Multiple Fault Diagnosis in Complex Physical Systems

Matthew Daigle, Xenofon Koutsoukos, and Gautam Biswas
Institute for Software Integrated Systems (ISIS)
Department of Electrical Engineering and Computer Science
Vanderbilt University
Nashville, TN 37235
{matthew.j.daigle,xenofon.koutsoukos,gautam.biswas}@vanderbilt.edu

Abstract

Multiple fault diagnosis is a challenging problem because the number of candidates grows exponentially in the number of faults. In addition, multiple faults in dynamic systems may be hard to detect, because they can mask or compensate each other's effects. The multiple fault problem is important, since the single fault assumption can lead to incorrect or failed diagnoses when multiple faults occur. We present an approach to simultaneous and cascaded multiple fault diagnosis in dynamical systems. Our approach is based on the TRANSCEND fault isolation scheme, where fault effects are represented as qualitative fault signatures. A notion of multiple fault diagnosability is introduced with respect to most likely minimal candidates. The online fault isolation algorithm explores the candidate space in increasing candidate size to generate minimal candidates. A mobile robot example demonstrates the approach.

1 Introduction

Fault detection and isolation (FDI) is a key component of any safety-critical system. When faults and degradations occur, it is important to quickly identify the fault that occurred so corrective actions can be taken in a timely manner and catastrophic situations can be avoided. In general, a number of different failures can happen in complex systems, and the likelihood of multiple faults occurring increases in harsh operating environments. FDI schemes that do not take into account multiple faults run the risk of generating incorrect diagnoses or even failing to find a diagnosis after faults occur.

Our approach focuses on multiple fault diagnosis in complex physical systems. It is based on the TRANSCEND framework [Mosterman and Biswas, 1999; Manders *et al.*, 2000], which employs a qualitative approach for analysis of fault transient behavior. The diagnosis model is used to generate fault signatures, which represent magnitude and higher order effects of faults on the measurements.

Multiple fault diagnosis is a difficult problem in dynamical systems because interactions among fault effects can obscure the fault signatures. In this paper, we provide a systematic scheme for generation of multiple fault signatures from the single fault signatures. We analyze the multiple fault signatures to define the notion of n-diagnosability, which defines diagnosability with respect to most likely minimal fault sets, where n is the maximum allowed fault multiplicity. We then present an extension to the online fault isolation algorithm of TRANSCEND such that it finds the most likely minimal fault set that is consistent with the observed measurement deviations. If a system is n-diagnosable for some n, the algorithm will isolate a unique multiple fault candidate, if n or less faults occur.

Previous work in multiple fault diagnosis has concentrated mostly on static systems. The approach in [de Kleer and Williams, 1987] is based on conflict recognition and candidate generation. The system, GDE, utilizes the notion of minimal candidates, and chooses the next best measurements to make based on *a priori* fault probabilities. In our approach, measurements must be selected at design time, and they are used to generate and refine fault hypotheses when deviations from nominal behavior are observed. The GDE approach parallels the consistency-based diagnosis approach of [Reiter, 1987], an extension of which is presented in [Ng, 1990] to handle diagnosis of devices whose behavior changes over time. The changes are modeled by a set of qualitative simulation states. A similar approach that handles behavioral modes is presented in [Subramanian and Mooney, 1996]. In contrast, our approach applies to continuous-time models and can handle both additive and multiplicative faults. A control theory-based approach based on residual structures is described in [Gertler, 1998]. A residual structure is derived to meet the desired isolation properties. Our approach to multiple fault representation is somewhat analogous, although our residuals map to a richer feature set.

The paper is organized as follows. Section 2 describes the TRANSCEND approach to qualitative fault isolation and presents the example model. Section 3 formulates the representation of multiple faults and a notion of multiple fault diagnosability based on the representation. Section 4 extends the fault isolation algorithm of TRANSCEND to account for multiple faults. Section 5 demonstrates our approach to multiple fault diagnosis. Section 6 concludes the paper.

2 Background

TRANSCEND [Mosterman and Biswas, 1999] is a well-developed methodology for diagnosis of abrupt faults in complex physical systems with continuous dynamics. It employs a qualitative model-based approach for fault isolation. System models are constructed using bond graphs [Karnopp *et al.*, 2000]. Faults are modeled as abrupt and persistent changes in parameter values of components in the bond graph model of the system.

Fault isolation in TRANSCEND is based on a qualitative analysis of the transient dynamics caused by abrupt faults. Deviations in measurement values after a fault occurrence constitute a fault signature, where predicted deviations in magnitude and higher order derivative values are mapped to $\{+, 0, -\}$ symbols, which correspond to a deviation above normal, no deviation, and a deviation below normal, respectively.

Fault isolation in TRANSCEND utilizes a Temporal Causal Graph (TCG) representation, which can be derived directly from the bond graph model of the system. The TCG captures the causal and temporal relations between system variables. It specifies the signal flow graph of the system in a form where edges are labeled with single component parameter values or direct or inverse proportionality relations.

Fault signatures are generated using a forward-propagation algorithm on the TCG to predict qualitative effects of faults on measurements. The qualitative effect of a fault, + or −, is propagated to all measurement vertices in the TCG to determine fault signatures for each measurement. We denote the set of all faults as $F = \{f_1, f_2, \ldots, f_\kappa\}$ and the set of all measurements as $M = \{m_1, m_2, \ldots, m_\lambda\}$. For $f \in F$ and $m \in M$, $\sigma_{f,m}$ is the fault signature for measurement m given fault f has occurred. Two faults $f_i, f_j \in F$ are distinguishable using fault signatures if $(\exists m \in M)\ \sigma_{f_i,m} \neq \sigma_{f_j,m}$.

