the ARRs whose support include C_j are always violated, which means that the observed signature cannot change value over time, hence proving the proposition. ■

Let us call $\mathcal{M}$*-ARRs* the ARRs that are directly obtained from a given complete matching $\mathcal{M}$. The number of $\mathcal{M}$*-ARRs* is the same for any existing complete matching, as it only depends on the redundancy degree of the observed system [12]. Given the set of $\mathcal{M}$*-ARRs*, also called *primary ARRs*, additional ARRs can be obtained by combining $\mathcal{M}$*-ARRs*, i.e. by using one ARR to substitute for a common variable in another ARR. These additional *combined ARRs* correspond to those that would be obtained from the other complete matchings existing on the system structure.

Definition 1 Given a set $\mathcal{A}$ of ARRs, we define $\mathcal{A}_b$ as a *basis* of $\mathcal{A}$ if all the ARRs of $\mathcal{A}$ can be obtained by combining[5] two or several ARRs of $\mathcal{A}_b$ and this is not true for any $\mathcal{A}_b' \subset \mathcal{A}_b$.

Proposition 2 Given a system $\Sigma=(E,X\cup O)$ and $\mathcal{M}$ a complete matching between E and X, then the set of $\mathcal{M}$*-ARRs* is a basis for the whole set of ARRs of Σ.

Proof of proposition 2. An ARR is a constraint deduced from the system model that contains only observed variables. Using graph theory and the search for a complete matching is a way of implementing unknown variable structural elimination, leading to $\mathcal{M}$*-ARRs* after analytical calculations are performed. Now, any combination of $\mathcal{M}$*-ARRs* is also an ARR, hence the proposition. ■

Cordier *et al.* [7] showed that by construction, combined ARRs are clearly redundant when considered just as equations but that for fault isolation purposes, one has to consider their support *Supp(ARR)*, i.e. one has to consider component-supported ARRs. They proved that a combined ARR, say $\mathrm{ARR_j}$, is a *logical consequence* of a set of ARRs, say $\mathrm{ARR_i}$s, iff $Supp(ARR_j) \supseteq \cup_i Supp(ARR_i)$. ARR_j is then said *logically redundant.*

The methods proposed by the FDI community for generating the ARRs do not take into account the ARR supports, for example the $\mathcal{M}$*-ARRs* may be simply combined in an exhaustive manner. This is even the case for efficient methods as [2] based on searching all the paths within AND-OR graphs.

Conversely, our method for generating ARRs (cf. Section III.C) takes into account that ARRs should be considered together with their component supports and this is just what makes it suitable as input for a diagnosability analysis.

Our method only makes use of structural properties of primary model relations and tracks structural changes along the combinations. In other word, ARRs' are characterized by their structure but their analytical expression is not generated. However, the method handles the *validity conditions* analytical expressions that may be associated to the model relations, characterising specific operating modes of the system, possibly fault modes, or singularities related to the analytical expressions.

Note that the fault signatures of a multiple operating mode system are dynamic, in the sense that the set of ARRs that take place in the fault signature is not the same across time, i.e. when the system is in different operating modes. Our method keeps track of these changes as well.

III. A Structural Approach for Diagnosability

This section first explores the diagnosability properties of a system, leading to the concepts of strong/weak diagnosability, and full/partial diagnosability. These concepts are bridged to the fault detectability properties proposed in the literature [13]. It then provides a recursive characterization of the sets of ARRs for successively included sensor sets.

A. Diagnosability characterization

We consider a system Σ, a set of sensors $S=\{S_i\}$, and a set of faults $F=\{F_i\}$ (single or multiple), and represent it by the triple (Σ, S, F). Let us define by OBS_j the value tuple returned by the sensors at some time point t_j corresponding to the set of observed variables O[6]. OBS is used instead of OBS_j when the time label is not relevant. Let us also define by $\mathrm{OBS_{Fi}}$ the set of all possible tuples consisting of observed variable values (regardless of time) under the fault F_i.

Then we first have the following definition:

Definition 2 *(Diagnosis candidate)*

Given the triple (Σ, S, F), $F_i \in F$ is a diagnosis candidate at time point t_j if and only if $OBS_j \in \mathrm{OBS_{Fi}}$. F_i is a minimal

[4] As opposed to weakly detectable [13] for which there must exist one time point t at which the residual takes a non zero value (but is not required to hold).

[5] "combining" is understood as performing variable substitution.

[6] In the case of a dynamic system, we consider that OBS_j is an observation tuple that leads to a stabilized observed signature. Noise and decision strategies in perturbated environnements are out of the scope of this paper.

diagnosis candidate if $\forall F_k \subset F_i$, F_k is not a diagnosis candidate.

The diagnosability properties of a triple *(Σ, S,F)* arises from the discriminability properties of each pair of individual faults in *F* [20]. Different levels of discriminability can be exhibited, hence different levels of diagnosability.

Definition 3 *(Discriminability)*

1. Two faults F_i and F_j are said to be *strongly discriminable* if and only if for any *OBS*, when F_i is among the diagnosis candidates, F_j never is, and conversely. In other words, $OBS_{Fi} \cap OBS_{Fj} = \emptyset$.
2. Two faults F_i and F_j are said to be *non discriminable* if and only if for any *OBS*, when F_i is among the diagnosis candidates, then F_j also is, and conversely. In other words, $OBS_{Fi} = OBS_{Fj}$.
3. A fault F_i is said to be *weakly discriminable* from a fault F_j if and only if, when F_i is among the diagnosis candidates, there exists at least one OBS^k, such that F_j is not and at least one OBS^l with $OBS^l \neq OBS^k$, such that F_j also is. In other words, $OBS_{Fi} \backslash OBS_{Fj} \neq \emptyset$. If F_i is weakly discriminable from F_j or F_j is weakly discriminable from F_i, the pair of faults (F_i, F_j) is said to be weakly discriminable.

Note that non discriminability and strong discriminability are symmetric relations, whereas weak discriminability is not. However, if F_i is weakly discriminable from F_j or F_j is weakly discriminable from F_i, then F_i and F_j are neither strongly discriminable nor non discriminable.

The situations of Figure 2 illustrate definition 3.

Note that detectability can be defined as discriminability from the no fault mode.

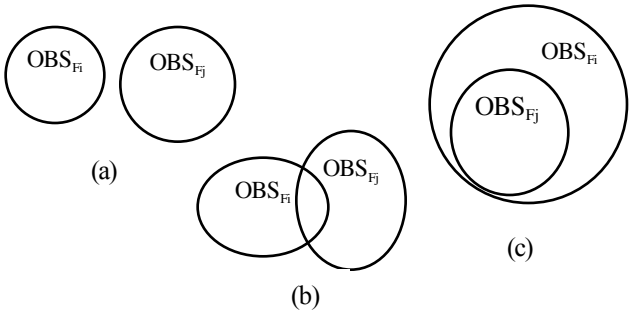

Fig. 2. (a) Two strongly discrimable faults; (b) F_i is weakly discriminable from F_j, and conversely; (c) F_i is weakly discriminable from F_j, F_j is not weakly discriminable from F_i but it is not non discriminable neither.

The definitions of *Strong Diagnosability* and *Weak diagnosability* can now be provided.

Definition 4 *(Strong diagnosability)*. A triple *(Σ, S,F)* is *strongly diagnosable* if and only if for any pair $(F_i, F_j) \in F \times F$ where $F_i \neq F_j$, $OBS_{Fi} \cap OBS_{Fj} = \emptyset$, i.e. any pair of faults is strongly discriminable[7].

Definition 5 *(Weak diagnosability)*. A triple *(Σ,S,F)* is *weakly diagnosable* if and only if for any pair $(F_i, F_j) \in F \times F$ where $F_i \neq F_j$, $OBS_{F1} \backslash OBS_{F2} \neq \emptyset$ or $OBS_{F2} \backslash OBS_{F1} \neq \emptyset$, i.e. any pair of faults is either strongly discriminable or weakly discriminable.

[7] *Strong diagnosability* meets *diagnosability* as defined in [5].

On the other hand, the non-discriminability relation is an equivalence relation which allows us to classify the faults: two faults are in the same *D-class* if and only if they are non discriminable. Note that two faults from different classes may be strongly or weakly discriminable.

Definition 6 *(Discrimination level)*. Given a triple *(Σ, S,F),* its *discrimination level* $\mathcal{D}_S$ is defined as the number of D-classes obtained for the set of sensors *S*.

Then, we can provide the definition of the *Diagnosability Degree.*

Definition 7 *(Diagnosability Degree)*. Given a triple *(Σ, S,F),* its *diagnosability degree* d_S is defined as the quotient of the number of *D-classes* by the number of faults in *F*, i.e. $d_S = \mathcal{D}_S / card(F)$.

Full diagnosability and *Partial diagnosability* can now be defined as follows.

Definition 8 *(Full and partial diagnosability)*. A triple *(Σ, S,F)* is fully diagnosable if and only if its diagnosability degree is equal to 1; it is *partially diagnosable* otherwise.

Proposition 3 The number of *D-classes* of a fully diagnosable system is equal to the number of faults, *card(F)*.

Proof of Proposition 3. Trivial. ■

Proposition 4 Under the ARR-based exoneration assumption and given a triple *(Σ,S,F),* two faults of *F* are either strongly discriminable or non discriminable.

Proof of Proposition 4. This results follows directly from *Proposition 1.*

Corollary 1 Under the ARR-based exoneration assumption a fully diagnosable triple *(Σ, S,F)* is strongly diagnosable.

Proof of Corollary 1. This results follows from Definition 8 and Proposition 4.

The above definitions and results apply to a system and a given set of existing sensors. Let us now rise the question the way around: which sensors to achieve best diagnosability properties? Let us recall that, in the rest of the paper, the working assumption is that hardware redundancy in the instrumentation is not permitted, i.e. only one sensor per variable can be in use.

