*Transactions of the Institute of Measurement and Control* 32, 5 (2010) pp. 487–510

# An event-based approach to integrated parametric and discrete fault diagnosis in hybrid systems

Matthew J. Daigle[1], Xenofon D. Koutsoukos[2] and Gautam Biswas[2]

[1]University of California, Santa Cruz, NASA Ames Research Center, Moffett Field, CA, USA
[2]Institute for Software Integrated Systems, Dept. of EECS, Vanderbilt University, Nashville, TN, USA

Fault diagnosis is crucial for ensuring the safe operation of complex engineering systems. These systems often exhibit hybrid behaviours, therefore, model-based diagnosis methods have to be based on hybrid system models. Most previous work in hybrid systems diagnosis has focused either on parametric or discrete faults. In this paper, we develop an integrated approach for hybrid diagnosis of parametric and discrete faults by incorporating the effects of both types of faults into our event-based qualitative fault signature framework. The framework allows for systematic design of event-based diagnosers that facilitate diagnosability analysis. Experimental results from a case study performed on an electrical power distribution system demonstrate the effectiveness of the approach.

**Key words:** discrete-event systems; electrical power distribution systems; hybrid bond graphs; hybrid systems; model-based diagnosis.

## Nomenclature

| | |
|---|---|
| $f$ | A fault |
| $F$ | The set of faults |
| $m$ | A measurement |
| $M$ | The set of measurements |

**Address for correspondence:** Matthew J. Daigle, NASA Ames Research Center, M/S 269-1, Moffett Field, CA 94025, USA. E-mail: matthew.j.daigle@nasa.gov

 10.1177/0142331208097840

| | |
|---|---|
| $q$ | A mode |
| $Q$ | The set of modes |
| $c$ | A candidate |
| $C$ | The set of all candidates |
| $d$ | A diagnosis |
| $\sigma$ | An event |
| $\Sigma$ | A set of events |
| $\lambda$ | A trace of events |
| $\lambda_{f,M,q}$ | A fault trace for fault $f$, measurements $M$, and mode $q$ |
| $L_{f,M,q}$ | The fault language for fault $f$, measurements $M$, and mode $q$ |
| $\mathcal{L}_{f,M,q}$ | The fault model for fault $f$, measurements $M$, and mode $q$ |
| $\lambda_{c,q}$ | A candidate trace for candidate $c$ starting in mode $q$ |
| $L_{c,M,q}$ | The candidate language for candidate $c$, measurements $M$, starting in mode $q$ |
| $\mathcal{D}_{F,M,Q}$ | A diagnoser for faults $F$, measurements $M$, and modes $Q$ |

## 1. Introduction

Fault diagnosis is crucial for ensuring the safe and reliable operation of complex engineering systems. Quick identification of faults and degradations leads to corrective actions and timely maintenance that avoids catastrophic situations. Most real-world, embedded systems exhibit hybrid, ie, mixed continuous and discrete behaviours, therefore, hybrid models have to be employed for designing correct tracking and diagnosis procedures. A comprehensive hybrid system fault diagnosis scheme must address parametric faults in components, such as changes in resistance and inductance values, and discrete faults that alter system configuration, such as when relays become stuck.

Most previous approaches do not develop a combined, unified approach to hybrid diagnosis. The application-specific approach of (Zhao *et al.*, 2005) is suitable only for simple hybrid automata. A parity relations approach is given in (Cocquempot *et al.*, 2004), but this scheme does not easily extend to non-linear systems with multiplicative faults. Other methods have addressed either discrete faults (Travé-Massúyes *et al.*, 2002; Dearden and Clancy, 2002; Hofbaur and Williams, 2002; Koutsoukos *et al.*, 2003; Wang *et al.*, 2007) or parametric faults (McIlraith *et al.*, 2000; Narasimhan and Biswas, 2007). In discrete approaches, fault modes are added to the nominal system model for each discrete fault. Modeling faults as separate modes typically reduces the discrete diagnosis task to a mode estimation problem that is solved by using particle filters or a combination of continuous state and mode observers (Dearden and Clancy, 2002; Hofbaur and Williams, 2002; Koutsoukos *et al.*, 2003; Wang *et al.*, 2007). Parametric approaches have used qualitative techniques to produce parametric fault candidates and have assumed only controlled mode changes (McIlraith et al., 2000) or allowed both controlled and autonomous mode changes (Narasimhan and Biswas, 2007).

In contrast, we present an integrated model-based approach diagnosing both parametric and discrete faults in hybrid systems. We establish a compact, integrated hybrid modelling framework, using hybrid bond graphs (Mosterman and Biswas, 1998), that can represent both parametric and discrete faults. Our diagnosis approach extends the Hybrid TRANSCEND (Mosterman and Biswas, 1999; Narasimhan and Biswas, 2007) methodology, which diagnoses parametric faults based on efficient qualitative analysis of fault transients. By including discrete faults, we develop a unified hybrid diagnosis methodology. We also extend our event-based diagnosis framework for continuous systems (Daigle, *et al.*, 2007b) to hybrid systems, and this allows us to establish and verify notions of diagnosability for hybrid systems within our framework. We demonstrate the effectiveness of the approach using a real hybrid system, the Advanced Diagnostics and Prognostics Testbed (ADAPT) (Poll *et al.*, 2007), deployed at NASA Ames Research Center, which is functionally representative of a spacecraft's electrical power system. We provide diagnosability analysis for a subset of ADAPT, and demonstrate our techniques with experiments performed on the actual testbed.

The paper is organized as follows. Section 2 outlines our diagnosis approach and architecture. Section 3 presents our modelling framework. Section 4 describes our integrated fault isolation algorithms. Section 5 establishes notions of diagnosability for hybrid systems and develops the event-based diagnoser. Section 6 describes ADAPT as a case study and presents experiments performed on the actual testbed. Section 7 presents conclusions and future work.

## 2. Hybrid diagnosis approach

Given a hybrid system, we first define a set of parametric and discrete faults, $F = \{f_1, f_2, \ldots, f_n\}$. Faults are formally defined in Section 3. Second, we define a set of measurements $M = \{m_1, m_2, \ldots, m_p\}$, which are time-varying signals obtained from the available sensors. Last, we denote the set of modes as $Q = \{q_1, q_2, \ldots, q_r\}$. In general, mode changes can be controlled, which are known, or autonomous, which depend on internal system variables.

The fault isolation procedure is initiated at the time of fault detection, $t_d$, where $t_d \geq t_f$, the actual time of fault occurrence. Because mode changes may occur after $t_f$, these need to be accounted for because faults may affect the measurements differently in different modes of operation (Narasimhan and Biswas, 2007). We adopt the following definition of a *candidate*.

*Definition 1* (Candidate). A *candidate* $c$ is defined as $c = (f, q)$, where $f \in F$ is a hypothesized fault, and $q \in Q$ is a hypothesized current mode. The set of all candidates is denoted as $C$.

When faults occur, they produce deviations in the measurements from their nominal values. Our diagnosis model expresses these deviations as ordered event sequences, which are matched against observed deviations to determine which

candidates are consistent with the observations. For a point in time after fault occurrence, a *diagnosis* is a collection of candidates that are consistent with the observations up to that point.

*Definition 2* (Diagnosis). At time $t \geq t_f$, a diagnosis $d \subseteq C$ is a set of candidates consistent with the observations made on the system during the interval $[t_f, t]$.