Relative measurement orderings [Daigle *et al.*, 2005] are an extension to the original TRANSCEND algorithm. The extended algorithm uses predicted temporal orders of measurement deviations to discriminate between faults. This is extended for multiple fault diagnosis. Like fault signatures, measurement orderings are derived systematically from the TCG. They are based on common subpaths in the model. A measurement ordering is denoted as $m_1 \prec_f m_2$, meaning that if fault f occurs, measurement m_1 will deviate before measurement m_2. We denote the set of such orderings as Ω_{f_i} for fault $f_i \in F$. Two faults are distinguishable using orderings if their ordering sets are in *temporal conflict*.

Definition 1 (Temporal Conflict). *Ω_{f_i} is in temporal conflict with Ω_{f_j} if $(\exists m_i, m_j \in M) m_i \prec_{f_i} m_j \wedge m_j \prec_{f_j} m_i$.*

Fault isolation starts with a backward propagation of an observed symbolic deviation to identify initial fault candidates. Once candidate hypotheses are identified, a forward propagation algorithm generates the fault signatures and measurement orderings, i.e., the effects of each hypothesized fault on measurements. Then observed deviations are compared to predictions using a progressive monitoring scheme to discriminate between the fault hypotheses.

Throughout the paper we focus on a mobile robot as an example system. Details of the system model and TCG for this system are described in [Daigle *et al.*, 2006] and very briefly here. The bond graph is shown in Figure 1. The robot model consists of inertia, capacitor, and resistor elements modeling

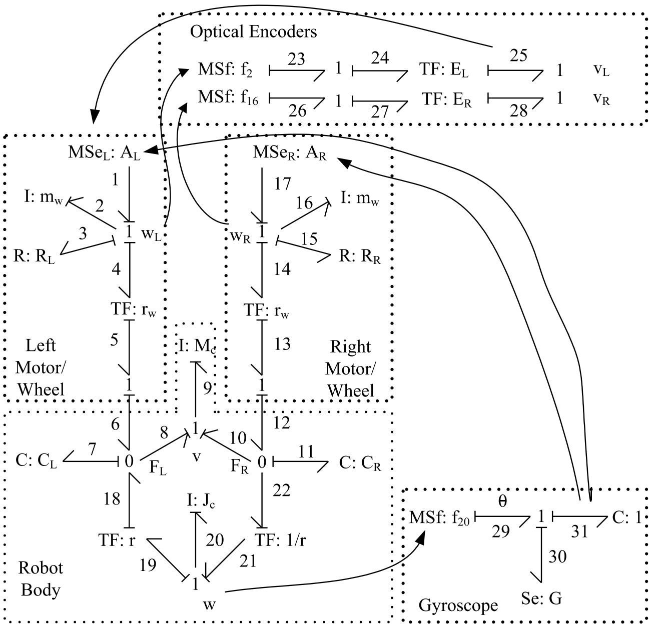


Figure 1: Mobile robot bond graph

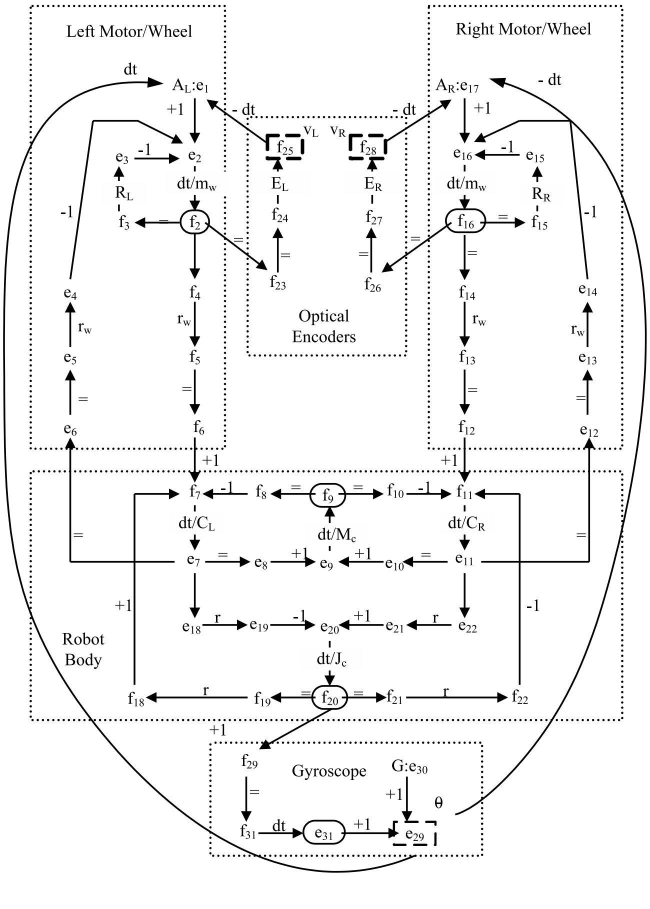


Figure 2: Mobile robot TCG

Fault	v_L	v_R	θ	Measurement Orderings
A_L^-	$0-$	$0\star$	$0+$	$v_L \prec_{A_L^-} v_R, v_L \prec_{A_L^-} \theta$
A_R^-	$0\star$	$0-$	$0-$	$v_R \prec_{A_R^-} v_L, v_R \prec_{A_R^-} \theta$
E_L^-	$-+$	$0\star$	$0-$	$v_L \prec_{E_L^-} v_R, v_L \prec_{E_L^-} \theta$
E_R^-	$0\star$	$-+$	$0+$	$v_R \prec_{E_R^-} v_L, v_R \prec_{E_R^-} \theta$
G^+	$0+$	$0-$	$+-$	$\theta \prec_{G^+} v_L, \theta \prec_{G^+} v_R$
G^-	$0-$	$0+$	$-+$	$\theta \prec_{G^-} v_L, \theta \prec_{G^-} v_R$

Table 1: Fault signatures for a robot system

masses and inertias, mechanical stiffness, and energy dissipation in the system, respectively. The 1-junctions represent the common velocity points, and the 0-junctions common force points. The TCG is given in Figure 2. State variables are circled and measured variables boxed. Edges with a dt specifier imply an integration effect. All other edges are instantaneous.