Definition 9 *(Minimal Additional Sensor Sets)*. Consider a partially diagnosable triple *(Σ,S,F),* a set of potential additional sensors $\mathcal{S}^+$, and define as $F^+ = F \cup \mathcal{S}_f^+$ the updated set of faults where $\mathcal{S}_f^+ \subseteq \mathcal{S}^+$ is the (sub)set of additional sensors that may be faulty (the remaining ones being assumed fully reliable). Then an Additional Sensor Set is defined as a set of sensors $\mathcal{S} \subseteq \mathcal{S}^+$ such that $(\Sigma, S \cup \mathcal{S}, F^+)$ has maximal diagnosability degree, i.e. $d_{S \cup \mathcal{S}} \geq d_{S \cup \mathcal{S}'}$ for all $\mathcal{S}' \subseteq \mathcal{S}^+$ and $\mathcal{S}' \neq \mathcal{S}$. A Minimal Additional Sensor Set is an additional sensor set $\mathcal{S}$ such that $\forall\, \mathcal{S}'' \subset \mathcal{S}$, $\mathcal{S}''$ is not an additional sensor set.

For a partially diagnosable triple *(Σ, S,F)*, the number of D-classes is a monotonically increasing function of the number of sensors chosen in $\mathcal{S}^+$, i.e. $\mathcal{D}_{S \cup \mathcal{S}} \subseteq \mathcal{D}_{S \cup \mathcal{S}'}$ for $\mathcal{S} \subseteq \mathcal{S}'$, but this is not true for the diagnosability degree $d_{S \cup \mathcal{S}} = \mathcal{D}_{S \cup \mathcal{S}} / card(F^+)$. Indeed, $card(F^+)$ depends itself of $\mathcal{S}$ if sensor faults are considered. The information provided by the diagnosability

degree is complementary to the information given by the discrimination level alone. Indeed, it evaluates the number of D-classes in relation with the number of faults. Now we have the following result.

Proposition 5 Given a partially diagnosable triple (Σ, S, F) and a set of potential additional sensors $\mathcal{S}^+$, consider a set of sensors $S \subseteq \mathcal{S}^+$.

1. When $card(S)$ increases, an increase of $d_{S\cup S}$ implies an increase of $\mathcal{D}_{S\cup S}$.
2. If the set of faults F^+ is limited to single faults and if the diagnosability degree $d_{S\cup S}$ is maximal, then the discrimination level $\mathcal{D}_{S\cup S}$ is also maximal.

Proof of Proposition 5. Consider a set of additional sensors $S \subseteq \mathcal{S}^+$ and assume that S is increased by one sensor resulting in S_{new}. Then we have four possible cases:

Case1: $card(F^+_{new})=card(F^+)$ and $\mathcal{D}_{S\cup S_{new}}=\mathcal{D}_{S\cup S}$, which implies that $d_{S\cup S_{new}}=d_{S\cup S}$.

Case2: $card(F^+_{new})=card(F^+)$ and $\mathcal{D}_{S\cup S_{new}}>\mathcal{D}_{S\cup S}$, which implies that $d_{S\cup S_{new}}>d_{S\cup S}$.

Case3: $card(F^+_{new})>card(F^+)$ and $\mathcal{D}_{S\cup S_{new}}>\mathcal{D}_{S\cup S}$, and we get $d_{S\cup S_{new}}= [\mathcal{D}_{S\cup S}+a]/[card(F^+)+b]$ where a and b are positive integers, which can be put in the form: $d_{S\cup S_{new}}=[\mathcal{D}_{S\cup S}/card(F^+)]+\Delta$,
where $\Delta=[a - b\times\mathcal{D}_{S\cup S}/card(F^+)]/[card(F^+)+b)]$,
which leaves the sign of $[d_{S\cup S_{new}}-d_{S\cup S}]=\Delta$ undetermined, this latter being the same as the sign of $[a/b]- [\mathcal{D}_{S\cup S}/card(F^+)]$.

Case4: $card(F^+_{new})>card(F^+)$ and $\mathcal{D}_{S\cup S_{new}}=\mathcal{D}_{S\cup S}$, which implies that $d_{S\cup S_{new}}<d_{S\cup S}$.

Hence, $d_{S\cup S_{new}}>d_{S\cup S}$ (which is only true in cases 2 and possibly case 3) implies $\mathcal{D}_{S\cup S_{new}}>\mathcal{D}_{S\cup S}$, proving 1.

Under the single fault assumption, Case 1, 2 and 4 remain the same but Case 3 must be instanciated by a=1 and b=1. Since $\mathcal{D}_{S\cup S_{new}}>\mathcal{D}_{S\cup S}$, (Σ, S, F^+) is partially diagnosable and we have $\mathcal{D}_{S\cup S}/card(F^+)\leq 1$. This implies that the sign of $[a/b]-[\mathcal{D}_{S\cup S}/card(F^+)]$ is always positive, hence $d_{S\cup S_{new}}>d_{S\cup S}$. Therefore, from Case 2 and Case 3, $\mathcal{D}_{S\cup S_{new}}>\mathcal{D}_{S\cup S}$ implies $d_{S\cup S_{new}}>d_{S\cup S}$ (which was not true in the general case). We hence have $\mathcal{D}_{S\cup S}$ increases iff $d_{S\cup S}$ increases, proving 2. ■

Remark: Note that if S is such that $d_{S\cup S}$ is maximal, S is not necessarily minimal in the sense that it may exist $S' \subset S$ such that $d_{S\cup S'} = d_{S\cup S}$.

Proposition 6 Given a partially diagnosable triple (Σ, S, F) and a set of potential additional sensors $\mathcal{S}^+$, consider a set of sensors $S\subseteq \mathcal{S}^+$ and the corresponding updated set of faults F^+ partitioned as $F^+ = F_S^+ \cup F_M^+$, where F_S^+ is the set of single faults and F_M^+ the set of multiple faults. Then, under the ARR-based exoneration assumption, if S is such that $(\Sigma, S\cup S, F_S^+)$ has maximal discrimination level, then $(\Sigma, S\cup S, F^+)$ has also maximal discrimination level.

Proof of Proposition 6. Consider that S is such that $(\Sigma, S\cup S, F_S^+)$ has maximal discrimination level $\mathcal{D}_{S\cup S}$ and denote by $\mathcal{A}_{S\cup S}$ the set of ARRs when the sensors of $S\cup S$ are in use, then there exists no $S' \subseteq \mathcal{S}^+\backslash S$ such that $\mathcal{D}_{S\cup S\cup S'} > \mathcal{D}_{S\cup S}$. Therefore, for any new ARR arising from the addition of sensors of S', there exists at least one ARR of $\mathcal{A}_{S\cup S}$ having the same support with respect to F_S^+. Let's call ARR_{new} one of such new ARRs and ARR_{old} one "old" ARR having the same support.

Under the ARR-based assumption, the fault signature γ_K of any multiple fault F_K, where K is a proper subset of {1,2,…,n}, is obtained from the single fault signatures γ_α, $\alpha \in \{1, 2, \ldots, n\}$, of single faults $F_1, F_2, \ldots, F_n$ as follows: the entry γ_{iK} is non zero if at least one entry $\gamma_{i\alpha}$, $\alpha \in \{1, 2, \ldots, n\}$ is non zero. Hence, ARR_{new} has the same support with respect to F_M^+ as ARR_{old}, and therefore it has the same support with respectto the whole set F^+. In consequence the fault signatures arising from the new ARRs are not different from existing ones, meaning that $(\Sigma, S\cup S, F^+)$ has maximal discrimination level. ■

The above proposition is important from a practical point of view since it means that for maximizing the discrimination level of a system, it is enough to consider single faults when the ARR-based exoneration holds.

B. Recursive characterization of ARRs corresponding to successively included sensor sets

Our method is based on hypothesizing sensors and producing the corresponding *hypothetical component supported ARRs* (H-ARR). The starting point assumption is that *all* the unknown variables have an hypothetical sensor, i.e. the system is assumed to be *fully sensored*, and we define the full set of hypothetical sensors as S_X. Subsequent hypotheses proceed to the retraction of sensors, providing at the same time the changes on the analytical redundancies by redefining the set of ARRs. This section shows that if the ARR supports are traced for, full diagnosability analysis can be achieved from ARR generation for the fully sensored system only.

Proposition 7 Let us consider the system model $\Sigma=(E, X\cup O)$, where E is partitioned into behavioral relations E_{beh} and observation relations E_{obs}. Assume that Σ is fully sensored, i.e. every variable in X is sensored ($card(X)=card(O)$) and there is one observation relation e^i_{obs} in E_{obs} of the form $o_i=f_i(x_i)$ for every couple (x_i, o_i), then the set of primary behavioral relations E_{beh} instanciated by $x_i=f_i^{-1}(o_i)$[8] constitutes a basis $\mathcal{A}_b$ of the whole set $\mathcal{A}$ of ARRs of Σ.

Proof of Proposition 7. Since there is one observation relation e^i_{obs} in E_{obs} of the form $o_i=f_i(x_i)$ for every couple (x_i, o_i), a complete matching $\mathcal{M}_{E\leftrightarrow X}$ between E and X is obtained by associating the unknown variable x_i to its corresponding observation relation e^i_{obs}. Therefore, the relations in E_{beh} are all redundant relations and the corresponding ARRs, obtained by replacing the unknown variables by their observation

[8] We consider that all sensor models of the form $o_i=f(x_i)$ are invertible, i.e. f^{-1} exists (which seems reasonable for sensors).

expression $x_i=f_i^{-1}(o_i)$ form a basis of the whole set of ARRs of Σ. ■

Consider a set of relations Γ linking a set of variables $\{v_i\}$, then we note $\Gamma(v_i)$ the subset of relations that contain the variable v_i.

Proposition 8 Consider the system $\Sigma'=(E', X\cup O')$ for which $E' = E_{beh} \cup E'_{obs}$ and X is the same as in Proposition 7 and assume that one sensor $S(x_i)$ has been retracted, resulting in $E'_{obs} = E_{obs} \setminus \left\{e^i_{obs}\right\}$ and $O'=O\setminus\{o_i\}$. Then a basis $\mathcal{A}_b$' of the whole set $\mathcal{A}$ ' of ARRs of Σ' is given by $\{\mathcal{A}_b\setminus\mathcal{A}_b(o_i) \cup Comb(\mathcal{A}_b(o_i))\}$, in which $Comb(\mathcal{A}_b(o_i))$ is the set of combined ARRs obtained from $\mathcal{A}_b(o_i)$ by eliminating variable o_i, i.e. extracting o_i from one of the relations in $\mathcal{A}_b(o_i)$ and substituting it in the other relations, assuming that the substitution is possible from an analytical point of view.

