Sensor Placement for Fault Diagnosis

Mattias Krysander and Erik Frisk

***Abstract*—An algorithm is developed for computing which sensors to add to meet a diagnosis requirement specification concerning fault detectability and fault isolability. The method is based only on the structural information in a model, which means that possibly large and nonlinear differential–algebraic models can be handled in an efficient manner. The approach is exemplified on a model of an industrial valve where the benefits and properties of the method are clearly shown.**

***Index Terms*—Fault diagnosis, fault isolation, sensor placement.**

I. Introduction

FAULT diagnosis and process supervision are an increasingly important topic in many industrial applications and also in an active academic research area. The nature of a model-based diagnosis system is highly dependent on the type of model that is used. For works based on continuous differential/difference-equation-based models (see, e.g., see [1] and [2] and the references therein for discrete-event models [3], [4] and for diagnosis of hybrid systems [5]). To be able to perform model-based supervision, some redundancy is needed, and this redundancy is typically provided by sensors mounted on the process. Scientific attention has mainly been devoted to design a diagnosis system given a model of a process equipped with a set of sensors. Not much attention has yet been devoted to deciding which sensors to include in the process.

Deciding where to put sensors correctly, which makes it possible to meet a given diagnosis performance specification, is the topic of this paper. There are many types of performance measures in diagnosis, for example, detection performance, false-alarm probabilities, time to detection, etc. In this paper, sensors are placed such that maximum isolability is possible, i.e., faults in different components should, as far as possible and desired, be able to be isolated from each other. Since sensor placement is often done early in the design phase, possibly before a reliable process model can be developed, the method developed in this paper is based on a structural process model. This is a coarse model description that can be obtained early and without major engineering efforts. Also, this means that large and nonlinear differential–algebraic models can be handled in an efficient manner. The drawback with structural methods is that only best case results are obtained, (see [6] for a more in-depth discussion on this).

The main objective of this paper is to develop an algorithm that, from a given model and a specified detectability and isolability performance specification, computes a characterization of all possible sets of sensors, which makes it possible to meet the requirement specification.

This paper is organized as follows. A formal problem formulation is presented in Section II. Section III gives a background of the theoretical tools used in the development of the method in Section IV. The algorithm[1] is then summarized in Section V and thoroughly exemplified on an industrial valve model in Section VI. Relations to other published related works are discussed in Section VII, and some conclusions are given in Section VIII.

II. Problem Formulation

Before the main objective of this paper is formally presented, a small example is discussed that illustrates the fundamental problems in sensor placement for fault diagnosis. The example is modeled by a fifth-order linear system of ordinary differential equations. This example will be used throughout this paper, although the results will be equally applicable to large-scale nonlinear differential–algebraic models. The model consists of the following

$$\begin{aligned}
e_1: &\quad \dot{x}_1 = -x_1 + x_2 + x_5\\
e_2: &\quad \dot{x}_2 = -2x_2 + x_3 + x_4\\
e_3: &\quad \dot{x}_3 = -3x_3 + x_5 + f_1 + f_2\\
e_4: &\quad \dot{x}_4 = -4x_4 + x_5 + f_3\\
e_5: &\quad \dot{x}_5 = -5x_5 + u + f_4
\end{aligned}$$

where x_i are the state variables, u is a known control signal, and f_i are the faults that we want to detect and isolate. Since there are no specified sensors, there is no redundancy, and the faults are not detectable.

In this example, faults are modeled by fault signals that are included in the model equations, and $f_i \neq 0$ indicates a fault. A more general way to include faults is to assign assumptions, or support, to the equations. This type of fault modeling can also easily be used with the approach that will be presented later, but for the sake of simplicity, fault signal modeling will be used in this paper. Also, from now on, only single faults will be considered, and f_i will then be used to denote both the fault signal and the fault mode.

Let F denote the set of faults. A detectability performance specification is then a set $F_{\text{det}} \subseteq F$ specifying the detectability

Manuscript received January 8, 2007; revised February 13, 2007, June 28, 2007, and May 7, 2008. Current version published October 20, 2008. This paper was recommended by Associate Editor G. Biswas.

M. Krysander and E. Frisk are with the Department of Electrical Engineering, Linköping University, 581 83 Linköping, Sweden (e-mail: matkr@isy.liu.se; frisk@isy.liu.se).

Digital Object Identifier 10.1109/TSMCA.2008.2003968

[1] A Matlab implementation, released under GNU General Public License, of the algorithm presented in this paper is available at http://www.fs.isy.liu.se/Software/SensPlaceTool/.

1083-4427/$25.00 © 2008 IEEE

requirement, and an isolability requirement is a set I of ordered pairs $(f_i, f_j) \in F_{\text{det}} \times \mathcal{F}_{\text{det}}$, meaning that f_i is isolable from f_j. Note that we assume that all faults that are included in the isolability specification I are also required to be detectable.

Since the fault isolability capability always increases when adding new sensors, there are minimal elements in the family of sensor sets that achieves a certain level of fault isolability. Therefore, we define *minimal sensor set* as a minimal set of sensors to add to achieve a specified performance specification.

Definition 1 (Minimal Sensor Set): Let $\mathcal{S}$ be the set of possible sensor locations, i.e., the set of measurable variables, and let S be a multiset defined on $\mathcal{S}$. Then, S is a minimal sensor set, with respect to a given detectability and isolability specification, if adding the sensors in S fulfills the specification and all proper subsets of S do not.

Note that S is a multiset, which is similar to a set but allows multiple instances of a member. Generalizations of the standard set operations like union and intersection are straightforward. Multisets are used instead of regular sets since it may be necessary to add more than one sensor measuring the same variable.

Returning to the example, the first question is: What are the minimal sensor sets achieving detectability of all faults? Here, it is assumed that sensors measure a state variable or a function thereof. It can be verified, using conditions for fault detectability in linear systems [7], that $\{x_1\}$, $\{x_2\}$, $\{x_3, x_4\}$ are minimal sensor sets achieving detectability. This means that, by adding any of these sensor sets, it is possible to generate a residual, i.e., a signal used as a fault indicator, that is sensitive to all the faults.

The second step is to require not only detectability but also isolability properties. Here, isolability refers to isolability as it is commonly used in FDI and the AI community. Formally defining isolability involves many aspects (see, e.g., [8]). In the context of this paper, it is sufficient to know that if fault f_i is isolable from fault f_j, then there exists a residual that is sensitive to f_i but not f_j. For details on how isolability is formally defined in this paper, see Sections III and IV. In the example, it can be verified that there are five minimal sensor sets that achieve maximal fault isolation: $\{x_1, x_3\}$, $\{x_1, x_4\}$, $\{x_2, x_3\}$, $\{x_2, x_4\}$, and $\{x_3, x_4\}$. Thus, adding sensors measuring the variables in any of these sets, or a superset of the variables, achieves maximum fault isolability.

Now, it is of course the case that the new sensors may also become faulty. If we want also faults in the new sensors to be isolable from the other faults, we may have to add additional sensors. In this case, if maximum fault isolability is desired also for faults in the new sensors, there are nine minimal sensor sets where one sensor set is two sensors measuring x_1 and one for x_3, i.e., the multiset $S = \{x_1, x_1, x_3\}$ is a minimal sensor set.

Based on this introductory example, the problem formulation of this paper can now be stated as follows.

> Given a model, possible sensor locations, and a detectability/isolability performance specification, find all minimal sensor sets with respect to the required specification.

The methods developed in the sections that follow aim at addressing this problem for general nonlinear differential-algebraic models. Doing this analytically has been difficult since inference concerning solutions to the model equations has been needed. Instead, a method based on utilizing only the structure of the model is employed. This gives generic results that hold in a best case situation. An advantage is that very large models can be handled in an efficient manner. See Section III for some further results on the relation between structural and analytical properties of a model. See also [6] for an in-depth discussion on this topic.

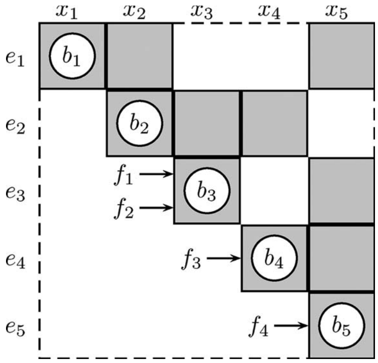

Fig. 1. Structure of the linear example in Section II. Gray areas indicate nonzero elements.

III. Theoretical Background

The sensor placement problem will be solved here using a structural representation of the model. The structural representation of a set of equations M with unknown variables X's is a bipartite graph, with variables and equations as node sets. The known variables are, in this paper, omitted in the structure because they will not be needed for the analysis. There is an edge in the graph between a node representing an equation $e \in M$ and a node representing an unknown variable $x \in X$ if the variable x is contained in e. For notational convenience, we will denote the node representing an equation e or a variable x simply by the equation name e and the variable name x, respectively. A bipartite graph can be described by a biadjacency matrix where the rows and columns correspond to the node sets, and the position (i, j) is one if there is an edge between nodes i and j; otherwise, it is zero.