Table 1 shows fault signatures for actuator (left: A_L^-, right: A_R^-), encoder (left: E_L^-, right: E_R^-), and gyroscope (positive bias: G^+, negative bias: G^-) faults in the mobile robot system. The measurements include velocity of the left wheel, v_L, velocity of the right wheel, v_R, and heading, θ. The first symbol indicates a predicted magnitude change (discontinuity) and the second symbol indicates the first nonzero slope symbol in this measurement. A $\star$ indicates an indeterminate effect. It is indistinguishable from a + or − because it could manifest as either effect. For example, from the TCG we cannot determine whether A_L^- causes a $0+$ or a $0-$ effect on v_R. Relative measurement orderings are also listed in the table.

3 Multiple Fault Diagnosability

Single faults are isolated by comparing predicted to actual measurement deviations. The predictions depend on which measurements are selected in the system, because different measurements provide different discriminatory information. If the prediction models (fault signatures and measurement orderings) of two faults differ, we say that these two faults are distinguishable.

Definition 2 (Single Fault Distinguishability)**.** *Two faults $f_i, f_j \in F$ are distinguishable if $(\exists m \in M)\ \sigma_{f_i,m} \neq \sigma_{f_j,m}$ or $(\exists m_i, m_j \in M) m_i \prec_{f_i} m_j \wedge m_j \prec_{f_j} m_i$.*

Definition 3 (Single Fault Diagnosability)**.** *A system is single fault diagnosable if $(\forall f_i, f_j \in F)$ f_i and f_j are distinguishable.*

For single faults, the isolation procedure compares the observed measurement deviations over time to those predicted by the fault signatures and measurement orderings. If the system is diagnosable, then there exists a unique fault which is consistent with these deviations.

We expand our fault isolation procedure to deal with multiple fault candidates.

Definition 4 (Candidate)**.** *A candidate is a set of faults $c \subseteq F$ that is consistent with the observations. The set of all candidates is denoted as $C = \mathcal{P}(F)$ and of all candidates of size $\leq n$ as $C(n)$.*

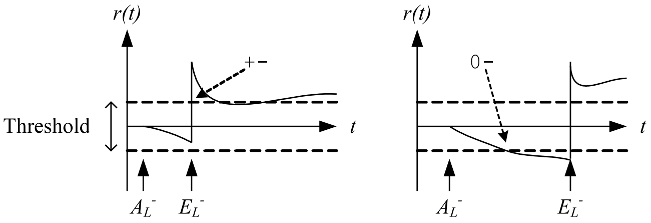


Figure 3: Effect of fault occurrence times on symbol generation of residual $r(t)$

Multiple fault diagnosis algorithms are more complex than single fault diagnosis algorithms for two reasons. First, the effects of a fault could be masked or compensated by the effects of another fault. For example, A_L^- may occur, causing deviations of $0-$ on v_L, $0-$ on v_R, and $0+$ on θ. Clearly, these observations are consistent with only A_L^- occurring. However, if A_R^- also occurred, but with a smaller magnitude so that the effects of A_L^- dominate, the fault sets $\{A_L^-\}$ and $\{A_L^-, A_R^-\}$ cannot be distinguished. So, we seek to define diagnosability with respect to most likely minimal candidates.

The second complication in multiple fault diagnosis is that the same multiple fault can manifest in different ways. For example, A_L^- with E_L^- could either produce a $0-$ effect or a $-+$ effect on v_L, depending on which fault occurs first, and on the fault propagation delays in the system. If E_L^- occurs first, we will see $-+$ because discontinuities are observed at the point of fault occurrence. However, if A_L^- occurs first, we may see either $0-$ or $-+$ depending on how soon E_L^- occurs after A_L^-. Figure 3 illustrates this point. If E_L^- occurs close enough to A_L^-, the deviation caused by A_L^- may not be detected. The symbol generation on the measurement residual could compute either effect. The second change is also not helpful because it could either be caused by a new fault or the dynamics of the original fault.

3.1 Representing Multiple Faults

Taking into account these issues, we represent the effects of multiple faults on a single measurement as the union of predicted single fault effects. For example, the fault set $\{A_L^-, E_L^-\}$ could manifest either $0-$ or $-+$ on v_L, $0-$ or $0+$ on v_R, and $0-$ or $0+$ on θ.

A multiple fault signature for a set of faults $F' \subseteq F$, denoted by $\sigma_{F',m}$, is an element of the set of possible fault signatures for the faults in F', i.e., $\Sigma_{F',m} = \{\sigma_{f,m} | f \in F'\}$. We define a *complete fault signature* as follows.

Definition 5 (Complete Fault Signature)**.** *A complete fault signature for fault $f \in F$, denoted σ_f, is a tuple $(\sigma_{f,m_1}$, σ_{f,m_2}, ..., $\sigma_{f,m_\lambda})$ consisting of the signatures for f on each measurement. A complete multiple fault signature for fault set $F' \subseteq F$ is an element of the set of complete fault signatures $\Sigma_{F'}$, where an element is denoted as $\sigma'_F = (\sigma_{F',m_1}, \sigma_{F',m_2}, \ldots, \sigma_{F',m_\lambda})$, such that $(\forall \sigma_{F'} \in \Sigma_{F'})(\forall \sigma_{F',m_i} \in \sigma_{F'})\ \sigma_{m_i} \in \Sigma_{F',m_i}$.*

Informally, a complete multiple fault signature for F' is a complete signature which can be constructed by choosing and

	v_L	v_R	θ	Realizable?
1	0−	0−	0−	no
2	0−	0−	0+	yes
3	0−	0+	0−	no
4	0−	0+	0+	yes
5	−+	0−	0−	yes
6	−+	0−	0+	no
7	−+	0+	0−	yes
8	−+	0+	0+	no

Table 2: The complete signatures of $\Sigma_{\{A_L^-,E_L^-\}}$ and their physical realizability

combining signatures for single measurements from faults in the fault set F'. As an example, Table 2 shows $\Sigma_{\{A_L^-,E_L^-\}}$.