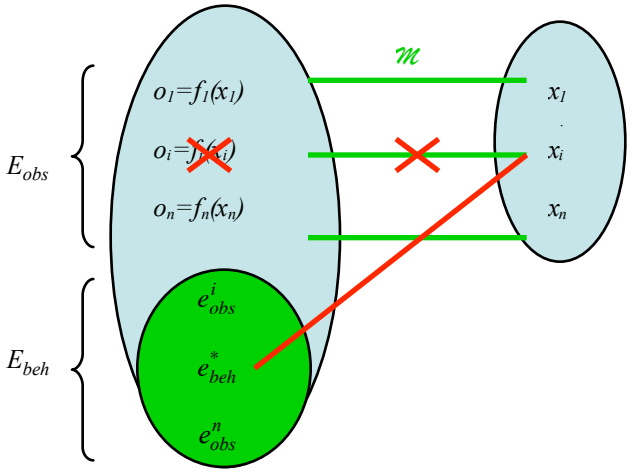

Fig. 3. Retracting one sensor

Proof of Proposition 8. Let us construct a complete matching $\mathcal{M}_{E'\leftrightarrow X}$ between E' and X from $\mathcal{M}_{E\leftrightarrow X}$. All the matches are still possible except the one for x_i since e^i_{obs} is not in E' anymore. x_i must hence be matched to a relation e^*_{beh} of $E_{beh}(x_i)$ (relations in E_{beh} that contain x_i). These relations are just the ones that correspond to $\mathcal{A}_b(o_i)$. The ARRs in $\mathcal{A}_b(o_i)$ are not ARRs of Σ' and new ARRs must be added, namely $Comb(\mathcal{A}_b(o_i))$ which is obtained by using e^*_{beh} to provide x_i, then substituted in $E_{beh}(x_i)\setminus\{e^*_{beh}\}$, also substituted for the other o_j, $j\neq i$. Equivalently, considering $\mathcal{A}_b(o_i)$, o_i is extracted from the ARR that corresponded to e^*_{beh} and substituted in the other ARRs of $\mathcal{A}_b(o_i)$. $\mathcal{A}_b(o_i)$ is hence replaced by $Comb(\mathcal{A}_b(o_i))$ and there is a one to one correspondance between the operation of retracting a sensor and the operation of combining ARRs, as illustrated on Figure 3. ■

Note that when sensors must be considered as fault candidates and taken into account in the ARRs supports, the support of a combined ARR_k, obtained by combining ARR_i and ARR_j through the measured variable o_i as output of $S(x_i)$, is always such that:

$$Supp(ARR_i)\cup Supp(ARR_j)\not\subset Supp(ARR_k) \qquad (1)$$

implying that ARR_k is not logically redundant w.r.t. ARR_i and ARR_j. This comes from the fact that:

$$Sens\text{-}Supp(ARR_k)=$$
$$(Sens\text{-}Supp(ARR_i) \cup Sens\text{-}Supp(ARR_j))\setminus S(x_i) \qquad (2)$$

Consider $\Sigma=(E,X\cup O)$ as in Proposition 7. Given the full set of sensors $S_X=\{S(x_1), \ldots, S(x_n)\}$ and a set $I\in\mathcal{P}(X)$ where $\mathcal{P}(X)$ is the power set of X, define $S^I=S_X\setminus\{S(x_i)/i\in I\}$ and the resulting observed variable set $O^I = O\setminus\{o_i / i\in I\}$ and observation relation set $E^I_{obs} = E_{obs}\setminus\{e^i_{obs} / i\in I\}$. Define now the systems $\Sigma^I = (E^I, X\cup O^I)$, where $E^I = E_{beh}\cup E^I_{obs}$. Let us denote by $\mathcal{A}^I_b$ an ARR base corresponding to $\Sigma^I = (E^I, X\cup O^I)$.

Corollary 2 Consider $\Sigma=(E,X\cup O)$ and the systems $\Sigma^I = (E^I, X\cup O^I)$ as defined above, then the whole set of ARRs of the fully sensored system Σ is given by:

$$\mathcal{A} = \bigcup_{I\in\mathcal{P}(X)} \mathcal{A}^I_b \qquad (3)$$

Proof of Corollary 2. By Proposition 8, there is a one to one correspondance between the operation of retracting a sensor and the operation of combining ARRs. To obtain $\mathcal{A}$, all the possible combinations of sensors, generated from the power set of X, to be retracted must be considered. This result is illustrated on Figure 4 where $\mathcal{A}^{(i)}_b = \left\{\mathcal{A}^I_b / card(I) = i\right\}$. ■

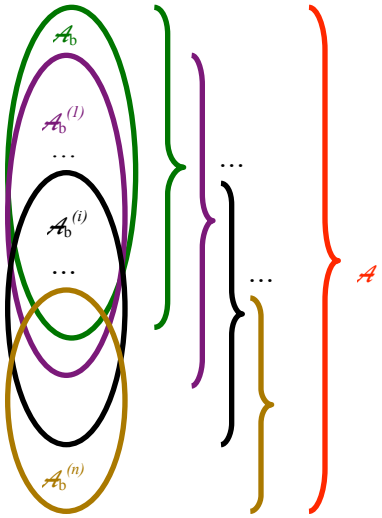

Fig. 4. The whole set of ARRs $\mathcal{A}$ of the fully sensored system as union of ARR bases of successive partially sensored systems.

C. Generating ARRs and tracing their supports

The results presented in section III.B provide the basis of a method to derive the H-ARRs associated to the different situations of sensors.

Proposition 7 states that for the fully sensored system (all the unknown variables have an hypothetical sensor) the basis[9] of ARRs $\mathcal{A}_b$ is just given by the model primary relations.

Then, Proposition 8 shows that considering the successive retraction of sensors is equivalent to successively combining the ARRs by substituting for the variables that correspond to retracted sensors. This results in generating the set of bases $\mathcal{A}_b^I$, for $I \in \mathcal{P}(X)$.

ARR generation of course includes tracing the sensor-supports, i.e. generating the new ARRs sensor-supports from their parents' sensor supports. This is key to the further diagnosability analysis.

In addition to sensor-supports, other important attributes also need to be traced to avoid irrelevant combinations. In particular, we have to avoid combining a combined ARR with one of its own ascendant ARRs. This is possible tracing the ARRs *relation-support*, i.e. the support in terms of the underlying primary relations.

Another issue concerns the conditions under which a variable substitution can be actually performed. A given relation (primary or not) can be interpreted in a causal way, hence determining one or the other of the involved variables. Substitution can be performed when the two considered ARRs have consistent causalities, as explained below. This requires to trace the *causal-relation-supports* as well. The possible causal interpretations may be submitted to validity conditions depending on the corresponding mathematic analytical form. For example, the relation $x = y \times z$ has three causal interpretations, i.e. three associated causal relations: $x=y\times z$ itself in which x is considered as causally dependent on y and z; $y=x/z$ in which y is considered as causally dependent on x and z, under validity condition $z \neq 0$; and $z=x/y$ in which z is considered as causally dependent on x and y, under validity condition $y \neq 0$.

If invertible, every algebraic relation hence gives rise to a set of causal-relations with their associated validity conditions. A differential relation in canonical form, i.e. with one single derivative appearing on the left hand side of the relation, has one single causal interpretation, in which the derivative depends on the other variables (differential causality). The variable under the derivative depends on its derivative (integral causality) [16].

In a given relation, the variable that is causally dependent on the others is defined as the *causally downstream variable*. Two causal relations are said to have *consistent causalities* with respect to a common variable v iff v is the causally downstream variable in one of the relations but not in the other.

Note that primary relations may have validity conditions that are not related to their causal interpretation; this is the case when the modelled system has multiple operational modes (nominal or faulty). In the general case, the validity condition of an ARR is hence given by a logical formula whose truth value depends on algebraic propositions over the set of variables involved in the ARR.

[9] In this case, the basis is unique.

The validity condition of a combined ARR is given by the conjunction of the ascendant ARRs validity conditions.

In consequence, the following concepts are associated to an H-ARR:

- the *relation-support,* noted *Rel-Supp(.)*, which indicates the primary relations which underlie the H-ARR;
- the *causal-relation-support,* noted *CRel-support(.)*, which indicates the primary causal relations which underlie the H-ARR;
- the *validity condition*, noted *VC(.),* arising from the conjunction of the validity conditions of its ascendant H-ARRs;
- the *structure*, noted *Struct(.)*, which indicates by a non zero entry the observed variables which appear in the H-ARR or one of its ascendants: the observed variables which have been substituted for are marked "#"; the H-ARR *causal interpretation* is indicated as follows:
  - the variable that is causally downstream is marked "⊗";
  - the remaining variables are marked "×";
- the *component-support,* noted *Comp-Supp(.)*, which indicates the components whose models underlie the H-ARR;
- the *sensor-support*, noted *Sens-Supp(.)*, that indicates the sensors whose models underlie the H-ARR.

The above attributes can be summarized in an H-ARR table composed of as many fields.

In summary, the procedure principle is as follows. The generation of H-ARR starts from hypothesizing all the possible sensors, i.e. for any variable i, $i=1,\ldots, n$, we have $S(x_i)$: $o_i=f_i(x_i)$. In this situation, every primary relation r_i provides a set of *primary H-ARRs* r_i^j, $j=1,\ldots, p_i$, whose cardinality is the number of causal interpretations. The total number of primary H-ARRs is hence given by $\Theta_0 = \sum_{i=1}^{n} p_i$. The sensors are then retracted in sequence and the new H-ARRs that result from the retraction are generated. The new H-ARRs arise from substituting the variable whose sensor has been retracted from one H-ARR into another, under consistent causality conditions. In doing so, all the H-ARR attributes are traced for.