The structure of the example formulated in Section II is shown in Fig. 1 as a biadjacency matrix of the bipartite graph. The position (e_i, x_j) is one if x_j or any time derivative appears in equation e_i. This structural representation of dynamical systems has been used in, for example, [9] and [10]. There exist other structural representations of dynamical systems, but the one used here is a compact representation suitable for the sensor placement problem [6].

A. Dulmage–Mendelsohn Decomposition

The objective of this section is to introduce notation and a basic theoretical tool, the Dulmage–Mendelsohn decomposition [11], regarding structural models and bipartite graphs that will be used in the coming sections. The decomposition is shown in Fig. 2.

The decomposition defines a partition $(M_0, M_1, \ldots, M_n, M_\infty)$ of the set of equations M, a similar partition of the set of

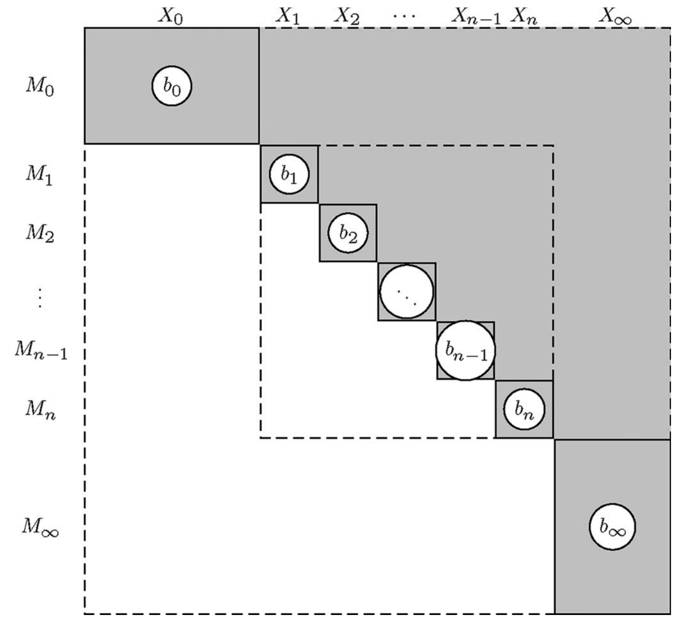

Fig. 2. Dulmage–Mendelsohn decomposition.

unknowns X's, and a partial order on sets M_i. If the rows and columns are rearranged according to this order, the biadjacency matrix has the upper block triangular form shown in Fig. 2. There are zero entries in the white parts of the matrix, and there might be ones in the gray-shaded parts.

Three main parts of M can be identified in the partition, namely, M_0 is called the structurally underdetermined part, $\cup_{i=1}^{n} M_i$ is the structurally just-determined part, and M_∞ is the structurally overdetermined part. Not all parts may be present in a given model, for example, the structure in Fig. 1 only contains a just-determined part.

In Fig. 2, each pair (M_i, X_i) is related to a block which is denoted by b_i. The blocks $b_i,\ i = 1, \ldots, n$ in the just-determined part are called strongly connected components. In this paper, the just-determined and overdetermined parts are of particular interest. This is because, in the overdetermined part, there are more equations than unknown variables, which implies that there exists some degree of redundancy, and this is the part of the model that is useful for monitoring the process. It will thus be convenient to define an operator $(\cdot)^+$ that extracts the overdetermined part of a set of equations, i.e., $M^+ = M_\infty$. Faults that influence the just-determined part are not detectable, and the structure of the just-determined part will be instrumental in determining the minimal sensor sets defined in Section II. A characteristic property of the just-determined part is that there are equally many equations as unknown variables in blocks $b_1, \ldots, b_n$. To prove some of the results in this paper, a formal definition of the decomposition is needed, and a brief description is included in the Appendix.

B. Structural Formulation of Fault Diagnosis

In this section, we will give structural characterizations of fault diagnosis properties. By doing this, the sensor placement problem can be formulated as a graph theoretical problem.

Let M and F denote a set of equations and a set of single faults, respectively. Without loss of generality, it is possible to assume that a single fault can only violate one equation. If a fault signal f appears in more than one equation, we simply replace f in the equations with a new variable x_f and add equation $f = x_f$ which will then be the only equation violated by this fault. An example of this procedure is also given in the example in Section VI. Let $e_f \in M$ be the equation that might be violated by a fault $f \in F$. For the example introduced in Section II, $e_{f_1} = e_{f_2} = e_3$, $e_{f_3} = e_4$, and $e_{f_4} = e_5$.

A fault f is detectable if there exists an observation that is consistent with fault mode f and inconsistent with the no-fault mode. This means that a detectable fault can violate a monitorable equation in the model describing the fault-free behavior. Since an equation is, in the generic case, monitorable if it is contained in the structurally overdetermined part of M, structural detectability can be defined as follows [1].

Definition 2: A fault f is structurally detectable in a model M if $e_f \in M^+$.

Returning to the example and illustrating the correspondence between detectable faults and structurally detectable faults, assume that a sensor y measuring x_4 has been added to the process and included in the model by $e_6 : y = x_4$. Faults f_3 and f_4 are the detectable faults, and a residual that is capable of detecting them is

$$r = 20y + 9\dot{y} + \ddot{y} - u = 5f_3 + \dot{f}_3 + f_4$$

which is, in fact, the only residual generator for this model modulo postfiltering. Thus, faults f_1 and f_2 are not detectable.

The structurally overdetermined part M^+ of the model $M = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ is equal to $\{e_4, e_5, e_6\}$. The equations $e_{f_3} = e_4$ and $e_{f_4} = e_5$ corresponding to the detectable faults f_3 and f_4 belong to M^+, but not the equations corresponding to the other faults. This implies, according to Definition 2, that the detectable faults f_3 and f_4 are the structurally detectable faults in M which is in accordance with the analytical result earlier.

Detection is a special case of isolation, i.e., a fault is detectable if the fault is isolable from the no-fault mode. By noting this similarity, it holds that a fault f_i, isolable from f_j, can violate a monitorable equation in the model describing the behavior of the process having a fault f_j. The set of equations valid with a fault f_j is $M \setminus \{e_{f_j}\}$, and the monitorable part of these equations is, in the generic case, equal to $(M \setminus \{e_{f_j}\})^+$. This motivates the following structural characterization of isolability.

Definition 3: A fault f_i is structurally isolable from f_j in a model M if

$$e_{f_i} \in \left(M \setminus \{e_{f_j}\}\right)^+. \tag{1}$$

The structural detectability and isolability definitions will next be used in a structural approach for solving the sensor placement problem.

IV. Structural Approach

Theoretical results and an algorithm outline to solve the problem posed in Section II are formulated here using the theory in Section III. A complete description of the algorithm is then given in Section V.

A general assumption of the approach is that the model does not contain any underdetermined part. This is not a restrictive assumption since any complete physical model will, given an initial condition, have a unique solution and thereby no underdetermined part. Without loss of generality, it is also assumed that no fault affects more than one equation and that possible sensors measure a function of one unknown variable. In case there are possible sensors that measure some function h of more than one unknown variable, include a new equation $x_{new} = h(x)$ in the model.

A. Sensor Placement for Detectability

A basic building block in the final algorithm will be to find minimal sensor sets that achieve structural detectability of faults in an exactly determined set of equations. This section will be devoted to solving this subproblem by first outlining the solution for the example system from Section II and then formally proving the solution. Although the example is given by analytical equations, all results in this section are based on the structural model only.

The example model is, without any additional sensors, an exactly determined set of equations with five equations and five unknown variables x_i, i.e., all faults are undetectable. Consider first the fault f_3. To make this fault detectable, according to Definition 2, an additional sensor is needed such that equation $e_{f_3} = e_4$ becomes a member of the overdetermined part of the model.

It is straightforward to verify that f_3 becomes detectable if and only if any of the variables $\{x_1, x_2, x_4\}$ are measured. For example, measuring x_4 makes the new measurement equation, together with equations e_4 and e_5, an overdetermined set of equations. For this set of equations, a residual generator which is sensitive to fault f_3 can easily be derived. A similar line of reasoning can be made when measuring x_1 or x_2.

Then, why are x_1, x_2, and x_4 exactly those measurements that give detectability of f_3? The explanation can be seen in Fig. 1 where it can be noted that block b_1 is connected to b_2 via a nonzero element in position (1, 2) and that b_2 is connected to b_4 in a similar fashion. Thus, there is a connection between variables x_1, x_2, and x_4, which is precisely the variable in block b_4 including fault f_3. Measuring x_3, i.e., the variable in b_3, does not give detectability of f_3 since there is no connection between b_3 and block b_4.