A complete multiple fault signature can be created by choosing single signatures from 1 to $|F'|$ faults, where $|F'|$ is the size of the fault set F'. As a result, a complete multiple fault signature set will consist of all those complete signatures of the individual faults it contains. Therefore, fault effects due to fault masking and compensation are included. In general, for $F'' \subseteq F'$, we have $\Sigma_{F''} \subseteq \Sigma_{F'}$. This is evidenced in Table 2, e.g., $\{A_L^-, E_L^-\}$ can produce (−+,0+,0−), and according to Table 1, so can E_L^- by itself. The double fault $\{A_L^-, E_L^-\}$ may occur, but the observed deviations may be consistent with A_L^- or E_L^- occurring by themselves.

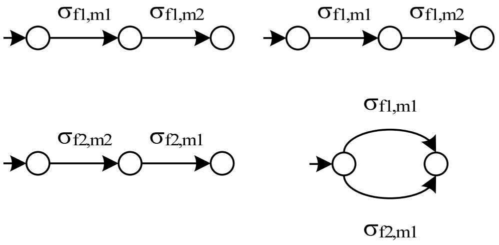
3.2 Physically Realizable Fault Signatures

Not all signatures in $\Sigma_{F'}$ may physically manifest in the system behavior, determined by the fault propagation times inherent in the system. The set $\Sigma_{F'}$ can be constrained by using temporal information in the system model. The resulting set is called the set of physically realizable fault signatures.

Definition 6 (Physical Realizability)**.** *A physically realizable complete fault signature for a fault set F', denoted $\Sigma^R_{F'}$, is the set of multiple fault signatures for F' that is consistent with the TCG model of system behavior.*

Whether some $\sigma_{F'} \in \Sigma_{F'}$ belongs in $\Sigma^R_{F'}$ can be determined using relative measurement orderings. Consider E_L^- and G^+. Both faults produce discontinuities (−+ or +−) on some measurement. Because discontinuities manifest at the point of fault occurrence, it is not possible for both faults to occur and not observe a discontinuity. We must either observe −+ on v_L, +− on θ, or both. Therefore, (0+,0−,0−), for example, should not be in $\Sigma^R_{\{E_L^-,G^+\}}$.

This notion can be formalized with relative measurement orderings. Essentially, single fault orderings should be obeyed with respect to single fault signatures. If some fault f_i produces a deviation on a measurement, m_i, before another measurement, m_j, and another fault f_j produces a deviation on m_j before m_i, then if both faults occur, we cannot observe f_i's effect on m_j together with f_j's effect on m_i as the first effects on m_i and m_j[1]. To see f_i's effect on m_j, we would have had to observe its effect on m_i first. Similarly, to see f_j's effect on m_i, we would have had to observe its effect on m_j first.



(a) Constraint 1 (b) Constraint 2

Figure 4: Realizability constraint representations

For simplicity, we express this constraint in terms of two faults and two measurements. An automata representation is given as Figure 4(a). The top automaton represents the ordering $m_1 \prec_{f_1} m_2$ and the bottom $m_2 \prec_{f_2} m_1$. If f_1 effects m_1 first (event $\sigma_{f1,m1}$) and f_2 effects m_2 first (event $\sigma_{f2,m2}$), then we cannot observe both f_1's effect on m_2 and f_2's effect on m_1 as the first deviations on m_1 and m_2. If these are the only two measurements, then if f_1 and f_2 occur together, we must observe f_1's effect on m_1 or f_2's effect on m_2 as the first deviation on the respective measurements. This property is expressed by the synchronous composition of the two automata, and stated formally as the following lemma.

Lemma 1 (Realizability Constraint 1)**.** *For two faults $f_i, f_j \in F$ and two measurements $m_i, m_j \in M$, if $m_i \prec_{f_i} m_j$ and $m_j \prec_{f_j} m_i$, then $(\forall \sigma_{\{f_i,f_j\}} \in \Sigma_{\{f_i,f_j\}})$, $\sigma_{\{f_i,f_j\}} \notin \Sigma^R_{\{f_i,f_j\}}$ if $\sigma_{\{f_i,f_j\},m_i} = \sigma_{f_j,m_i} \neq \sigma_{f_i,m_i}$ and $\sigma_{\{f_i,f_j\},m_j} = \sigma_{f_i,m_j} \neq \sigma_{f_j,m_j}$.*

A related constraint evolves from this information. Consider again the fault set $\{A_L^-, E_L^-\}$. Orderings predict that both faults manifest in v_L first. Therefore, if v_L deviates as 0−, then A_L^- will propagate to the rest of the measurements before E_L^- does, so we will not see any effects inconsistent with A_L^-, e.g., we will not see 0− on θ. This is because E_L^- cannot propagate from v_L to θ any faster than A_L^- can.

The physical reasoning behind this constraint is that the ordering $m_i \prec_{f_i} m_j$ implies that the fastest way to reach m_j is through m_i given f_i has occurred. So if some other fault reaches m_i first, it will traverse this same path to m_j, and cause m_j to deviate from its effect propagating on this path (or from some faster path f_j to m_j). Therefore when f_i finally reaches m_i, it cannot propagate to m_j any faster than f_j had, so we cannot observe its effect on m_j.

