A Qualitative Event-based Approach to Multiple Fault Diagnosis in Continuous Systems using Structural Model Decomposition

Matthew J. Daigle[a,1,*], Anibal Bregon[b,2], Xenofon Koutsoukos[c], Gautam Biswas[c], Belarmino Pulido[b,2]

[a]*NASA Ames Research Center, Moffett Field, CA, 94035, USA*
[b]*Departamento de Informática, Universidad de Valladolid, Valladolid, 47011, Spain*
[c]*Institute for Software Integrated Systems, Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, 37235, USA*

Abstract

Multiple fault diagnosis is a difficult problem for dynamic systems, and, as a result, most multiple fault diagnosis approaches are restricted to static systems, and most dynamic system diagnosis approaches make the single fault assumption. Within the framework of consistency-based diagnosis, the challenge is to generate conflicts from dynamic signals. For multiple faults, this becomes difficult due to the possibility of fault masking and different relative times of fault occurrence, resulting in many different ways that any given combination of faults can manifest in the observations. In order to address these challenges, we develop a novel multiple fault diagnosis framework for continuous dynamic systems. We construct a qualitative event-based framework, in which discrete qualitative symbols are generated from residual signals. Within this framework, we formulate an online diagnosis approach and establish definitions of multiple fault diagnosability. Residual generators are constructed based on structural model decomposition, which, as we demonstrate, has the effect of reducing the impact of fault masking by decoupling faults from residuals, thus improving diagnosability and fault isolation performance. Through simulation-based multiple fault diagnosis experiments, we demonstrate and validate the concepts developed here, using a multi-tank system as a case study.

Keywords: fault diagnosis, model-based diagnosis, multiple faults, diagnosability, structural model decomposition, discrete-event systems

1. Introduction

Safety-critical systems require quick and robust fault diagnosis mechanisms to improve performance, safety, and reliability, and enable timely and rapid intervention in response to adverse conditions so that catastrophic situations can be avoided. However, complex systems can fail in many different ways, and the likelihood of multiple faults occurring increases in harsh operating environments. Diagnosis methodologies that do not take into account multiple faults may generate incorrect diagnoses or even fail to find a diagnosis when multiple faults occur.

Multiple fault diagnosis in static systems has been addressed previously [1–3], where the inherent complexity of the problem has been well-demonstrated; the diagnosis space becomes exponential in the number of faults, and this complicates the diagnosis task. Furthermore, in dynamic systems, the problem is even more challenging, as the effects

*Corresponding author.
Email addresses: matthew.j.daigle@nasa.gov (Matthew J. Daigle), anibal@infor.uva.es (Anibal Bregon), xenofon.koutsoukos@vanderbilt.edu (Xenofon Koutsoukos), gautam.biswas@vanderbilt.edu (Gautam Biswas), belar@infor.uva.es (Belarmino Pulido)

[1]M. Daigle's work has been partially supported by the NASA System-wide Safety and Assurance Technologies (SSAT) project.
[2]A. Bregon and B. Pulido's work has been supported by the Spanish MINECO grant DPI2013-45414-R.
[3]The authors acknowledge the *8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS 2012)*, August 29-31, 2012, Mexico City, Mexico, for recommending the symposium version of this paper for publication in the IFAC Journal on Engineering Applications of Artificial Intelligence.

of multiple faults may mask one another, thus making it difficult to differentiate between multiple fault diagnoses [4–6]. Due to fault masking, multiple faults can produce a variety of different observations, and this adds uncertainty, which, in turn, reduces the discriminatory ability of the diagnosis algorithms. Moreover, the more faults considered, the more possible ways in which their effects can interleave, making it less likely that the fault diagnoses can be uniquely isolated given a set of observations.

Due to its complexity, multiple fault diagnosis of dynamic systems has not been sufficiently addressed in the literature. In [7], changes are modeled by a set of qualitative simulation states. Later, [8] integrated the model-based diagnosis approach in [1] and the qualitative reasoning approach in [7], to multiple fault diagnosis for dynamic systems using behavioral modes with a priori probabilities. In a related approach, semi-quantitative simulation is used [4], changing the configuration of the model every time a fault appears. However, in these kinds of approaches, the qualitative modeling framework quantizes the state space and specifies qualitative relations between the quantized states, which can result in a large number of states, i.e., such approaches can suffer from the state explosion problem.

In control theory-based diagnosis approaches (known as fault detection and isolation, or FDI approaches), the proposal in [9] is based on the analysis of residual structures. In [5], the authors integrate residual-based and consistency-based approaches that can automatically handle multiple faults in dynamic systems. However, these approaches use only binary signatures (effect or no effect), and so it becomes very difficult to distinguish between different potential multiple faults.

In contrast, our previous work in multiple fault diagnosis for continuous systems [6, 10], is based on a qualitative fault isolation (QFI) framework [11]. It describes how multiple faults manifest in the system measurements and provides algorithms for fault isolation. By using qualitative information defined with respect to a nominal reference, the state explosion of qualitative simulation approaches is avoided. Unlike other FDI approaches, diagnostic information is enhanced using qualitative symbols, instead of binary effect/no effect information, and by including the sequence of observations.

The QFI approach was based on using residuals (the difference between observed and expected system behavior) computed from a global system model. Since faults affect all residuals that have a causal path from the fault to the residual, fault masking can have a significant, adverse impact on multiple fault diagnosability when the number of residuals affected by a fault is large. To avoid this problem, in [12], we explored the idea of using structural model decomposition to improve diagnosability, by deriving local submodels that decouple faults from residuals, so that each fault affects only a small set of residuals [9, 13]. This decreases the possibility of masking, and, as such, leads to improvements in multiple fault diagnosability.

In this paper, we extend the previous work in event-based QFI of single faults [14] to develop an online multiple fault diagnosis approach for dynamic systems that takes advantage of structural model decomposition. In this framework, diagnostic observations take the form of symbolic *traces* representing sequences of qualitative effects on the residuals. First, we develop a systematic approach for predicting the possible traces that can be produced by multiple faults, based on a specific composition of those produced by the constituent faults. Second, we develop an online fault isolation algorithm that maps observed traces to the set of minimal diagnoses that could have produced that trace. Third, we introduce definitions of diagnosability to characterize the potential fault isolation performance for different residual sets, and show how structural model decomposition can significantly improve diagnosability in the multiple-fault case. Fourth, using a multi-tank system as a case study, and over a comprehensive set of simulation-based experiments, we provide offline diagnosability results and online multiple fault isolation results to (*i*) demonstrate and validate the overall approach, (*ii*) illustrate the improvement in performance obtained through the use of structural model decomposition, and (*iii*) show the performance improvement over approaches that use binary fault signatures without temporal information. The multi-tank system is also used as a running example throughout the paper.

The paper is organized as follows. Section 2 presents our modeling background and formulates the multiple fault diagnosis problem. Section 3 overviews the structural model decomposition approach, and develops the qualitative fault isolation methodology for multiple faults, which predicts the possible traces that can be produced by a set of faults. Section 4 presents the online multiple fault isolation approach, which determines the set of faults that can produce an observed trace. Section 5 formalizes our definitions of distinguishability and diagnosability in order to characterize the fault isolation performance of a system using our approach. Section 6 presents the results for the case study. Section 7 describes related work in multiple fault diagnosis. Section 8 concludes the paper.

2. Problem Formulation

In this work, we consider the problem of multiple fault diagnosis in continuous systems. We first overview our system modeling approach, followed by a definition of the multiple fault diagnosis problem.

2.1. System Modeling

In our framework, a model is defined as a set of variables and a set of constraints among the variables [13]:

Definition 1 (Constraint)**.** A constraint c is a tuple (ε_c, V_c), where ε_c is an equation involving variables V_c.

Definition 2 (Model)**.** A *model* $\mathcal{M}$ is a tuple $\mathcal{M} = (V, C)$, where V is a set of variables, and C is a set of constraints among variables in V. V consists of five disjoint sets, namely, the set of state variables, X; the set of parameters, Θ; the set of inputs, U; the set of outputs, Y; and the set of auxiliary variables, A.

The set of output variables, Y, corresponds to the (measured) sensor signals. Parameters, Θ, include explicit model parameters that are used in the model constraints. Auxiliary variables, A, are additional variables that are algebraically related to the state, parameter, and input variables, and are used to reduce the structural complexity of the equations. The set of input or exogenous variables, U, is assumed to be known.

In this paper, we use a multi-tank system as a case study. The system consists of n tanks connected serially, as shown in Fig. 1. For each tank i, where $i \in [1, 2, \ldots, n]$, u_i denotes the input flow, m_i denotes the liquid mass, p_i denotes the tank pressure, q_i denotes the mass flow out of the drain pipe, K_i denotes the tank capacitance, and Re_i denotes the drain pipe resistance. For adjacent tanks i and $i+1$, $q_{i,i+1}$ denotes the mass flow from tank i to tank $i+1$ through the connecting pipe, and $Re_{i,i+1}$ is the connecting pipe resistance. The constraints for tank i are as follows:

$$
\begin{aligned}
\dot{m}_i &= u_i + q_{i-1,i} - q_i - q_{i,i+1}, \\
m_i &= \int_{t_0}^{t} \dot{m}_i \, dt, \\
p_i &= \frac{1}{K_i} m_i, \\
q_i &= \frac{1}{Re_i} p_i, \\
q_{i-1,i} &= \frac{1}{Re_{i-1,i}} (p_{i-1} - p_i), \\
q_{i,i+1} &= \frac{1}{Re_{i,i+1}} (p_i - p_{i+1}).
\end{aligned}
$$

For tank 1, $q_{0,1} = 0$, and for tank n, $q_{n,n+1} = 0$. The complete set of possible measurements in the system corresponding to p_i, q_i, and $q_{i,i+1}$ are p_i^*, q_i^*, and $q_{i,i+1}^*$, described by the following constraints:

$$
\begin{aligned}
p_i^* &= p_i, \\
q_i^* &= q_i, \\
q_{i,i+1}^* &= q_{i,i+1}.
\end{aligned}
$$

Here, the * superscript is used to denote a measured value of a physical variable, e.g., p_i is pressure and p_i^* is the measured pressure.[4]

As a running example to explain and illustrate the concepts throughout the paper, we use a standard three-tank system in which, unless otherwise specified, the pressures are measured.

[4] Since p_i is used to compute other variables, it cannot belong to Y and a separation of the variables is required.

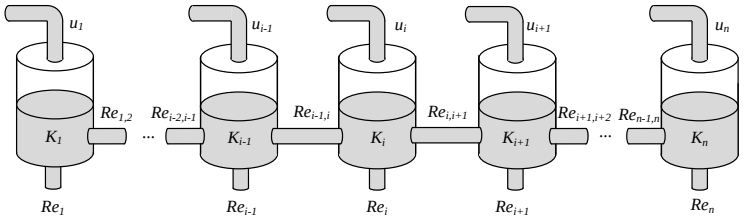

Figure 1: Tank system schematic.

Example 1. For the three-tank system, the model $\mathcal{M}$ is represented by the variable sets $X = \{m_1, m_2, m_3\}$, $\Theta = \{K_1, K_2, K_3, Re_1, Re_2, Re_3, Re_{1,2}, Re_{2,3}\}$, $U = \{u_1, u_2, u_3\}$, $Y = \{p_1^*, p_2^*, p_3^*\}$, and $A = \{\dot{m}_1, \dot{m}_2, \dot{m}_3, p_1, p_2, p_3, q_1, q_2, q_3\}$; and the set of constraints $C = \{c_1, c_2, \ldots, c_{17}\}$, which are given as follows:

$$\dot{m}_1 = u_1 - q_1 - q_{1,2}, \tag{c_1}$$
$$\dot{m}_2 = u_2 + q_{1,2} - q_2 - q_{2,3}, \tag{c_2}$$
$$\dot{m}_3 = u_3 + q_{2,3} - q_3, \tag{c_3}$$
$$m_1 = \int_{t_0}^{t} \dot{m}_1 \, dt, \tag{c_4}$$
$$m_2 = \int_{t_0}^{t} \dot{m}_2 \, dt, \tag{c_5}$$
$$m_3 = \int_{t_0}^{t} \dot{m}_3 \, dt, \tag{c_6}$$
$$p_1 = \frac{1}{K_1} m_1, \tag{c_7}$$
$$p_2 = \frac{1}{K_2} m_2, \tag{c_8}$$
$$p_3 = \frac{1}{K_3} m_3, \tag{c_9}$$
$$q_1 = \frac{1}{Re_1} p_1, \tag{c_{10}}$$
$$q_2 = \frac{1}{Re_2} p_2, \tag{c_{11}}$$
$$q_3 = \frac{1}{Re_3} p_3, \tag{c_{12}}$$
$$q_{1,2} = \frac{1}{Re_{1,2}} (p_1 - p_2), \tag{c_{13}}$$
$$q_{2,3} = \frac{1}{Re_{2,3}} (p_2 - p_3), \tag{c_{14}}$$
$$p_1^* = p_1, \tag{c_{15}}$$
$$p_2^* = p_2, \tag{c_{16}}$$
$$p_3^* = p_3. \tag{c_{17}}$$

In our context, a fault is the cause of an unexpected, persistent deviation of the system behavior from the acceptable nominal behavior. Specifically, in our framework, we link faults to the set of parameters Θ in $\mathcal{M}$. More formally, a fault is defined as follows.

Definition 3 (Fault)**.** A *fault*, denoted as f, is a persistent deviation of exactly one parameter $\theta \subseteq \Theta$ of the system model $\mathcal{M}$ from its nominal value.