The aforementioned reasoning indicates that some order between the strongly connected components is needed, and (15) in the Appendix formally defines such an ordering. Fig. 3 shows the Hasse diagram of the partial order of the strongly connected components for the example. Thus, for example, $b_5 \leq b_3$, and there is no order between b_3 and b_4. This ordering makes it possible to state exactly which equations, in an exactly determined model, that become overdetermined when adding a sensor.

The following lemma formalizes the previous discussion. This also gives, according to Definition 2, which faults that become detectable as a result of adding a sensor. Before the lemma can be stated, some notation is needed. Each block b_i is directly related to an equation set M_i, as defined in Section III-A, and therefore, an order is implicitly defined on the sets M_i. See the Appendix for a formal definition.

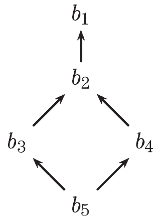

Fig. 3. Hasse diagram of the partial order over the set of strongly connected components B.

Lemma 1: Let M be an exactly determined set of equations, b_i be a strongly connected component in M with equations M_i, and $e \notin M$ be an equation corresponding to measuring any variable in b_i. Then

$$(M \cup \{e\})^+ = \{e\} \cup (\cup_{M_j \leq M_i} M_j). \tag{2}$$

Proof: The proper overdetermined part $(M \cup \{e\})^+$ is defined by the minimal subset of $M \cup \{e\}$ with maximum surplus. The maximum surplus of all subsets of M is zero. By adding one equation e, we know that the maximum surplus of any subset of $M \cup \{e\}$ is at most one. Since $\mathrm{var}(\{e\}) \subseteq \mathrm{var}(M)$, it follows that $\varphi(M \cup \{e\}) = 1$. Hence, the minimal set with surplus one is the proper overdetermined part of $M \cup \{e\}$. Any such set contains e since all other sets have surplus less than or equal to zero. This means that the sought set can be written as $E \cup \{e\}$, where $E \subseteq M$. Since the surplus of $E \cup \{e\}$ is one and the surplus of E can be at most zero, it follows that the surplus of E is zero. Let $\mathcal{L}$ be a sublattice of the subset lattice of M defined similar to the set defined in (12). This means that $E \in \mathcal{L}$. Furthermore, $\varphi(E \cup \{e\}) = 1$ only if $\mathrm{var}(\{e\}) \subseteq \mathrm{var}(E)$. This implies that $M_i \subseteq E$. The minimal set E in $\mathcal{L}$ such that $M_i \subseteq E$ is according to (14) $E = \cup_{M_j \leq M_i} M_j$, and this completes the proof. ■

Achieving detectability of one fault affecting a strongly connected component immediately implies detectability of all faults affecting the same component. Therefore, it makes sense to define an equivalence relation on the set of faults, where all faults influencing the same strongly connected component are equivalent. A set of equivalent faults is denoted as $[f_i]$, where f_i is one element in the equivalence class. Now, based on Lemma 1, it is clear that measuring a variable in a block ordered higher than the block where the fault enters achieves detectability. Now, let $P \subseteq X$ be a set of possible sensor locations and introduce the set

$$D([f_i]) = \{x | b_i \leq b_j, x \in X_j \cap P\} \tag{3}$$

where X_j is the set of variables corresponding to block b_j according to Section III-A and b_i is the block that is influenced by the faults in $[f_i]$. The set $D([f_i])$ is thus the set of variables such that measuring *any* variable in the set achieves detectability of all faults in the equivalence class $[f_i]$. Note that $D([f_i])$ is also a function of the specification P. However, since the specification is fixed, for notational convenience, this dependence will not be explicitly stated.

Returning to the example and utilizing the previous result, one can see that detectability of f_4 comes automatically when adding sensors to achieve detectability of either the faults in $[f_1]$ or $[f_3]$. This is because b_5 is less than or equal to both b_3

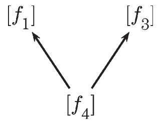

Fig. 4. Hasse diagram of the partial order for the linear example over the set of fault classes. In Fig. 1, it can be seen that $[f_1] = [f_2]$. Classes $[f_1]$ and $[f_3]$ are the maximal elements of the partial order.

and b_4, and according to Lemma 1, block b_5 is automatically included in any overdetermined set of equations when $[f_1]$ or $[f_3]$ are made detectable. This means that it is only necessary to ensure detectability for a subset of the fault classes to ensure detectability of all faults. To illustrate exactly which classes, introduce an order on the equivalence classes of F, defined as $[f_i] \leq [f_j]$ if $b_i \leq b_j$, where b_k is the block where the faults in $[f_k]$ enter the model. Fig. 4 shows the Hasse diagram of the partial order for the example model. Here, one can see that, in the example, it is necessary and sufficient to ensure detectability of the maximal elements of the partial order. In the example, the set of possible sensor locations is X, but with a P that is a proper subset of X, one might have the case where a maximal fault class is not detectable regardless of which sensors in P are added. In such a case, one needs to consider the maximal elements among the detectable fault classes.

The following theorem proves the general result and summarizes the discussion of this section.

Theorem 1: Let M be an exactly determined set of equations, F be the corresponding set of faults, $P \subseteq X$ be the set of possible sensor locations, and M_S be the equations corresponding to adding a set of sensors S. Then, maximal detectability of F in $M \cup M_S$ is obtained if and only if S has a nonempty intersection with $D([f])$ for all $[f] \in F_m$, where F_m is the set of maximal fault classes among the fault classes with $D([f]) \neq \emptyset$.

Proof: First, note that faults in fault classes with $D([f]) = \emptyset$ cannot be made detectable with any of the available sensor locations. Therefore, let F_m be, among the fault classes with $D([f]) \neq \emptyset$, the set of maximal elements with respect to the partial order. Then, maximal fault detectability is obtained if and only if the fault classes in F_m are detectable. This follows from Lemma 1 and Definition 2 which state that if a sensor is added such that a fault in a higher ordered fault class is detected, detectability for the lower ordered fault classes is also obtained.

Furthermore, Lemma 1 also states that a fault f in F becomes detectable if and only if we measure at least one unknown variable in blocks that are greater or equal than the block that includes the fault equation, i.e., if we measure a variable in $D([f])$. A sensor addition that makes all faults in F detectable must thus have a nonempty intersection with $D([f])$ for all $[f] \in F_m$. ■

The previous result can be summarized in an algorithm that, given a model M, a set of faults F and a set of possible sensor locations P compute the family of detectability sets $\mathcal{D}$.

```
function $\mathcal{D}$ = Detectability($M, F, P$)
    Compute block and fault class orders using $M$;
    $F_m$ = set of maximal fault classes among $[f]$ s.t. $D([f]) \neq \emptyset$;
    $\mathcal{D} = \{D([f]) | [f] \in F_m\}$;
```

Our objective was not to compute the set of detectability sets $\mathcal{D}$, but rather minimal sensor sets. For this, note that a hitting set for a family of sets is a set that has nonempty intersection with each set in the family. Thus, a minimal hitting set algorithm [12], [13] applied to the family of sets $\mathcal{D}$ can be used to find all minimal sensor sets.

For the example model, as previously noted, the maximal fault classes are $[f_1]$ and $[f_3]$, and the corresponding detectability sets are

$$D([f_1]) = \{x_1, x_2, x_3\} \quad D([f_3]) = \{x_1, x_2, x_4\}.$$

Theorem 1 gives that the minimal sensor sets that achieve detectability of all faults are $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$, which are the same sensor sets as was determined in Section II.

B. Sensor Placement for Isolability of Detectable Faults

This section describes the basic ideas of how to find the minimal sensor sets such that maximum single-fault isolability is obtained under the assumption that all faults are structurally detectable. In the next section, this assumption will be removed.

The problem of achieving maximum isolability of the set of single faults F can be divided into $|F|$ subproblems, one for each fault, as follows. For each fault $f_j \in F$, find all measurements that make the maximum possible number of faults $f_i \in F \setminus \{f_j\}$ isolable from f_j. The solution to the isolability problem will then be obtained by combining the results from all subproblems.

Each subproblem can be formulated as a detectability problem, as will be shown next. Assume that M is a model, including sensors such that all faults are detectable, and M_S represents a set of equations describing an additional sensor set S. Given the sensor set S, a fault f_i is isolable from f_j in the model $M \cup M_S$ if

$$e_{f_i} \in \left((M \setminus \{e_{f_j}\}) \cup M_S\right)^+ \tag{4}$$

according to Definition 3. By introducing $M' = M \setminus \{e_{f_j}\}$, this can be written as

$$e_{f_i} \in (M' \cup M_S)^+ \tag{5}$$

which, according to Definition 2, means that f_i is structurally detectable in $M' \cup M_S$. Hence, the maximum possible number of faults $f_i \in F \setminus \{f_j\}$ is isolable from f_j in $M \cup M_S$ if the maximum possible number of faults $f_i \in F \setminus \{f_j\}$ is structurally detectable in the model $(M \setminus \{e_{f_j}\}) \cup M_S$. This shows that each subproblem can be formulated as a detectability problem.