The goal is to find the diagnosis that includes all possible candidates consistent with the observed measurement deviations and observed mode changes. We adopt the architecture shown in Figure 1. A hybrid observer, implemented as a switched extended Kalman filter, computes the expected behaviour of the plant and hypothesizes autonomous mode changes, based on inputs $\mathbf{u}(t)$ and controlled mode change commands $\sigma_q$. The difference between observed outputs, $\mathbf{y}(t)$, and expected outputs, $\hat{\mathbf{y}}(t)$, defines the residual, $\mathbf{r}(t)$. Both parametric and discrete faults will cause statistically significant differences between the observed and expected values of measurements, because they represent a difference in the behaviours of the actual system and its model. The fault detector employs a statistical test and a sliding window that tracks the mean and variance of measured signals to robustly determine if the residual is non-zero (Biswas *et al.*, 2003). The fault isolation method, described in Section 4, uses a symbol generation scheme to generate symbolic values of the magnitude and slope of the deviation, and the event-based diagnoser uses the sequence of deviations and controlled mode changes to isolate faults. In this paper, we allow both controlled and autonomous mode changes to occur prior to fault isolation, but during fault isolation, assume that only controlled mode changes occur. Including autonomous mode changes makes the framework more complex, and is not necessary to develop and demonstrate the main contributions of this paper, or for the faults studied in ADAPT. The handling of autonomous mode changes that occur during fault isolation can be found in (Narasimhan and Biswas, 2007; Daigle, 2008).

## 3. Modeling hybrid systems

We develop component-based models of hybrid physical systems using hybrid bond graphs (HBGs) (Mosterman and Biswas, 1998). HBGs extend bond graphs

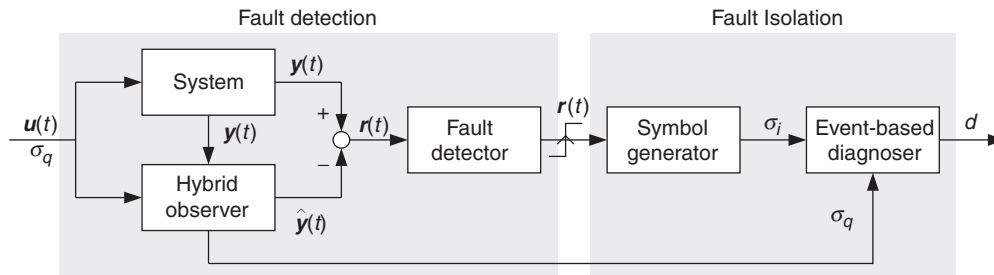


**Figure 1** Event-based diagnosis architecture

Karnopp et al. 2000), which define an energy-based, multi-domain, topological modelling scheme for dynamic systems. They are particularly suitable for diagnosis of dynamic systems representing physical processes because they incorporate causal and temporal information required for deriving and analysing fault transients. Further, they compactly represent the state equations of the system by capturing the topological relations between components, processes and system behaviour.

## 3.1 Hybrid bond graphs

In bond graphs, vertices represent components, and bonds, drawn as half arrows, represent ideal energy connections between them. Associated with each bond are two variables: *effort* and *flow*, denoted by $e_i$ and $f_i$, respectively, where $i$ is the bond number, and the product $e_i \cdot f_i$ defines the rate of energy transfer through the bond. In the electrical domain, these variables map to voltage and current, respectively. 1-junctions represent series connections (where all $f$ are equal and $\sum e = 0$), and 0-junctions represent parallel connections (where all $e$ are equal and $\sum f = 0$). Other bond graph elements model energy dissipation as resistances ($R$, where $e = Rf$), energy storage as capacitances ($C$, where $\dot{e} = \frac{1}{C}f$) and inductances ($I$, where $\dot{f} = \frac{1}{I}e$), and energy sources as sources of flow ($Sf$, where $f = u$ for input $u$) and effort ($Se$, where $e = u$ for input $u$). The constituent equations of the bond-graph elements form a set of differential algebraic equations that describe the continuous system behaviour.

Hybrid bond graphs extend bond graphs by introducing *switching junctions* (Mosterman and Biswas, 1998). Switching junctions enable a junction to be in either the on or off mode. Off 1-junctions behave as sources of zero flow, so they impose $f = 0$ on all their bonds. Similarly, off 0-junctions act as sources of zero effort. In either case, the off junction inhibits the flow of energy in the incident bonds. When on, switching junctions behave as normal junctions. As an example, consider the circuit given in Figure 2. The switch is effectively a series connection that turns on or off, so it is represented by a switching 1-junction, denoted by the dashed arrow in Figure 3. In addition to the difference in energy flow paths, the difference in behaviour between the two modes is made explicit through *causality*, ie, the computational dependencies between the variables of the system, denoted by the causal stroke on one end of the bond. In Figure 3(a), the junction is on, so one bond imposes effort and the other bond imposes flow.

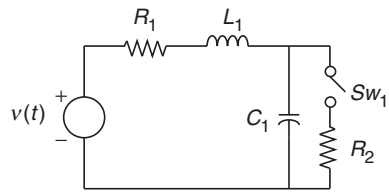


**Figure 2** Switched circuit schematic

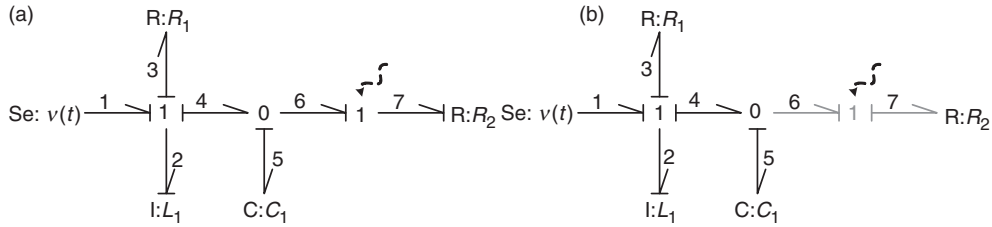


**Figure 3** Switched circuit HBGs: (a) HBG with the switch on, (b) HBG with the switch off

In Figure 3(b), the junction is off, so both bonds impose zero flow. Causality can be derived automatically from the bond-graph model (Karnopp *et al.*, 2000).

The switching behaviour of a controlled junction is defined by a *control specification* (CSPEC), modelled as a finite automaton (Mosterman and Biswas, 1998; Narasimhan and Biswas, 2007). A CSPEC defines a finite number of states. The state transitions are attributed to controlled and autonomous events. The output of the CSPEC determines whether the junction is on or off. So, the system mode is defined implicitly by the individual states of all the CSPECs, providing a concise representation of the hybrid system model. A single mode change may correspond to multiple junctions switching mode. Therefore, events may be shared over different CSPECs. Given an event $\sigma$ and the current system mode $q$, the new system mode $q'$ is given by $q' = \mu(\sigma, q)$, where the system mode transition function, $\mu$, simply applies $e$ to all CSPECs, and obtains the new CSPEC state for each controlled junction. Associated with each mode is a continuous bond graph for which the computational model can be derived automatically (Karnopp *et al.*, 2000). The number of distinct modes of the system is exponential in the number of switching junctions, but an important advantage of the HBG framework is that the continuous model in a new mode can be easily derived, and all modes of the system need not be pre-enumerated.

## 3.2 Modelling faults

We focus on the diagnosis of single, abrupt, persistent faults in hybrid systems. We model faults as unobservable events and we classify them as *parametric* and *discrete* faults.

Parametric faults manifest as unexpected changes in system parameter values in the model, and represent partial failures or degradations in system components that correspond to HBG model parameters. In this work, we define parametric faults as an abrupt step change in a component parameter value.