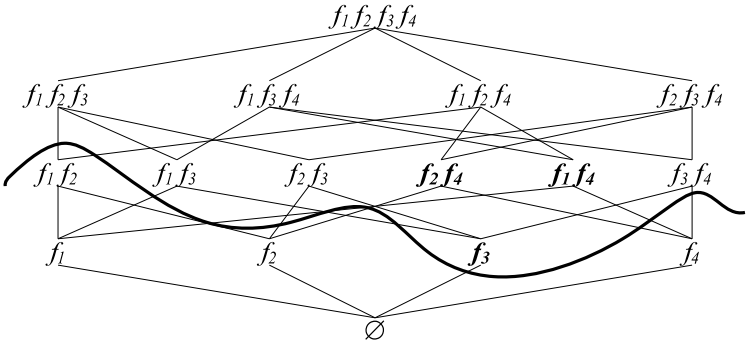

Figure 2: Lattice representation of the candidate space.

A fault is named by the associated parameter and its direction of change, i.e., θ^+ (resp., θ^-) denotes a fault defined as an increase (resp., decrease) in the value of parameter θ. In general, we use F to denote a set of faults.

Example 2. In the three-tank system, the complete fault set is $F = \{K_1^-, K_1^+, K_2^-, K_2^+, K_3^-, K_3^+, Re_1^-, Re_1^+, Re_{1,2}^-, Re_{1,2}^+, Re_2^-, Re_2^+, Re_{2,3}^-, Re_{2,3}^+, Re_3^-, Re_3^+\}$.

2.2. Problem Definition

Fault isolation proceeds as a cycle of observation and hypothesis generation. In multiple fault diagnosis, a diagnostic hypothesis, or *diagnosis*, for short, is defined as a set of faults that is consistent with the observations.

Definition 4 (Diagnosis)**.** For a given fault set F, a *diagnosis* $d \subseteq F$ is a set of faults that is consistent with a sequence of observations λ.

Intuitively, a diagnosis represents a single potential explanation for observed faulty behavior.

Example 3. The diagnosis $\{K_1^-, R_3^+\}$ (in shorthand, we write $K_1^- R_3^+$) means that K_1^- and R_3^+ together produce symptoms that are consistent with the observations.

A set of diagnoses is denoted as D. For a set of single faults F, there are $2^{|F|}$ unique diagnoses (including the empty set), i.e., $|F|$ single faults, $\binom{|F|}{2}$ double faults, $\binom{|F|}{3}$ triple faults, and so on. Clearly, the space of diagnoses is exponential in the number of faults. It can be represented using a lattice structure [1]; Fig. 2 shows the lattice structure for a system where $F = \{f_1, f_2, f_3, f_4\}$.

In dynamic systems, fault masking can manifest when the effects of one fault dominate the effects of another fault, so that effects of the second fault are not directly observed. As a result, the following property holds.

Lemma 1. *For $d, d' \subseteq F$, if d is a diagnosis and $d \subset d'$, then d' is a diagnosis.*

Lemma 1 also holds in static diagnosis, e.g., as in [1]. As observations are made, we eliminate certain diagnoses and form a cut across the lattice, displayed as a bold line in the figure, such that everything below the cut has been eliminated, while everything above the cut is a diagnosis.

From this property, the concept of a minimal diagnosis manifests.[5]

Definition 5 (Minimal Diagnosis)**.** A diagnosis d is *minimal* if there is no diagnosis d' where $d' \subset d$.

By Lemma 1, we can represent the complete set of diagnoses concisely by the set of minimal diagnoses. Therefore, we need only to generate minimal diagnoses because all diagnoses can be generated from the minimal diagnosis set. From the diagnosis space, we can define two sets, the minimal diagnosis set, and the maximal diagnosis set.

[5] In [15], a diagnosis is by definition minimal. Here, to be more general, we define a diagnosis to be any consistent set of faults, and explicitly define the notion of a minimal diagnosis.

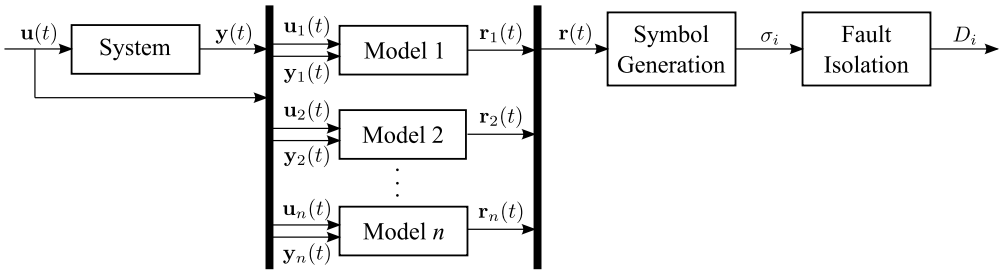

Figure 3: Computational architecture for multiple fault diagnosis based on structural model decomposition.

Definition 6 (Minimal Diagnosis Set)**.** The *minimal diagnosis set* D^- is the set of minimal diagnoses.

Definition 7 (Maximal Diagnosis Set)**.** The *maximal diagnosis set* D^+ is the set of all diagnoses.

In Fig. 2, the minimal diagnosis set consists of the candidates indicated in bold. The maximal diagnosis set consists of all the minimal diagnoses and all possible supersets of the minimal diagnoses, i.e., all candidates above the line. Formally, we can generate D^+ from D^- by adding to D^- all diagnoses d' which are a strict superset of any $d \in D^-$. From a practical standpoint, maintaining only the set of minimal diagnoses is more efficient; further, the probability of some set of faults d occurring is always higher than some $d' \supset d$ occurring if we assume that faults are independent, so the minimal diagnoses are also more likely.

Further practical considerations may also warrant an assumption on the size of diagnoses to consider.

Assumption 1 (Fault Cardinality)**.** At most l faults occur together in the system.

This assumption does not limit the generality of our approach, since we can always set l to $|F|$. However, in practice, usually l is set to 1 or 2, with the implication being that the probability of any set of faults of size greater than l occurring is negligible. This can also result in a reduction of computational complexity, because it limits the portion of the diagnosis space that needs to be explored. The multiple fault diagnosis problem then becomes the following.

Problem (Multiple Fault Diagnosis)**.** Given a system model, $\mathcal{M}$, with a set of faults, F, and a cardinality limit l, the *multiple fault diagnosis problem* is to find the subset of the minimal diagnosis set D^- of candidates with cardinality $\leq l$ for a given sequence of observations, λ.

Our proposal for solving this problem is described primarily in Sections 3 and 4, and the computational architecture is summarized in Fig. 3. Given a system model, we define a set of residuals based on structural model decomposition (Section 3.1). The system produces outputs $\mathbf{y}(t)$ given inputs $\mathbf{u}(t)$, which get organized into local inputs and outputs for the submodels, i.e., $\mathbf{u}_i(t)$ and $\mathbf{y}_i(t)$ for model i, which computes residuals $\mathbf{r}_i(t)$ (Section 3.2). A symbol generation algorithm [16] computes from these residuals observed qualitative effects, σ_i. We then develop a method to systematically determine what the sequences of observed qualitative effects of a set of faults will be on these residuals (Section 3.3). Based on this, for online diagnosis, the fault isolation algorithm matches the observed sequence to the minimal diagnosis set that could have produced it, excluding fault sets with cardinality above the limit (Section 4). Like classical diagnosis approaches [1, 15], our approach is model-based, and so can be applied to any system modeled as a set of ordinary differential equations.

3. Qualitative Fault Isolation Framework

We adopt an event-based qualitative fault isolation framework, extending the single-fault framework presented in [14]. In this section, we describe the methodology that determines the possible observations that can be produced by a set of faults that occur. In this diagnosis paradigm, we generate discrete observations based on the analysis of *residuals*.

Definition 8 (Residual)**.** A *residual*, r_y, is a time-varying signal computed as the difference between an output, $y \subseteq Y$, and a predicted value of the output y, denoted as $\hat{y}$.

In order to generate a residual, we require a dynamic model to generate predicted values for each y. If the model is correct, then when a fault occurs, it will produce significant, observable differences between y and $\hat{y}$. Our fault isolation framework is based on an analysis of these differences, rooted in transient analysis [11]. We apply signal processing algorithms to transform these differences into sequences of qualitative observations from which to perform diagnostic reasoning [16].

In the following subsections, we first describe how to compute residuals using the concept of structural model decomposition. We then describe the form that observations and observation sequences take in our approach. Following that, we describe how we determine the observation sequences that multiple faults can produce.

3.1. Structural Model Decomposition

In order to compute a residual r_y for an output y, we must compute a predicted value of the output, $\hat{y}$. To do this, we require a notion of computational causality for a model. A *causal assignment* specifies the computational causality for a constraint c, by defining which $v \in V_c$ is the dependent variable in equation ε_c.

Definition 9 (Causal Assignment)**.** A *causal assignment* α to a constraint $c = (\varepsilon_c, V_c)$ is a tuple $\alpha = (c, v_c^{out})$, where $v_c^{out} \in V_c$ is assigned as the dependent variable in ε_c.

We write a causal assignment of a constraint using its equation in a causal form, with := to explicitly denote the causal (i.e., computational) direction. To compute the variables of a model, we require each constraint to have a causal assignment, and that the set of causal assignments is consistent, i.e., (*i*) input and parameter variables cannot be the dependent variables in the causal assignment, (*ii*) an output variable cannot be used as the independent variable, and (*iii*) every variable, which is not an input or parameter, is computed by only one (causal) constraint. An algorithm for finding a consistent causal assignment to a model is given in [17].

Example 4. For the three-tank system, the constraints given in Example 1 are written such that the causal assignment should be made where the = sign is replaced by :=, i.e., the variable on the left-hand side is the independent variable in each of the constraints.

We can use the global model of the system with a consistent set of causal assignments to compute residuals. Given the inputs, the set of causal constraints can be used to compute predictions of the outputs and form $\hat{y}$ for each member of Y. As an alternative, we can instead, through structural model decomposition, define a set of *local* submodels, each with its own set of local outputs $Y_i \subseteq Y$. The advantage of this approach is that each local residual responds only to the subset of the faults included in that submodel, in contrast to a global model residual that will potentially be sensitive to all faults. The decoupling property of the local submodel residuals translates to fewer opportunities for faults to mask each other, and we will see later how that translates to improved diagnosability and fault isolation performance.

Different structural model decomposition methods have been proposed to decompose a system model into minimal over-determined submodels that are sufficient for fault diagnosis [13, 18–20]. In this work, we will use the decomposition framework proposed in [13]. In [13], a model decomposition algorithm is provided that, given a model $\mathcal{M}$, a set of consistent causal assignments, a set of potential local input variables, and a set of desired output variables, finds a minimal submodel that computes the desired outputs using only the provided inputs. In this context, a submodel can be defined as follows.

Definition 10 (Submodel)**.** A *submodel* $\mathcal{M}_i$ of a model $\mathcal{M} = (V, C)$ is a tuple $\mathcal{M}_i = (V_i, C_i)$, where $V_i \subseteq V$ and $C_i \subseteq C$.

For the purposes of residual generation, we want to find submodels that compute some y, i.e., this is the submodel output. For inputs, we can use the global model inputs U, and also measured values from the sensors, so variables in Y (excluding the output variable for the submodel). By using measured values of sensors as inputs, we require only a subset of the model constraints in order to compute any given variable. The model decomposition algorithm is straightforward; it starts at the desired output variables and propagates backwards through the causal constraints, modifying causal assignments when a potential input variable can be used. Additional details on this approach and the structural model decomposition algorithms can be found in [13].

Example 5. Using this approach on the three-tank system for the output set $Y = \{p_1^*, p_2^*, p_3^*\}$, we find the set of submodels (one for each measured variable) given in Table 1. For example, the second submodel computes p_2^* using the measured values of p_1^*, p_3^*, and u_2. Because p_1^* and p_3^* are provided as inputs, p_2^* can be computed with m_2 as the only state variable, and only the subset of constraints involving the second tank.

Table 1: Submodels for the global model, $\mathcal{M}$, of the three-tank system with $Y = \{p_1^*, p_2^*, p_3^*\}$.

States (X_i)	**Parameters** (Θ_i)	**Inputs** (U_i)	**Outputs** (Y_i)	**Causal Assignments** ($\mathcal{A}_i$)
m_1	$K_1, Re_1, Re_{1,2}$	p_2^*, u_1	p_1^*	$p_1^* := p_1$
				$p_1 := m_1/K_1$
				$m_1 := \int_{t_0}^{t} \dot{m}_1 \, dt$
				$\dot{m}_1 := -q_1 - q_{1,2} + u_1$
				$q_1 := p_1/Re_1$
				$q_{1,2} := (p_1 - p_2)/Re_{1,2}$
				$p_2 := p_2^*$
m_2	$K_2, Re_{1,2}, Re_2, Re_{2,3}$	p_1^*, p_3^*, u_2	p_2^*	$p_2^* := p_2$
				$p_2 := m_2/K_2$
				$m_2 := \int_{t_0}^{t} \dot{m}_2 \, dt$
				$\dot{m}_2 := q_{1,2} - q_2 - q_{2,3} + u_2$
				$q_{1,2} := (p_1 - p_2)/Re_{1,2}$
				$q_{2,3} := (p_2 - p_3)/Re_{2,3}$
				$p_1 := p_1^*$
				$q_2 := p_2/Re_2$
				$p_3 := p_3^*$
m_3	$K_3, Re_{2,3}, Re_3$	p_2^*, u_3	p_3^*	$p_3^* := p_3$
				$p_3 := m_3/K_3$
				$m_3 := \int_{t_0}^{t} \dot{m}_3 \, dt$
				$\dot{m}_3 := q_{2,3} - q_3 + u_3$
				$q_{2,3} := (p_2 - p_3)/Re_{2,3}$
				$q_3 := p_3/Re_3$
				$p_2 := p_2^*$

3.2. Residual Analysis

In this section, we describe how we analyze residual signals and transform them into a discrete set of qualitative observations upon which to perform diagnostic reasoning.