Next, we use the example formulated in Section II to outline the solution of one subproblem before formally proving the solution. Assume that we have added sensors measuring $\{x_3, x_4\}$ such that all faults are detectable. Furthermore, assume that these sensors can be faulty and denote these faults f_5 and f_6, respectively. A permuted row structure of the obtained model $M = \{e_1, e_2, \ldots, e_7\}$ is shown in Fig. 5.

Consider the subproblem associated with fault f_1. The set M' in (5) is equal to $M \setminus \{e_{f_1}\} = M \setminus \{e_3\}$. The subproblem is, given the model $M \setminus \{e_3\}$, to find the minimal additional

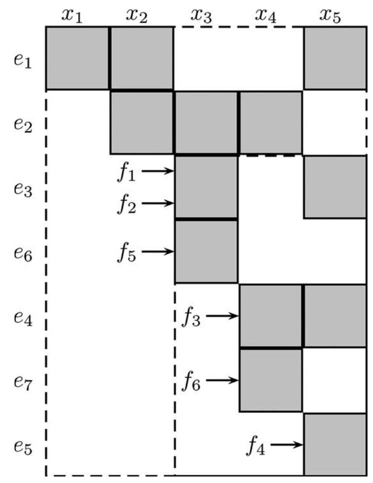

Fig. 5. Block structure of the example in Section II extended with measurements of x_3 and x_4.

sensor sets S's such that as many of the faults $f_2, f_3, \ldots, f_6$ as possible become detectable in $M' \cup M_S$.

The faults can, depending on which equations they violate, be divided into the following three types: faults that do not violate any equation in M', faults that violate equations in the structurally overdetermined part $(M')^+$, and faults that violate other equations in M', i.e., $M' \setminus (M')^+$. In the example, we have that $(M')^+ = \{e_4, e_5, e_7\}$ and $M' \setminus (M')^+ = \{e_1, e_2, e_6\}$ which is equal to the structurally just-determined part of M'. This implies that f_2 is not included in M'; f_3, f_4, and f_6 belong to the structurally overdetermined part; and f_5 belongs to the structurally just-determined part. Fault f_2 is not included in M' and cannot be structurally detectable in $M' \cup M_S$ for any sensor set S. This implies that f_2 is not isolable from f_1 with any sensor addition, and this also follows from the fact that these two faults violate the same equation. Faults f_3, f_4, and f_6 in the structurally overdetermined part $(M')^+$ are, according to Definition 2, structurally detectable in M' and require no additional measurements. Fault f_5 in the just-determined part is not detectable, but f_5 can become detectable in $M' \cup M_S$ if S is appropriately selected.

Sufficient and necessary requirements on S can be computed by the function Detectability described in Section IV-A. By applying this function to the structurally just-determined part of M', i.e., the subgraph of M' defined by the node sets $\{e_1, e_2, e_6\}$ and $\{x_1, x_2, x_3\}$, we get that $D([f_5]) = \{x_1, x_2, x_3\}$. Hence, one of the variables in the detectability set $\{x_1, x_2, x_3\}$ must be measured to make the faults $F \setminus \{f_1, f_2\}$ detectable in $M' \cup M_S$, and this implies that all faults in $F \setminus \{f_1, f_2\}$ are isolable from f_1 in $M \cup M_S$. The solution to the subproblem related to fault f_1 will be the computed detectability set. The next lemma formalizes the solution of a subproblem like the one discussed previously.

Theorem 2: Let M be a set of equations with no structurally underdetermined part, F be a set of structurally detectable faults in M, $P \subseteq X$ be the set of possible sensor locations, and M_S be the equations added by adding the sensor set S. For an arbitrary fault f_j, assume that M^0 is the just-determined part of $M \setminus \{e_{f_j}\}$, F^0 is the set of faults contained in M^0, and $\mathcal{D} = \mathtt{Detectability}(M^0, F^0, P)$. Then, the maximum possible number of faults $f_i \in F \setminus \{f_j\}$ is structurally isolable from f_j in $M \cup M_S$ if and only if S has a nonempty intersection with all sets in $\mathcal{D}$.

Proof: Given a sensor set S, a fault f_i is structurally isolable from f_j if (4) holds according to (1). This is equivalent to saying that f_i is structurally detectable in $(M \setminus \{e_{f_j}\}) \cup M_S$. Since all faults are structurally detectable, it follows that $e_{f_j} \in M^+$. This implies that the underdetermined part of $M \setminus \{e_{f_j}\}$ is empty. The faults in the structurally overdetermined part of $M \setminus \{e_{f_j}\}$ are, according to Definition 2, structurally detectable. From Theorem 1, maximal detectability of faults F^0 in the structurally just-determined part M^0 of $M \setminus \{e_{f_j}\}$ is obtained if and only if S has a nonempty intersection with all detectability sets contained in $\mathcal{D} = \mathtt{Detectability}(M^0, F^0, P)$. ■

The result of the theorem can be summarized in a function that, given a model M, a set of detectable faults F in M, a set of possible sensor locations P, and a fault $f \in F$, computes the family of detectability sets $\mathcal{D}$ that solves the isolability subproblem for f.

```
function $\mathcal{D} = \mathtt{IsolabilitySubProblem}(M, F, P, f)$
  $M^0 =$ just-determined part of $M \setminus \{e_f\}$;
  $F^0 =$ the set of faults $F$ included in $M^0$;
  $\mathcal{D} = \mathtt{Detectability}(M^0, F^0, P)$;
```

An additional sensor set that maximizes the set of fault pairs (f_i, f_j) such that f_i is structurally isolable from f_j must have a nonempty intersection with all detectability sets found in all subproblems.

```
function $\mathcal{D} = \mathtt{Isolability}(M, F, P)$
  $\mathcal{D} = \varnothing$;
  for $f_i \in F$
    $F' = F \setminus \{f_i\}$;
    $\mathcal{D} = \mathcal{D} \cup \mathtt{IsolabilitySubProblem}(M, F', P, f_i)$;
  end
```

The minimal sensor sets that maximize the isolability can be found by applying a minimal hitting set algorithm to the sets in the output $\mathcal{D}$.

For the example shown in Fig. 5, the families of detectability sets of the different subproblems are

$$\begin{aligned}&\{\{x_1, x_2, x_3\}\} \text{ for } f_1, f_2, \text{ and } f_5\\&\{\{x_1, x_2, x_4\}\} \text{ for } f_3 \text{ and } f_6\\&\varnothing \text{ for } f_4.\end{aligned} \tag{6}$$

We have found two distinct detectability sets, and the minimal hitting sets are $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$. These sets are the minimal additional measurements that achieve maximum single-fault isolability.

C. Sensor Placement for Both Detectability and Isolability

We have shown how isolability can be achieved in a model where all faults are structurally detectable. Next, we will extend the presented solution to models where faults may not be structurally detectable in the original model.

The solution is first outlined for the example described in Section II. The faults in this model are not detectable, and we want to find all minimal sensor sets that maximize fault detectability and isolability. We have shown in Section IV-A that the minimal sets of measurements to achieve full detectability are $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$. If we add, for example, a sensor measuring x_1 described by an equation e_s, we get a new model $M \cup \{e_s\}$ where all faults are detectable. Since all faults are detectable, the previously described method to achieve maximum isolability can be applied to the model $M \cup \{e_s\}$. The minimal sensor sets that solve this problem are $\{x_3\}$ and $\{x_4\}$. By combining this result with the fact that a sensor measuring x_1 has been added to obtain detectability, it follows that $\{x_1, x_3\}$ and $\{x_1, x_4\}$ are two possible sensor sets that achieve maximum detectability and isolability. To compute all minimal sensor sets that achieve maximum isolability, we also have to investigate the solutions when we choose to measure $\{x_2\}$ or $\{x_3, x_4\}$ to obtain full detectability. By solving one isolability problem for each of the minimal sensor sets that achieves full detectability, we get that the minimal sensor sets are $\{x_1, x_3\}$, $\{x_1, x_4\}$, $\{x_2, x_3\}$, $\{x_2, x_4\}$, and $\{x_3, x_4\}$ which are the same sets as in Section II.