For simplicity, we express this constraint also in terms of two faults and two measurements. An automata representation is given as Figure 4(b). The top automaton represents the ordering $m_1 \prec_{f_1} m_2$ and the bottom represents the constraint that we will only observe the effect on a measurement from one fault. If f_2 effects m_1 first, then we cannot observe f_1's effect on m_2. This property is expressed by the synchronous composition of the two automata, and stated formally as the following lemma.

[1] We are only interested in the first observed measurement deviation since that is what the symbol generator provides.

Lemma 2 (Realizability Constraint 2)**.** *For two faults $f_i, f_j \in F$ and two measurements $m_i, m_j \in M$, if $m_i \prec_{f_i} m_j$, then $(\forall \sigma_{\{f_i,f_j\}} \in \Sigma_{\{f_i,f_j\}})$, $\sigma_{\{f_i,f_j\}} \notin \Sigma^R_{\{f_i,f_j\}}$ if $\sigma_{\{f_i,f_j\},m_i} = \sigma_{f_j,m_i} \neq \sigma_{f_i,m_i}$ and $\sigma_{\{f_i,f_j\},m_j} = \sigma_{f_i,m_j} \neq \sigma_{f_j,m_j}$.*

Table 2 lists the set of physically realizable signatures based on these constraints for $\{A_L^-, E_L^-\}$. Signatures 1, 3, 6, and 8 are not realizable due to the second constraint.

An additional constraint that we impose is to only allow certain combinations of faults, as this will also limit the number of complete multiple fault signatures. It does not make sense to allow fault sets consisting of multiple changes of the same parameter because we assume fault effects are persistent. Therefore, examples such as $\{G^+, G^-\}$ are not valid candidates.

We also employ practical knowledge about systems to limit the size of allowable fault candidate sets. The assumption is that candidates with a large number of faults are highly unlikely, therefore, we assume that the maximum candidate size is $\leq n$. The set of all fault signatures for fault sets of size $\leq n$ is denoted as $\Sigma(n) = \{\sigma_{F'} \in \Sigma_{F'} | F' \subseteq F, |F'| \leq n\}$. The set of all physically realizable fault signatures for fault sets of size $\leq n$ is denoted as $\Sigma^R(n) = \{\sigma_{F'} \in \Sigma^R_{F'} | F' \subseteq F, |F'| \leq n\}$.

The realizability constraints can be extended to multiple faults and measurements. A general way to describe the constraints is by using the automata representation. For a given fault set, we can describe its possible set of event trajectories (and thus physically realizable fault signatures) by taking the synchronous product of all the single fault orderings and the two-state automata that represent a measurement being effected by only one fault. To compute $\Sigma^R(n)$ from this, we need only restrict the trajectories to those including events from at most n faults.

We can also define the measurement orderings that can be created by multiple faults as $\Omega_{\{F_i,F_j\}} = \Omega_{F_i} \cap \Omega_{F_j}$, for $F_i, F_j \subseteq F$. That is, only shared measurement orderings will be consistent with both faults occurring in any order. This can be seen in the automata representation of the orderings.

3.3 n-diagnosability

Based on the set of physically realizable multiple fault signatures and relative measurement orderings for multiple faults, we can define the notion of distinguishability between candidates for multiple faults.

Definition 7 (Multiple Fault Distinguishability)**.** *Two fault sets F_i and F_j are distinguishable if $\Sigma^R_{F_i} \cap \Sigma^R_{F_j} = \emptyset$ or Ω_{F_i} is in temporal conflict with Ω_{F_j}.*

Informally, two fault sets are distinguishable if it is not possible for them to manifest in the system measurements in the same way. We do not, however, define multiple fault diagnosability using this definition. We described previously how, due to fault masking and compensation, a fault set and a superset may manifest in the same way. If so, then for $F' \subseteq F''$, $\Sigma^R_{F'} \subseteq \Sigma^R_{F''}$, and $\Omega_{F'} \subseteq \Omega_{F''}$. We, therefore, consider diagnosability only with respect to minimal candidates.

Definition 8 (Minimal Candidate)**.** *A candidate c is minimal if there does not exist a candidate c' such that $c' \subset c$.*

In addition to using minimal candidates, we also consider the likelihood of fault occurrence. The assumption is that all faults are equally likely, so candidates of smaller size are more likely than those of larger size. Therefore, the ultimate goal of the fault isolation procedure is in isolating the minimal candidate of smallest size. In general, $\{f_1, f_2\}$ and $\{f_3\}$ may both be minimal candidates, because one is not a subset of the other. We consider $\{f_3\}$ to be the simpler explanation because it is of smaller size. Therefore, the fault isolation procedure does not have to consider less likely candidates when more likely candidates exist.

The main reason for operating with most likely candidates is that fault masking and compensation may prevent us from isolating the true set of faults that has occurred. We do not wish to classify a system as undiagnosable because we cannot distinguish between a candidate a superset. Like other work, we assume the principle of parsimony [Reiter, 1987] and consider a diagnosis as the simplest explanation given the observed measurement deviations. The assumption is further supported, in general, by the fact that the probability of failure occurrence decreases significantly as fault size increases. A diagnosis only represents a *best effort* result. A diagnosis of $\{f_1, f_2\}$, for example, means that at least f_1 and f_2 must have occurred, but does not mean that some other fault f_3 has not also happened, rather, it only implies that f_3 could not have occurred by itself.

Definition 9 (Fault Isolation Procedure)**.** *Given a candidate size limit $n > 0$ and the set of measurement orderings, the fault isolation procedure is a function $I : \Sigma^R(n) \rightarrow \mathcal{P}(C(n))$.*

Fault isolation operates in a progressive fashion as new measurements deviate. Because only physically realizable fault signatures for candidates of size $\leq n$ are given as input, this function will always return a nonempty set of candidates. Multiple fault diagnosability is defined in terms of the fault isolation procedure and the given candidate size limit.