Ideally, in the nominal situation, residual signals are zero, hence, any deviation from zero indicates a fault. Because reasoning over the continuous residual signals is difficult and computationally demanding, we abstract a residual into a symbolic form (see Fig. 3). Observations are produced once a deviation in a residual is detected. The transient in the residual signal at this time is abstracted using qualitative `+`, `-`, and `0` values in the signal magnitude and slope. Consequently, the interpretation for these qualitative values for the signal magnitude is: a `0` means the observation is within the nominal thresholds, i.e., $-T < r_y < T$ for threshold T; a `+` means the observation y is above the predicted output $\hat{y}$ plus the threshold T, i.e., $r_y > T$; and a `-` means the observation is below the predicted output minus the threshold, i.e., $r_y < -T$. For the slope, the interpretation is the same, with r replaced by $\dot{r}$, and with a different threshold value specific to the slope. The threshold T can be computed using robust statistical techniques, and, in general, may change over time [16].[6]

So, in our context, an *observation* is defined as follows.

Definition 11 (Observation)**.** An *observation* for a residual r, denoted σ_r, is a pair of symbols $s_1 s_2$ representing qualitative changes in magnitude and slope of r, respectively.

[6] In theory, higher-order changes can also be used as diagnostic information. In practice, however, it is difficult to reliably extract higher-order changes from a signal, and so we do not typically use that information for diagnosis [21].

As residuals deviate due to faults, we obtain an *observation sequence*.

Definition 12 (Observation Sequence)**.** For a set of residuals R, an *observation sequence*, denoted λ_R, is a sequence of observations $\sigma_{r_1}\sigma_{r_2}\ldots\sigma_{r_n}$, where $1 \leq n \leq |R|$, and $r_1 \neq r_2 \neq \ldots \neq r_n$.

In this work, only the first deviation of a residual is meaningful, hence the requirement that an observation sequence for a set of residuals contains at most one observation for each residual.

3.3. Event-based Fault Modeling

The goal of qualitative fault isolation is to determine which diagnoses can produce a given observation sequence. The basis of this approach is the *fault signature*.

Definition 13 (Fault Signature)**.** A *fault signature* for a fault f and residual r, denoted by $\sigma_{f,r}$, is a pair of symbols $s_1 s_2$ representing potential qualitative changes in magnitude and slope of r caused by f at the point of the occurrence of f. The set of fault signatures for f and r is denoted as $\Sigma_{f,r}$.

The complete set of possible fault signatures for a residual that we consider here is $\{+-, -+, 0+, 0-, +0, -0\}$. A fault signature on residual r_y for output y is written as $r_y^{s_1 s_2}$, e.g., $r_{p_1^*}^{+-}$.

For an initial observation σ_r, we must find all f for which $\sigma_{f,r} = \sigma_r$. As more observations are obtained, the problem becomes more complex, because we are then concerned with *sequences* of fault signatures. The sequence of fault signatures produced by a fault is constrained by the system dynamics, and these constraints are captured using the concept of *relative residual orderings* [22]. They are based on the intuition that the effects of a fault will manifest in some parts of the system (i.e., some residuals) before others. For a given model (or submodel), the relative ordering of the residual deviations can be computed based on analysis of the transfer functions from faults to residuals, as proven in [22].

Definition 14 (Relative Residual Ordering)**.** A *relative residual ordering* for a fault f and residuals r_i and r_j, is a tuple (r_i, r_j), denoted by $r_i \prec_f r_j$, representing that f always manifests in r_i before r_j. The set of all residual orderings for f in R is denoted as $\Omega_{f,R}$.

Note that in this definition, we are referring specifically to deviations in the residuals caused by faults. In this paper, to make the approach as general as possible, we assume that fault signatures and relative residual orderings are given as inputs. In practice, this information can be generated by manual analysis of the system model, by simulation, or automatically from certain types of models, e.g., as presented in [10, 11].

Example 6. Table 2 shows the predicted fault signatures and residual orderings for the global model of a three-tank system with $F = \{K_1^-, K_2^-, K_3^-, Re_1^+, Re_2^+, Re_3^+, Re_{1,2}^+, Re_{2,3}^+\}$, $Y = \{p_1^*, p_2^*, p_3^*\}$, and $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$. For example, consider K_1^-. An abrupt decrease in K_1 would cause an abrupt increase in p_1 (see c_7), and thus, an abrupt increase in p_1^* (see c_{15}). The increase in p_1 would also cause an increase in the flow to the second tank, through which the integration manifests as a first-order increase in p_2 and p_2^* (resulting in $r_{p_2^*}^{0+}$). Similarly the increase in p_2 causes a second-order increase in p_3 and p_3^* (resulting in $r_{p_3^*}^{0+}$). The first-order increase in p_2 also causes a second-order decrease in p_1 and p_1^* (resulting in $r_{p_1^*}^{+-}$). Because of the integrations, the abrupt change in $r_{p_1^*}$ is observed first, followed by the change in $r_{p_2^*}$ and then $r_{p_3^*}$, resulting in the residual orderings $r_{p_1^*} \prec r_{p_2^*}$, $r_{p_1^*} \prec r_{p_3^*}$, and $r_{p_2^*} \prec r_{p_3^*}$.

Example 7. Table 3 shows the predicted fault signatures and residual orderings for the minimal submodels of a three-tank system with $F = \{K_1^-, K_2^-, K_3^-, Re_1^+, Re_2^+, Re_3^+, Re_{1,2}^+, Re_{2,3}^+\}$, $Y = \{p_1^*, p_2^*, p_3^*\}$, and $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$. Because some faults appear only in a subset of the submodels (see Table 1), some residuals do not respond to some faults. For example, K_1^- will cause a deviation only in $r_{p_1}^*$. Because any two residuals in this residual set are computed independently (i.e., from a different submodel), we cannot derive any residual orderings among these two residuals. The only orderings we can define are for those in which, for a given fault, it causes a response in one residual but no response in another.

For a set of faults, given potential fault signatures and residual orderings, we can describe what potential sequences of fault signatures may be produced by any combination of faults. Such a sequence is termed a *fault trace*.

Table 2: Fault Signatures and Relative Residual Orderings for the Global Model, $\mathcal{M}$, of the Three-tank System.

Fault	$r_{p_1^*}$	$r_{p_2^*}$	$r_{p_3^*}$	Relative Residual Orderings
K_1^-	+−	0+	0+	$r_{p_1^*} \prec r_{p_3^*}, r_{p_1^*} \prec r_{p_2^*}, r_{p_2^*} \prec r_{p_3^*}$
K_2^-	0+	+−	0+	$r_{p_2^*} \prec r_{p_3^*}, r_{p_2^*} \prec r_{p_1^*}$
K_3^-	0+	0+	+−	$r_{p_3^*} \prec r_{p_1^*}, r_{p_3^*} \prec r_{p_2^*}, r_{p_2^*} \prec r_{p_1^*}$
Re_1^+	0+	0+	0+	$r_{p_1^*} \prec r_{p_3^*}, r_{p_1^*} \prec r_{p_2^*}, r_{p_2^*} \prec r_{p_3^*}$
$Re_{1,2}^+$	0+	0−	0−	$r_{p_2^*} \prec r_{p_3^*}$
Re_2^+	0+	0+	0+	$r_{p_2^*} \prec r_{p_3^*}, r_{p_2^*} \prec r_{p_1^*}$
$Re_{2,3}^+$	0+	0+	0−	$r_{p_2^*} \prec r_{p_1^*}$
Re_3^+	0+	0+	0+	$r_{p_3^*} \prec r_{p_1^*}, r_{p_3^*} \prec r_{p_2^*}, r_{p_2^*} \prec r_{p_1^*}$

Table 3: Fault Signatures and Relative Residual Orderings for the Minimal Submodels of the Three-tank System.

Fault	$r_{p_1^*}$	$r_{p_2^*}$	$r_{p_3^*}$	Residual Orderings
K_1^-	+−	00	00	$r_{p_1^*} \prec r_{p_3^*}, r_{p_1^*} \prec r_{p_2^*}$
K_2^-	00	+−	00	$r_{p_2^*} \prec r_{p_3^*}, r_{p_2^*} \prec r_{p_1^*}$
K_3^-	00	00	+−	$r_{p_3^*} \prec r_{p_2^*}, r_{p_3^*} \prec r_{p_1^*}$
Re_1^+	0+	00	00	$r_{p_1^*} \prec r_{p_3^*}, r_{p_1^*} \prec r_{p_2^*}$
$Re_{1,2}^+$	0+	0−	00	$r_{p_1^*} \prec r_{p_3^*}, r_{p_2^*} \prec r_{p_3^*}$
Re_2^+	00	0+	00	$r_{p_2^*} \prec r_{p_3^*}, r_{p_2^*} \prec r_{p_1^*}$
$Re_{2,3}^+$	00	0+	0−	$r_{p_3^*} \prec r_{p_1^*}, r_{p_2^*} \prec r_{p_1^*}$
Re_3^+	00	00	0+	$r_{p_3^*} \prec r_{p_2^*}, r_{p_3^*} \prec r_{p_1^*}$

Definition 15 (Fault Trace)**.** A *fault trace* for a set of faults F over residuals R, denoted by $\lambda_{F,R}$, is a sequence of fault signatures that can be observed given the occurrence of the faults.

We group the set of all fault traces into a *fault language*.

Definition 16 (Fault Language)**.** The *fault language* for a set of faults F with residual set R, denoted by $L_{F,R}$, is the set of all fault traces for F over the residuals in R.

For diagnosis, we are given some observation sequence λ_R, and we must find all F such that there is some $\lambda_{F,R} \in L_{F,R}$ where $\lambda_{F,R} = \lambda_R$. So, we must determine the fault languages for every potential set of faults up to the fault cardinality limit l. Constructing the fault language for single faults is straightforward. For fault f, given the set of possible fault signatures $\Sigma_{f,r}$ for each $r \in R$, and the set of relative residual orderings $\Omega_{f,R}$, we can construct the fault language as the set of all traces of length $\leq |R|$, that includes, for every $r \in R$ that will deviate due to f, a fault signature $\sigma_{f,r}$, such that the sequence of fault signatures satisfies $\Omega_{f,R}$. One way to compute this is through synchronization of the signatures and orderings [14].

Example 8. Given $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$ from the global model, for fault K_2^-, from Table 2 we see that the fault effects will appear first on $r_{p_2^*}$, and then it is unknown whether $r_{p_1^*}$ or $r_{p_3^*}$ will deviate next. Hence, there are two possible fault traces: $r_{p_2^*}^{+-} r_{p_1^*}^{0+} r_{p_3^*}^{0+}$ and $r_{p_2^*}^{+-} r_{p_3^*}^{0+} r_{p_1^*}^{0+}$. On the other hand, for Re_3^+, there is only one possible fault trace, $r_{p_3^*}^{0+} r_{p_2^*}^{0+} r_{p_1^*}^{0+}$.

Example 9. Given $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$ from the local submodels, for fault K_2^-, from Table 3 we see that the fault effects will appear first on $r_{p_2^*}$, and then, since the fault is not included in the other submodels (see Table 1), no other residuals will deviate. Thus, we have the orderings $r_{p_2^*} \prec r_{p_1^*}$ and $r_{p_2^*} \prec r_{p_3^*}$. So, there is only one possible fault trace in $L_{K_2^-,R}$, $r_{p_2^*}^{+-}$. On the other hand, for $Re_{2,3}^+$, there are two residuals that will deviate, $r_{p_2^*}$ and $r_{p_3^*}$, as the fault appears in both of the corresponding submodels. There are then two possible fault traces in $L_{Re_{2,3}^+,R}$, $r_{p_2^*}^{0+} r_{p_3^*}^{0-}$ and $r_{p_3^*}^{0-} r_{p_2^*}^{0+}$.

For multiple faults, however, an observation sequence will consist of some fault signatures from one fault, and some fault signatures from another fault. Each fault manifests in its own way (i.e., its own single-fault trace). When

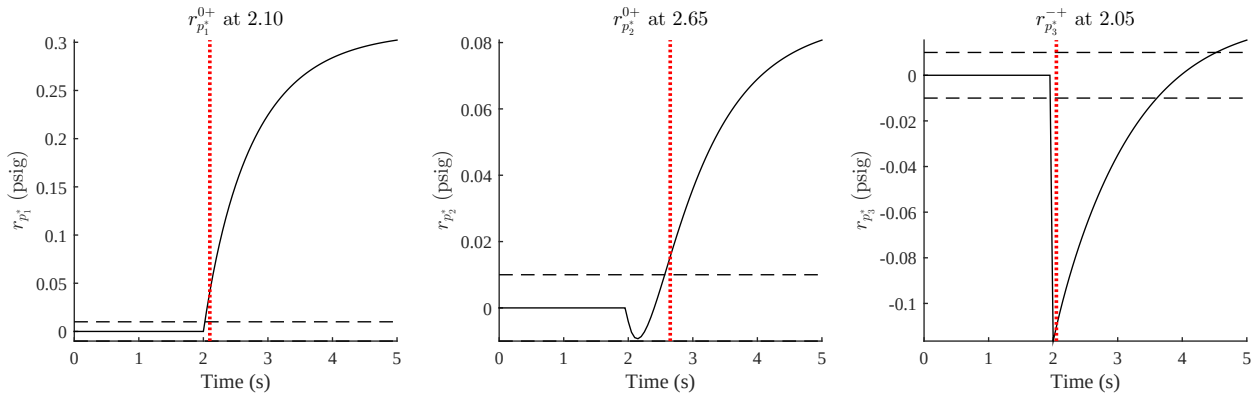

Figure 4: Observations for the fault $Re_1^+ K_3^+$, with K_3 doubling at 2 s and Re_1 doubling at 2.05 s, resulting in $r^{-+}_{p_3^*} r^{0+}_{p_1^*} r^{0+}_{p_2^*}$.