The following description summarizes the suggested algorithm that, given a model M with no structurally underdetermined part, a set of faults F, and a set of possible sensor locations P, computes the family $\mathcal{S}$ of all minimal sensor sets that achieve maximum isolability. In the algorithm, the join operation of two multisets A and B will be used. The join operation is denoted by $A \uplus B$ and is defined as a multiset containing all elements in $A \cup B$ with a multiplicity equal to the sum of the multiplicities in A and B. For example, $\{x_1, x_2\} \uplus \{x_1\} = \{x_1, x_1, x_2\}$.

```
function $\mathcal{S} = \texttt{SensorPlacement}(M, F, P)$
  $\mathcal{S} = \varnothing$;
  $M^0 =$ just-determined part of $M$;
  $F^0 =$ the set of faults $F$ included in $M^0$;
  $\mathcal{D} = \texttt{Detectability}(M^0, F^0, P)$;
  $\mathcal{S}_d =$ MinimalHittingSets $(\mathcal{D})$;
  for $S_i \in \mathcal{S}_d$
    Create the extended model $M_e = M \cup M_{S_i}$;
    $F_e =$ the faults included in $M_e$;
    $\mathcal{D} = \texttt{Isolability}(M_e, F_e, P)$;
    $\mathcal{S}_i =$ MinimalHittingSets$(\mathcal{D})$;
    $\mathcal{S} = \mathcal{S} \cup \{S_i \uplus S' | S' \in \mathcal{S}_i\}$;
  end
  Delete nonminimal sensor sets in $\mathcal{S}$;
```

D. Efficiency Improvements

All operations in the algorithm are polynomial except for the minimal hitting set algorithm which is NP-hard. This means that a worst case might be intractable. However, well-formed models of physical systems typically have a structure which makes the computations less demanding. Also, and this is maybe the most important aspect, the number of measurable signals in the specification P is a prime indicator of the complexity. Thus, it is not primarily the number of equations or the number of faults but rather the user-specified sensor specification that controls the complexity.

In addition, the basic algorithm can be made more efficient by avoiding multiple computations of some detectability sets. In (6), we can see that the detectability sets in several of the subproblems coincided, and therefore, one might suspect that the function `Isolability` can be improved in terms of efficiency. In this section, we will investigate the properties of structural isolability that will be used to reduce the computational complexity of the function `Isolability`.

Consider the model in Fig. 5. An example of two subproblems that resulted in the detectability set $\{x_1, x_2, x_4\}$ are the subproblems related to f_3 and f_6. This is not a coincidence, and the reason why this happens will be explained next.

First, note that f_3 is not isolable from f_6 in the structure in Fig. 5 because

$$e_{f_3} \notin (M \setminus \{e_{f_6}\})^+ = M^+ \setminus \{e_{f_3}, e_{f_6}\} \tag{7}$$

and vice versa. Hence, we need to find a detectability set for making f_3 isolable from f_6, and the one for achieving that f_6 becomes isolable from f_3. These detectability sets are equal to $\{x_1, x_2, x_4\}$, and one might suspect that f_3 is isolable from f_6 if and only if f_6 is isolable from f_3. This symmetry of the isolability relation will next be shown to hold for detectable faults in general. To do this, a partition of an overdetermined part will first be defined.

A key property in the determination of structural isolability is the set $(M \setminus \{e_{f_j}\})^+$ which is determined by the result of the combined operation of removing an equation and then computing the overdetermined part. The resulting set of the combined operation has been studied in [14] and can be characterized as follows. There exists a partition $(M_1, M_2, \ldots, M_p)$ of the overdetermined part M^+ such that, for any equation $e \in M_k$, it holds that

$$(M \setminus \{e\})^+ = M^+ \setminus M_k \tag{8}$$

By comparison of (7) and (8), we get that $\{e_{f_3}, e_{f_6}\}$ is one set in the partition of M^+ in the example. Both faults f_3 and f_6 violate equations in the same set of the partition, and none of these faults is isolable from the other fault, i.e., these faults are indistinguishable. Next, we prove that this holds in general.

Theorem 3: Given a model M, let f_i and f_j be two structurally detectable faults in M. Fault f_i is structurally isolable from f_j if and only if e_{f_i} and e_{f_j} belong to different sets in the partition defined in (8).

Proof: Fault f_i is structurally isolable from f_j if and only if (1) holds according to Definition 3. By using (8), (1) can be expressed as

$$e_{f_i} \in M^+ \setminus M_k \tag{9}$$

where M_k is the set in the partition such that $e_{f_j} \in M_k$. Since f_i is structurally detectable, i.e., $e_{f_i} \in M^+$, it follows that (9) is equivalent to $e_{f_i} \notin M_k$, and this completes the proof. ■

A result of this theorem is that the isolability relation is symmetric on the set of detectable faults, and this is the reason for obtaining the detectability set $\{x_1, x_2, x_4\}$, both when finding sensors for making f_3 isolable from f_6 and vice versa. This implies that we can reduce the isolability analysis to ordered pair of faults. That is, given an enumeration of the faults $F = \{f_1, \dots f_n\}$, line 4 in the function `Isolability` should be replaced by $F' = \{f_j | j > i\}$. Another straightforward improvement is to compute the corresponding subproblem for at most one fault entering the same equation.

For the example, the subproblem for f_2 need not be solved if the subproblem for f_1 is solved, since these problems have the same solution. By considering the order of faults, we get the measurements needed to distinguish f_1 and f_5 in the subproblem related to f_1, and measurements needed to distinguish f_3 and f_6 in the subproblem related to f_3. All other subproblems return the empty family of detectability sets. Hence, in this example, the detectability sets are found only once.

E. Adding Sensors With Faults

Sensors might have corresponding sensor faults. When adding a sensor, it is possible that a new fault is introduced into the model, and in this section, it is shown how these additional sensor faults can be handled in algorithm `SensorPlacement` described in Section IV-C.

Consider again the example introduced in Section II and assume now that we want to find all minimal sensor sets that maximize the fault isolability when all sensors introduce new possible faults. To do this, we will follow the algorithm `SensorPlacement` and describe how some of the lines should be modified to cope with additional sensor faults.

The additional sensors that have a corresponding sensor fault have to be specified in the algorithm. This is done by introducing an additional input set $P_f \subseteq P$ where sensors measuring variables in P_f may become faulty and the other sensors may not.

The purpose of lines 5 and 6 is to compute all sensor sets that achieve full detectability. In Section IV-A, it was shown that $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$ are the minimal sensor sets that make faults $f_1, \dots, f_4$ detectable. No subset of these sensor sets is therefore a solution to the extended problem concerning also sensor faults. To determine if one of these sensor sets is a solution also to the extended problem, assume that x_3 and x_4 are measured. If the measurement equations are called e_6 and e_7, respectively, we obtain the structure in Fig. 5. Let the sensor faults corresponding to the measurements of x_3 and x_4 be denoted by f_5 and f_6 and include these faults in the model as done previously. The structurally overdetermined part $\{e_3, e_4, e_5, e_6, e_7\}$ of the model in Fig. 5 includes the sensor equations, and it follows that the additional sensor faults f_5 and f_6 are detectable and require no additional sensors. Thus, in the example, all sensor faults become detectable, and this holds, in general, according to the following result.

Theorem 4: Let M be a model with no underdetermined part, and let $x \in \mathrm{var}(M)$ be measured with a sensor described by an equation $e \notin M$. Then, a sensor fault violating e will be structurally detectable in $M \cup \{e\}$.

Proof: The sensor fault is structurally detectable if $e \in (M \cup \{e\})^+$. Since there is no underdetermined part in M, it follows that $\varphi(M)$ is equal to the maximal surplus for any set contained in M. The maximal surplus of any set in $M \cup \{e\}$ is $\varphi(M) + 1$. Any set with surplus $\varphi(M) + 1$ has to include e and, particularly, the minimal set of the maximal surplus $\varphi(M) + 1$. This implies that $e \in (M \cup \{e\})^+$, which was to be proved. ■

The result of this theorem is for the example that $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$ are the minimal sensor sets that make all faults, including the new faults introduced by the added sensors, detectable. Hence, lines 5 and 6 in `SensorPlacement` described in Section IV-C do not need to be changed at all.

On line 7, a minimal sensor set S_i that achieves full detectability is selected, and on line 8, the equations M_{S_i} are added to the original model to form the extended model M_e. If the sensors may become faulty, i.e., if $s \in S_i$ belongs to P_f, then these faults must be added to the model as done in Fig. 5. These faults and the original faults in F are then stored on line 9 in F_e.

The purpose of lines 10–11 is to, given the extended model M_e, find the family $\mathcal{S}_i$ of all minimal additional sensor sets S''s achieving maximum isolability among both the faults in F_e and the sensor faults associated with the additional sensors S''s. The next result states that if S' achieves maximum isolability among the faults F_e, then S' also achieves the maximum isolability among all faults, including the faults introduced by the sensors in S'.

Theorem 5: Let M be a model with no underdetermined part and F be a set of structurally detectable faults in M. Furthermore, let M_S be an equation set describing additional sensors and F_S be the associated set of sensor faults. Then, for any sensor fault $f \in F_S$ and for any fault $f' \in (F \cup F_S) \setminus \{f\}$, it holds that f is isolable from f' and that f' is isolable from f in $M \cup M_S$.