Unfortunately, different combinations may lead to the same H-ARR. Hence before adding a new H-ARR, it is checked for equality against the existing ones. Note that two H-ARRs are said to be equal (noted $=_f$) when they correspond to the same algebraic variety. Equivalently, $\text{H-ARR}_i =_f \text{H-ARR}_j$ if and only if all their attributes but the causal-relation-support are equal.

Let us call Θ the total number of resulting H-ARRs, composing the final H-ARR table.

Let us mention that the set of resulting H-ARRs is not sensitive to the order of the retraction because the children ARRs remain in the list even when combined ones are generated.

The ideas behind the H-ARR generation algorithm are first explained with an example, then the algorithm is given.

Example 1:

Consider the simple example of an adder (A) connected to an inverter (I) given in figure 5.

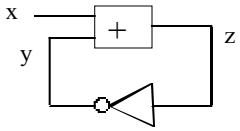

Fig 5. An adder connected to an inverter

The model $\Sigma=(E,X\cup O)$ of such system is given by:

$$\begin{array}{ll} \textit{Set } E_{beh}\textit{:} & \textit{Set } E_{obs}\textit{:} \\ \text{A: } e_1\text{: } z=x+y \quad (4) & S(x)\text{: } \quad x^*=x \quad (5) \\ \text{I: } \ e_2\text{: } y=-z & S(y)\text{: } \quad y^*=y \\ & S(z)\text{: } \quad z^*=z \end{array}$$

where $X=\{x, y, z\}$ is the set of unknown variables and $O=\{x^*, y^*, z^*\}$ is the set of observations given by the hypothetical sensors. When all three sensors are available, then from Proposition 7, an ARRs basis is simply given by the primary relations in which every variable is replaced according to its observation model.

$$\begin{array}{ll} ARR_1^0\text{: } z^*=x^*+y^* & Supp(ARR_1^0)=\{\text{A},S(x),S(y),S(z)\} \\ ARR_2^0\text{: } y^*=-z^* & Supp(ARR_2^0)=\{\text{I}, S(y), S(z)\} \end{array} \quad (6)$$

and $\mathcal{A}_b=\{ARR_1^0, ARR_2^0\}$.

ARR_1^0 and ARR_2^0 can be interpreted causally, resulting in:

$$\begin{array}{ll} ARR_1\text{: } z^*=x^*+y^* & Supp(ARR_1)=\{\text{A},S(x),S(y),S(z)\} \\ ARR_2\text{: } x^*=z^*-y^* & Supp(ARR_2)=\{\text{A},S(x),S(y),S(z)\} \\ ARR_3\text{: } y^*=-x^*+z^* & Supp(ARR_3)=\{\text{A},S(x),S(y),S(z)\} \\ ARR_4\text{: } y^*=-z^* & Supp(ARR_4)=\{\text{I},S(y),S(z)\} \\ ARR_5\text{: } z^*=-y^* & Supp(ARR_5)=\{\text{I},S(y),S(z)\} \end{array} \quad (7)$$

Formulas in (7) constitute the primary H-ARRs. Note that in this example $\Theta_0=5$ and no validity conditions are needed.

TABLE I. PRIMARY H-ARR TABLE

	Relation		Causal Relation					VC	Structure			Component		Sensor Support		
	e1	e2	r1	r2	r3	r4	r5		x	y	z	A	I	S(x)	S(y)	S(z)
1	x		x						x	x	⊗	x		x	x	x
2	x			x					⊗	x	x	x		x	x	x
3	x				x				x	⊗	x	x		x	x	x
4		x				x				⊗	x		x		x	x
5		x					x			x	⊗		x		x	x

Let us now consider the system $\Sigma'=(E',X\cup O')$ corresponding to the situation in which only $S(x)$ and $S(z)$ are available but not $S(y)$, i.e. retraction of $S(y)$ (marked # in the 6th line of Table II). ARR_1^0 and ARR_2^0 are not ARRs anymore and one new ARR comes from combining one ARR from $\{ARR_1, ARR_2, ARR_3\}$ with one ARR from $\{ARR_4, ARR_5\}$, since they share variable y^*. Any combination causally consistent with respect to y^* leads to the same ARR. This is illustrated below for two particular combinations:

Causal-relation-support	H-ARR	Support
ARR_1 , ARR_4	$z^*=x^*-z^*$	A, I, $S(x)$, $S(z)$
ARR_3, ARR_4	$-z^*=-x^*+z^*$	A, I, $S(x)$, $S(z)$

The fact that the two obtained H-ARRs above are equal illustrates that two H-ARRs are equal when all their attributes are equal but their causal-relation-support. The only resulting ARR is hence:

$$ARR_3^0\text{: } x^*=2z^* \quad Supp(ARR_3^0)=\{\text{A, I, } S(x), S(z)\} \quad (8)$$

This results in $\mathcal{A}_b^{\{y\}}=\{ARR_3^0\}$. ARR_3^0 has two causal interpretations leading to ARR_6 and ARR_7 in Table II. Note that $S(y)$ is not in the support of ARR_3^0 (nor in that of ARR_6 and ARR_7).

Let us now consider the situation in which we retract $S(z)$. Combining one ARR from $\{ARR_1, ARR_2, ARR_3\}$ with one ARR from $\{ARR_4, ARR_5\}$ with respect to z^* provides ARR_4^0:

$$ARR_4^0\text{: } x^*=-2y^* \quad Supp(ARR_4^0)=\{\text{A, I, } S(x), S(y)\} \quad (9)$$

which corresponds to the situation in which only $S(z)$ has been retracted, hence $\mathcal{A}_b^{\{z\}}=\{ARR_4^0\}$. ARR_4^0 has two causal interpretations leading to ARR_8 and ARR_9 in Table II. The situation in which both $S(y)$ and $S(z)$ have been retracted corresponds to a non monitorable system (no ARRs), i.e. $\mathcal{A}_b^{\{y,z\}}=\varnothing$. Indeed, although ARR_3^0 shares variable z^* with ARR_1^0 and ARR_2^0, ARR_3^0 cannot be combined with ARR_1^0 and ARR_2^0, which are the own ascendants of ARR_3^0.

Finally, considering the retraction of $S(x)$ does not give rise to any combination, and hence does not provide any new ARR.

TABLE II. FINAL H-ARR TABLE

	Relation Support		Causal Relation Support					VC	Structure			Component Support		Sensor Support		
	e1	e2	r1	r2	r3	r4	r5		x	y	z	A	I	S(x)	S(y)	S(z)
1	x		x						x	x	⊗	x		x	x	x
2	x			x					⊗	x	x	x		x	x	x
3	x				x				x	⊗	x	x		x	x	x
4		x				x				⊗	x		x		x	x
5		x					x			x	⊗		x		x	x
6	x	x	x			x			x	#	⊗	x	x	x		x
7	x	x			x	x			⊗	#	x	x	x	x		x
8	x	x	x			x			x	⊗	#	x	x	x	x	
9	x	x			x		x		⊗	x	#	x	x	x	x	

The main steps to generate the H-ARR are provided in the algorithm below. The elements of the "Structure" field sub-table are noted $s(\alpha,\beta)$, where α corresponds to the α^{th} H-ARR and β corresponds to the β^{th} observed variable column, $\beta=1, \ldots, n$.

/*H-ARR generation algorithm*/

$S_X=\{S(x_1),\ldots, S(x_n)\}$

$h=\Theta_0$

For i=1, ..., n do

(a) Select $S(x_i)\in S_X$

(b) Define the set $J=\{j/o_i \in Struct(H\text{-}ARR_j)\}$
If card $J \leq 1$, then go to (a)
Else $\forall (p,q) \in J^2, p \neq q$
If $Rel\text{-}Supp(H\text{-}ARR_p) \cap Rel\text{-}Supp(H\text{-}ARR_q) = \varnothing$
*/*the above condition avoids substitutions of H-ARRs with the same relation support*/*
and $\neg\exists\beta$ such that
$(s(p,\beta)\in\{\times,\otimes\} \wedge s(q,\beta)=\#) \vee (s(p,\beta)=\# \wedge s(q,\beta)\in\{\times,\otimes\})$
*/*the above condition avoids irrelevant substitutions due to structure inconsistency*/*
and $(s(p,i)=\otimes \vee s(q,i)=\otimes)$
and $\neg(s(p,i)=\otimes \wedge s(q,i)=\otimes)$
*/*these conditions guarantee consistent causal interpretations*/*
and $(VC(H\text{-}ARR_p) \wedge VC(H\text{-}ARR_q))$ is satisfiable.
*/*the above condition guarantees consistent validity conditions*/*
then build $H\text{-}ARR_{h+1}$ such that
$Rel\text{-}Supp(H\text{-}ARR_{h+1}) = Rel\text{-}Supp(H\text{-}ARR_p) \cup Rel\text{-}Supp(H\text{-}ARR_q)$
$CRel\text{-}Supp(H\text{-}ARR_{h+1}) = CRel\text{-}Supp(H\text{-}ARR_p) \cup CRel\text{-}Supp(H\text{-}ARR_q)$
$VC(H\text{-}ARR_{h+1}) = VC(H\text{-}ARR_p) \wedge VC(H\text{-}ARR_q)$
$Comp\text{-}Supp(H\text{-}ARR_{h+1}) = Comp\text{-}Supp(H\text{-}ARR_p) \cup Comp\text{-}Supp(H\text{-}ARR_q)$
$Sens\text{-}Supp(H\text{-}ARR_{h+1}) = Sens\text{-}Supp(H\text{-}ARR_p) \cup Sens\text{-}Supp(H\text{-}ARR_q) \backslash S(x_i)$
$Struct(H\text{-}ARR_{h+1})$ is given by :
$s(h+1,i) = \#$
/ # marks the variable o_i as substituted */*
For $\beta \neq i$
If $s(p, \beta)=\times$ or $s(q, \beta)=\times$, then $s(h+1, \beta)=\times$
If $s(p, \beta)=\#$ or $s(q, \beta)=\#$, then $s(h+1, \beta)=\#$
If $s(p, \beta)=\otimes$ or $s(q, \beta)=\otimes$, then $s(h+1, \beta)=\otimes$
End for
*/*the above conditions update the status of the variables appearing in the H-ARR structure*/*
If $H\text{-}ARR_{h+1} \neq_f H\text{-}ARR_i \ \forall i<h+1$ then add $H\text{-}ARR_{h+1}$
*/*the above condition avoids adding an H-ARR which is equal to an existing one*/*
h=h+1
End if
End if
End else
End for

The H-ARR algorithm removes sensors one by one and combines H-ARRs consequently. First a sensor $S(x_i) : o_i=f_i(x_i)$ is selected and a set J of H-ARR's indexes is defined as the set of H-ARR in which o_i is involved. Then the pairs of H-ARRs to be combined are taken in J. Obviously, another sensor is chosen if the cardinality of J is ≤ 1.