Definition 10 (n-diagnosability)**.** *Given a candidate size limit n, a system is n-diagnosable if after all measurements have deviated, $(\forall \sigma_{F'} \in \Sigma^R(n))$ $|I(\sigma_{F'})| = 1$.*

Informally, a system is n-diagnosable if given any physically realizable multiple fault signature for candidates of size $\leq n$, a single minimal candidate of smallest size $\leq n$ is isolated. We next describe our fault isolation procedure based on this notion of multiple fault diagnosability.

4 Diagnosing Multiple Faults

We follow the conflict-based approach of [de Kleer and Williams, 1987], where a conflict is defined as a set of assumptions which cannot all be true, and thus support a symptom (e.g., $\overline{a_1 \wedge a_2 \wedge a_3}$). In TRANSCEND, the TCG is used to create a direct mapping from faults to symptoms, i.e., fault signatures and measurement orderings. Instead of using conflicts, we refer to a *hypothesis set*, which represents all possible faults which can explain a particular symptom.

Definition 11 (Hypothesis Set)**.** *A hypothesis set is a set of faults, at least one of which must have occurred given a particular set of measurement deviations that have occurred.*

A hypothesis set is equivalent to a conflict, in that it represents a set of negated assumptions (an assumption being that a certain parameter is not faulty), at least one of which must be true (e.g., a conflict $\overline{a_1 \wedge a_2 \wedge a_3} \equiv \overline{a_1} \vee \overline{a_2} \vee \overline{a_3} \equiv f_1 \vee f_2 \vee f_3$, a hypothesis set).

Hypothesis sets can be generated directly from the fault signature matrix and measurement orderings. Given a measurement deviation, we construct the hypothesis set to be the set of faults consistent with the deviation. For example, given a 0− for v_L and using only fault signatures produces the hypothesis set $\{A_L^-, A_R^-, E_R^-, G^-\}$. Any of these faults occurring, or combinations of them, support the symptom.

Candidate generation proceeds similar to [de Kleer and Williams, 1987]. As new measurements deviate, new hypothesis sets are generated. These hypothesis sets restrict the possible candidate space and result in a new set of minimal candidates. Given a new hypothesis set, new candidates are formed by adding a single fault from the new hypothesis set. Since a hypothesis set is a set of faults consistent with an observation, these new candidates will also be consistent with the new observation as well as all old observations covered by the base candidate.

Because n-diagnosability only requires isolating a unique candidate of the smallest size, we introduce a candidate size limit into our procedure. As long as we have a candidate at our current size level, we do not explore candidates of larger size. Further, we only perform this analysis if we eliminate all candidates at the current level.

To illustrate the general approach, consider the fault set $\{A_L^-, A_R^-, E_L^-, E_R^-\}$. The candidate space, which can be represented as a lattice of C, is shown in Figure 5. The candidate size limit is given as $n = 2$, and the starting size level is $n = 1$. Given the first measurement deviation −+ for v_L generates the hypothesis set $\{E_L^-\}$, because only that fault can produce that deviation on v_L given v_R and θ have not yet deviated. We now know that this fault must have occurred. At a later time point, we are given the deviation 0− for v_R. This generates the hypothesis set $\{A_R^-, E_L^-\}$, because only these faults can cause v_R to deviate that way given θ has not yet deviated. A_L^- is not included in this hypothesis set because it did not cause v_L to deviate, so we can't see its effect on v_R (this relates to the second realizability constraint). At this point, we still have a candidate of size 1, so we do not yet consider any of size 2. If we were to consider the complete fault set, then a deviation of +− for θ would rule out the possibility that E_L^- by itself occurred, and we now consider candidates of size 2. If the system is 2-diagnosable, a unique candidate of size 2 will be identified.

The pseudocode for the online diagnosis algorithm is shown as Algorithm 1. It works as follows. As new measurements deviate, hypothesis sets are formed and the candidate set refined by eliminating inconsistent candidates. This follows the TRANSCEND approach. Eliminated candidates are saved for later analysis. If a single unique candidate is found during this procedure, the candidate is returned as the most likely minimal candidate, barring any future measurement deviations.

When faults at the candidate size level l are all eliminated, the discarded minimal candidates are used to produce new minimal candidates of size $l + 1$ using the hypothesis sets gathered. This procedure is given as Function 2. For each eliminated candidate, new candidates of size $l + 1$ are formed using the hypothesis set which caused it to be eliminated. Since the hypothesis set caused the elimination, the hypothesis set and the eliminated candidate have no common fault, so a candidate of size l cannot be constructed. Since new candidates are formed by adding exactly one fault from the hypothesis set, only candidates of size $l + 1$ are formed.

Each new candidate formed is then checked for consistency with hypothesis sets that were recorded after its base candidate was eliminated. If the new candidate is consistent with all of these, it is added to the current candidate list. If not, it is added to the eliminated candidates list, because applying a new hypothesis set would form a candidate of size $l + 2$, which we are not considering at that time. If no new candidates are found then the level is increased and the process repeated. If the size limit is reached, then an unmodeled fault or a fault combination of size $> n$ has occurred.