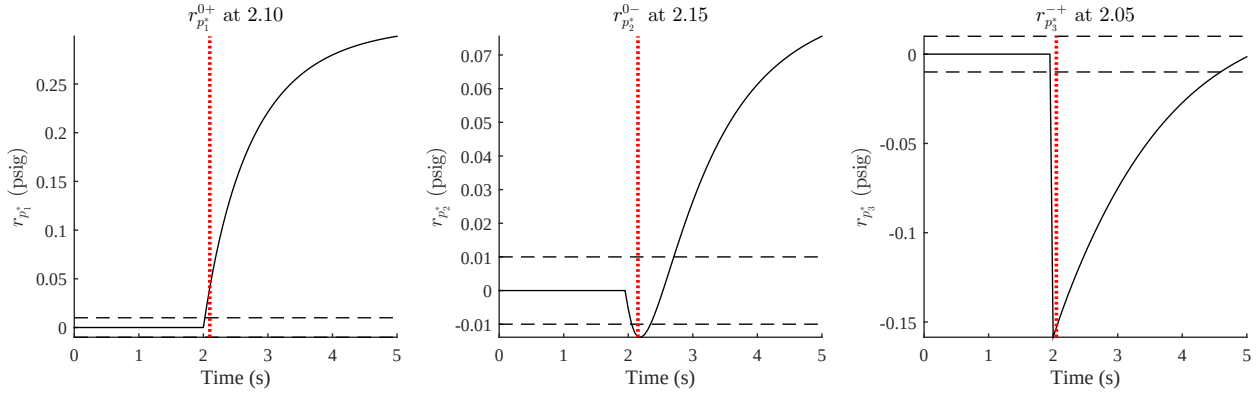

Figure 5: Observations for the fault $Re_1^+ K_3^+$, with K_3 doubling at 2 s and Re_1 tripling at 2.05 s, resulting in $r^{-+}_{p_3^*} r^{0+}_{p_1^*} r^{0-}_{p_2^*}$.

they occur together, the trace associated to the multiple-fault will be some merging of the traces of the constituent faults. How the individual faults come together to produce a single observed trace depends on the relative fault magnitudes and the relative times of occurrence. At the extreme, one fault in the fault set can either be (*i*) much larger than all the other faults, or (*ii*) occur earlier than all the other faults, such that the observed trace may be consistent with only that one fault occurring by itself. That is, it may completely mask all other faults. In the other extreme, we could observe a fault trace where each observed constituent signature is being produced by a different fault.

Example 10. For example, consider the fault $Re_1^+ K_3^+$, with $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$. The fault language for Re_1^+ consists of the single trace $r^{0+}_{p_1^*} r^{0+}_{p_2^*} r^{0+}_{p_3^*}$, and the fault language for K_3^+ consists of the single trace $r^{-+}_{p_3^*} r^{0-}_{p_2^*} r^{0-}_{p_1^*}$. When these two faults occur together, we must see some kind of composition of these two traces. The actual trace observed will depend on relative fault magnitudes and fault occurrence times. Fig. 4 shows one scenario, with K_3 doubling at 2 s and Re_1 doubling at 2.05 s. First, we observe $r^{-+}_{p_3^*}$ from K_3^+, followed by $r^{0+}_{p_1^*}$ from Re_1^+. We can see that $r_{p_2^*}$ begins to decrease (from K_3^+) but before crossing the threshold increases due to Re_1^+, and is observed as $r^{0+}_{p_2^*}$. If K_3^+ is larger, as in Fig. 5, then the decrease in $r_{p_2^*}$ may be larger and get detected instead, resulting in an observation of $r^{0-}_{p_2^*}$ instead. If Re_1^+ instead occurs first, as in Fig. 6, we may see $r^{0+}_{p_1^*}$ before $r^{-+}_{p_3^*}$.

To begin to formalize this concept, we first address the question of what the combined observed effect of two faults is on a single residual. There are three cases to consider. Either (*i*) no fault affects that residual, in which case no observation will be made for that residual; (*ii*) exactly one fault affects that residual, in which case the observed signature must be the same as for that fault occurring by itself; or (*iii*) both faults affect that residual, in which case

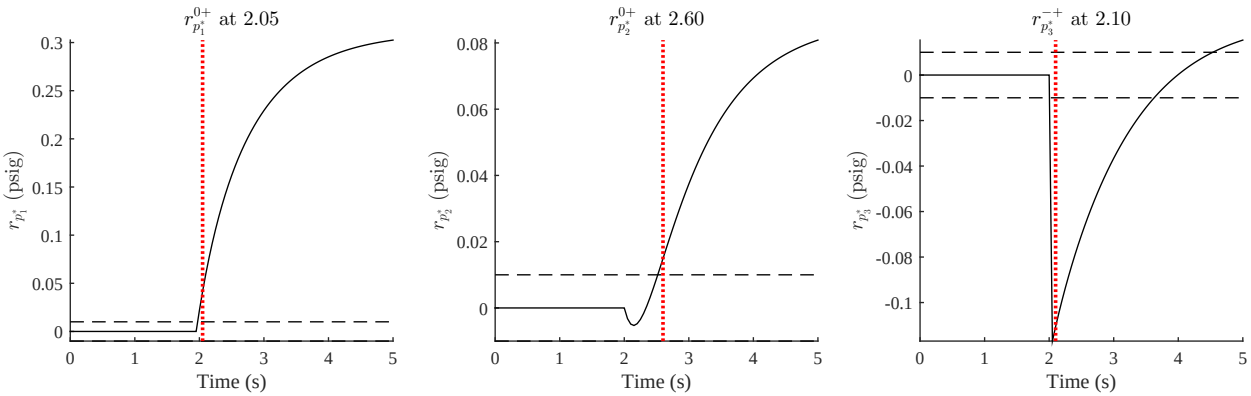

Figure 6: Observations for the fault $Re_1^+ K_3^+$, with Re_1 doubling at 2.05 s and K_3 doubling at 2.05 s, resulting in $r^{0+}_{p^*_1} r^{-+}_{p^*_3} r^{0+}_{p^*_2}$.

the observed signature must be some combination of the predicted signatures for the two faults. The third case can manifest in one of two ways: (*i*) one fault completely masks the other, either by occurring early enough or having a large enough magnitude, in which case that fault's signature is observed on the residual, or (*ii*) one fault does not mask the other, in which case we observe some combination of the individual signatures. Regarding this final case, we make the following assumption.

Assumption 2 (Signature Combination)**.** For residual r, if f_i produces $\sigma_{f_i,r}$ and f_j produces $\sigma_{f_j,r}$, where $\sigma_{f_i,r} \neq \sigma_{f_j,r}$, then when f_i and f_j both occur either $\sigma_{f_i,r}$ or $\sigma_{f_j,r}$ will be observed.

That is, we assume that we must observe a complete fault signature (both magnitude and slope) for one of the faults; we cannot observe some combination of their fault signatures (magnitude from one and slope from the other) or some other novel signature not predicted by either fault by itself.[7] Embedded in this assumption also is that the effects from either fault cannot perfectly cancel, i.e., that eventually the effect from one fault will dominate and be observed. Also embedded in this assumption is that the given fault signatures are valid at all operating points of the system, i.e., that f_i will produce only signatures in Σ_{f_i,r_i} does not change given that some f_j has occurred.

Given Assumption 2, we can obtain the following lemma, which summarizes all the cases mentioned above.

Lemma 2. *Given two faults, f_i and f_j, and some residual r, an observation σ_r when the faults both occur must belong to $\Sigma_{f_i,r} \cup \Sigma_{f_j,r}$.*

Further, we can claim the following.

Lemma 3. *Given residuals R and faults f_i and f_j, $\Omega_{\{f_i,f_j\},R} = \Omega_{f_i,R} \cap \Omega_{f_j,R}$.*

That is, the residual orderings for a multiple fault are given by the intersection of the individual orderings. If two faults would alone produce conflicting orderings, then when they occur together we cannot make any statement about which residual will deviate first. If two faults would alone produce the same ordering, then when occurring together we must observe the same ordering. This is derived from the main theorem behind residual orderings [10, 22].

Not only must the composed traces be consistent with the intersection of the orderings, but the actual signatures observed must be consistent, as stated in the following.

Lemma 4. *If $r_i \prec_f r_j \in \Omega_{f,R}$, then some $\sigma_{f,r_j} \in \Sigma_{f,r_j}$ cannot be observed until some signature is observed on r_i.*

[7] In some practical circumstances, this assumption may not hold. It can be easily dropped if we consider magnitude and slope effects on a residual as two distinct observations, rather than a single observation, where we have the additional temporal constraint in observation sequences that the magnitude effect for some residual must be observed before its slope effect. When considering these as two separate observations then the framework we present here is still valid, however tackling that more general case is beyond the scope of this paper.

Algorithm 1 $L_{ij} \leftarrow$ `ComposeTraces`$(\lambda_{i,R}, \lambda_{j,R})$

1: $L \leftarrow \{\epsilon\}$
2: $L_{ij} \leftarrow \varnothing$
3: **while** $|L| > 0$ **do**
4: $\quad \lambda \leftarrow \mathtt{pop}(L)$
5: $\quad \lambda_i^* \leftarrow \lambda\lambda^1_{i,R-R_\lambda}$
6: $\quad \lambda_j^* \leftarrow \lambda\lambda^1_{j,R-R_\lambda}$
7: $\quad$ **if** $\lambda_i^* = \lambda$ **and** $\lambda_j^* = \lambda$ **then**
8: $\quad\quad L_{ij} \leftarrow L_{ij} \cup \{\lambda\}$
9: $\quad$ **end if**
10: $\quad$ **if** $\lambda_i^* \neq \lambda$ **then**
11: $\quad\quad L \leftarrow L \cup \{\lambda_i^*\}$
12: $\quad$ **end if**
13: $\quad$ **if** $\lambda_j^* \neq \lambda$ **then**
14: $\quad\quad L \leftarrow L \cup \{\lambda_j^*\}$
15: $\quad$ **end if**
16: **end while**

For example, if we have $r_1 \prec_{f_1} r_2$ and $r_2 \prec_{f_2} r_1$, we cannot observe some σ_{f_1,r_2} followed by some σ_{f_2,r_1}. Residual orderings for f_1 require that r_1 must deviate before we see the effect from f_1 on r_2. In other words, what this lemma says is that if we get some trace resulting from some $f_i f_j$ and we project out any observations that were not the result of f_i, then the resulting trace λ_{f_i} must belong to $L_{f_i,R_{\lambda_{f_i}}}$, where for some trace λ, R_λ denotes the set of residuals included in the signatures of λ.

Together, these lemmas establish how to define a composition operation for traces, $\oplus$. First, though, we require the definition of a prefix of a trace.

Definition 17 (Prefix)**.** A trace λ_i is a *prefix* of trace λ_j, denoted by $\lambda_i \sqsubseteq \lambda_j$, if there is some (possibly empty) sequence of events λ_k that can extend λ_i s.t. $\lambda_i\lambda_k = \lambda_j$.

Definition 18 (Trace Composition)**.** A trace $\lambda_{f_if_j,R} = \sigma_1\sigma_2, \ldots, \sigma_n$ is a *composition* of traces $\lambda_{f_i,R}$ and $\lambda_{f_j,R}$, i.e., $\lambda_{f_if_j,R} \in \lambda_{f_i,R} \oplus \lambda_{f_j,R}$, if for every $\sigma_i \in \lambda_{f_if_j,R}$, $\sigma_i \sqsubseteq \lambda_{f_i,R-R_{\sigma_1,\ldots,\sigma_{i-1}}}$ or $\sigma_i \sqsubseteq \lambda_{f_j,R-R_{\sigma_1,\ldots,\sigma_{i-1}}}$.

Essentially, this means that if we want to construct a composition of two traces, the signatures in the new trace must come from either of the two original traces (Lemma 2), and the residual orderings must be respected (Lemmas 3 and 4). This follows from the lemmas above. Note that $\lambda_{f_i,R} \subseteq \lambda_{f_i,R} \oplus \lambda_{f_j,R}$ and $\lambda_{f_j,R} \subseteq \lambda_{f_i,R} \oplus \lambda_{f_j,R}$.

The algorithm to find all compositions of two traces $\lambda_{i,R}$ and $\lambda_{j,R}$ is given as Algorithm 1. We have a working set of traces L and a set of completed traces L_{ij}. Initially, we start with the empty trace ϵ. We then try to extend it with the first signature of $\lambda_{i,R}$ and $\lambda_{j,R}$, where for a trace λ, λ^i refers to the ith signature. These get added to L. We continue examining traces in L. A trace λ in L is replaced with an extension via $\lambda_{i,R-R_\lambda}$ and/or $\lambda_{j,R-R_\lambda}$. The extended trace is placed back into L. If the trace was not extended, this means that the trace is complete and goes in the set of completed traces L_{ij}.

Example 11. As an example, consider the fault $Re_1^+K_3^+$. A fault trace for Re_1^+ is $r^{0+}_{p_1^*} r^{0+}_{p_2^*} r^{0+}_{p_3^*}$, and a fault trace for K_3^+ is $r^{-+}_{p_3^*} r^{0-}_{p_2^*} r^{0-}_{p_1^*}$. Obviously, when both faults occur together, either $r^{0+}_{p_1^*}$ or $r^{-+}_{p_3^*}$ will have to be observed first. In Algorithm 1, lines 5 and 6 create these initial traces and they are added to L. Then, depending on relative magnitudes and fault occurrence times, if $r^{0+}_{p_1^*}$ is observed first (from Re_1^+) we will see either $r^{0+}_{p_2^*}$ (from Re_1^+) or $r^{-+}_{p_3^*}$ (from K_3^+). If instead $r^{-+}_{p_3^*}$ is observed first (from K_3^+), we will next see either $r^{0-}_{p_2^*}$ (from K_3^+) or $r^{0+}_{p_1^*}$ (from Re_1^+). This will be followed by an observation on the last residual (one consistent with either of the faults). The composition of the individual fault traces is then $\{r^{0+}_{p_1^*} r^{0+}_{p_2^*} r^{0+}_{p_3^*}, r^{0+}_{p_1^*} r^{0+}_{p_2^*} r^{-+}_{p_3^*}, r^{0+}_{p_1^*} r^{-+}_{p_3^*} r^{0-}_{p_2^*}, r^{0+}_{p_1^*} r^{-+}_{p_3^*} r^{0+}_{p_2^*}, r^{-+}_{p_3^*} r^{0-}_{p_2^*} r^{0-}_{p_1^*}, r^{-+}_{p_3^*} r^{0-}_{p_2^*} r^{0+}_{p_1^*}, r^{-+}_{p_3^*} r^{0+}_{p_1^*} r^{0-}_{p_2^*}, r^{-+}_{p_3^*} r^{0+}_{p_1^*} r^{0+}_{p_2^*}\}$.