Two H-ARR selected in the set J can be combined under several conditions :

- Their relation-supports are disjoined ($Rel\text{-}Supp(H\text{-}ARR_p) \cap Rel\text{-}Supp(H\text{-}ARR_q) = \varnothing$) else you combine two H-ARR issued from the same relation.
- An already substituted variable from one of the selected H-ARRs cannot be reintroduced to generate a new H-ARR ($\neg\exists\beta$ such that $(s(p, \beta) \in \{\times,\otimes\} \wedge s(q, \beta) = \#) \vee (s(p, \beta)=\# \wedge s(q, \beta) \in \{\times,\otimes\})$).
- One of the selected H-ARRs has the variable o_i marked as causally downstream variable but not both ($(s(p,i)=\otimes \vee s(q,i)=\otimes)$ and $\neg(s(p,i)=\otimes \wedge s(q,i)=\otimes)$), otherwise underlying causalities are not consistent.
- The combination of validity conditions must be consistent ($(VC(H\text{-}ARR_p) \wedge VC(H\text{-}ARR_q))$ is satisfiable).

If all the conditions are fulfilled, then a new H-ARR can be generated. Note that the chosen sensor $S(x_i)$ is not in the sensor-support and the associated variable o_i is marked as substituted (#) in the generated H-ARR structure. The generated H-ARR is added to the table if it is not equal to any existing one.

D. Hypothetical Fault Signature Matrix.

In the following, let us denote by $\mathcal{S}^*$ the set of potential sensors, specified at the system's design stage. The choice of $\mathcal{S}^*$ may be guided by the set of faults F to be considered (sensoring variables which only appear in the model relations of the components that cannot be faulty is of no use) and by technological constraints. If a given set of sensors is available anyway these are necessarily included in $\mathcal{S}^*$.

Definition 10 *(Hypothetical Fault Signature (HFS) Matrix)*. The Hypothetical Fault Signature Matrix is defined as the set of fault signatures that would result from the availability of any combination of sensors in $\mathcal{P}(\mathcal{S}^*)$, where $\mathcal{S}^* \subseteq S_X$ and $\mathcal{P}(\mathcal{S}^*)$ denotes the power set of $\mathcal{S}^*$.

The HFS matrix can easily be obtained from the H-ARRs table. Indeed, it corresponds, after simple manipulations, to the sub-table given by the *component-support* together with the *sensor-support* fields, complemented by the corresponding validity condition field. The manipulations to be performed on the H-ARR table are given in the following algorithm:

/*HFS generation algorithm*/

Step 1- Compact all the H-ARRs which do not differ in the component-support, sensor-support or validity-condition field in the same equivalence class;

Step 2- Discard the columns corresponding to those components which are not included in the set of faults to be considered; sensors have a different status and all sensor-support columns must remain as they provide information to be used for diagnosibility assessment.

Step 3- Discard those H-ARRs whose sensor-supports include a sensor which is not in S*.

Step 4- Remove all the table sub-fields but the component-support, sensor-support and validity condition sub-fields.

The HFS matrix makes the correspondence between the sensors, the resulting H-ARRs and the components that support these ARRs while taking into account the different validity conditions of the H-ARRs, i.e. operating regions of the system and analytical singularity conditions. Note that HFS is actually the FS matrix of $(\Sigma, \mathcal{S}^*, F)$.

IV. Diagnosability Assessment

The HFS matrix summarises all the required information to perform a complete diagnosability assessment, i.e. to provide all the Minimal Additional Sensor Sets (MASSs) that guarantee a desired discrimination level.

A. Alternative Fault Signature Matrices

Given the HFS matrix, every possible combination of sensors gives rise to a corresponding FS matrix. Considering the set of all possible combinations of sensors hence leads to a set of *Alternative Fault Signature Matrices*, as defined in Definition 11.

Definition 11 *(Alternative Fault Signature Matrices)* The Alternative Fault Signature (AFS) Matrices are given by all the FS matrices corresponding to all the possible sensor sets $S \in \mathcal{P}(\mathcal{S}^*)$.

Note that if we start with a non empty set of already available sensors S_a, then the AFS matrices are obtained for all the possible sensor sets $S=S_a \cup S'$, where $S' \in \mathcal{P}(\mathcal{S}^* \setminus S_a)$.

B. Analysing an FS matrix

Given a system (Σ, S, F) with $S \subseteq \mathcal{S}^*$, its FS matrix is obtained from the HFS matrix by removing all the columns corresponding to hypothetical sensors not included in S and all the *H-ARRs* such that $sens\text{-}Supp(H\text{-}ARR_i) \not\subset S$.

The FS matrix can be analyzed from the rows (H-ARR supports) or from the columns (fault signatures) point of view. The two following results correspond to the former and the latter, respectively.

Proposition 9 Under the ARR-based exoneration assumption, two H-ARRs that have the same support have the same *fault sensitivity*, i.e. they have a non zero value for the same fault situations.

Proof of Proposition 9. Trivial. ■

According to Proposition 9, the H-ARRs can be grouped into equivalence classes corresponding to the same support, i.e. the same rows in the HFS matrix, under equal validity conditions. In the following and when not ambiguous, the term "H-ARR" is used to denote such equivalence classes.

Proposition 10 Under the ARR-based exoneration assumption and given a partially diagnosable triple (Σ, S, F):

1. The number of *D-classes* $\mathcal{D}_S$ of (Σ, S, F) is given by the number of different fault signatures (column vectors) of its FS matrix.
2. Consider a structured matrix M like the FS matrix and assume that the non zero values are given the value 1, then the number of different column vectors is given by its analytical column rank, noted *rank(M)*.

Proof of Proposition 10. Under the ARR-based exoneration assumption, and by proposition 4, if two faults F_i and F_j are discriminable, they are strongly discriminable, i.e $OBS_{Fi} \cap OBS_{Fj} = \emptyset$, which is equivalent to having different fault signatures. Hence, the number of different fault signatures in $FS(\Sigma, S, F)$ provides the number of *D-classes* of (Σ, S, F), proving 1.

The proof of the second part of the proposition comes trivialy from the fact that two column vectors whose values are 0 or 1 are linearly independent if and only if they are not equal. ■

From the above result and definition 7, one can also derive the diagnosability degree of a system (Σ, S, F).

C. Minimal additional sensor sets

This section characterizes the minimal additional sensor sets (MASS), which guarantee maximal discrimination level.

Corollary 3 Under the ARR-based exoneration assumption and given a system (Σ, S, F) and a set of possible sensors $\mathcal{S}^*$, i.e. $S \subseteq \mathcal{S}^*$, the maximal achievable discrimination level $\mathcal{D}_{MAX}$ is equal to the number of D-classes $\mathcal{D}_{\mathcal{S}^*}$ of $(\Sigma, \mathcal{S}^*, F)$, which is by Proposition 10 equal to the number of different fault signatures of the HFS matrix, *rank(HFS)*.

As seen in section III.A, in the general case, the number of D-classes is a monotone function of the number of sensors in S, i.e. $\mathcal{D}_S \subseteq \mathcal{D}_{S'}$ for $S \subseteq S'$, but this is not true for the diagnosability degree $d_S = \mathcal{D}_S / card(F)$. Indeed, $card(F)$ depends itself on the considered set of sensors. Neverthless, the diagnosability degree is obviously a monotone function of the number of sensors in S in the two particular cases:

1. Only sensors in S_a can undergo faults.
2. All sensors are reliable and cannot undergo faults (which is a particular case of 1 for $S_a = \emptyset^{10}$).

Corollary 4 Under the ARR-based exoneration assumption and assuming that only the sensors in S_a can be faulty, consider a system (Σ, S, F) and a set of possible sensors $\mathcal{S}^*$, i.e. $S \subseteq \mathcal{S}^*$, then the number of faults remains constant (since the sensors that can be added are assumed fully reliable) and the maximal achievable diagnosability degree d_{MAX} is equal to $d_{\mathcal{S}^*}$ which is in turn equal to $rank(HFS)/card(F)$.

Corollary 5 Under the same assumptions as Corollary 4, an AFS matrix with the same rank as HFS corresponds to a MASS. The MASS is obtained as the union of the *sensor-supports* associated to the H-ARRs of this AFS matrix, excluding the already available sensors S_a.

An operational method for deriving the MASSs is directly obtained from Corollary 5: derive all the AFS matrices and retain the ones with maximal rank. However, this method can be significantly improved because lower bounds for the number of H-ARRs and for the number of sensors can be determined *a priori* and this cuts down the combinatorial issue.

Given that a binary number of x bits allows for 2^x-1 non null different combination binary vectors, we have the following result:

Proposition 11 *(Minimum number of ARRs)* Under the ARR-based exoneration assumption, the minimum number of ARRs needed to discriminate a set of faults F of cardinal n_F, i.e. to obtain a FS matrix of full rank, is $x = Ceil[log(n_F+1)]$, where *log* is the base 2 logarithm and *Ceil[y]* is the minimal integer value greater than or equal to y.

Proof of Proposition 11. Since x ARRs allow for a maximum of 2^x-1 different fault signatures, the upper bound for the number of faults that can be discriminated with x ARRs is given by $n_F \leq 2^x$-1, and hence the result. ■

Proposition 12 *(Minimum number of sensors)* The number of sensors can be bounded by s_{MIN}, where s_{MIN} equals the minimal cardinal of any of the H-ARRs sensor-supports.