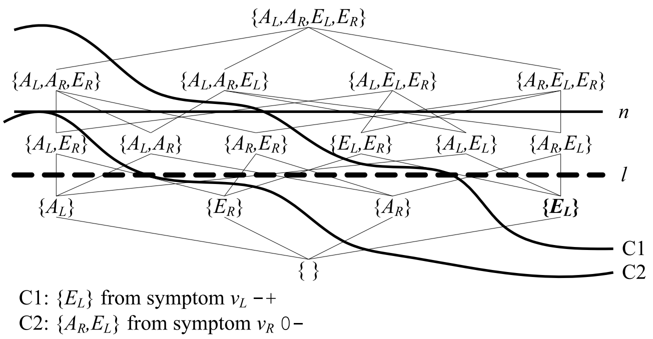


Figure 5: Candidate lattice for fault set $\{A_L^-, A_R^-, E_L^-, E_R^-\}$

Algorithm 1 Fault Isolation

```
Input: maximum candidate size $n$
Variables: current candidates list, hypothesis sets list,
eliminated candidates list
When a new measurement deviates:
Form the conflict and record it
Eliminate inconsistent candidates
if no candidates are left then
    Expand eliminated candidates to the next size
end if
if one candidate is left then
    Return the candidate
end if
```

Theorem 1. *Algorithm 1 will return a unique most likely minimal candidate if the system is n-diagnosable and a fault combination of size $l \leq n$ occurs.*

Proof. The algorithm never eliminates consistent candidates. The algorithm also only considers larger candidates when no smaller candidate can explain the observations. Therefore, the algorithm will find the smallest set of candidates at any level. If the system is n-diagnosable, then a unique candidate will exist of size $\leq n$. If so, at the lowest possible level the

Function 2 Expand Candidates

```
Input: maximum candidate size $n$
if candidate size limit is exceeded then
  Return failure
end if
for all eliminated candidates of the previous size do
  Construct new candidates using the conflict that caused its elimination
end for
Eliminate candidates inconsistent with the recorded conflicts
if no candidates are left then
  Expand eliminated candidates to the next size
else
  Return candidates
end if
```

algorithm will find a unique candidate. □

If n is fixed, the computational complexity of the algorithm is polynomial in the number of single faults, because $O(|F|^n)$ multiple faults are considered. If n is left unspecified, we are limited to a fault multiplicity of $|F|$. In this case the algorithm is exponential in the number of single faults.

In the single fault algorithm, as soon as a single fault is isolated, it is declared as the true fault, and future measurements deviating can be ignored. In the case of multiple faults, a single isolated fault does not necessarily indicate the true fault. It only indicates the current simplest diagnosis, given the deviations observed thus far. So, future measurement deviations may result in a better understanding of what faults actually occurred in the system. If there is a unique candidate at any point, the algorithm will return it. Because more measurement deviations can only expand this candidate, the current unique candidate is partially correct. Future deviations may or may not provide a more exact diagnosis.

5 Mobile Robot Example

In this section, we go through a detailed example execution of Algorithm 1. First, however, we must analyze the diagnosability of the system to ensure we will get unique results. We let $n = 2$ for our analysis.

Table 3 lists some of the physically realizable fault signatures for the robot system. There are several points to make here. First, the signature (0+,0−,0+) is absent. This is because it violates the realizability constraints. There are several double faults which contain this signature in their signature set. However, this signature is not physically realizable for any of them. Take for example, $\{A_L^-, A_R^-\}$. Only A_R^- can produce 0+ on v_L. Because A_L^- causes v_L to deviate first, this means that A_R^- will affect θ first, however only A_L^- can produce 0+ on θ. Thus, this signature violates the second realizability constraint for this double fault.

We also see from Table 3 that the system is not 2-diagnosable. If θ deviates first, observing either (0−,0−,+−) or (0−,0−,−+) cannot be explained by a single fault, but two double faults are consistent with each. For example, consider observing (0−,0−,+−) with θ deviating first. If then both wheels start slowing down, this cannot be explained by G^+ by itself. However, given that both velocities are below nominal, we cannot determine which actuator fault caused it, because only θ allows us to discriminate between them in this case. Orderings do not help either, because even if we see v_L or v_R deviate next, we do not know if that deviation was due to G^+ propagating or an actuator fault appearing. Although we cannot distinguish which actuator fault occurred with G^+, we still know that G^+ must have occurred, and that some actuator fault has also occurred. This can sometimes be helpful.

$\Sigma^R(2)$	Smallest minimal candidates
(0−, 0−, 0−)	$\{A_R^-\}$ (v_R first) or $\{A_L^-, A_R^-\}$ (v_L first)
(0−, 0−, 0+)	$\{A_L^-\}$ (v_L first) or $\{A_L^-, A_R^-\}$ (v_R first)
(0−, 0−, +−)	$\{A_L^-, G^+\}, \{A_R^-, G^+\}$ (θ first) or $\{A_L^-, G^+\}$ (v_L first) or $\{A_R^-, G^+\}$ (v_R first)
(0−, 0−, −+)	$\{A_L^-, G^-\}, \{A_R^-, G^-\}$ (θ first) or $\{A_L^-, G^-\}$ (v_L first) or $\{A_R^-, G^-\}$ (v_R first)
(0+, 0−, 0−)	$\{A_R^-\}$ (v_R first)
(0+, 0−, +−)	$\{G^+\}$ (θ first) or $\{A_R^-, G^+\}$ (v_R first)
⋮	⋮
(−+, −+, 0−)	$\{E_L^-, E_R^-\}$ (v_L or v_R first)
(−+, −+, 0+)	$\{E_L^-, E_R^-\}$ (v_L or v_R first)

Table 3: 2-Diagnosability analysis for the mobile robot

We now consider a double fault which is distinguishable, and demonstrate the execution of the algorithm. Table 4 illustrates the approach for $\{E_L^-, G^+\}$ occurring. First, v_L deviates with a −+. Only an encoder fault of the left wheel can produce such a deviation on v_L given that no other measurements have deviated, thus the hypothesis set is $\{E_L^-\}$ which becomes our first candidate. Next, v_R deviates with a 0−. Given that θ has not yet deviated, the hypothesis set becomes $\{A_R^-, E_L^-\}$. G^+ is not included in this hypothesis set because we would have seen θ deviate if it had occurred (constraint 1), and neither is A_L^-, because to observe its effect on v_R would mean we would have seen its effect on v_L (constraint 2). Since $\{E_L^-\}$ is consistent with this hypothesis set, it remains a candidate. Next, θ deviates with a +−. The hypothesis set is $\{G^+\}$ since only G^+ can cause θ to deviate in that way. Since $\{E_L^-\}$ is not consistent with this hypothesis set, it is eliminated. We now have to expand our eliminated candidates to explain the observations. Since the hypothesis set $\{G^+\}$ eliminated $\{E_L^-\}$, we form the new candidate $\{E_L^-, G^+\}$. Since all measurements have deviated, we can be sure that this is our smallest minimal candidate. Since $\{E_L^-, G^+\}$ is distinguishable from all other double faults, the algorithm gives a unique result.