Discrete faults manifest as discrepancies between the actual and expected mode of a switching component. In HBGs, mode changes are modelled using switching junctions, and discrete faults are captured as unexpected changes in CSPEC state. We model discrete faults as unobservable events in the CSPEC that result in these unexpected

mode changes. Since mode changes may correspond to many junctions changing mode, fault events may be shared among the different CSPECs of the component. The linking of discrete faults to fault events gives, as with parametric faults, a one-to-one mapping between model entities and faults. This leads to the following definition of the CSPEC to include discrete faults.

*Definition 3* (Control Specification). A *control specification* is a tuple $\mathcal{M} = (S, E, t, o, s_0)$, where $S$ is the finite set of states, $E = E_o \cup E_u$ is the set of observable and unobservable (fault) events, $t : S \times E \to S$ is the transition function, $o : S \to \{on, off\}$ is the output function, and $s_0 \in S$ is the initial state.

### 3.3 Temporal causal graphs

We capture the qualitative effects of faults on the measurements using the temporal causal graph (TCG), derived automatically from the bond graph in a given system mode (Mosterman and Biswas, 1999). The TCG captures the causal and temporal relations between system variables and parameters, so it can be used to predict the qualitative effects of faults on the measurements. The vertices of the TCG are the system variables. The labelled edges represent the qualitative relationships between the variables, ie, equality (=), direct (+1) or inverse (−1) proportionality, integration ($dt$), and parametric relations (eg, $1/R_1$). The directionality of these edges is determined by causality.

We augment the standard TCG defined in (Mosterman and Biswas, 1999; Narasimhan and Biswas, 2007) to capture the effect of discrete faults on the system variables by creating a new vertex in the TCG for each discrete fault event. We create new edge types linked to the appropriate junction variables (ie, efforts and flows), and introduce a vertex for the junction mode, $p_i$, for each switching junction $i$, which the discrete fault will also affect. Discrete fault events are connected to the junction variables that they affect, if they are plausible in the current expected mode. The new edge labels are $Z$, implying that the fault causes the variable value to go to zero instantaneously (for a junction turning off), and $N$, implying that the fault causes the variable value to go from zero to non-zero instantaneously (for a junction turning on).

The TCGs for the circuit example are given in Figure 4. In the circuit, we define the measurements as $M = \{i_1, v_2, i_3\}$, ie, the current through $L_1$, the voltage across $C_2$, and the current through $R_2$, respectively. Consider the mode where the relay is closed. To correctly associate the fault event with the junction variables, we need to determine which variables are immediately affected by the change in the junction mode. The closed switch creates a configuration where a voltage is imposed on $R_2$, determining the current flow. An open switch, however, imposes zero current on $R_2$, and, therefore, determines its voltage. When a junction turns off as the result of a fault, the determining flow, $f_7$, will be immediately affected because it goes to zero. Since $e_7$ is related to $f_7$ by an algebraic gain, it too will be immediately affected. It is necessary to directly relate the fault to the voltage $e_7$, because otherwise the effect of the discrete

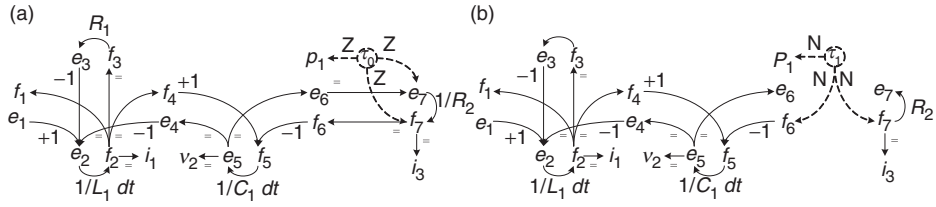


**Figure 4** Switched circuit TCGs (a) TGG for the on mode, (b) TGG for the off mode

fault on that variable may not be correctly predicted. The TCG for this mode is shown in Figure 4(a). The $\tau_0$ fault will immediately affect $e_7$, $f_7$ and $p_1$. Details for more general cases are provided in (Daigle, 2008).

When a switch is expected to be off, a non-zero current can only be explained by the switch state being on. In the HBG, if the junction is off, then its flows (for a 1-junction) or efforts (for a 0-junction) will no longer be zero, so these variables will be affected at the point of fault occurrence. The TCG in this mode is shown in Figure 4(b). The $\tau_1$ fault will immediately affect $f_6$, $f_7$ and $p_1$.

## 4. Qualitative fault isolation

The transients caused by abrupt fault occurrences are represented as symbolic magnitude and slope values, called *fault signatures*, that form the basis for qualitative fault isolation. Assuming that the system output is continuous and continuously differentiable except at the points of fault occurrence and mode changes, the transient response after fault occurrence can be approximated by a Taylor series expansion, which defines the change in terms of magnitude and higher order derivatives (Mosterman and Biswas, 1999). We abstract these changes using the qualitative values $+$, $-$, and 0, which imply an increase, decrease or no change from the nominal behaviour, respectively. We represent signatures using two symbols, a predicted immediate change in magnitude (implying a discontinuity), and the predicted change in slope. If a discontinuity is produced, the magnitude symbol is $+$ or $-$, otherwise it will be 0. Details may be found in (Mosterman and Biswas, 1999).

We augment fault signatures to include information directly indicative of discrete faults. Discrete faults cause mode changes at junctions, and, as a result, variable values linked to a junction may go from non-zero to zero abruptly (for a junction turning off) or go from zero to non-zero abruptly (for a junction turning on). Measuring variables affected in this manner provides additional discriminatory information, because finite changes in parameter values for parametric faults cannot cause this behaviour. Therefore, we include additional symbols N, Z and X, implying zero to non-zero, non-zero to zero, or no discrete change behaviour in the measurement from the estimate. Given a measurement $m$, deviation $d$, and discrete change $c$, we write a signature as an event using $m^{d,c}$, eg, $m_1^{+-,X}$.

*Definition 4* (Fault Signature). A *fault signature* for a fault $f$ and measurement $m$ in mode $q$ is the qualitative magnitude, slope and discrete change in $m$ caused by the occurrence of $f$, and is denoted by $\sigma_{f,m,q}$. We denote all possible signatures for $f$ and $m$ in $q$ as $\Sigma_{f,m,q}$, and denote the set of all fault signatures for fault $f$ and measurements $M$ in mode $q$ as $\Sigma_{f,M,q}$, where $\Sigma_{f,M,q} = \bigcup_{m \in M} \Sigma_{f,m,q}$.

In addition to fault signatures, we also capture the temporal order of measurement deviations, termed *relative measurement orderings*, which refer to the intuition that fault effects will manifest in some parts of the system before others. Energy storage elements in the path between two measured variables impose a delay in the progression of the transient responses from one measurement to the other. If there are no energy storage elements, the relation between the two transients is algebraic and no delay will be observed. This is based on an analysis of the transfer functions from faults to measurements, and details are given in (Daigle *et al.*, 2007a).

*Definition 5* (Relative Measurement Ordering). If a fault $f$ manifests in measurement $m_i$ before measurement $m_j$ then we define a *relative measurement ordering* between $m_i$ and $m_j$ for fault $f$, denoted by $m_i \prec_f m_j$. We denote the set of all measurement orderings for a fault $f$ and measurement set $M$ in mode $q$ as $\Omega_{f,M,q}$.