Using this algorithm, we can construct traces for faults of any size. A trace $\lambda_{F,R}$ for $F = \{f_1, f_2, \ldots, f_n\}$ is a fault trace if $\lambda_{F,R} \in \lambda_{f_1,R} \oplus \lambda_{f_2,R} \oplus \ldots \oplus \lambda_{f_n,R}$. That is, multiple-fault traces are constructed as compositions of the traces of the constituent faults. Note that every fault trace for every $f \in F$ will also be a fault trace for F. To construct the fault

Algorithm 2 $L_{ij,R} \leftarrow$ `ComposeLanguages`$(L_{i,R}, L_{j,R})$

```
1: L_{ij,R} ← ∅
2: for all λ_{i,R} ∈ L_{i,R} do
3:     for all λ_{j,R} ∈ L_{j,R} do
4:         L_{ij,R} ← L_{ij,R} ∪ ComposeTraces(λ_{i,R}, λ_{j,R})
5:     end for
6: end for
```

language, we need simply to find all compositions of the fault traces for the constituent faults. This can be done in a constructive manner, where we first find the compositions for f_1 and f_2 in F, then composing those traces with the fault traces for f_3, and so on.

Composing two languages can be accomplished through Algorithm 2. For every pair of traces in the two languages, `ComposeTraces` is called to obtain all compositions of those two traces, and these are added to the composed language. To obtain the fault language for $F = \{f_1, f_2, \ldots, f_n\}$, we first compose $L_{f_1,R}$ with $L_{f_2,R}$ to obtain $L_{f_1f_2,R}$, then compose that with $L_{f_3,R}$ to obtain $L_{f_1f_2f_3,R}$, and so on.

Example 12. As an example, consider the fault set $Re_1^+K_3^+$. Since each fault contains only the single trace in its language, the set of composed traces for this fault set, as computed in Example 11, is also the fault language for $Re_1^+K_3^+$. The observed traces in Figs. 4–6 can be found within this language.

Example 13. Consider now the fault $Re_1^+K_3^+$, but with the submodel-based residual set. The fault language for Re_1^+ contains only $r_{p_1^*}^{0+}$ and the fault language for K_3^+ contains only $r_{p_3^*}^{-+}$. Therefore, the fault language is simply $\{r_{p_3^*}^{-+}r_{p_1^*}^{0+}, r_{p_1^*}^{0+}r_{p_3^*}^{-+}\}$.

A single-fault language grows as $O(|R|!)$, because in the worst case all possible interleavings of the residuals can occur. One benefit of using structural model decomposition is that each fault, on average, affects a smaller number of residuals. In fact, as the number of tanks grows, the number of residuals a fault affects in this case is at most 2, compared to n for the global model residuals. Therefore, the fault languages are much smaller when using structural model decomposition; they contain smaller traces and fewer traces, compared to those based on the global model residuals. So, on average, the computational complexity reduces significantly when structural model decomposition is used.

Given all the fault languages, fault isolation is, in theory, a trivial problem; we can simply search all the fault languages and a fault set F is a diagnosis if its language contains the observation sequence. However, it should be clear now that a fault language can be quite large. Not only does the size of a fault language grow exponentially with the number of residuals, but the number of languages to consider is, in general (i.e., without the fault cardinality limit l), exponential too, since there is an exponential number of diagnoses. Therefore, the naive approach to multiple fault diagnosis, in which we generate all fault languages and search them online, is not feasible in practice. An online approach, in which diagnoses are found incrementally as observations are received, is presented in the next section.

4. Multiple Fault Diagnosis

In Section 3, the problem addressed was, given a set of faults, to find all the potential traces it can produce, i.e., find the fault language. In this section, we consider the inverse problem, which is, given an observed trace, determine which fault sets are *consistent* with an observed trace, i.e., which are diagnoses.

In this framework, we follow the approach of consistency-based diagnosis [1, 15]. In this approach, fault isolation is based on *conflicts*, which are related to a set of correctness assumptions for the model that are not consistent with current observations from the system. In [15], a conflict is defined as a set of components for which all of them being nonfaulty is inconsistent with the model and the observations. Generalizing, we can say that a conflict is a set of correctness assumptions (e.g., a fault has not occurred) that cannot all be true, given the model and the observations.[8]

[8] For the component-based, static diagnosis problems in [1, 15], the correctness assumptions directly take the form $\neg AB(c)$ (meaning component c is not faulty) or $OK(c)$ (meaning component c is nominal). Here, since faults are not directly associated with components, but rather with model parameters, the correctness assumptions directly take the form of $\neg f$, i.e, that a fault has not occurred.

For example, a conflict of assumptions a_1, a_2, a_3 means that $\neg a_1 \vee \neg a_2 \vee \neg a_3$, i.e., either a_1 or a_2 or a_3 are not true. In this work, our correctness assumptions are that the parameter values in Θ are nominal, e.g., a_1 means that f_1 has not occurred, so a conflict is equivalent to a set of single faults that can explain an observation, i.e., $\neg a_1 \vee \neg a_2 \vee \neg a_3$ is $f_1 \vee f_2 \vee f_3$. So, a conflict is a set of faults, e.g., $\{f_1, f_2, f_3\}$, any one of which is consistent with a given observation.

In order to derive a conflict for a given observation, we must answer the question, which faults can produce the observation? In our framework, an observation is the deviation of some residual, i.e., a fault signature. The single-fault languages describe which signatures a single fault can produce, and in what sequence relative to other signatures. So, if a given fault signature is observed, we can check which fault can produce that signature, and since orderings must still be respected, it must be produced as the first signature in some fault trace, ignoring signatures for residuals that have already deviated (Lemma 4). Specifically, a conflict in our framework is defined as follows.

Definition 19 (Conflict)**.** Given a set of potential faults F, a set of residuals R, an observation sequence λ, and a new observation σ, a conflict C is a set of faults $C \subseteq F$, where for each $f \in C$, there is some $\lambda' \in L_{f,R-R_\lambda}$ such that $\sigma \sqsubseteq \lambda'$.

That is, given an observation sequence, for a fault to be able to explain a new observation, and be included in the conflict, it must be able to produce that observation as the first observation in some trace of its reduced fault language. The fault language must be reduced to the residual set $R - R_\lambda$, because it could be that the residuals for which we have observed signatures in λ were produced by other faults.

Example 14. Consider the global model residual set $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$ and the fault set $\{K_1^-, K_2^-, K_3^-, Re_1^+, Re_{1,2}^+, Re_2^+, Re_{2,3}^+, Re_3^+\}$. Say the first observation is $r_{p_1^*}^{+-}$. Then, the conflict is $\{K_1^-\}$, as that is the only fault that may produce that particular signature (see Table 2). Say the next observation is $r_{p_2^*}^{0+}$. Now, the conflict for that observation is $\{K_1^-, Re_1^+, Re_2^+, Re_{2,3}^+\}$, as these are the only faults that could produce this observation given that $r_{p_1^*}$ has already deviated. Note that K_3^- and Re_3^+ are not included, as they require that $r_{p_3^*}$ would have already deviated to be included in the conflict.

The diagnosis process proceeds incrementally, as new observations are made [1]. The initial diagnosis set is $\varnothing$. After the first observation, we obtain a conflict, and this simply becomes the new diagnosis set. After the next observation, we have a new conflict, and the new minimal diagnosis set is computed from the previous minimal diagnosis set and the conflict. Diagnosis proceeds in this way.

The incremental multiple fault isolation procedure is given as Algorithm 3. The algorithm is given as inputs the previous diagnosis D_i, the previous observation sequence λ_i, the new observation σ_{i+1}, and the candidate cardinality limit l. First, the conflict C is generated according to Defn. 19. Then, for each current diagnosis, we extend it once for each fault in the conflict to create an initial new diagnosis set D. This may produce diagnoses that are not minimal, i.e., for some $d \in D$ there may be some $d' \in D$ for which $d \subseteq d'$, in which case d' can be removed from D. Also, using the candidate cardinality limit l, we want to remove any diagnoses that are greater than the limit. This pruning step is done to produce the new diagnosis D_{i+1}. This method, without the fault size limit, produces equivalent results to the pruned hitting set tree approach proposed in [15].

Although we employ a fault cardinality limit of l, we are actually limited in what we can distinguish by the number of residuals, $|R|$. When a new observation is received, each diagnosis d in the diagnosis set is either consistent with that new observation, in which case it remains in the minimal diagonosis set, or it is inconsistent, in which case some new fault must be added to the diagnosis. In the latter case, for each fault in $f \in (C - d)$ we create a new diagnosis $d \cup \{f\}$. Each new diagnosis is created by extending with only one fault, since we want the minimal diagnoses only. Thus, if we can only have at most $|R|$ observations, the size of any diagnosis that is generated cannot exceed $|R|$.

Example 15. Consider again the fault and residual sets in Example 14, with $l = 2$. Say that we observe first $r_{p_1^*}^{+-}$, then the conflict is $\{K_1^-\}$ and the initial diagnosis set is also $\{K_1^-\}$. Next we observe $r_{p_2^*}^{0-}$, so the conflict is $\{Re_{1,2}^+\}$, as this is the only fault that can produce this observation. Then, the new diagnosis set is $\{K_1^- Re_{1,2}^+\}$, i.e, we know that both faults must have occurred. Next we observe $r_{p_3^*}^{0-}$, then the conflict is $\{Re_{1,2}^+, Re_{2,3}^+\}$, and so the new diagnosis set remains $\{K_1^- Re_{1,2}^+\}$. Note that we generate the candidate $K_1^- Re_{1,2}^+ Re_{2,3}^+$, however it is not minimal and is covered by the other candidate, and so not included in the minimal diagnosis set. That is, we are unsure as to whether the $r_{p_3^*}^{0-}$ observation came from $Re_{1,2}^+$, which we already know must have occurred, or $Re_{2,3}^+$ for which we are unsure that it has occurred. In fact, it is less likely that the triple fault occurred rather than the double fault.

Algorithm 3 $D_{i+1} \leftarrow$ `FaultIsolation`$(D_i, \lambda_i, \sigma_{i+1}, l)$

$D \leftarrow \varnothing$
$D_{i+1} \leftarrow \varnothing$
$C \leftarrow \{f \in F : \exists \lambda \in L_{f,R-R_{\lambda_i}}$ **where** $\sigma_{i+1} \sqsubseteq \lambda\}$
for all $d \in D_i$ **do**
  for all $f \in C$ **do**
    $D \leftarrow D \cup \{d \cup \{f\}\}$
  end for
end for
for all $d \in D$ **do**
  if $|d| \leq l$ **and** d is minimal **then**
    $D_{i+1} \leftarrow D_{i+1} \cup \{d\}$
  end if
end for

Example 16. Consider the same scenario, but using the local submodel residual set (see Table 3). First, we observe $r^{+-}_{p^*_1}$, and as before $\{K^-_1\}$ is the conflict and the initial diagnosis set. Next, we observe $r^{0-}_{p^*_2}$, and again the conflict is $\{Re^+_{1,2}\}$, and the new diagnosis set is $\{K^-_1 Re^+_{1,2}\}$. Next, we observe $r^{0-}_{p^*_3}$. Here, the conflict is only $\{Re^+_{2,3}\}$, because $Re^+_{1,2}$ is now independent of this residual. Thus, the new diagnosis set is $\{K^-_1 Re^+_{1,2} Re^+_{2,3}\}$, i.e., we know for certain that all three faults have occurred. If it was in fact only $K^-_1 Re^+_{1,2}$ that had occurred then we would not observe any deviation in r_{p_3} and the diagnosis would be correct as well.

These examples demonstrate the power of the structural model decomposition-based residual set. Here, with the same fault scenario, we obtain two different minimal diagnosis sets with the two different residual sets. With the global model residuals, if $K^-_1 Re^+_{1,2} Re^+_{2,3}$ occurs, it can also look like only $K^-_1 Re^+_{1,2}$ has occurred. But, with the local submodel residuals, we will be able to distinguish between the two cases.

The naive approach to multiple fault isolation has poor space complexity, because it needs to compute all the languages for all possible fault sets. However, time complexity for online isolation is fast, i.e., using an efficient data structure like a hash table, observed traces can be quickly mapped to consistent fault sets. On the other hand, the incremental approach has good space complexity, because only single fault information has to be captured; fault languages for multiple faults do not need to be generated. The traces themselves do not need to be directly represented, but instead only the signatures and residual orderings for each fault ($O(|R|)$ signatures and $O(|R|^2)$ orderings). When a new observation is obtained, we must search through all $|F|$ single faults to produce the conflict and update the previous diagnosis set to obtain the new diagnosis set. The smaller the conflict, the less the work done in creating the new diagnosis set.

Structural model decomposition provides an advantage in both approaches. Primarily, the advantages derive from the decoupling of faults from residuals. Due to this, each single fault responds to less than $|R|$ residuals, on average. On the other hand, in the global model, typically all residuals in R are affected by a single fault. So, fault traces and fault languages are smaller, on average, especially for multiple faults, because there are fewer ways in which faults can interact and the resulting fault languages are smaller. With the naive approach then, space complexity reduces. With the online approach, since the conflicts are, on average, smaller (since only a subset of the faults can produce any given observation in a residual), and so the complexity of producing a new diagnosis with each new observation is smaller.