Proof: By assumption, the faults in F are detectable, and the faults in F_S are detectable according to Theorem 4. Since both f' and f are structurally detectable, it is sufficient to show that f' is structurally isolable from f in $M \cup M_S$ according to Theorem 3.

First, assume that $f' \in F$. All faults in F are structurally detectable, and it follows that f' is structurally detectable, i.e.,

$$e_{f'} \in M^+. \tag{10}$$

From the fact that $M \subseteq (M \cup M_S) \setminus \{e_f\}$, it follows that $M^+ \subseteq ((M \cup M_S) \setminus \{e_f\})^+$. This and (10) imply that $e_{f'} \in ((M \cup M_S) \setminus \{e_f\})^+$, i.e., f' is structurally isolable from f according to Definition 3.

Finally, assume that $f' \in F_S \setminus \{f\}$. From Theorem 4, we get that f' is structurally detectable in $M \cup \{e_{f'}\}$, i.e.,

$$e_{f'} \in (M \cup \{e_{f'}\})^+ \tag{11}$$

From the fact that $M \cup \{e_{f'}\} \subseteq (M \cup M_S) \setminus \{e_f\}$, it follows that $(M \cup \{e_{f'}\})^+ \subseteq ((M \cup M_S) \setminus \{e_f\})^+$. This and (11) imply that $e_{f'} \in ((M \cup M_S) \setminus \{e_f\})^+$, i.e., f' is structurally isolable from f according to Definition 3, and this completes the proof. ■

The theorem shows that once sensors and sensor faults have been added to the original model on line 8, the minimal additional sensor sets to achieve maximum isolability can be computed exactly as before, i.e., lines 10–14 need not be changed. In conclusion, the only difference in the function `SensorPlacement` when considering sensor faults is to add the additional input P_f that should be used in the creation of the extended model M_e on line 8.

A difference in the result from the case when not considering sensor faults is that the solution might include two sensors measuring the same variable. For the example, the minimal sensor sets when considering sensor faults are $\{x_1, x_1, x_3\}$, $\{x_1, x_1, x_4\}$, $\{x_1, x_2, x_3\}$, $\{x_1, x_2, x_4\}$, $\{x_1, x_3, x_4\}$, $\{x_2, x_2, x_3\}$, $\{x_2, x_2, x_4\}$, $\{x_2, x_3, x_4\}$, and $\{x_3, x_3, x_4, x_4\}$. These are the nine sensor sets from Section II. Any of these sets is a superset of some solutions obtained in Section IV-C when not considering sensor faults.

V. Algorithm Summary

The problem formulation in Section II included a performance specification concerning isolability and detectability properties, and this was not covered by the algorithms in Section IV. It turns out that only minor modifications are necessary, and the algorithms, including the modifications on how to handle such specifications, will be summarized in this section.

In the algorithm, we will use a convenient but not fully general representation of a detectability and isolability specification. In this representation, a detectability and isolability specification is given by a family $I = \{F_1, F_2, \ldots, F_n\}$ of disjoint sets $F_i \subseteq F$, specifying that the faults in $\cup_{F_i \in I} F_i$ should be detectable and that the pair of faults included in different sets F_i should be isolable from each other. In addition to the information given by P_f of which sensors that have sensor faults, we also have to include how additional sensor faults should be included in the specification I. Assume that all information about all additional sensor faults is included in an object A.

The specification I will replace the input fault set F, and the information about sensor faults A will replace the input P_f in the algorithm `SensorPlacement` presented in Section IV-C. The resulting algorithm will then, given a model M, a detectability and isolability specification I, a set of possible sensor locations P, and the isolability information about sensor faults A, compute all minimal sensor sets that achieve the detectability and isolability specification I. If the detectability and isolability specification I is not attainable with any sensor addition, the minimal sensor sets maximizing the desired properties specified by I are computed.

There are two changes caused by including the specification I in `SensorPlacement` described in Section IV-C. First, the set of faults that should be detectable according to the specification I has to be computed as $F = \cup_{F_i \in I} F_i$. Second, instead of computing just the set F_e of faults included in the extended model M_e on line 9, we need to compute an updated version I_e of the isolability specification I by using the sensor fault information A.

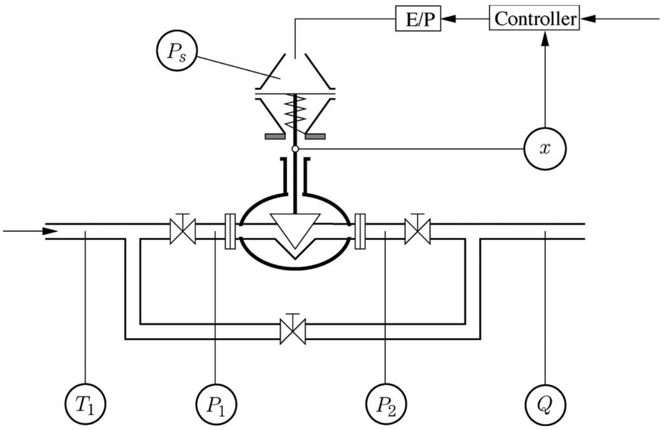

Fig. 6. DAMADICS valve.

```
function $\mathcal{S}$ = SensorPlacement($M, I, P, A$)
  $\mathcal{S} = \varnothing$;
  $F = \cup_{F_i \in I} F_i$;
  $M^0$ = just-determined part of $M$;
  $F^0$ = the set of faults $F$included in$M^0$;
  $\mathcal{D}$ = Detectability($M^0, F^0, P$);
  $\mathcal{S}_d$ = MinimalHittingSets($\mathcal{D}$);
  for $S_i \in \mathcal{S}_d$
    Create the extended model $M_e = M \cup M_{S_i}$ using $A$;
    $I_e$ = updated $I$ with new sensor faults according to$A$;
    $\mathcal{D}$ = Isolability($M_e, I_e, P$);
    $\mathcal{S}_i$ = MinimalHittingSets($\mathcal{D}$);
    $\mathcal{S} = \mathcal{S} \cup \{S_i \uplus S' | S' \in \mathcal{S}_i\}$;
  end
  Delete nonminimal sensor sets in $\mathcal{S}$;
```

The function `Isolability` called on line 11 has to cope with an isolability specification instead of the extended fault set F_e, and this is done as follows. The function `Isolability` computes inputs to the isolability subproblems. Instead of computing the needed measurements for all ordered pairs, it is sufficient to do this only for all ordered pairs including faults from different sets $F_i \in I$. The isolability subproblem for a fault $f \in F_i$ is then to compute the detectability sets for making the maximum possible number of faults $\cup_{j:j>i} F_j$ isolable from f. This is implemented in a new `Isolability` function as follows:

```
function $\mathcal{D}$ = Isolability($M, I, P$)
  $\mathcal{D} = \varnothing$;
  for $F_i \in I$
    for $f \in F_i$
      $F' = \cup_{j:j>i} F_j$;
      $\mathcal{D} = \mathcal{D} \cup$ IsolabilitySubProblem($M, F', P, f$);
    end
  end
```

VI. Example

The example used to illustrate the results is an industrial valve. A schematic figure of the valve is shown in Fig. 6 and consists of three main components: the control valve, a bypass valve, and a spring-and-diaphragm pneumatic servomotor to operate the valve plug. The figure also shows an internal control loop that is used to increase the accuracy of the valve plug

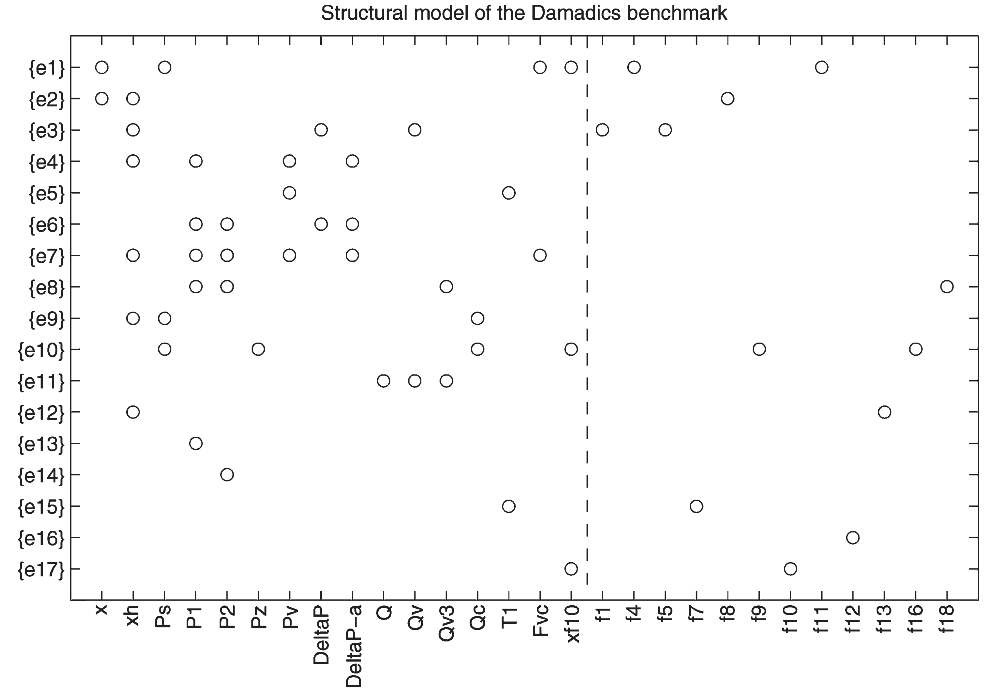

Fig. 7. Structure of the DAMADICS valve model.