For hypothesized fault $f$ in mode $q$, and using the TCG, we perform a forward propagation of the fault effects to the measured variables, producing the fault signatures and measurement orderings for the fault (Daigle, 2008; Mosterman and Biswas, 1999). For discrete faults, we use the TCG for the mode induced by the fault, which is the correct mode in which to make predictions. To deal with the effects of discrete faults, we propagate qualitative increasing/decreasing values, but also the discrete change symbol. The `Z` or `N` symbol will continue to propagate from a fault event to other variables in the TCG related by a gain, ie, until an edge that is not labelled by $=$ or a parameter (without integration). From that point on, the `X` symbol will be propagated along the current path.

Fault signatures combined with relative measurement orderings provide a framework for event-based diagnosis, where significant measurement deviations are symbolically abstracted to events. Because measurement orderings define only a partial order, and the fault signature for a given $f$, $m$ and $q$ may not be unique, then the combination of all fault signatures and relative measurement orderings may yield a number possible ways a fault can manifest within a particular mode. We denote each of these possibilities as a *fault trace*.

*Definition 6* (Fault Trace). A *fault trace* for a fault $f$ over measurements $M$ in mode $q$, denoted by $\lambda_{f,M,q}$, is a string of length $\leq |M|$ that includes, for every $m \in M$ that will deviate due to $f$, a fault signature $\sigma_{f,m,q}$, such that the sequence of fault signatures satisfies $\Omega_{f,M,q}$.

Note that the definition implies that fault traces are of maximal length, ie, a fault trace includes deviations for all measurements affected by the fault. We group the set of all fault traces into a *fault language*. The *fault model*, defined by a *finite automaton*, concisely represents the fault language.

*Definition 7* (Fault Language). The *fault language* of a fault $f \in F$ with measurement set $M$ in mode $q$, denoted by $L_{f,M,q}$, is the set of all fault traces for $f$ over measurements $M$.

*Definition 8* (Fault Model). The *fault model* for a fault $f \in F$ with measurement set $M$ in mode $q$, is the finite automaton that accepts exactly the language $L_{f,M,q}$, and is given by $\mathcal{L}_{f,M,q} = (S, s_0, \Sigma, \delta, A)$ where $S$ is a set of states, $s_0 \in S$ is an initial state, $\Sigma$ is a set of events, $\delta : S \times \Sigma \to S$ is a transition function, and $A \subseteq S$ is a set of accepting states.

The fault models can be derived systematically from the signatures and orderings by representing them as finite automata and performing a synchronization operation (Daigle *et al.*, 2007b).

## 4.1 Symbol generation

Symbol generation symbolically abstracts the residual signals of deviating measurements into qualitative features representing observed fault signatures. Once a fault has been detected, the initial measurement deviation is abstracted to produce an initial measurement deviation event, $\sigma_0$, for the diagnoser. As further measurements deviate, additional events (ie, $\sigma_1, \sigma_2, \ldots, \sigma_p$, where $p$ is the number of measurements) may be produced that correspond to the deviations in these measurements. Once a fault is detected, we stop the observer from tracking the state trajectory, otherwise it may compensate for the faulty behaviours and symbol generation will be incorrect. Future estimated values provided for symbol generation are computed using only the model of the system.

In order to represent the dynamics of the deviation, two features are extracted from the residual, and these are abstracted to $+$, `0` and $-$ symbols, representing above, at, and below nominal, respectively. The two features capture (i) the sign of a discontinuity if it occurred (otherwise represented by 0), and (ii) the slope of the residual signal. For example, a positive discontinuity, or jump, followed by an immediate drop in signal magnitude is represented as $+-$, and no discontinuity with a positive signal slope represented as `0`$+$. A robust method is employed for symbol generation by combining a sliding window technique with the Z-test to determine non-zero residual values (Biswas *et al.*, 2003).

In addition to deriving the $+$, $-$ and `0` values, we extend symbol generation in Hybrid TRANSCEND by introducing the `N`, `Z`, and `X` symbols. These symbols can also be computed robustly using the Z-test by taking a small window of samples after fault detection and determining if the estimated and measured values are non-zero or zero over the window (Daigle, 2008). If the estimate is non-zero and the measurement is zero, we report `Z`, and if the estimate is zero and the measurement is non-zero, we report `N`, otherwise, we report `X`.

## 4.2 Hypothesis generation and refinement

Symbol generation produces qualitative values for measurement residuals when they become non-zero. Hypothesis generation uses the initial deviation, $\sigma_0$, and the

hypothesized system mode at the time of fault detection, $\hat{q}(t_d)$, to produce a consistent set of fault candidates. Refinement prunes the set as additional measurements deviate by matching against predicted traces.

At the time of fault occurrence, $t_f$, the system is in some mode $q(t_f^-)$. For parametric faults, the assumption is that $q(t_f^-) = q(t_f^+)$, but for discrete faults, a mode change is induced, ie, $q(t_f^-) \neq q(t_f^+)$. For a discrete fault event $f$, $q(t_f^+) = \mu(f, q(t_f^-))$. We assume that we can accurately track the system mode under nominal conditions, and that only controlled mode changes occur after fault occurrence. Therefore, we hypothesize the faults in the current expected mode that could have occurred to produce the observed deviation. The procedure for handling the presence of unpredicted, but nominal autonomous mode changes during isolation is described in (Daigle, 2008; Narasimhan and Biswas, 2007).

Given a hypothesized system mode after fault detection, $\hat{q}(t_f^-) \in Q_{t_f}$, and given a measurement deviation, $\sigma_0$, we perform a backward propagation in the TCG starting from the observed measurement deviation back to possible changes in parameter values and the occurrence of discrete fault events (Mosterman and Biswas, 1999; Daigle, 2008). When a fault event is encountered, we check if that fault would have caused the hypothesized change in the junction variable, eg, if the flow of a 1-junction decreased and the junction was on, a discrete fault causing the junction to turn off would be consistent with the observation.

Given the initial fault candidates, we refine them by comparing their predicted fault traces to those provided by symbol generation. Hypothesis refinement prunes the current diagnosis as additional measurements deviate. In hybrid systems, refinement needs to deal with controlled mode change events $\sigma \in \Sigma_Q$ as well as measurement deviation events $\sigma \in \Sigma_M$. In new modes of operation, the predicted effects of a fault may change. If a fault candidate $(f, q)$ is consistent with all previous measurement deviations, it will also be consistent with a new deviation if the new deviation matches the predictions of $f$ in $q$, with all previously deviated measurements projected out. The measurement projection operation is defined as follows.

*Definition 9* (Measurement Projection). A *measurement projection* for a trace over measurements $M_i$, $P_{M_i}$, is defined as

$$\begin{aligned} P_{M_i}(\epsilon) &= \epsilon \\ P_{M_i}(\sigma_m) &= \begin{cases} \epsilon & m \notin M_i \\ \sigma_m, & m \in M_i \end{cases} \\ P_{M_i}(\lambda\sigma_m) &= P_{M_i}(\lambda)P_{M_i}(\sigma), \end{aligned}$$

where $\epsilon$ is the empty trace. The *measurement projection* for a fault language $L$ is defined as $P_{M_i}(L) = \{\lambda : \lambda = P_{M_i}(\lambda') \wedge \lambda' \in L\}$.

Essentially, $P_{M_i}(\lambda)$ returns a version of $\lambda$ without the events for measurements in $M - M_i$. Note also that $L_{f,M-M_i} = P_{M-M_i}(L_{f,M})$, ie, the fault language computed with a subset of the measurements, $M - M_i$, is the same as that fault language with the measurements in $M_i$ projected out.