Further, since conflicts are smaller, fewer new diagnoses will be generated and the diagnosis results will have less ambiguity. This result is captured formally through diagnosability analysis, described in the next section.

5. Diagnosability

Diagnosability describes, for a given system model and diagnosis scheme, how well faults can be isolated. Such a metric is useful during system design, and is the basis of sensor selection approaches [23–25].

Diagnosability is founded on the notion of *distinguishability*, which is concerned with whether, if some fault set F_i occurs, can it produce the same observation sequence as some other fault set F_j. If so, then if that observation

sequence occurs, the fault isolation algorithm will not be able to determine whether it is F_i or F_j that has occurred. In our framework, distinguishability of faults is derived from the fault languages, and can be defined as follows.

Definition 20 (Distinguishability)**.** With residuals R, a fault set F_i is distinguishable from a fault set F_j, denoted by $F_i \nsim_R F_j$, if there is no $\lambda_i \in L_{F_i,R}$ where for some $\lambda_j \in L_{F_j,R}$, $\lambda_i \sqsubseteq \lambda_j$.

If a fault set F_i produces a trace that is a prefix of a trace that may be produced by another fault set F_j, then, when that trace occurs, both F_i and F_j will be consistent and will in the (maximal) diagnosis set. Since F_i will produce no other observation, F_j cannot be eliminated, so we can never confirm that F_j has not actually occurred in this case.

Example 17. Consider the single faults K_1^- and Re_3^+ with the global model residual set $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$. Here, $L_{K_1^-,R} = \{r_{p_1^*}^{+-} r_{p_2^*}^{0+} r_{p_3^*}^{0+}\}$, and $L_{Re_3^+,R} = \{r_{p_3^*}^{0+} r_{p_2^*}^{0+} r_{p_1^*}^{0+}\}$. Clearly, these two faults are distinguishable from each other, because the first observations will always be different. However, the faults K_1^- and $K_1^- Re_3^+$ are not distinguishable from each other. In practice, this is due to fault masking.

Since structural model decomposition decouples faults from residuals, it can eliminate some masking possibilities and, thus, improve diagnosability.

Example 18. Consider again the single faults K_1^- and Re_3^+ but with the submodel-based residual set $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$. Here, $L_{K_1^-,R} = \{r_{p_1^*}^{+-}\}$, and $L_{Re_3^+,R} = \{r_{p_3^*}^{0+}\}$. Clearly, these two faults are distinguishable. Also, $K_1^- Re_3^+$ is distinguishable from both K_1^- and Re_3^+. However, the converse is still not true, i.e., K_1^- is not distinguishable from $K_1^- Re_3^+$, and Re_3^+ is not distinguishable from $K_1^- Re_3^+$. If K_1^- occurs, then we see $r_{p_1^*}^{+-}$, which so far, is consistent with both the single and the double fault. We then have to wait infinitely long to ensure that $r_{p_3^*}^{0+}$ does not occur and confirm that K_1^- has occurred by itself, and so we say they are not distinguishable.

With distinguishability defined, we can now begin to define diagnosability. Diagnosability is defined as a metric that expresses the number of distinguishable pairs of faults, for a given diagnosis model and a set of possible observations. Thus, a higher diagnosability is better. Here, in the diagnostic context, the model takes the form of the fault languages, and the observations are based upon the available residuals. An ideal fault isolation algorithm could, at best, do as well as established by diagnosability.

Definition 21 (l-Diagnosability)**.** For set of fault sets $\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$, where for $F_i \in \mathcal{F}$ $|F_i| \leq l$, and with fault languages $L_{\mathcal{F},R} = \{L_{F_1,R}, L_{F_2,R}, \ldots, L_{F_n,R}\}$ the *l-diagnosability* of $\mathcal{F}$ is the number of fault set pairs, $(F_i, F_j \in \mathcal{F})$ $F_i \neq F_j$ where $F_i \nsim_R F_j$.

Here, the set of fault sets does not have to be the full powerset of single faults, e.g., it may include only single faults, single faults and double faults, etc.

For $|\mathcal{F}|$ possible fault sets, the worst (i.e., minimum) possible diagnosability is 0 and the best (i.e, maximum) is $|\mathcal{F}|(|\mathcal{F}| - 1)$. Recall that distinguishability is not a symmetric property, so for two F_i and F_j we count both $F_i \nsim_R F_j$ and $F_j \nsim_R F_i$. The normalized diagnosability metric is expressed as the fraction of actual diagnosability over the best possible diagnosability.

Example 19. Consider the single faults $F = \{K_1^+, K_1^-, K_2^+, K_2^-, K_3^+, K_3^-, Re_1^+, Re_1^-, Re_2^+, Re_2^-, Re_3^+, Re_3^-, Re_{1,2}^+, Re_{1,2}^-, Re_{2,3}^+, Re_{2,3}^-\}$ with the global model residuals $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$. There are 16 single faults, so at best the 1-diagnosability is 240. Here, we obtain 100% diagnosability, i.e., all single-fault pairs can be distinguished. Consider now the single faults and the double faults; there are 112 double faults (excluding those of the form $\theta^+\theta^-$), for a total of $|\mathcal{F}| = 128$, and diagnosability is at most $16,256$. Here, we obtain 72.27% diagnosability.

Different sets of residuals provide different diagnostic information, and, hence, different diagnosability.

Example 20. Consider the same single fault set as in the previous example, but with the local submodel residuals $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$. Here, diagnosability is 96.67%. Diagnosability is not perfect, because for some fault set pairs we have to wait infinitely long to distinguish them. For example, If Re_1^+ occurs, we observe $r_{p_1^*}^{0+}$. This observation can be due also to $Re_{1,2}^+$ (see Table 3), therefore, the fault isolation algorithm will include both faults in the diagnosis set, and since no other residuals will deviate, $Re_{1,2}^+$ cannot be eliminated. Consider now the single and double faults.

Now, diagnosability is 90.22%, which is a significant improvement over using the global model residuals. Although structural model decomposition decreases diagnosability slightly in the single-fault case, when considering double-faults, the advantage is quite clear.

Diagnosability can be improved further by including the combined residual sets from both the global model and the local submodels.

Example 21. Consider the same fault set as the previous example, but now with both the global model and local submodel residual sets. Now, diagnosability is 91.44%.

Defn. 21 defines diagnosability with respect to the maximal diagnosis set. But, note that, as described in Section 3, if $F \subseteq F'$, then $L_{F,R} \subseteq L_{F',R}$. Therefore, F will never be distinguishable from F'. So, if F occurs, it will *always* produce some trace that could have also been produced by F'. However, most often we are interested only in whether we have a unique diagnosis in the *minimal* diagnosis set. For example, in [15], the term *diagnosis* is explicitly defined to be the minimal explanations for faulty behavior, via the principle of parsimony. In such a case, we do not need to distinguish between some fault F and some other fault F' if $F \subset F'$, because if F is consistent then F' will not be included in the minimal diagnosis set. This leads to an alternative definition of diagnosability.

Definition 22 (Minimal l-Diagnosability)**.** For set of faults $\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$, where for $F_i \in \mathcal{F}$ $|F_i| \leq l$, and with fault languages $L_{\mathcal{F},R} = \{L_{F_1,R}, L_{F_2,R}, \ldots, L_{F_n,R}\}$ the *minimal l-diagnosability* of $\mathcal{F}$ is the number of fault pairs, $(F_i, F_j \in \mathcal{F})$ $F_i \nsubseteq F_j$ where $F_i \nsim_R F_j$.

Example 22. Consider the single faults $F = \{K_1^+, K_1^-, K_2^+, K_2^-, K_3^+, K_3^-, Re_1^+, Re_1^-, Re_2^+, Re_2^-, Re_3^+, Re_3^-, Re_{1,2}^+, Re_{1,2}^-, Re_{2,3}^+, Re_{2,3}^-\}$ with the global model residuals $R = \{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$. With minimal 2-diagnosability, the best score reduces to $16,023$, and the diagnosability is 74.67%, which is a bit better than with the previous diagnosability definition. For the local submodel residuals, minimal 2-diagnosability is 90.56%, again a bit better than with the standard definition. With the combined residual sets it is 91.32%. Since the set of residuals depends on the selected sensors, including more sensors or considering a different sensor set can also impact diagnosability. Measuring the flows, we see an increase in diagnosability for the global model residuals, but a decrease for the local submodel residuals, since the the amount of decoupling provided by structural model decomposition is reduced with this sensor set. Diagnosability results are summarized in Table 4.

Table 4: 2-Diagnosability for the Three-tank System.

	Pressures			Flows		
	Global	Local	Combined	Global	Local	Combined
Maximal	72.27%	90.22%	91.44%	76.77%	88.66%	90.22%
Minimal	74.67%	90.56%	91.32%	79.24%	89.12%	90.10%

It is also interesting to investigate how diagnosability changes with the size of the system. As the system size increases, there are more faults, and hence more ways for the faults to interact and reduce distinguishability. For n single faults, there are $\binom{n}{2}$ double faults, and for $|\mathcal{F}|$ fault sets, there are $(|\mathcal{F}|)(|\mathcal{F}| - 1)$ fault set pairs for which to check diagnosability. So, as the system size increases, the best possible diagnosability increases very fast, and we want the actual diagnosability to grow at least that fast. For example, if diagnosability is 80% and the system size is increased, we want the number of new distinguishable pairs to be at least 80% of the new potentially distinguishable pairs. Relative to the system size, we want diagnosability to not decrease.

Example 23. Consider the tank system, where the number of tanks is increased. Diagnosability is shown in Fig. 7. Each new tank adds 6 new single faults, and the total number of single faults is $|F^1| = 4 + 6(n - 1)$ for n tanks. There are $\binom{|F^1|}{2} - |F^1|/2$ total double faults (the $|F^1|/2$ is to eliminate fault sets containing both θ^+ and θ^- for some fault parameter θ), and so $|\mathcal{F}| = |F^1|/2 + \binom{|F^1|}{2}$ total fault sets. So the best possible diagnosability for 2 tanks is 2450, for 3 tanks is $16,256$, and for 4 tanks is $58,322$. However, considering the local submodel residuals with minimal 2-diagnosability, only 376 fault set pairs are indistinguishable for 2 tanks, 1310 for 3 tanks, $2,798$ for 4 tanks, and so

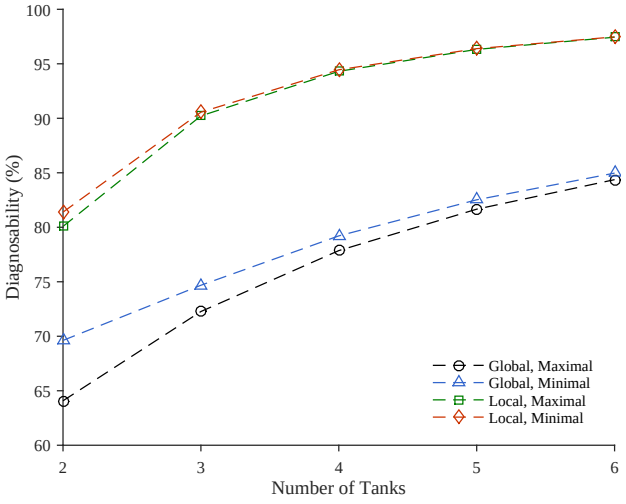

Figure 7: 2-diagnosability as a function of the number of tanks.

on. Relative to the best possible diagnosability, this number does not grow as fast. In fact, it grows relatively slower, so diagnosability is increased relative to the best possible diagnosability, as shown in Fig. 7. This occurs for both the global model and local submodel residual sets for both diagnosability definitions.

6. Experimental Results

In this section, we evaluate the multiple fault diagnosis approach online using a simulated three-tank system. The overall diagnosis approach is described in [26], except that we use the multiple fault isolation approach developed in this paper. Residuals are generated as described in Section 4. Observations (fault signatures) are generated from the residuals using a signal processing technique involving the Z-test [16]; for the purposes of this paper, we assume that observed fault signatures are correctly generated (initial progress on dropping this assumption for the single-fault case is described in [27]).

For all fault scenarios, we consider the fault set $\{K_1^-, K_2^-, K_3^-, Re_1^+, Re_{1,2}^+, Re_2^+, Re_{2,3}^+, Re_3^+\}$, and consider residuals $\{r_{p_1^*}, r_{p_2^*}, r_{p_3^*}\}$ for both the global model and local submodels. To demonstrate the improvements offered by using qualitative fault signatures and residual orderings, we consider also the local submodel residual set without orderings and using only binary fault signatures (effect/no effect).

As a first scenario, we consider a double fault in which K_3^- first occurs at $t = 5$ s, followed by Re_1^+ at $t = 5.05$ s. Consider first diagnosis with the global model residual set, shown in Fig. 8. At $t = 5$ s $r_{p_3^*}^{+-}$ is observed, resulting in a conflict of $\{K_3^-\}$ and initial diagnosis set $\{K_3^-\}$. At $t = 5.05$ s, $r_{p_2^*}^{0+}$ is observed, resulting in a conflict of $\{K_3^-, Re_2^+, Re_3^+, Re_{2,3}^+\}$. The minimal diagnosis set remains as $\{K_3^-\}$. At $t = 5.1$ s, $r_{p_1^*}^{0-}$ is observed. The conflict is $\{K_3^-, Re_3^+, Re_2^+, Re_{2,3}^+, Re_1^+, Re_{1,2}^+, K_2^-\}$. Still, the minimal diagnosis set remains as $\{K_3^-\}$. So, even though two faults occurred, with the global residual set the observations are consistent with only K_3^- occurring by itself.