We next consider a case where, although the signature is realizable for a single fault, can only be explained by a double fault. The signature (0−, 0−, 0−) is realizable for A_R^-,

Observation	Hypothesis set	Candidates	Eliminated
1. v_L −+	$\{E_L^-\}$	$\{E_L^-\}$	$\emptyset$
2. v_R 0−	$\{A_R^-, E_L^-\}$	$\{E_L^-\}$	$\emptyset$
3. θ +−	$\{G^+\}$	$\emptyset$	$\{E_L^-\}$
	Apply (3)	$\{E_L^-, G^+\}$	$\emptyset$

Table 4: Algorithm execution example 1

Observation	Hypothesis set	Candidates	Eliminated
1. v_L 0−	$\{A_L^-\}$	$\{A_L^-\}$	$\emptyset$
2. v_R 0−	$\{A_L^-, A_R^-\}$	$\{A_L^-\}$	$\emptyset$
3. θ 0−	$\{A_R^-\}$	$\emptyset$	$\{A_L^-\}$
	Apply (3)	$\{A_L^-, A_R^-\}$	$\emptyset$

Table 5: Algorithm execution example 2

however if v_R does not deviate first it cannot be only A_R^- which has occurred. However, this signature is realizable for $\{A_L^-, A_R^-\}$, and we show how the algorithm derives this result.

Table 5 summarizes the algorithm execution for this case. First, we see v_L deviate with 0−. Only A_L^- is consistent with v_L deviating first with this effect, thus the hypothesis set is $\{A_L^-\}$. Next, we observe v_R deviate with 0−. Given θ has not yet deviated, $\{A_L^-, A_R^-\}$ is the hypothesis set for the new observation. E_L^- is not included because to observe its effect on v_R would mean we would have seen its effect on v_L (constraint 2). Next, we see θ deviate with 0−. Only A_R^- can cause this (and not E_L^- for the previous reason). Therefore $\{A_L^-\}$ is eliminated, and we expand the candidate into $\{A_L^-, A_R^-\}$. Again, we have a unique result.

6 Conclusions

Multiple fault diagnosis in dynamical systems is complex due to fault masking, compensation, and the many ways multiple faults can manifest. We have presented here an approach to qualitative isolation of multiple faults as an extension of the TRANSCEND approach. We described a notion of multiple fault diagnosability defined over smallest minimal candidates, and presented an algorithm to isolate multiple faults based on this notion. We then discussed the 2-diagnosability analysis of a mobile robot system, and illustrated the algorithm on distinguishable double faults.

Future work will address the scalability of the approach to larger systems and exploring conditions which satisfy n-diagnosability for a specific n. The notion of dealing with only the smallest l value and moving to the next l value may also be relaxed by taking into account a priori fault probabilities for the different component parameters, for which more efficient candidate generation strategies will be explored, such as conflict-directed A* [Williams and Ragno, to appear]. Exploring fault identification and fault-adaptive control in the presence of multiple faults is also an open area of research.

Acknowledgment

This work was supported in part by NSF CNS-0452067 and NSF CNS-0347440.

References

[Daigle *et al.*, 2005] M. Daigle, X. Koutsoukos, and G. Biswas. Relative measurement orderings in diagnosis of distributed physical systems. In *43rd Annual Allerton Conference on Communication, Control, and Computing*, pages 1707–1716, September 2005.

[Daigle *et al.*, 2006] M. Daigle, X. Koutsoukos, and G. Biswas. Distributed diagnosis of coupled mobile robots. In *Proceedings 2006 IEEE International Conference on Robotics and Automation*, pages 3787–3794, May 2006.

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

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

[Karnopp *et al.*, 2000] D. C. Karnopp, D. L. Margolis, and R. C. Rosenberg. *Systems Dynamics: Modeling and Simulation of Mechatronic Systems*. John Wiley & Sons, Inc., New York, 3rd edition, 2000.

[Manders *et al.*, 2000] E.-J. Manders, S. Narasimhan, G. Biswas, and P.J. Mosterman. A combined qualitative/quantitative approach for fault isolation in continuous dynamic systems. In *SafeProcess 2000*, volume 1, pages 1074–1079, Budapest, Hungary, June 2000.

[Mosterman and Biswas, 1999] P.J. Mosterman and G. Biswas. Diagnosis of continuous valued systems in transient operating regions. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 29(6):554–565, 1999.

[Ng, 1990] H. T. Ng. Model-based, multiple fault diagnosis of time-varying, continuous physical devices. In *Sixth Conference on Artificial Intelligence Applications*, volume 1, pages 9–15, May 1990.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. In Matthew L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, pages 352–371. Morgan Kaufmann, Los Altos, California, 1987.

[Subramanian and Mooney, 1996] S. Subramanian and R. J. Mooney. Qualitative multiple-fault diagnosis of continuous dynamic systems using behavioral modes. In *The 1996 13th National Conference on Artificial Intelligence*, pages 965–970, August 1996.

[Williams and Ragno, to appear] B. C. Williams and R. Ragno. Conflict-directed A* and its role in model-based embedded systems. *Special Issue on Theory and Applications of Satisfiability Testing, Journal of Discrete Applied Math*, to appear.