When a controlled mode change event $\sigma \in \Sigma_Q$ occurs, the candidates are updated to the new mode by applying the mode change to the candidate mode, ie, for $c = (f, q)$, the updated candidate is $c' = (f, \mu(\sigma, q))$. If the candidate is a discrete fault, it may not change to the expected mode. For any future measurement deviations, a candidate must be consistent with the predictions for the new mode, with previously deviated measurements projected out. When a new measurement deviation $\sigma \in \Sigma_M$ occurs, the current candidates are checked to be consistent, ie, for each $c = (f, q)$ in the current diagnosis $d$, that $\sigma \sqsubseteq \lambda \in L_{f, M-M_i, q}$, where $M_i$ is the set of previously deviated measurements, and $\sqsubseteq$ defines the prefix operator on traces. If inconsistent, the candidate is dropped, otherwise, it is retained. A unique fault may be identified before a maximal trace is encountered, in which case, fault isolation can terminate early.

## 5. The Event-based Diagnoser

From the fault models, we systematically construct an event-based diagnoser that can be used for diagnosability analysis. We first establish notions of diagnosability and then describe the diagnoser design procedure.

### 5.1 Diagnosability

Diagnosability is an important property of a system, because it enables us to make guarantees about the unique isolation of faults. We first provide definitions of *distinguishability* and *diagnosability* and then describe how these notions are captured in our event-based framework.

In general, distinguishability refers to the notion that two candidates will always produce different effects. For hybrid systems, we must define this with respect to an initial expected mode at the point of fault occurrence, ie, that which will be used as the reference mode in computing deviations from expected behaviour.

*Definition 10* (Distinguishability). For an expected mode $q \in Q$ at the point of fault occurrence, a candidate $c_i$ is distinguishable from a candidate $c_j$, denoted by $c_i \not\sim_q c_j$, if for any possible empty or non-empty sequence of controlled mode changes, $c_i$ always eventually produces effects on the measurements that $c_j$ cannot.

Under our framework, the effects of a fault on the measurements take the form of a trace, which consists of measurement deviation events and controlled mode change events. We define a candidate trace, which represents all possible event sequences consistent with a candidate, as follows.

*Definition 11* (Candidate Trace). An event trace $\lambda = \sigma$ is a *candidate trace* for $c = (f_i, q_i)$ and initial mode of fault occurrence $q_0$, if $\sigma \sqsubseteq \lambda' \in L_{f_i, M, q_i}$ where $q_i = \mu(f_i, q_0)$. An event trace $\lambda = \lambda_i \sigma_{i+1}$ is a *candidate trace* for $c = (f_i, q_{i+1})$ and initial mode of fault occurrence $q_0$, if $\lambda_i$ is a candidate trace for $(f_i, q_i)$, and if $\sigma_{i+1} \in \Sigma_Q$ then $\mu(\sigma_{i+1}, q_i) = q_{i+1}$, or if $\sigma_{i+1} \notin \Sigma_Q$ then $q_i = q_{i+1}$ and $\sigma_{i+1} \sqsubseteq \lambda' \in L_{f_i, M-M_i, q_{i+1}}$. A candidate trace for $c$ with initial mode $q_0$ is denoted as $\lambda_{c, q_0}$.

Clearly, there may be an infinite number of candidate traces because controlled mode changes may keep occurring indefinitely. However, in specifying distinguishability, we are only concerned with *maximal* traces, ie, those for which all measurements that will deviate in the candidate mode already have deviated.

*Definition 12* (Maximal Candidate Trace). A candidate trace $\lambda_{c,q_0}$ for $c = (f_i, q_i)$ is *maximal* if $L_{f,M-M_i,q_i} = \emptyset$, where $M_i$ is the set of deviated measurements for $\lambda_{c,q_0}$.

Now, we can define the language of a candidate $c$ with respect to an initial mode of fault occurrence $q_0$, $L_{c,M,q_0}$ as the set of maximal candidate traces for $c$ starting in $q_0$.

*Definition 13* (Candidate Language). The *candidate language* for candidate $c$, measurements $M$, and initial mode of fault occurrence $q_0$, denoted as $L_{c,M,q_0}$, is the set of all maximal candidate traces $\lambda_{c,q_0}$.

Based on the language of a candidate, we can formally establish distinguishability in our framework.

*Lemma 1. For an expected mode $q_0 \in Q$ at the point of fault occurrence, a candidate $c_i$ is distinguishable from a candidate $c_j$ given measurements $M$ and possible modes $Q$, if there does not exist a pair of candidate traces $\lambda_{c_i,q_0} \in L_{c_i,M,q_0}$ and $\lambda_{c_j,q_0} \in L_{c_j,M,q_0}$ such that $\lambda_{c_i} \sqsubseteq \lambda_{c_j}$.*[1]

So, if a maximal candidate trace, which is a sequence of controlled mode change events and measurement deviation events, for some candidate is a prefix of a trace for a second candidate, then if the first candidate occurs and produces that trace, the first candidate cannot be distinguished from the second because no more measurements will deviate with which to distinguish them (since the trace is maximal).

Since candidate traces include mode change events, the candidate languages cover all possible sequences of controlled mode change events interleaved with measurement deviations. In many cases, these languages may be extremely large, but in order to verify distinguishability of faults, this cannot be avoided. In online diagnosis, however, we do not need to enumerate all traces because we can construct the candidates which match the partial traces efficiently using the hypothesis generation and refinement algorithms discussed in the previous section.

In our framework, a *system* can be defined as follows.

*Definition 14* (System). A *system* $\mathcal{S}$ is defined as $(F, M, Q, L_{F,M,Q})$, where $F = \{f_1, f_2, \ldots, f_n\}$ is a set of faults, $M = \{m_1, m_2, \ldots, m_p\}$ is a set of measurements, $Q = \{q_1, q_2, \ldots, q_r\}$ is a set of modes, and $L_{F,M,Q}$ is the set of fault languages for each fault in each mode, ie, $L_{F,M,Q} = \{L_{f,M,q} : f \in F, q \in Q\}$.

Based on distinguishability, we obtain the following notion of diagnosability for a system.

*Definition 15* (Diagnosability). A system $\mathcal{S} = (F, M, Q, L_{F,M,Q})$ is *diagnosable* if for all $c_i$ and $c_j$ and possible modes of fault occurrence $q_0 \in Q$, if $c_i \not\sim_{q_0} c_j$.

If the system is diagnosable, then for every two candidates, for any set of controlled mode changes, are distinguishable using the measurements in $M$. So, each sequence of

[1]Proofs can be found in the Appendix.

measurement deviations and controlled mode changes we observe can be eventually linked to a single, unique candidate. Hence, we can uniquely isolate all candidates of interest. If the system is not diagnosable, then ambiguities may remain when fault isolation terminates. Diagnosability can tell us when fault isolation may terminate, ie, when a unique result is obtained, or analysis shows that no further events can help to distinguish the remaining faults.

The definition of diagnosability allows making guarantees about fault isolation. Since the diagnoser has no control over which controlled mode change events are issued, we cannot, in general, make any restrictions about when a mode change event will be issued. So, diagnosability is conservative. It may be possible, however, to avoid ambiguous diagnosis results if certain mode changes are blocked or executed. We define this as $Q$-diagnosability, which implies some interaction between the diagnoser and the controller.