Consider now diagnosis with the local submodel residual set, shown in Fig. 9. At $t = 5$ s, $r_{p_3^*}^{+-}$ is observed, resulting in a conflict and initial minimal diagnosis set of $\{K_3^-\}$. At $t = 5.1$ s, $r_{p_1^*}^{0+}$ is observed, resulting in a conflict of $\{Re_{1,2}^+, Re_1^+\}$. Unlike with the global model residuals, with the local submodel residual set, K_3^- has no effect on this residual, so here we are confident that a second fault has occurred. In this case, then, the minimal diagnosis set is $\{K_3^- Re_{1,2}^+, K_3^- Re_1^+\}$. Although it is ambiguous as to which fault occurred with K_3^-, at least it is known that a double fault has definitely occurred.

Consider now diagnosis with the local submodel residual set, but without residual orderings and using binary fault signatures. At $t = 5$ s, $r_{p_3^*}^{+-}$ is observed, resulting in a conflict and initial minimal diagnosis set of $\{K_3^-, Re_{23}^+, Re_3^+\}$. At

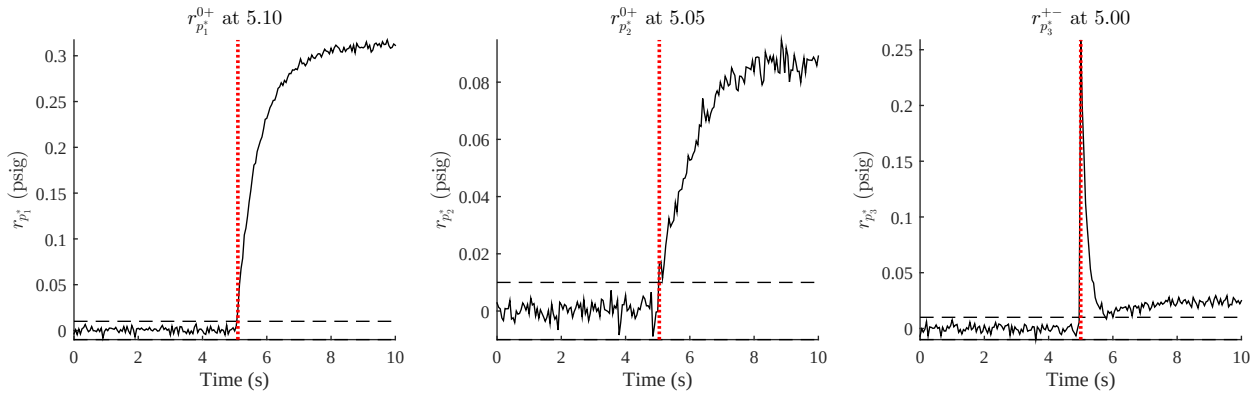

Figure 8: Observations for the candidate $K_3^- Re_1^+$ with the global model residuals.

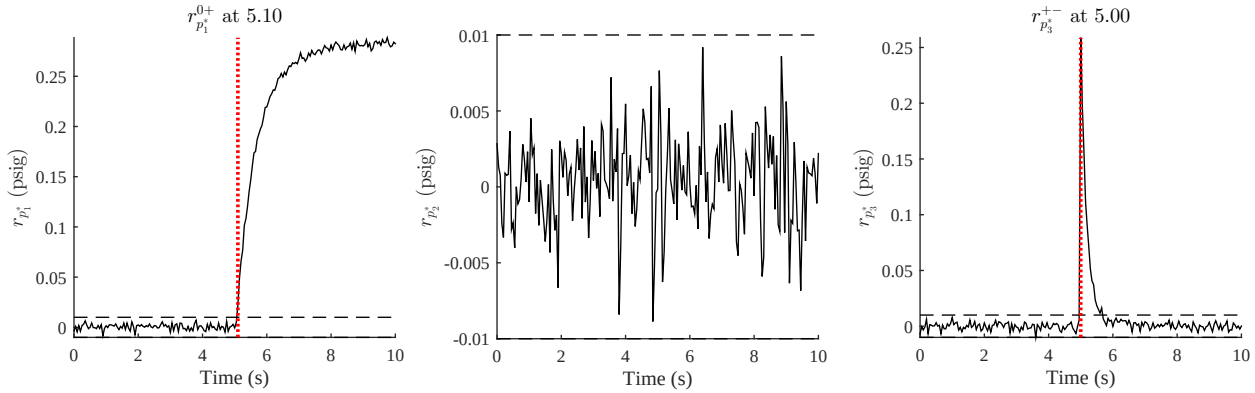

Figure 9: Observations for the candidate $K_3^- Re_1^+$ with the local submodel residuals.

$t = 5.1$ s, $r^{0+}_{p^*_1}$ is observed, resulting in a conflict of $\{K_1^-, Re_{1,2}^+, Re_1^+\}$. The minimal diagnosis set is then $\{K_1^- K_3^-, K_3^- Re_1^+, K_3^- Re_2^+, K_1^- Re_{23}^+, Re_1^+ Re_{23}^+, Re_{12}^+ Re_{23}^+, Re_1^+ Re_3^+, Re_{12}^+ R_3^+\}$. Clearly, although the actual double fault is included in the minimal diagnosis set, the reduction in diagnostic information, as expected, results in a significant loss in precision. A double fault is known to have occurred, but there are 8 consistent diagnoses.

As a second scenario, consider a triple fault, with K_1^- occurring at $t = 5$ s, $Re_{1,2}^+$ occurring at $t = 5$ s, and Re_3^+ occurring at $t = 5.05$ s. Consider first diagnosis with the global model residual set, shown in Fig. 10. At $t = 5$ s, $r^{+-}_{p^*_1}$ is observed, which can be due only to K_1^-, and so the conflict is $\{K_1^-\}$ and the initial diagnosis set is $\{K_1^-\}$. At $t = 5.05$ s, $r^{0+}_{p^*_2}$ is observed, so the conflict is $\{K_1^-, Re_{2,3}^+, Re_1^+, Re_2^+\}$, but the minimal diagnosis set remains as $\{K_1^-\}$. At $t = 5.10$ s, $r^{0+}_{p^*_3}$ is observed, and so the conflict is $\{K_1^-, K_2^-, Re_1^+, Re_2^+, Re_3^+\}$. Still, the minimal diagnosis is only $\{K_1^-\}$, i.e., the observation sequence is still consistent with only a single fault occurring and we cannot say for sure whether a second (or third) fault has also occurred.

Now, consider diagnosis with the local submodel residual set, shown in Fig. 11. At $t = 5$ s, $r^{+-}_{p^*_1}$ is observed, which can be due only to K_1^-, and so the conflict is $\{K_1^-\}$ and the initial diagnosis set is $\{K_1^-\}$. At $t = 5.05$ s, $r^{0-}_{p^*_2}$ is observed, so the conflict is $\{Re_{1,2}^+\}$, so the minimal diagnosis set is $\{K_1^- Re_{1,2}^+\}$, i.e., we know that definitely two faults have occurred, and the two faults that must have occurred are unambiguous. This is due to the decoupling from the local submodels; now, K_1^- cannot mask the effects of $Re_{1,2}^+$ as seen with the global model residuals. At $t = 5.10$ s, $r^{0+}_{p^*_3}$ is observed, and so the conflict is $\{Re_3^+\}$, and the minimal diagnosis set is $\{K_1^- Re_{1,2}^+ R_3^+\}$, i.e., we know that three faults have occurred and they are known accurately. The more the decoupling, the more accurate multiple fault diagnosis can be, because of the decreased chances for fault masking.

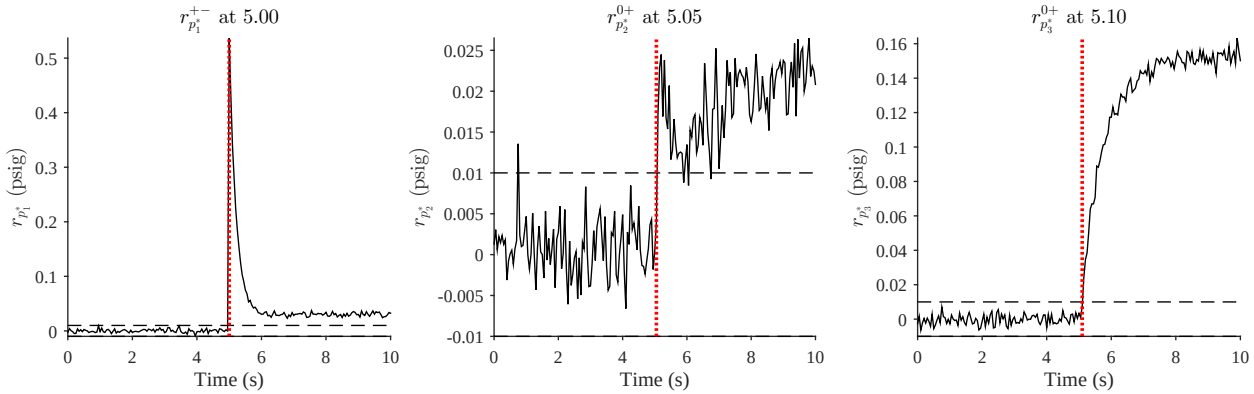

Figure 10: Observations for the candidate $K_1^- Re_{1,2}^+ Re_3^+$ with the global model residuals.

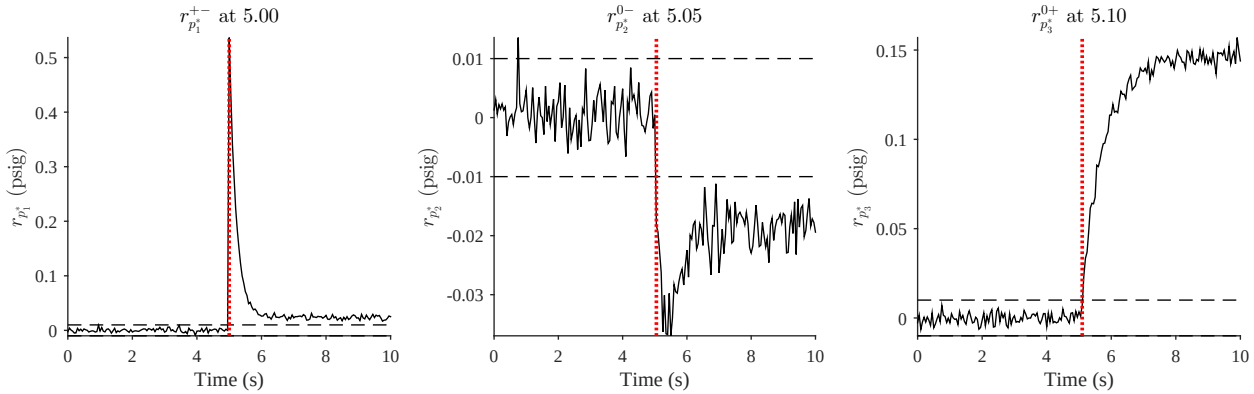

Figure 11: Observations for the candidate $K_1^- Re_{1,2}^+ Re_3^+$ with the local submodel residuals.

Consider now diagnosis with the local submodel residual set, but again without residual orderings and only binary fault signatures. Since deviations are observed in all three residuals, all combinations of faults that cover the three residuals are diagnoses. This would include some double faults (e.g., $K_1^- Re_{23}^+$), so the minimal diagnosis set would not include the true triple fault, whereas in the previous situation the exact triple fault was diagnosed.

6.1. Summary of Results

In order to characterize the multiple fault diagnosis performance, and to compare the global model and local submodel residual sets, we present a comprehensive set of experimental results performed in simulation. We consider both single- and double-fault scenarios, where the first fault is always injected at $t = 5.00$ s, and the second fault also at $t = 5.00$ s or a little later at $t = 5.05$ s. All fault combinations are considered, and $l = 3$ is assumed.

We consider a set of metrics to evaluate the performance of diagnosis. We consider both the maximal and minimal diagnosis sets. We determine (*i*) whether or not the true fault is found in the diagnosis set D, (*ii*) the length of the diagnosis set $|D|$, and (*iii*) the accuracy, as measured by $1/|D|$ if the true fault is in D or 0 otherwise. Since diagnosability is not perfect, we expect at least that the true fault is in D, and we desire D to have at best only one diagnosis, the true fault set.

For the global model residuals, we find that for the maximal diagnosis set, the true fault is always included. On average, $|D| = 6.39$, and accuracy is 30%. Accuracy is rather low, since the size of the diagnosis set is often large. For the minimal diagnosis set, the true fault is only included in the diagnosis set 42% of the time. Here, since the true fault is always in the maximal diagnosis set, it can only not be in the minimal diagnosis set in one case, which is when a double fault occurs with a trace that is consistent with a single fault. On average, $|D| = 1.84$, and accuracy is 28%.

Accuracy is likely a bit lower since sometimes the true fault is not in the minimal diagnosis set, even though $|D|$ is, on average, smaller.

For the local submodel residuals, we find that for the maximal diagnosis set, the true fault is, like with the global model residuals, always included. On average, $|D| = 3.35$, a littler more than half of that for the global model residuals, and accuracy is 56%, nearly double that for the global model residuals. For the minimal diagnosis set, the true fault is included in the diagnosis set 83% of the time, again about twice that for the global model residuals. On average, $|D| = 1.57$ and accuracy is 59%, again much better than with the global model residuals.

For the local submodel residuals without orderings and with binary fault signatures, we find that for the maximal diagnosis set, the true fault is, as expected, always included. On average, $|D| = 14.33$, which is significantly worse than the previous two approaches. Accuracy is 7%, which is also a significant decrease in performance. For the minimal diagnosis set, the true fault is included in the diagnosis set 63% of the time. This is better than using all diagnostic information on the global model residuals due to the benefit of using structural model decomposition, but worse than using all diagnostic information with the local submodel residuals. On average, $|D| = 7.58$ and accuracy is 8%, which are both worse than the previous two approaches. Overall, these results are expected, because when using less information for diagnosis, the diagnosis sets will be less accurate and precise.

Clearly, the use of structural model decomposition greatly improves the performance of multiple fault isolation. These results are expected, given the diagnosability analysis. With the metrics used here, we find that using the local submodel residuals offers about twice the performance as using the global model residuals. This is mainly due to the fact that conflicts are, on average, smaller, so the diagnosis sets are smaller and, as such, accuracy improves. Using both residual sets, performance would improve further.