Proof of Proposition 12. Consider a system (Σ, S, F) and a set of possible sensors $\mathcal{S}^*$, i.e. $S \subseteq \mathcal{S}^*$ such that $card(S)$ is lower than the cardinal of any of the H-ARRs sensor-supports, then (Σ, S, F) is non monitorable, i.e. has no ARR, hence proving the proposition. ■

For a given system $(\Sigma, \mathcal{S}^*, F)$, the combinatorial procedure to obtain the set of MASSs, denoted as $\mathcal{MASS}$, can hence start with the AFS matrices having a minimum number x of H-

ARRs (rows) and a minimum number of supporting sensors s_{MIN}.

The above procedure determines all the MASSs. If one is interested in the MASS with maximal cardinality s_{MAX}, the procedure stops when the AFS matrices for sets of sensors of cardinality s_{MAX} have been derived and tested.

Note that the different MASS achieve a maximal discrimination level but their corresponding *D-classes* may be different. This choice is left to the user.

Instead of exploring all the combinations, the problem of determining the MASS that achieve a given discrimination level (*MASS problem*) can be formulated as an optimisation problem, according to a given cost criterion.

Let us notice that in its general form the MASS problem is not easy, in particular because the set of faults which are asked to be discriminated may vary with the proposed set of sensors when some sensors can themselves be faulty. Also, because there may be several minima.

Finally, let us notice that the ultimate general formulation of the MASS problem would be in a multiple operational modes context. The solution to this problem means not only to find the optimal set of sensors but to determine the optimal set of operating modes in which the system must be operated.

V. Application

This section illustrates our method first with the well-known polybox example and second with an industrial actuator device.

A. Polybox example

Let us consider the system taken from [7] known as the polybox and given in Figure 6. This example is interesting to illustrate the H-ARR generation procedure more than for the subsequent diagnosability analysis.

The set of unknown variables is X = {*a, b, c, d, e, f, g, x, y, z*}, and the set of observations is O = {a_{obs}, b_{obs}, c_{obs}, d_{obs}, e_{obs}, f_{obs}, g_{obs}}. Table III provides the system model.

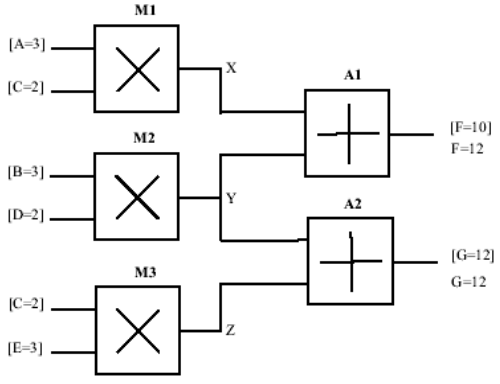

Fig. 6. The POLYBOX example

To build the H-ARR table, we start with 15 primary H-ARRs, as given by Table IV. Indeed, every primary relations has three causal interpretations, some of which has restricted validity. Successive combination of these H-ARRs provides a final H-ARR table with 113 H-ARRs.

TABLE III. POLYBOX MODEL

	Set E_{beh}		Set E_{obs}
M1	$e_1 : x=a*c$	$S(a)$	$a_{obs}=a$
M2	$e_2 : y=b*d$	$S(b)$	$b_{obs}=b$
M3	$e_3 : z=c*e$	$S(c)$	$c_{obs}=c$
A1	$e_4 : f=x+y$	$S(d)$	$d_{obs}=d$
A2	$e_5 : g=y+z$	$S(e)$	$e_{obs}=e$
		$S(f)$	$f_{obs}=f$
		$S(g)$	$g_{obs}=g$

Let us suppose the standard situation for the polybox: *x*, *y*, and *z* cannot be sensored and the sensors for the other variables are available, i.e. S_a={*a, b, c, d, e, f, g*}. The HFS matrix, which is actually the FS matrix for the given situation, is obtained from the H-ARR table by applying the algorithm of section III.D. It contains 24 ARRs.

Finally, the FS matrix is compacted according to the H-ARRs which are equivalent in terms of component-supports and sensor-supports. Only one ARR is retained for each class. When the ARRs in a class differ in their validity condition, the least constrained relation is selected. The final FS matrix is given in Table V. It contains 5 ARRs.

The three first ARRs are the standard ones [7]:

$ARR_1 : f_{obs} - a_{obs}.c_{obs} - b_{obs}.d_{obs} = 0$
Comp-Supp(ARR_1) : {A1, M1, M2}

$ARR_2 : g_{obs} - b_{obs}.d_{obs} - c_{obs}.e_{obs} = 0$
Comp-Supp(ARR_2) : {A2, M2, M3}

$ARR_3 : f_{obs} - g_{obs} - c_{obs}.(a_{obs} - e_{obs}) = 0$
Comp-Supp(ARR_3) : {A1, A2, M1, M3}

The two other ARRs have a validity condition related to a singular point (division by e_{obs} and a_{obs} respectively).

ARR_4: $[a_{obs}(g_{obs}-b_{obs}d_{obs})]/e_{obs}+b_{obs}d_{obs}=f_{obs}$
ARR_5: $b_{obs}d_{obs}+[(f_{obs}-b_{obs}d_{obs})e_{obs}]/a_{obs}=g_{obs}$

These are not classical and as a matter of fact, they are not useful when sensor faults are not considered. Indeed, their support includes the five components of the polybox.

B. Damadics-Actuator Device

1) The DAMADICS benchmark actuator

The application example deals with an industrial smart actuator consisting of a flow servo valve driven by a smart positioner; it is used as benchmark in the context of the European DAMADICS project. Faults in these actuators or more generally in final control elements appear relatively often in the industrial practice. The faults cause long-term process disturbances and may even trigger the installation shut down, which may influence the final product quality.

[10] Let's recall that S_a is the set of available sensors.

TABLE IV. INITIAL H-ARR TABLE

	Relation Support					Causal relation Support														Validity Condition					Structure												Component Support					Sensor Support									
	e1	e2	e3	e4	e5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	vc1	vc2	vc3	vc4	vc5	a	b	c	d	e	f	g	x	y	z	M1	M2	M3	A1	A2	S(a)	S(b)	S(c)	S(d)	S(e)	S(f)	S(g)	S(x)	S(y)	S(z)
1	x					x																				x		x					⊗			x					x		x					x		
2	x						x														x					⊗		x					x			x					x		x					x		
3	x							x														x				x		⊗					x			x					x		x					x		
4		x							x																	x		x					⊗				x					x		x					x	
5		x								x													x			⊗		x					x				x					x		x					x	
6		x									x													x		x		⊗					x				x					x		x					x	
7			x									x															x		x				⊗				x					x		x		x				x
8			x										x							x						⊗		x				x				x					x		x		x				x
9			x											x							x						x		⊗				x				x					x		x		x				x
10				x											x															⊗			x	x												x		x	x	
11				x												x														x			⊗	x												x		x	x	
12				x													x													x			x	⊗												x		x	x	
13					x													x													⊗		x	x												x		x	x
14					x														x													x	⊗	x												x		x	x
15					x															x													x		x ⊗												x		x	x

TABLE V. FINAL HFS MATRIX

	Relation Support					Causal relation Support														Validity Condition					Structure												Component Support					Sensor Support									
	e1	e2	e3	e4	e5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	vc1	vc2	vc3	vc4	vc5	a	b	c	d	e	f	g	x	y	z	M1	M2	M3	A1	A2	S(a)	S(b)	S(c)	S(d)	S(e)	S(f)	S(g)	S(x)	S(y)	S(z)
1	x	x		x		x			x						x											x	x	x	x		⊗		#	#		x	x		x		x	x	x	x		x				
2		x	x		x				x			x						x									x	x	x	x		⊗		#	#		x	x		x		x	x	x	x		x				
3	x		x	x	x	x						x			x				x							x		x		x	⊗	x	#	#	#	x		x	x	x		x			x						
4	x	x	x	x	x	x			x				x		x					x					x	x	#	x	x	⊗	x	#	#	#	x	x	x	x	x	x	x			x		x				
5	x	x	x	x	x			x	x			x				x		x				x				x	x	#	x	x	⊗	#	#	#	x	x	x	x	x	x	x			x		x				

The benchmark actuator, simply named *actuator* in the following, interacts with the controlled process. It is used in the evaporation station of a sugar factory in Poland [1][19]. The position set-point of the actuator is the output of the process controller (flow or level controller) and the actuator modifies the position of the valve permitting a direct effect on the primary variable (flow or level) in order to follow the flow or level set point. In this example, the actuator is used to control the flow on the valve outlet F (cf. Fig. 7).

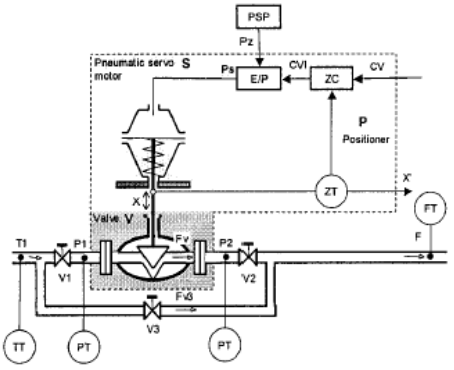

Fig. 7. The actuator scheme

As shown on Figure 7, the components of the actuator system are the following[11]:

1- A *Control valve or hydraulics (H)* which prevents, allows and/or limits the flow of fluids through the control system. Changing the state of the control valve is performed by a pneumatic servomotor.

2- A *pneumatic servomotor or mechanics (M)* can be defined as a compressible (air) fluid powered device in which the fluid acts upon the flexible diaphragm, to provide linear motion of the servomotor stem.

3- A *Positioner* device which eliminates the control-valve-stem miss-positions produced by external or internal causes such as friction, pressure unbalance, hydrodynamic forces *etc*. It consists in an inner loop with a proportional controller within a cascade control structure (ZC), including the output signal of the outer loop of the flow controller and the inner loop of the position controller. It also contains the pressure supplier system (PSP) that generates a constant supply pressure P_z. An electro-pneumatic transducer (E/P) adds to P_z the pressure obtained from converting the control current provided by the position controller. Thus, the output P_s of E/P is injected in the servomotor's chamber, controlling the rod displacement X which is fed back to the position controller ZC through a displacement transducer (ZT).