*Definition 16* ($Q$-diagnosability). A system $\mathcal{S} = (F, M, Q, L_{F,M,Q})$ is *$Q$-diagnosable* if for all candidates $c_i$ and $c_j$ and possible modes of fault occurrence $q_0 \in Q$ and $c_i$ is not distinguishable from $c_j$ (ie, $c_i \sim_{q_0} c_j$), then for every maximal candidate trace $\lambda_{c_i,q_0}$ where $\lambda_{c_i,q_0}$ is a prefix of some candidate trace $\lambda_{c_j,q_0}$, either (i) there exists some sequence of controlled mode changes $\lambda_Q$ where the trace $\lambda_{c_i,q_0}\lambda_Q$ is not maximal for any candidate, or (ii) for every trace $\lambda_{c_k}$ such that $\lambda_{c_k}\lambda_Q$ equals $\lambda_{c_i,q_0}$ (where $\lambda_Q$ is a sequence of controlled mode changes), $\lambda_{c_k}$ is not maximal for any candidate.

If the system is $Q$-diagnosable, then for any trace that violates diagnosability, there is some sequence of controlled mode changes that can be applied such that the new trace is no longer maximal, ie, more measurement deviations will occur, or for every partial trace that can become the violating trace via a sequence of controlled mode changes, the partial trace is not maximal. The first case corresponds to executing controlled mode changes to ensure more measurement deviations will occur. The second case corresponds to blocking a sequence of controlled mode changes such that we never encounter the violating trace in the first place.

## 5.2 Diagnoser design

We construct from our fault models an event-based diagnoser, which is an extended form of a finite automaton. If our system is diagnosable, we can construct a diagnoser that uniquely isolates all candidates. If not, the constructed diagnoser will give ambiguous results in some cases. We define formally a *diagnoser* in our framework.

*Definition 17* (Diagnoser). A *diagnoser* for a fault set $F$, measurements $M$ and modes $Q$, is defined as $\mathcal{D}_{F,M,Q} = (S, I, \Sigma, \delta, A, D, Y)$ where $S$ is a set of states, $I \subseteq S$ is set of initial states, $\Sigma$ is a set of events, $\delta : S \times \Sigma \to S$ is a transition function, $A \subseteq S$ is a set of accepting states, $D \subseteq 2^C$ is a set of diagnoses, and $Y : S \to D$ is a diagnosis map.

A diagnoser is a finite automaton extended by a set of diagnoses and a diagnosis map. The initial states correspond to possible starting modes at the point of fault occurrence. A diagnoser takes events as inputs, which correspond to measurement deviations and

controlled mode changes. From the current state, a measurement deviation event causes a transition to a new state. The diagnosis for that new state represents the set of candidates that are consistent with the sequence of events seen up to the current point in time, ie, it encodes the results that hypothesis generation and refinement would obtain.

The accepting states of the diagnoser correspond to a fault isolation result. We say that a diagnoser *isolates* a candidate if it accepts all possible valid traces for the candidate and the accepting states map to diagnoses containing the candidate.

*Definition 18* (Isolation). A diagnoser $\mathcal{D}_{F,M,Q}$ *isolates* a candidate $c$ if it accepts all $\lambda \in L_{c,M,q_0}$ for all nominal $q_0 \in Q$, and for each $s \in A$ that accepts a $\lambda \in L_{c,M,q_0}$, $c \in Y(s)$.

The notion of isolation gives us an indication of correctness of our diagnosers. If our diagnoser isolates all candidates, then it covers all possible observable fault traces, and, therefore, is constructed correctly. More importantly, we also would like to achieve unique isolation of candidates, which is a stronger notion of isolation.

*Definition 19* (Unique Isolation). A diagnoser $\mathcal{D}_{F,M,Q}$ *uniquely isolates* a candidate $c$ if it isolates $c$ and for each $s \in A$ that accepts some $\lambda_c \in L_{c,M,q_0}$, $\{c\} = Y(s)$.

For unique isolation, we require the diagnoser isolates the candidate, but also that the corresponding accepting states uniquely determine $c$. This means that the diagnoser will accept all valid fault traces, but also that each fault trace will uniquely identify a single candidate, which relates directly to diagnosability. If the system is not diagnosable, we can also use the diagnoser to determine which traces result in ambiguities, and if possible, be able to avoid those traces by permitting or prohibiting certain controlled mode changes during isolation, ie, determine if the system is $Q$-diagnosable.

So, we would like to systematically construct a diagnoser for a hybrid system $\mathcal{S}$ that isolates all possible candidates. Further, we would like to prove that if $\mathcal{S}$ is diagnosable, then this diagnoser *uniquely* isolates all candidates. To do this, we use individual diagnosers for each fault-mode pair, and provide a composition operator to simultaneously compose all the individual diagnosers to a global diagnoser that isolates all the valid candidates.

First, we construct a diagnoser, $\mathcal{D}^*_{\{f\},M,q}$ for each single fault $f$ within each mode $q$ from $\mathcal{L}_{f,M,q}$.

*Definition 20* ($\mathcal{D}^*_{\{f\},M,q}$). Given fault $f$ and mode $q$ for measurements $M$, with $\mathcal{L}_{f,M,q} = (S,\ s_0,\ \Sigma,\ \delta,\ A)$, $\mathcal{D}^*_{\{f\},M}$ is defined as $(S,\ s_0,\ \Sigma,\ \delta,\ A',\ \{\{(f,q)\}\},\ Y)$, where $Y(s) = \{(f,q)\}$ for all $s \in S$, and $A' = A$ if $S \neq \{s_0\}$, or $A' = \{s_0\}$ otherwise.

Since we want to use the diagnoser to analyze system diagnosability, we need to be sure that for any possible controlled mode behaviour after fault occurrence, faults can be isolated. Given modes $Q = \{q_0, q_1, \ldots, q_r\}$, we abstract the mode change events to $\Sigma_Q = \{\sigma_{q_0}, \sigma_{q_1}, \ldots, \sigma_{q_r}\}$, where $\sigma_{q_i}$ indicates an event that changes the system to mode $q_i$.[2] The sequence of the commands should be consistent with the discrete mode behaviour of the system, which can be described by the mode transition function $\mu$.

[2]Different events that both change the system to some mode $q_i$ are abstracted to the same event $\sigma_{q_i}$.

We simultaneously compose each of the individual diagnosers $\mathcal{D}_{\{f\},M,q}$. In incremental consistency checking, we project out measurements that have already deviated to obtain the set of consistent candidates for a new observation. For a diagnoser, the state-based form of the measurement projection operation on traces is formalized using *boundaries* and *boundary transition functions*.

*Definition 21* (Boundary). The *boundary* for a state $s$ and deviated measurements $M_i$, $B_{M_i}(s)$, is defined as the set of all states $\delta(\lambda, s)$ such that $P_{M-M_i}(\lambda) = \epsilon$.

The boundary for a state $s$ is basically the set of states that may be transitioned to from $s$ via a trace $\lambda$ consisting of only events for measurements that have already deviated, ie, measurements corresponding to the events for traces in the history of the state. Using the notion of a boundary, we define a *boundary transition function* with respect to a set of deviated measurements.

*Definition 22* (Boundary Transition Function). The *boundary transition function* for an event $\sigma$, state $s$ and set of deviated measurements $M_i$, denoted as $\delta_{M_i}(\sigma, s)$, is a transition function that maps $\sigma$ and $s$ to some state $s'$, where either (i) over all boundary states $s_B \in B_{M_i}(s)$, exactly one state can be reached via $\sigma$, in which case $s'$ is that unique state, or (ii) no states can be reached or more than one state can be reached from the boundary states, in which case $s' = \text{ø}$.

In other words, $\delta_{M_i}(\sigma, s)$ returns the unique state that can be reached from a boundary state of $s$ via $\sigma$, or ø if there are no states that can be reached or if the state is not unique. Because of the way the $\mathcal{D}^*_{\{f\},M,q}$ diagnosers are computed, the reachable state will always be unique or null, because traces with the same set of measurements map to the same state. In the following, we denote the measurements that have deviated in a state $s$ as $M(s)$.