With the combined residual set, for the maximal diagnosis, on average, $|D| = 3.33$ and accuracy is 56%, which are about the same as using only the local submodel residuals. For the minimal diagnosis set, the true fault is included in the diagnosis set 86% of the time, which is a small improvement. On average, $|D| = 1.9$ and accuracy is 59%. Here, although accuracy is the same as using only the local submodel residuals, $|D|$ is larger because it includes more triple faults, since there are 6 total residuals now, so there are more chances for double faults to be expanded to triple faults.

7. Related Work

Traditionally, multiple fault diagnosis solutions have concentrated mostly on static systems, e.g., [1, 2, 15] and, more recently, [3]. The approach in this paper is founded in the conflict recognition and candidate generation methodology in [1]. Their particular implementation, GDE, utilizes a notion of minimality equivalent to the one we use in this paper, although it only applies to static systems. In parallel, the consistency-based diagnosis approach of [15] also develops multiple fault diagnosis solutions for static systems. A key contribution of our work in applying this fundamental approach to dynamic systems is to generate conflicts for dynamic systems, based on the notion of fault traces.

These early works have been extended before to diagnosis of dynamic systems, by using qualitative simulation models [4, 7, 8], which suffer from the state explosion problem. In contrast, our approach uses a qualitative abstraction of the *residual space*, i.e., we require a reference only to nominal behavior, and, so, for each residual, we have a finite set of symbolic abstractions (i.e., fault signatures). From these, a finite set of fault traces can be constructed.

In control theory-based (FDI) diagnosis approaches, the proposal in [9] is, like our approach, based on the analysis of residual structures. The residual structure of [9] is derived offline to fulfill a set of desired isolation properties. Structural model decomposition allows us to achieve a similar structure, however our decomposition approach is more general. Further, FDI approaches use only binary signatures (effect or no effect), whereas we use a richer feature set defined by the fault signatures. Also, the use of relative residual orderings adds temporal information for diagnosis that improves the discriminatory power of the approach. Such information is also lacking from [5], which, like our approach, integrates residual-based and consistency-based approaches. In [28], the authors propose an efficient graph-theoretical algorithm for computing a set of testable submodels called Test Equation Supports (TES). Similarly to our approach, [28] structurally decomposes the system model into minimal submodels. The key difference with our approach is that the TESs in a direct way characterize the complete multiple fault isolability property of a model. However, isolability information in the proposed solution is only binary, and no information about the ordering in the residual deviation is used. A similar approach is followed in [29–31]. Since these approaches consider only binary information from residuals, our approach will always be more precise, as demonstrated in Section 6.

A more general approach based on binary tests is described in [32], which allows also for observation delay. Our multi-valued fault signatures can also be transformed to a binary format (e.g., one test would be if +− is observed in r_1, another would be if 0+ is observed on r_1, etc.), but this creates the problem that complex test expressions would be required (e.g., f_1 causes test 1 (r_1^{+-}) to be true *or* causes test 2 to be true (r_1^{+-})), which cannot be handled by that approach.

Some diagnosis approaches also use temporal information similar in concept to relative residual orderings. In [33], there is the concept of a fault influence path, which is similar to our notion of relative residual ordering, although the focus there is on hybrid systems. In [34], the model is decomposed into analytical redundancy relations, which are equivalent to our minimal submodels, and the order in the residuals deviations is also used for multiple fault isolation. However, in both these approaches, qualitative information in the residual is not used. In [35], directional residuals are computed for multiple fault distinguishability. However, the solutions proposed there only apply to double fault situations.

To make the approach as general as possible, we assumed in this work that qualitative fault signatures and relative residual orderings were given as inputs. In practice, many approaches can be found in the literature to generate this information automatically from a given model, and all of them can be integrated within the multiple fault diagnosis framework proposed in this paper [10, 36–39]. Additionally, such information can also be generated by manual analysis of the system model.

8. Conclusions

In this work, we have presented a qualitative, event-based framework for multiple fault isolation. Within this framework, we have developed a systematic approach for predicting the possible traces, called fault traces, that multiple faults can produce. Using diagnostic information provided by these fault traces, we have provided methods for online fault isolation of multiple faults, and for offline diagnosability analysis, which demonstrates how good an ideal fault isolation algorithm would perform given a diagnosis model in our framework. It was shown that using structural model decomposition can greatly improve fault diagnosis in the multiple-fault case, as it (*i*) reduces the possibility of fault masking due to the decoupling of faults and residuals, and thus improves diagnosability; and (*ii*) results in a reduction of the average computational complexity of the diagnosis problem, in both time and space. Using a tank system as a case study, it was shown that multiple fault diagnosis using structural model decomposition yields excellent performance.

The diagnosis approach presented in this paper can be used to quickly produce a set of diagnoses based on qualitative, event-based diagnostic information. Diagnosability is rarely perfect when considering multiple faults, and so ambiguity in diagnosis results can rarely be avoided. As such, it is important to follow qualitative fault isolation with a quantitative fault *identification* step [21]. For multiple faults, this becomes difficult because the dimension of the parameter estimation problem is increased. However, structural model decomposition provides an advantage here, because if the faults appear in different submodels, then instead of having one n-dimensional estimation problem, we have in the best case n 1-dimensional estimation problems. Future work will consider the problem of fault identification for the multiple-fault case.

A limitation of the framework is the assumption of correct observation of fault effects. If an observation is incorrect, this may lead to incorrect diagnoses being generated. Initial work on this topic for this diagnosis framework has been done in the single-fault case [27], however it should be extended also to the multiple-fault case. Here, the problem becomes more complex because an incorrect observation can be taken as evidence of a new fault when there is none, or mask the true presence of a new fault.

Here, we considered only continuous systems, but multiple fault isolation must also be considered for hybrid systems [10, 33, 40]. This is also a topic of future work. Initial work on extending our structural model decomposition approach to hybrid systems is presented in [17].

References

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

[2] P. Struss, O. Dressler, Physical negation: Integrating fault models into the general diagnostic engine, in: Proc. of the 11th International Joint Conference on Artifical Intelligence (IJCAI-89), Detroit, Michigan, USA, 1989, pp. 1318–1323.

[3] R. Abreu, A. van Gemund, Diagnosing multiple intermittent failures using maximum likelihood estimation, Artif. Intell. 174 (18) (2010) 1481–1497.
[4] D. Dvorak, B. Kuipers, Process monitoring and diagnosis: a model-based approach, IEEE Expert 6 (3) (1991) 67 –74.
[5] M. Nyberg, M. Krysander, Combining AI, FDI, and statistical hypothesis-testing in a framework for diagnosis, in: Proc. of IFAC Safeprocess'03, 2003, pp. 813–818.
[6] M. Daigle, X. Koutsoukos, G. Biswas, A qualitative approach to multiple fault isolation in continuous systems, in: Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, 2007, pp. 293–298.
[7] H. T. Ng, Model-based, multiple fault diagnosis of time-varying, continuous physical devices, in: Sixth Conference on Artificial Intelligence Applications, Vol. 1, 1990, pp. 9–15. doi:10.1109/CAIA.1990.89165.
[8] S. Subramanian, R. J. Mooney, Qualitative multiple-fault diagnosis of continuous dynamic systems using behavioral modes, in: The 1996 13th National Conference on Artificial Intelligence, 1996, pp. 965–970.
[9] J. Gertler, Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, New York, 1998.
[10] M. Daigle, A qualitative event-based approach to fault diagnosis of hybrid systems, Ph.D. thesis, Vanderbilt University (2008).
[11] P. Mosterman, G. Biswas, Diagnosis of continuous valued systems in transient operating regions, IEEE Transactions on Systems, Man and Cybernetics, Part A 29 (6) (1999) 554–565.
[12] M. Daigle, A. Bregon, G. Biswas, X. Koutsoukos, B. Pulido, Improving multiple fault diagnosability using possible conflicts, in: Proceedings of the 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, 2012, pp. 144–149.
[13] I. Roychoudhury, M. Daigle, A. Bregon, B. Pulido, A structural model decomposition framework for systems health management, in: Proceedings of the 2013 IEEE Aerospace Conference, 2013.
[14] M. J. Daigle, X. Koutsoukos, G. Biswas, A qualitative event-based approach to continuous systems diagnosis, IEEE Transactions on Control Systems Technology 17 (4) (2009) 780–793.
[15] R. Reiter, A theory of diagnosis from first principles, in: M. L. Ginsberg (Ed.), Readings in Nonmonotonic Reasoning, Morgan Kaufmann, Los Altos, California, 1987, pp. 352–371.
[16] M. Daigle, I. Roychoudhury, G. Biswas, X. Koutsoukos, A. Patterson-Hine, S. Poll, A comprehensive diagnosis methodology for complex hybrid systems: A case study on spacecraft power distribution systems, IEEE Transactions of Systems, Man, and Cybernetics, Part A 4 (5) (2010) 917–931.
[17] M. Daigle, A. Bregon, I. Roychoudhury, A structural model decomposition framework for hybrid systems diagnosis, in: 26th International Workshop on Principles of Diagnosis, 2015, pp. 201–208.
[18] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, Diagnosis and Fault-Tolerant Control, Springer, 2006.
[19] M. Cordier, P. Dague, F. Lévy, J. Montmain, M. Staroswiecki, 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. on Systems, Man, and Cybernetics. Part B: Cybernetics 34 (5) (2004) 2163–2177.
[20] B. Pulido, C. Alonso-González, Possible Conflicts: a compilation technique for consistency-based diagnosis, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34 (5) (2004) 2192–2206.
[21] E.-J. Manders, S. Narasimhan, G. Biswas, P.-J. Mosterman, A combined qualitative/quantitative approach for fault isolation in continuous dynamic systems, in: SafeProcess 2000, Vol. 1, Budapest, Hungary, 2000, pp. 1074–1079.
[22] M. J. Daigle, X. D. Koutsoukos, G. Biswas, Distributed diagnosis in formations of mobile robots, IEEE Transactions on Robotics 23 (2) (2007) 353–369.
[23] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, D. Teneketzis, Diagnosability of discrete-event systems, IEEE Transactions on Automatic Control 40 (9) (1995) 1555–1575.
[24] L. Travé-Massuyès, T. Escobet, X. Olive, Diagnosability analysis based on component supported analytical redundancy relations, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 36 (6) (2006) 1146–1160.
[25] S. Narasimhan, P. J. Mosterman, G. Biswas, A systematic analysis of measurement selection algorithms for fault isolation in dynamic systems, in: Proc. of the 9th International Workshop on Principles of Diagnosis, Cape Cod, MA USA, 1998, pp. 94–101.
[26] M. Daigle, I. Roychoudhury, A. Bregon, Qualitative event-based diagnosis applied to a spacecraft electrical power distribution system, Control Engineering Practice 38 (2015) 75–91.
[27] M. Daigle, I. Roychoudhury, A. Bregon, Qualitative event-based fault isolation under uncertain observations, in: Annual Conference of the Prognostics and Health Management Society 2014, 2014, pp. 347–355.
[28] M. Krysander, J. Åslund, E. Frisk, A structural algorithm for finding testable sub-models and multiple fault isolability analysis, in: 21st International Workshop on Principles of Diagnosis, Portland, Oregon, USA, 2010, pp. 79–86.
[29] I. Issury, D. Henry, C. Charbonnel, E. Bornschlegl, X. Olive, A boolean algebraic-based solution for multiple fault diagnosis: Application to a spatial mission, Aerospace Science and Technology 28 (1) (2013) 214 – 226. doi:http://dx.doi.org/10.1016/j.ast.2012.11.002.
[30] M. Bartyś, Diagnosing multiple faults with the dynamic binary matrix, in: 9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, Vol. 48, 2015, pp. 1297 – 1302.
[31] J. Koscielny, M. Bartys, M. Syfert, Method of multiple fault isolation in large scale systems, Control Systems Technology, IEEE Transactions on 20 (5) (2012) 1302–1310.
[32] A. Kodali, S. Singh, K. Pattipati, Dynamic set-covering for real-time multiple fault diagnosis with delayed test outcomes, IEEE Transactions on Systems, Man, and Cybernetics: Systems 43 (3) (2013) 547–562.
[33] G. K. Fourlas, Multiple Faults Diagnosability of Hybrid Systems, in: Proceedings of the 17th Mediterranean Conference on Control and Automation, Thessaloniki, Greece, 2009, pp. 365–370.
[34] N. Chatti, B. Ould-Bouamama, A.-L. Gehin, R. Merzouki, Signed bond graph for multiple faults diagnosis, Engineering Applications of Artificial Intelligence 36 (0) (2014) 134 – 147. doi:http://dx.doi.org/10.1016/j.engappai.2014.07.018.
[35] J. Koscielny, Z. Labeda-Grudziak, Double fault distinguishability in linear systems, International Journal of Applied Mathematics and Computer Science 23 (2) (2013) 1395–406.
[36] J. M. Kościelny, Fault isolation in industrial processes by the dynamic table of states method, Automatica 31 (5) (1995) 747–753.

[37] J. M. Kościelny, K. Zakroczymski, Fault isolation method based on time sequences of symptom appearance, in: Proceedings of IFAC SafeProcess 2000, 2000.
[38] V. Puig, F. Schmid, J. Quevedo, B. Pulido, A new fault diagnosis algorithm that improves the integration of fault detection and isolation, in: Proceedings of the 44th IEEE Conference on Decision and Control, 2005, pp. 3809–3814.
[39] E. Gelso, S. Castillo, J. Armengol, Structural analysis and consistency techniques for robust model-based fault diagnosis, Tech. Rep. 20, Institut d'Informática i Aplicacions, Universitat de Girona (2008).
[40] S. Narasimhan, L. Brownston, HyDE: A general framework for stochastic and hybrid model-based diagnosis, in: Proc. of the 18th Int. WS. on Principles of Diagnosis, 2007, pp. 186–193.