4- Connection components including a *Bypass Valve* (V3) and *Pipes* (PF).

The available sensors are given by S_a={$S(CV)$, $S(X)$, $S(F)$, $S(P_1)$, $S(P_2)$, $S(P_z)$, $S(T)$, $S(K_{V3})$}[12].

The mathematical equations describing the components behaviour come from [17]. The following table summarises the equations for each component in normal conditions. The sensor are assumed to be ideal, i.e. the sensor output variable is equal to the sensored variable.

The variables are defined as:

X' – hypothetical servomotor's rod displacement

X - servomotor's rod displacement

P_s - pressure in the servomotor's chamber

P_1 - pressure before valve

P_2 - pressure after valve

P_z - supply pressure (≈600 MPa)

P_v – water vapor pressure

ΔP - pressure difference across the valve ($P_1 - P_2$)

[11] Valves V1 and V2 are ignored in this example.

[12] These are the only available sensors on the real actuator although the DAMADICS benchmark did assume a few more sensors.

ΔP_{allow} – maximum allowable fluid pressure drop across the valve
F - volumetric flow on outlet actuator pipe
F_V - volumetric flow on the valve outlet
F_{V3} - volumetric flow on the manual valve (V3) outlet
Dm_a/dt – air mass flow
T – fluid temperature
F_{vc} – reaction force of valve plug
PV- process variable
CVI - control current (controller's output)
CV- process control value
K_{V3} – manual valve position

The notation $g_{NL\text{-}i}$ is used to denote non linear functions, k and k_i are constants.

TABLE VI. MATHEMATICAL EQUATION MODEL OF THE SYSTEM

Support	Equation	Val. Cond.
Pneumatic Servomotor M	$m\frac{d^2X'}{dt^2} = -k_v\frac{dX'}{dt} - k_x(k_1 + X') - F_{vc} + A_eP_s + mg$ $X = g_{NL-1}(X', D_b)$ $F_{vc} = k_{vc}(P_1 - P_2)/K_m$ $F_{vc} = k_{vc}(P_v - P_2)/K_m$, $K_m = g_{NL-2}(X)$ $P_v = g_{NL-3}(T_1)$	$0 \le X \le H_b$ $P_1 > P_v$ $P_1 \le P_v$
Control valve H	$F_V = kK_V\sqrt{\Delta P/\rho}$, $K_V = g_{NL-4}(X)$ $\Delta P_{allow} = K_m(P_1 - r_cP_v)$ $\Delta P = (P_1 - P_2)$ $\Delta P = (P_1 - r_cP_v)$	$\Delta P \ne 0$ $(P_1 - P_2) \le \Delta P_{allow}$ $(P_1 - P_2) > \Delta P_{allow}$
Electro-Pneumatic transducer E/P	$\frac{dP_s}{dt} = (P_s + P_a)\left(\frac{1}{m_a}\frac{dm_a}{dt} - \frac{A_e}{V_o + A_eX}\frac{dX}{dt}\right)$ $\frac{dm_a}{dt} = k_1g_{NL-5}(CVI)\sqrt{P_z - P_s}$ $\frac{dm_a}{dt} = k_1g_{NL-5}(CVI)\sqrt{P_z}$	 $CVI > 0$ $CVI \le 0$
Positioner feedback ZT	$PV = 0.5 + \frac{3}{\pi}a\sin(\frac{X}{H_b} - \frac{1}{2})$	
Controller ZC	$CVI = k_p(CV - PV)$	
Bypass valve V3	$F_{V3} = K_{V3}\sqrt{\Delta P}$	$\Delta P \ne 0$
Pipes Flow PF	$F = F_V + F_{V3}$	

2) Diagnosability degree

The results of sections III and IV are now applied to the actuator considering the operating mode defined by $P_1 > P_v$, $(P_1 - P_2) \le \Delta P_{allow}$ and $CVI > 0$. Then, the model for the pneumatic servomotor, the control valve and the electro-pneumatic transducer are brought back to one single behavioral relation. The complexity of the problem is given by 7 primary relations, 15 causal relations, 11 variables, and 7 components. Applying the H-ARR generation algorithm of section 3.C we obtain 220 H-ARR.

Let us consider the following set of components: {M, E/P,H, ZT, ZC, V3, FP} and the set of potential sensors S={$S(CV)$, $S(X)$, $S(F)$, $S(P_1)$, $S(P_2)$, $S(P_s)$, $S(P_z)$, $S(T)$, $S(CVI)$, $S(K_{V3})$}. All the components can be faulty but ZC and $S(CV)$.

Discrimination level with the available set of sensors. Let us consider that $\mathcal{S}^*$ is the set of actual sensors S_a, i.e. $\mathcal{S}^*$={ $S(CV)$, $S(X)$, $S(F)$, $S(P_1)$, $S(P_2)$, $S(P_z)$, $S(T)$, $S(K_{V3})$}. Applying the H-ARR algorithm (cf. Section III.C) and generating the corresponding FS matrix (cf. III.D), the number of *D-classes* $\mathcal{D}_{\mathcal{S}^*}$ is find to be equal to 4, as shown by the FS matrix in Table VII.

Hence the available sensors on the actuator allow one to discriminate between the following sets of components: {M, E/P, ZT, $S(P_z)$, S(T)}, {H, V3, PF, $S(F)$, $S(K_{V3})$}, $S(X)$,{$S(P_1)$, $S(P_2)$}; and the diagnosability degree is d=4/13.

TABLE VII. THE DAMADICS-ACTUATOR FS MATRIX

RR	M	E/P	H	ZT	V3	FP	$S(X)$	$S(P_z)$	$S(P_1)$	$S(P_2)$	$S(T)$	$S(F)$	$S(K_{V3})$
ARR_1			x		x	x	x		x	x		x	x
ARR_2	x	x		x			x	x	x	x	x		
ARR_3	x	x	x	x	x	x		x	x	x	x	x	x
ARR_4	x	x	x	x	x	x	x	x			x	x	x

Maximal discrimination level. Let us consider the possible sensors $\mathcal{S}^*$={$S(CV)$, $S(X)$, $S(F)$, $S(P_1)$, $S(P_2)$, $S(T)$, $S(P_z)$, $S(K_{V3})$, $S(P_s)$, $S(CVI)$, $S(PV)$} among which the added sensors $S(P_s)$, $S(CVI)$ and $S(PV)$ are considered as fully reliable. In this case, the number of H-ARR is 51, and the number of *D-classes* $\mathcal{D}_{\mathcal{S}^*}$ is 6: {M, $S(T)$}, {E/P, $S(P_z)$}, {H, V3, PF, $S(F)$, $S(K_{V3})$}, {ZT}, {$S(X)$}, and {$S(P_1)$, $S(P_2)$}. In the maximal discriminability situation, the system is hence partially diagnosable with d=6/13.

Notice that there is a lot of ambiguity with respect to components in the hydraulic subsystem: this is the result of little instrumentation in this subsystem.

MASS for achieving maximal discrimination level. We follow section IV.C and iterating on the set of n sensors $\mathcal{S}'$ such that $\{S_a \cup \mathcal{S}'\} \subseteq \mathcal{S}^*$, it is easy to determine that the MASS that provide the maximal discrimination level and diagnosability degree are {$S(P_s)$, $S(CVI)$} or {$S(P_s)$, $S(PV)$}.

VI. RELATED WORK

The method proposed in this paper builds on previous work [14][15]. Our work started with the method in [14] which was in the same spirit as [2]. It required to search for the different causal paths within the AND-OR graph obtained from the RPG graph augmented by the hypothetical sensors. Only one possible causal interpretation of the model was considered meaning no guarantee to produce all the relevant ARRs, and hence to assess the actual diagnosability capacity of the potential instrumentation.

It then appeared that hypothesizing sensor retractions could be a better alternative than sensor additions, since sensor retractions are simply implemented by combining previously derived ARRs when they share the sensored variable. Our method hence starts with the fully sensored system and relies

on the fact that the ARR set for this system includes the ARR sets for the successive partially sensored systems. Generating the whole set of ARRs of the fully sensored system can be achieved by simple table operations. The algorithm complexity remains however exponential in the number of variables.

The problem of generating ARRs from a system's equational model corresponds to the one of finding just over determinated subsystems of equations (constraints). These subsystems indeed correspond to ARRs, and equivalently to potential conflicts. Pulido and Alonso (2002) use a hyper-graph approach. They first look for *Minimal Evaluation Chains* which actually correspond to strictly over-determinated subsystems of equations that can be interpreted in at least one causal way. Each causal interpretation of a Minimal Evaluation Chain is called a *Minimal Evaluation Model*. This method suffers from the same intrinsic combinatorial problem as ours.

From the other hand, Krysander and Nyberg [11] propose a structural method for finding just over-determinated subsystems, called *Minimal Structurally Singular systems* (MSS). It differs from ours in the fact that a variable x and its derivative dx/dt are considered as different variables. This leads them to differenciate equations before being able to find a complete matching. The MSS subsystems are then found in a combinatorial way. The diagnosability problem is also approached. However, it is formulated as the problem of finding the best set of MSSs (equivalent to ARRs) which achieves a given discrimination level. The proposed criterion is the number of equations involved in the MSSs, as the authors claim that the lowest the number of equations, the more robust. This criterion is questionable and as a matter of fact the real MASS optimisation problem, which is in terms of sensors, is not formulated.

As mentioned in section IV, in its general form the MASS optimization problem is tricky. Some solutions exist in the litterature for simplified versions of this problem. [4] simplifies the problem in the three following aspects:

- It is assumed that only available sensors (sensors in S_a) can undergo faults, which means that the number of faults is considered constant.
- It is assumed that the system model includes differential relations only, resulting in a unique causal interpretation.
- It is not only required to discriminate the set of faults, i.e. to have different fault signatures, but to discriminate the faults along a specific pattern, i.e to have a diagonal fault signature matrix.

With the above restrictions, [4] brings back the solution to the search of a minimal cost matching in a weighted bipartite graph.