We now describe a composition operator, $\Pi$, that simultaneously combines the $\mathcal{D}^*_{\{f\},M,q}$ for each possible $(f, q)$ pair. We split the mode set $Q$ into nominal modes $Q_N$ and faulty modes $Q_F$.

*Definition 23* ($\Pi$ Composition). Given the set of all $k$ $(f, q)$ diagnosers, $\mathbb{D} = \{\mathcal{D}^*_{\{f\},M,q} : f \in F, q \in Q\}$, $\mathcal{D}^*_{F,M,Q} \triangleq \Pi(\mathbb{D})$, where

- $I = \{(s_{0,1}, s_{0,2}, \ldots, s_{0,k}, q, (\text{ø}, q)) : q \in Q_N\}$
- $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \cdots \cup \Sigma_k \cup \Sigma_Q$
- $\delta(\sigma, (s_{i,1}, s_{i,2}, \ldots, s_{i,k}, q_i, d_i)) = (s_{i+1,1}, s_{i+1,2}, \ldots, s_{i+1,k}, q_{i+1}, d_{i+1})$, where $\sigma \in \Sigma_Q$, all $s_{i+1,j} = s_{i+1,j}$, $q_{i+1} = \mu(\sigma, q_i)$, and $d_{i+1} = \{(f, \mu(\sigma, q)) : \mu(\sigma, q) \neq \text{ø} \wedge (f, q) \in d_i\}$
- $\delta(\sigma, (s_{i,1}, s_{i,2}, \ldots, s_{i,k}, q_i, d_i)) = (s_{i+1,1}, s_{i+1,2}, \ldots, s_{i+1,k}, q_{i+1}, d_{i+1})$, where $\sigma \in \Sigma_M$, $q_{i+1} = q_i$, $M_i = M((s_{i,1}, s_{i,2}, \ldots, s_{i,k}, q_i, d_i))$, $s_{i+1,j} = s_{i,j}$ if $\delta_{M_i,j}(\sigma, s_{i,j}) = \text{ø}$, or $\delta_{M_i,j}(\sigma, s_{i,j})$ otherwise, and $d_{i+1} = \{(f, q) \in d_i : \sigma \sqsubseteq \lambda \in L_{f,M-M_i,q}\} \neq \text{ø}$
- $S$ is the set of all $s$ reachable through $\delta$ from some $s_0 \in I$
- $A$ is the set of all $s_i = (s_{i,1}, s_{i,2}, \ldots, s_{i,k}, q_i, d_i) \in S$ where there exists some $s_{i,j} \in s_i$, with some $s_{B,j} \in B_{M(s_i)}(s_{i,j})$ where $s_{B,j} \in A_j$, such that $Y_j(s_{B,j}) \in Y(s_i)$
- $D$ is the set of all $d_i$ in each $(s_{i,1}, s_{i,2}, \ldots, s_{i,k}, q_i, d_i) \in S$
- $Y((s_{i,1}, s_{i,2}, \ldots, s_{i,k}, q_i, d_i)) = d_i$

*Theorem 1. The diagnoser $\mathcal{D}^*_{F,M,Q}$ isolates all valid candidates.*

Further, we can show that if the system $\mathcal{S}$ is diagnosable, then the diagnoser uniquely isolates all candidates.

*Theorem 2. A system $\mathcal{S} = (F, M, Q, L_{F,M,Q})$ is diagnosable if and only if $\mathcal{D}^*_{F,M,Q}$ uniquely isolates all valid candidates.*

## 6. Case study

We demonstrate the proposed diagnosis framework with experiments conducted on the ADAPT deployed at NASA Ames Research Center (Poll *et al.*, 2007). The testbed is functionally representative of a spacecraft's electrical power system, and consists of three subsystems: (i) power generation, which includes two battery chargers, (ii) power storage, which consists of three sets of lead-acid batteries, and (iii) power distribution, which consists of a number of relays and circuit breakers, two inverters, and various DC and AC loads.

We consider a subset of ADAPT to demonstrate our approach, which includes a lead-acid battery, two relays, and two DC loads. The battery is modelled by an electric circuit equivalent based on (Ceraolo, 2000) (Figure 5), and supplementary equations are given in Table 1. The battery capacity to store charge is modelled using a large capacitance, $C_0$. Other parameters model nonlinear, dissipative behaviours. The battery supplies voltage to the relays through a parallel connection, which in turn supply power to the two DC loads. The selected measurements are the battery voltage, $V_B(t)$, and the currents through the relays, $I_{L1}(t)$ and $I_{L2}(t)$, ie, $M = \{I_{L1}, I_{L2}, V_B\}$.

### 6.1 Modelling faults

We consider faults in the battery, loads, relays and sensors. Common battery faults include loss of charge and resistance increases brought about by battery use and age, which manifest as a side effect of the chemical reactions. Loss of charge capacity is

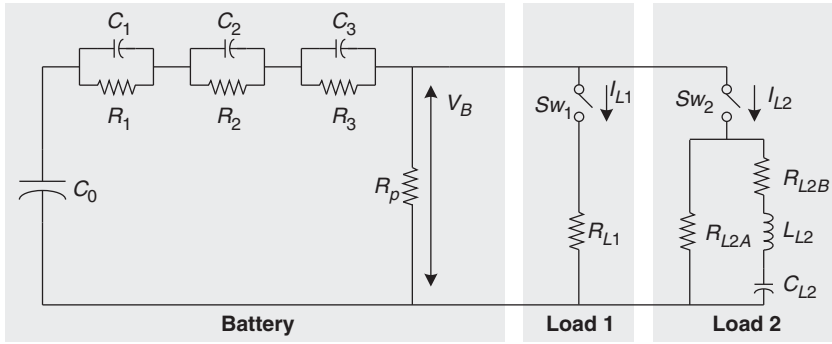


**Figure 5** Electrical circuit equivalent for the battery system

**Table 1** Supplementary battery equations

$$\dot{\theta}(t) = \frac{1}{C_\theta}\left(P_B - \frac{\theta(t) - \theta_a}{R_\theta}\right)$$

$$SOC = 1 - \frac{Q_{max} - q(t)}{K_c C_0^*\left(1 - (\theta/\theta_f)\right)^\epsilon}$$

$$DOC = 1 - \frac{Q_{max} - q(t)}{K_c C_0^*\left(1 - (\theta/\theta_f)\right)^\epsilon}\left(1 + (K_c - 1)\left(\frac{i(t)}{I^*}\right)^\delta\right)$$

$$R_1 = R_{10} + A_{11} SOC$$

$$R_2 = -R_{20}\ln(DOC)$$

$$R_3 = \frac{R_{30}\exp A_{31}(1 - SOC)}{1 + \exp A_{32} i(t)}$$

represented by a capacitance decrease, $C_0^-$, and an increase in internal losses by $R_1^+$. Faults can occur in the system loads, and these are represented by increases or decreases in their resistance values, $R_{L1}$ and $R_{L2A}$. For the sensors, we consider bias faults, which produce abrupt changes in the measured values manifesting as constant offsets. Sensor faults are labelled by the measured quantity they represent, eg, $V_B^+$ represents a bias fault in the battery voltage sensor. We represent discrete faults in $Sw_1$ and $Sw_2$ by fault events $\alpha$ and $\beta$, respectively, where a subscript of 0 indicates a stuck-off fault, and a subscript of 1 indicates a stuck-on fault.