positioning. The details of this model are not included in this presentation, and interested readers are referred to, e.g., [15] and the references therein. The structure of the model is derived in [9] and is shown in Fig. 7. Variables x and x_h are valve positioning variables; P_s, P_1, P_2, P_z, P_v, Δ_p, and Δ_{p-a} are pressures; Q, Q_v, Q_{v3}, and Q_c are fluid flows; T_1 is temperature; F_{vc} is a force; and all f_i are variables indicating which equations that the different faults influence. Fault f_{10} influences two equations, and therefore, the dummy variable x_{f10} has been introduced to ensure that the assumption that each fault only influences one equation holds. In this example, all unknown variables, *except* the dummy variable x_{f10}, are assumed to be possible sensor locations. Of course, no fault variables f_i can be measured.

The original model included a specified set of sensors, but since the objective here is to perform sensor placement analysis, almost all sensors have been removed. Three sensors have been kept, namely, measurements of the two ambient pressures P_1 and P_2 and the measurement of the valve position x that is used in the internal control loop. This leaves us with a model, which has no underdetermined part, consisting of 17 equations in 16 unknown variables and 12 different faults.

First, to determine which sensors are necessary to obtain detectability of all faults, the partial orders on the strongly connected components and the fault equivalence classes are computed. Fig. 8 shows the Hasse diagrams for both partial orders. In Fig. 8(b), it is clear that there are three maximal elements of the order, namely, $\{f_1, f_5\}$, $\{f_{18}\}$, and $\{f_9, f_{16}\}$. Thus, obtaining detectability of these faults will automatically provide detectability of all other faults. It is noted in Fig. 8(a) which strongly connected components that the maximal faults influence. Theorem 1 then gives that a sensor set achieving detectability has a nonempty intersection $D([f])$ for each maximal fault class. The variables that appear in each relevant strongly connected component are $X_1 = \{P_z\}$, $X_5 = \{Q\}$, $X_6 = \{Q_v\}$, $X_{11} = \{Q_{v3}\}$, and then

$$D(\{f_1, f_5\}) = \{Q, Q_v\} \quad D(\{f_{18}\}) = \{Q, Q_{v3}\}$$
$$D(\{f_9, f_{16}\}) = \{P_z\}$$

By computing minimal hitting sets for these three sets, one obtains two minimal sensor sets $\{P_z, Q\}$ and $\{P_z, Q_v, Q_{v3}\}$, and it can be verified using Definition 2 that all faults then become detectable.

Adding any of the aforementioned sets of sensors only achieves detectability of the faults and does not give full isolability. Running the algorithm from Section V, computing sensor sets that achieves maximum isolability also for faults in the new sensors gives eight minimal sensor sets. The minimal sensor sets has seven or eight sensors, and one minimal set is to add sensors measuring the variables $\{P_s, P_z, P_z, Q, Q, Q_{v3}, x\}$. Note here that we need to add two sensors each for variables P_z and Q. With these sensors, all faults are isolable from each other except for the pairs $\{f_4, f_{11}\}$, $\{f_1, f_5\}$, and $\{f_9, f_{16}\}$. This is because these faults cannot be isolated by adding more sensors measuring unknown variables since they appear in the same equation in the model. The only solution is to do further fault modeling [9] or, possibly rather unrealistic, to include a sensor that measures the fault signal directly as in [16].

VII. Related Work

Sensor placement for diagnosis and fault detection is a well-studied problem. Examples of previous works are [17] where sensor location for optimal detection performance is studied and [18] and [19] where an optimization problem related to sensor selection is studied. Another example is [20] where a PCA-based monitoring technique is optimized by suitable

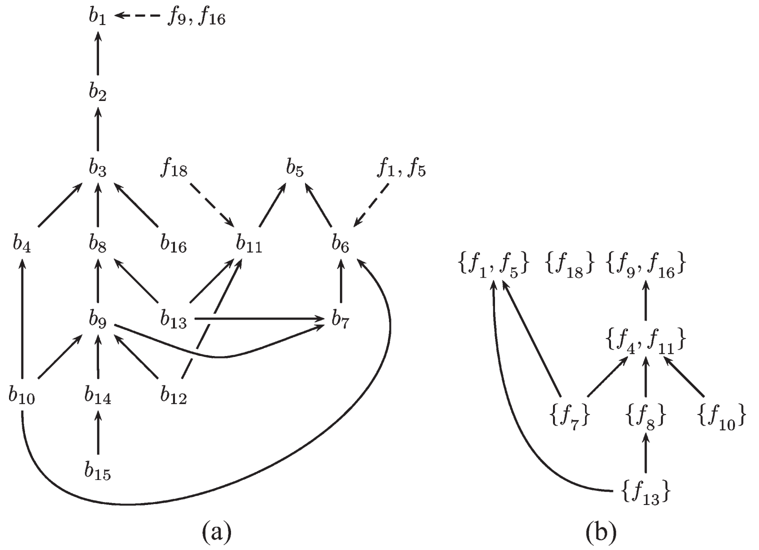

Fig. 8. Order among strongly connected components and faults for the DAMADICS valve model. (a) Order among strongly connected components. It is also noted, with dashed arrows, where some important faults appear in the model. (b) Order among the equivalence classes on the set of faults.

sensor selection. These papers, and other similar papers, have a rather different objective than our paper where optimal isolability properties are the objective. Therefore, this discussion on relations to other works will focus on papers that all have problem formulations with similarities to this paper.

In [16], the sensor placement problem is addressed using input–output separators in a graph-based representation of the system model. A main difference to our paper is that Commault *et al.* aim at adding sensors such that, in the linear case, it is possible to obtain a diagonal transfer matrix from faults to residual. This is often a rather unrealistic goal since this is only possible if there are more sensors than faults, and for example, if the added sensors may become faulty, it is generically not possible to solve the posed problem. In addition, it is, in the paper, assumed that fault signals can be measured, which is an unrealistic assumption.

The basic problem formulation in [21] is almost identical to our paper, but the model description is a little bit different. It is a graph-based description, and they do not allow cycles in the graph, and this results in loss of isolability performance in the solution. A drawback with their proposed solution is that their algorithm does not find all minimal sensor sets; the result does not even need to be minimal. However, it should be possible to use a minimal hitting set algorithm, instead of their greedy search, to obtain all minimal solutions to their posed problem. Another pair of differences is that they do not consider faults in the added sensors and also that faults entering in more than one equation are treated in a nonstandard way. For example, in their approach, it is not possible to add sensors such that the faults in the model

$$\dot{x} = Ax + \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} f$$

are isolable which is clearly possible.

A third related work is [22] where the problem is approached by hypothesizing sensors and then computing the set of analytical redundancy relations (ARRs), using all possible causalities, tracing the support of each ARR and then obtaining isolability properties of the model. Travé-Massuyès *et al.* assume exoneration, i.e., that a fault always makes the corresponding residuals to exceed their thresholds, which is not assumed in our paper since this is a rather unrealistic assumption. Our approach computes which sensors to add to obtain a certain isolability performance, while [22] does it the other way around, adding all possible sensors and then removing sensors until isolability performance decreases. One can expect severe complexity problems with such an approach since the number of ARRs is exponential in the redundancy of the model [14], and by adding all possible sensors, you obtain maximum redundancy. Another difference that is worth noting is that the performance measure in their paper is a scalar value, the diagnosability degree, which is equal to the quotient of the number of fault classes by the number of faults. However, different sensor setups may have different isolability properties and still have the same diagnosability degree. This is the reason why the complete isolability relation, rather than, e.g., the diagnosability degree, is used as a performance specification in our paper. Similar to our paper, that by Travé-Massuyès *et al.* also includes the case where the new sensors may also become faulty. However, this also typically means that you may have to add more than one sensor to a specific variable, and this is not covered in [22] indicating possibly incomplete results.

VIII. Conclusion

The sensor placement problem has been addressed in this paper. Since detectability and isolability performance is gained at the cost of sensor addition, the maximum possible isolability is not always the desired goal. Therefore, it is important that the desired isolability can be specified. Furthermore, there are often process variables that cannot be measured, and this information needs to be considered in a sensor placement analysis. New sensors may of course also become faulty, and these faults must also be included in the analysis. It has been shown that this typically means that more than one sensor has to be added measuring a specific signal.