[18] goes one step further and proposes to use genetic algorithms to perform the optimization task but still it makes the assumption that only available sensors can undergo faults.

Finally, let us mention the work by Frisk *et al.* [9] which propose an alternative way to achieve diagnosability. Instead of hypothesizing sensors, they propose to add analytical information about the faults. For example, they illustrate their idea with the case in which they can explicitely state that a fault derivative is constant. The idea is interesting but it is restricted to very specific types of faults and this approach has not the general scope of hypothesizing sensors.

VII. Conclusions

In this paper, we have presented a method for showing what gains in diagnosis can be made with which additional sensors. This is accomplished by analysing the system from the model based diagnosis viewpoint, given a set of faults that it is desirable to diagnose. The approach has been illustrated through an industrial actuator, taken as benchmark in the Damadics European project.

This work opens numerous perspectives in several directions. First of all, it calls for sophisticated optimisation methods to solve the MASS optimization problem in its general form. In its ultimate general form, when the choice spans over different operating modes, it bridges to the area of active diagnosis, i.e. the choice not only concerns the measure to be done but also the configuration in which the system must be put to perform the measure.

In the formulation adopted in this paper, the analysis has been performed with the ARR-exoneration assumption. It has been shown in [7] that this assumption can be removed within a fault signature matrix framework by considering that the non zero entries of the theoretical fault signatures can be matched to any observed truth value of the corresponding ARR. However, in this case, fault signatures cannot be analysed from the only syntaxic point of view. Indeed, two syntaxically equal signatures may permit a different instantiation and two syntaxically different signatures may permit an instantiation which makes them equal. The non exoneration case hence calls for further analysis.

The method is based on normal behaviour models in the general framework of systems with multiple operating modes, which are captured through the validity conditions associated to the ARRs. This feature should hence make the method easily extendable to the case when fault models are taken into account. This would extend the method not only to discriminate between components but to discriminate between different faults affecting the same component. More work is needed in this direction.

Finally, the working assumption through all the paper is that for economic reasons, hardware redundancy in the instrumentation is not permitted, i.e. only one sensor per variable can be considered. This assumption is acceptable, and even necessary, in many application domains like automobile, process industry, etc. However, for highly critical applications, hardware redundancy may be recommended to achieve a fault-tolerant architecture. We believe that our method can be useful to decide about the necessity of sensor redundancy. The first case is obviously when a given component fault and sensor fault fall into the same D-class; the only way to discriminate between these faults is by sensor redundancy. Another situation in which sensor redundancy may be recommended is

when a given sensor happens to be involved in a high number of ARR sensor-supports. The issue of considering diagnosability analysis in the context of both analytical and hardware redundancy seems promising and calls for further investigations.

REFERENCES

[1] Bartyœ, M., de las Heras, S. (2003). "Actuator simulation of the DAMADICS benchmark actuator system". In Proc. of IFAC Workshop SAFEPROCESS 2003. Washington, D.C., USA, pp. 963-968.

[2] Carpentier, T., (1999). Placement de capteurs pour la surveillance des processus complexes. Ph. D. Thesis in Université des Sciences et Technologies de Lille.

[3] Cassar, J.P. and Staroswiecki M. (1997). A structural approach for the design of failure detection and identification systems. In Proc IFAC, IFIP, IMACS Conference on Control of Industrial Systems, 329-334. Belfort, France.

[4] Commault C. and Dion J.M. (2003). Optimal sensor location for fault detection and isolation in linear structured systems, In Proc. of the European Control Conference ECC'03, Cambridge (UK).

[5] Console, L., Picardi, C. and Ribando, M. (2000). Diagnosis and Diagnosability Analysis using process algebra. In Proc. DX'2000. Morelia, Mexico.

[6] Cordier M-O., Dague P., Lévy F., Dumas M., Montmain J., Staroswiecki M. and Travé-Massuyès L. (2000). AI and automatic control approaches of model-based diagnosis: links and underlying. In Proc. of the IFAC Workshop SAFEPROCESS 2000. Budapest, Hungary.

[7] Cordier M-O., Dague P., Lévy F., Dumas M., Montmain J., Staroswiecki M. and Travé-Massuyès L. (2002). Conflicts versus analytical redundancy relations – A comparative analysis of the model based diagnosis approach from the AI and Automatic Control perspective. IEEE Trans. on SMC Part B, Vol. 34, n.5, pp.2163-2177.

[8] Dressler O., Struss P. (2003). A toolbox integrating model-based diagnosability analysis and automated generation of diagnostics. In Proc. of the 14th Int. Workshop on Principles of Diagnosis DX'03, June 11-14, Washington, D.C., USA, pp. 99-104.

[9] Frisk E. and Düstegör D. and Krysander M. and Cocquempot V. (2003). Improving fault isolability properties by structural analysis of faulty behavior models: application to the DAMADICS benchmark problem. In Proc. of IFAC Safeprocess'03. Washington, USA.

[10] Gissinger, G.L., Loung, M. and Reynaund, H.-F. (2000). Failure Detection and Isolation - Optimal design of instrumentation system. In Proc. IFAC Workshop SAFEPROCESS 2000. Budapest, Hungary.

[11] Krysander M., Nyberg M. (2002). Structural analysis utilizing MSS sets with application to a paper plant. In Proc. of the 13th Int. Workshop on Principles of Diagnosis DX'02, May 2-4, Semmering, Austria, pp. 51-57.

[12] Luong M., Maquin D., Ragot J. (1997). Sensor network design for failure detection and isolation. 3rd IFAC Symposium SICICA'97, Annecy, France.

[13] Nybert, M. (2002). Criterions for detectability and strong detectability of faults in linear systems", Int. J. Control, 2002, VOL. 75, 490-501.

[14] Travé-Massuyès, L., Escobet, T. and Milne, R. (2001). Model Based Diagnosability and Sensor Placement – Application to a Frame 6 Gas Turbine Subsystem. In Proc. of IJCAI'01, Seattle (US), August 2001.

[15] Travé-Massuyès, L., Escobet, T. and Spanache S. (2003). Diagnosability analysis based on component supported analytical redundancy relations. In Proc. of SAFEPROCESS'03, Washington, D.C. (US), June 2003.

[16] Travé-Massuyès, L. and P. Dague (2003). Raisonnement Causal en Physique Qualitative, Chap. 4 in "Modèles et Raisonnements Qualitatifs", Hermès, Traité IC2 Information, Commande, Communications, N°ISBN 2-7462-0744-3, 361p.

[17] Spanache, S., Escobet, T., de las Heras, S., (2002). Structural Analysis of the Damadics Benchmark. 4th DAMADICS Workshop in Bochum, Germany.

[18] Spanache, S., Escobet, T., Travé-Massuyès, L., (2004). Sensor optimization using genetic algorithms. Proc. Of the 15th Int. Workshop on Diagnosis Principles DX'04, Carcassonne, France.

[19] Syfert, M., Patton, R.J., Bartyœ, M., Quevedo, J (2003). "Development and application of methods for actuator diagnosis in industrial control systems (DAMADICS): A benchmark study", In Proc. of IFAC Workshop SAFEPROCESS 2003. Washington, D.C., USA, pp. 939-950.

[20] Struss P., Rehfus B., Brignolo R., Cascio F., Console L., Dague P., Dubois P., Dressler O., Millet D. (2002). Model-based tools for the integration of design and diagnosis into a common process – A project report, Proc. of DX'02, Semmering, Austria.

Louise Travé-Massuyès received a Ph.D. degree in control in 1984 and an Engineering Degree specialized in control, electronics and computer science in 1982, both from the *Institut National des Sciences Appliquées (INSA)* in Toulouse, France; Award from the *Union des Groupements d'Ingenieurs de la Region Midi-Pyrénées*. She received an « Habilitation à Diriger des Recherches » from Paul Sabatier University in 1998.

She is currently Research Director of *Centre National de Recherche Scientifique (CNRS),* working at *LAAS*, Toulouse, France, in which she is the scientific leader of the "Qualitative Diagnosis, Supervision and Control" Group for several years. Her main research interests are in Qualitative and Model-Based Reasoning and applications to dynamic systems Supervision and Diagnosis. She has been particularly active in bridging the AI and Control Engineering Model-Based Diagnosis communities, as leader of the BRIDGE Task Group of the MONET European Network. She has been responsible for several industrial and european projects and published more than 100 papers in scientific journals and international conference proceedings.

Dr. Travé-Massuyès current responsibilities include; member of the *IFAC Safeprocess* Technical Committee; co-leader of the Join Industry-Academic Laboratory AutoDiag, co-leader of the French CNRS Imalaia Group; member of the French Universities Council for the area "Computing, Control, and Signal Processing". She is a Senior Member of the *IEEE* Computer Society. Her e-mail address is louise@laas.fr.

Teresa Escobet received a Ph.D. in Industrial Engineering in 1997 and the Industrial Engineering Degree in 1989, both from the *Universitat Politècnica de Catalunya* in Barcelona, Spain.

She is currently Assistant Professor at *Universitat Politècnica de Catalunya,* lecturer in Automatic Control at *Escola Universitaria Politècnica de Manresa*, Spain. She is research member of the Advanced Control Systems research group of the Automatic Control Department in *Universitat Politècnica de Catalunya*. Her main research interests are in Dynamic System Modeling and Identification applied to Fault Detection, Isolation and Fault Tolerant Control. She has been involved in several European projects and networks and has published several papers in international conference proceedings and scientific journals.

Dr. Escobet current responsibilities include Assistant Director of the Automatic Control Department of *Universitat Politècnica de Catalunya*. Her e-mail address is teresa.escobet@upc.es.

Xavier Olive received a Ph.D. degree in control in 2003 from Paul Sabatier University in Toulouse, France, a D.E.A. in Control from the University of Montpellier and an Engineering Degree in Control from Ecole des Mines d'Alès, Alès, France, both in 2000.
He is currently working as an engineer at Alcatel Space Industries. His main research interests are in Model based Diagnosis, data handling systems and particularly Fault Detection, Isolation and Reconfiguration, planning and scheduling, applied to satellites' platform.