A key contribution is a new algorithm for sensor placement that copes with all aspects mentioned earlier. Given a model, an isolability specification, the possible sensor locations, and a specification of which sensors that may be faulty, the algorithm computes all minimal sensor sets that make, as far as possible and desired, faults isolable from each other. Typically, there is a cost associated with each type of sensor, for example, price, maintenance cost, reliability, etc. This means that the sensor set with the least number of sensors may not always be the best choice. Since the result of the algorithm contains *all* minimal sensor sets, it is straightforward to pose an optimality condition regarding cost to find the best choice of sensors to add.

All operations in the algorithm are polynomial except for a minimal hitting set computation which is NP-hard, and all known algorithms are, in our problem setting, exponential in the number of possible sensor locations. However, in many real applications, the control algorithms give a necessary requirement on sensors. Also, cost, space, weight, and other considerations give a restriction on possible sensor locations. Thus, the main variable that controls complexity is user controlled and, in real applications, is often limited in size.

The algorithm has been applied to a nontrivial industrial valve model with 17 equations and 15 possible sensor positions using a Matlab implementation of the algorithm that is available at http://www.fs.isy.liu.se/Software/SensPlaceTool/.

Appendix

This Appendix formally defines the Dulmage–Mendelsohn decomposition from Section III-A. These formal definitions are primarily used in the proofs of the results in Section IV. See, for example, [23] or [11] for a thorough treatment of the decomposition.

Let $|A|$ denote the cardinality of set A. Given a bipartite graph with node sets M and X, let the variables in $E \subseteq M$ be denoted by $\mathrm{var}(E)$ and the surplus of equation set E be defined by

$$\varphi(E) = |E| - |\mathrm{var}(E)|\,.$$

Given a model M, there is a family of subsets of M with the maximum surplus

$$\mathcal{L} = \{E \subseteq M | \varphi(E) \geq \varphi(E'),\ \forall E' \subseteq M\} \qquad (12)$$

Let $E_0 \supset E_1 \supset \cdots \supset E_{n-1} \supset E_n$ be any maximal descending chain of $\mathcal{L}$; then, the partition of M is defined as $M_0 = M \setminus E_0$, $M_i = E_{i-1} \setminus E_i$ for $i = 1, \ldots, n$, and $M_\infty = E_n$ (see [23]). The partition of X is defined as

$$X_i = \mathrm{var}(M_i) \setminus \mathrm{var}(E_i) \qquad (13)$$

for $i \in \{0, 1, \ldots, n\}$ and $X_\infty = \mathrm{var}(M_\infty)$. The partial order $\leq$ can be defined on sets M_i by

$$M_i \leq M_j, \qquad \text{if } \forall E \in \mathcal{L}(M_j \subseteq E \Rightarrow M_i \subseteq E). \qquad (14)$$

In Fig. 2, each pair (M_i, X_i) is related to a block which is denoted by b_i. Since there is a one-to-one correspondence between sets M_i and blocks b_i, we will also partially order blocks b_i in the same way

$$b_i \leq b_j, \qquad \text{if} \quad M_i \leq M_j \qquad (15)$$

There exist efficient algorithms to compute the Dulmage–Mendelsohn decomposition [24] from which also the ordering among strongly connected components is easy to extract. In Matlab, the decomposition is implemented in the `dmperm` command. This algorithm has time complexity $\mathcal{O}(\sqrt{n}\tau)$, where n is the number of variables and τ is the number of edges in the corresponding graph.

References

[1] M. Blanke, M. Kinneart, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*. New York: Springer-Verlag, 2003.

[2] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. New York: Marcel Dekker, 1998.

[3] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis, "Failure diagnosis using discrete-event models," *IEEE Trans. Control Syst. Technol.*, vol. 4, no. 2, pp. 105–124, Mar. 1996.

[4] J. Lunze and J. Schroeder, "Sensor and actuator fault diagnosis of systems with discrete inputs and outputs," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 2, pp. 1096–1107, Apr. 2004.

[5] S. Narasimhan and G. Biswas, "Model-based diagnosis of hybrid systems," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 3, pp. 348–361, May 2007.

[6] M. Krysander, "Design and analysis of diagnosis systems using structural methods," Ph.D. dissertation, Linköpings Universitet, Linköping, Sweden, Jun. 2006.

[7] M. Nyberg, "Criterions for detectability and strong detectability of faults in linear systems," *Int. J. Control*, vol. 75, no. 7, pp. 490–501, May 2002.

[8] M. Cordier, P. Dague, F. Levy, J. Montmain, M. Staroswiecki, and L. Travé-Massuyès, "Conflicts versus analytical redundancy relations: A comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 5, pp. 2163–2177, Oct. 2004.

[9] D. Dustegör, E. Frisk, V. Cocquempot, M. Krysander, and M. Staroswiecki, "Structural analysis of fault isolability in the DAMADICS benchmark," *Control Eng. Pract.*, vol. 14, no. 6, pp. 597–608, Jun. 2006.

[10] S. Ploix, M. Desinde, and S. Touaf, "Automatic design of detection tests in complex dynamic systems," in *Proc. 16th IFAC World Congr.*, Prague, Czech Republic, 2005.

[11] A. L. Dulmage and N. S. Mendelsohn, "Coverings of bipartite graphs," *Can. J. Math.*, vol. 10, pp. 517–534, 1958.

[12] R. Reiter, "A theory of diagnosis from first principles," *Artif. Intell.*, vol. 32, no. 1, pp. 57–95, Apr. 1987.

[13] J. de Kleer, "Diagnosing multiple faults," *Artif. Intell.*, vol. 32, no. 1, pp. 97–130, Apr. 1987.

[14] M. Krysander, J. Åslund, and M. Nyberg, "An efficient algorithm for finding minimal overconstrained subsystems for model-based diagnosis," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 1, pp. 197–206, Jan. 2008.

[15] M. Syfert, M. Bartys, J. Quevedo, and R. Patton, "Development and application of methods for actuator diagnosis in industrial control systems (DAMADICS): A benchmark study," in *Proc. IFAC Safeprocess*, Washington, DC, 2003, pp. 939–950.

[16] C. Commault, J. Dion, and S. Agha, "Structural analysis for the sensor location problem in fault detection and isolation," in *Proc. IFAC Safeprocess*, Beijing, China, 2006, pp. 949–954.

[17] M. Basseville, A. Benveniste, G. Moustakides, and A. Rougée, "Optimal sensor location for detecting changes in dynamical behavior," *IEEE Trans. Autom. Control*, vol. AC-32, no. 12, pp. 1067–1075, Dec. 1987.

[18] R. Debouk, S. Lafortune, and D. Teneketzis, "On an optimization problem in sensor selection," *Discret. Event Dyn. Syst.*, vol. 12, no. 4, pp. 417–445, Oct. 2002.

[19] T. Yoo and S. Lafortune, "NP-completeness of sensor selection problems arising in partially observed discrete-event systems," *IEEE Trans. Autom. Control*, vol. 47, no. 9, pp. 1495–1499, Sep. 2002.

[20] H. Wang, Z. Song, and H. Wang, "Statistical process monitoring using improved PCA with optimized sensor locations," *J. Process Control*, vol. 12, no. 6, pp. 735–744, Sep. 2002.

[21] R. Raghuraj, M. Bhushan, and R. Rengaswamy, "Locating sensors in complex chemical plants based on fault diagnostic observability criteria," *AIChE J.*, vol. 45, no. 2, pp. 310–322, Feb. 1999.

[22] L. Travé-Massuyès, T. Escobet, and X. Olive, "Diagnosability analysis based on component-supported analytical redundancy relations," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 36, no. 6, pp. 1146–1160, Nov. 2006.

[23] K. Murota, *Matrices and Matroids for System Analysis*. New York: Springer-Verlag, 2000.

[24] A. Pothen and C.-J. Fan, "Computing the block triangular form of a sparse matrix," *ACM Trans. Math. Softw.*, vol. 16, no. 4, pp. 303–324, Dec. 1990.

Mattias Krysander was born in Linköping, Sweden, in 1977. He received the M.Sc. degree in electrical engineering and the Ph.D. degree in electrical engineering from Linköping University, Linköping, Sweden, in 2000 and 2006, respectively.

He is currently with the Department of Electrical Engineering, Linköping University. His current research interests include model-based fault diagnosis using graph-theoretical and structural methods.

Erik Frisk was born in Stockholm, Sweden, 1971. He received the M.Sc. degree in electrical engineering and the Ph.D. degree in electrical engineering from Linköping University, Linköping, Sweden, in 1996 and 2001, respectively.

He is currently with the Department of Electrical Engineering, Linköping University. His current research interests include model-based fault detection and isolation in nonlinear large-scale systems and structural methods in fault diagnosis.