## 6.2 Diagnosability analysis

We denote the system mode as $q_{ij}$, where $i$ is the mode of $Sw_1$, and $j$ is the mode of $Sw_2$. For example, $q_{1\beta_0}$ is the mode where $Sw_1$ is on and $Sw_2$ is stuck off. We allow controlled mode events that switch the system from any one controlled mode to another, ie, $\Sigma_Q = \{\sigma_{q_{00}}, \sigma_{q_{01}}, \ldots, \sigma_{q_{11}}\}$. We restrict the occurrence of discrete faults to modes where a deviation would be produced. For example, $\alpha_1$ would not produce any deviations if it occurred in a mode where $Sw_1$ was already on.

The fault signatures and relative measurement orderings for the chosen faults are given in Table 2 for selected modes ($q_{**}$ indicates the signatures and orderings are valid for any mode). The non-linearities in the battery introduce ambiguity in the qualitative signatures, and this is denoted by the * symbol, eg, a signature of `0*` may manifest as `0+` or `0-`. Since the sensors are not part of any feedback loops in the system, sensor faults affect only the measurement provided by the sensor. The other measurements are not affected, and so the corresponding fault signatures are denoted by 00, indicating no change in the measurement from expected behaviour.

Selected fault models for ADAPT are shown in Figure 6. Consider the fault model $\mathcal{L}_{R_{L1}+,q_{11}}$, shown in Figure 6(a). From the orderings, the current through Load 1 must be the first to deviate, followed by the Load 2 current and battery voltage in any order.

**Table 2** Fault signatures and relative measurement orderings for the adapt subsystem

| Fault | $V_B$ | $I_{L1}$ | $I_{L2}$ | Measurement Orderings |
|---|---|---|---|---|
| $(V_B^+, q_{**})$ | +0,X | 00,X | 00,X | $V_B \prec I_{L1}, V_B \prec I_{L2}$ |
| $(V_B^-, q_{**})$ | −0,X | 00,X | 00,X | $V_B \prec I_{L1}, V_B \prec I_{L2}$ |
| $(I_{L1}^+, q_{**})$ | 00,X | +0,X | 00,X | $I_{L1} \prec V_B, I_{L1} \prec I_{L2}$ |
| $(I_{L1}^-, q_{**})$ | 00,X | −0,X | 00,X | $I_{L1} \prec V_B, I_{L1} \prec I_{L2}$ |
| $(I_{L2}^+, q_{**})$ | 00,X | 00,X | +0,X | $I_{L1} \prec V_B, I_{L2} \prec I_{L1}$ |
| $(I_{L2}^-, q_{**})$ | 00,X | 00,X | −0,X | $I_{L1} \prec V_B, I_{L2} \prec I_{L1}$ |
| $(C_0^-, q_{11})$ | +−,X | +−,X | +−,X | ø |
| $(R_1^+, q_{11})$ | 0-,X | 0-,X | 0-,X | ø |
| $(R_{L1}^+, q_{11})$ | 0*,X | −+,X | 0*,X | $I_{L1} \prec V_B, I_{L1} \prec I_{L2}$ |
| $(R_{L1}^-, q_{11})$ | 0*,X | +−,X | 0*,X | $I_{L1} \prec V_B, I_{L1} \prec I_{L2}$ |
| $(R_{L2A}^+, q_{11})$ | 0*,X | 0,X | −+,X | $I_{L2} \prec V_B, I_{L2} \prec I_{L1}$ |
| $(R_{L2A}^-, q_{11})$ | 0*,X | 0*,X | +−,X | $I_{L2} \prec V_B, I_{L2} \prec I_{L1}$ |
| $(\alpha_0, q_{\alpha_0 1})$ | 0*,X | −*,Z | 0*,X | $I_{L1} \prec V_B, I_{L1} \prec I_{L2}$ |
| $(\alpha_1, q_{\alpha_1 1})$ | 0*,X | +*,N | 0*,X | $I_{L1} \prec V_B, I_{L1} \prec I_{L2}$ |
| $(\beta_0, q_{1\beta_0})$ | 0*,X | 0*,X | −*,Z | $I_{L2} \prec V_B, I_{L2} \prec I_{L1}$ |
| $(\beta_1, q_{1\beta_1})$ | 0*,X | 0*,X | +*,N | $I_{L2} \prec V_B, I_{L2} \prec I_{L1}$ |

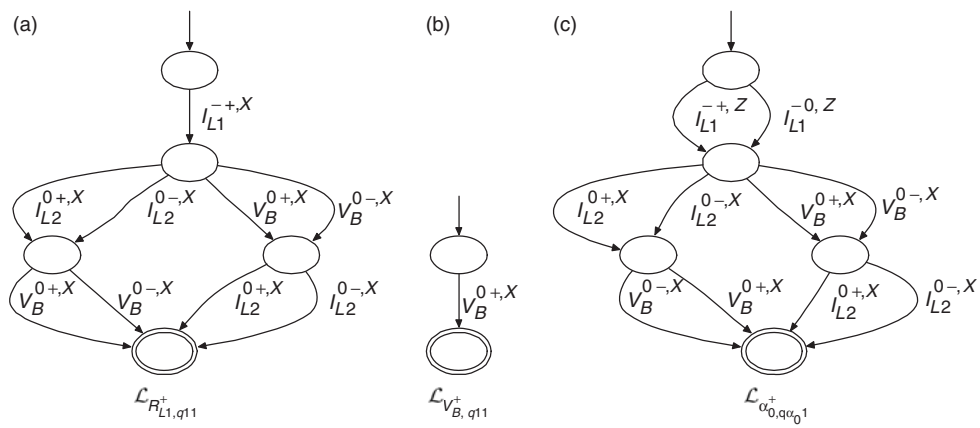


**Figure 6** Selected fault models for ADAPT

The direction of the changes in $I_{L2}(t)$ and $V_B(t)$ are unknown so both possibilities are represented.

Given any one mode, the system is diagnosable. After at most two measurement deviations, a unique candidate can be isolated. However, over all modes, the system is

not diagnosable. Figure 7 gives a partial diagnoser for the system that illustrates this property, with $F = \{C_0^-, R_{L1}^+\}$ and initial mode $q_{11}$ with $\sigma_{q_{01}}$ being the only controlled mode change event. We can see that if $I_{L1}^{+-,X}\sigma_{q_{01}}$ occurs, we reach an accepting state that corresponds to a diagnosis with multiple candidates. After that event, both $C_0^-$ and $R_{L1}^+$ are consistent. Since the state is accepting, it is possible that no new measurement deviations will occur to distinguish the faults. The resistance fault will have no visible effects on the rest of the measurements in this mode, because the source of the deviations is cut-off, so we would have to wait infinitely long to verify $R_{L1}^+$ was the true fault. Therefore, the system is not diagnosable. We can see, however, that the system is $Q$-diagnosable. If we prevent $\sigma_{q_{01}}$ from occurring, or change back to $q_{11}$ if it does occur, more measurements will deviate and we can distinguish the candidate uniquely.

## 6.3 Experimental results

We have performed experiments online on the ADAPT testbed. To demonstrate the diagnosis approach, we show the results obtained for a load fault and a switch fault. The measurements were sampled at 2 Hz for all the experiments. Extensive simulation experiments to evaluate the robustness of fault detection and symbol generation were also performed and are given in (Daigle, 2008), but omitted here for space. The nominal behaviour of the system is shown in Figure 8. As shown in the figure, the system parameters (Table 3) are fairly accurate, and the observer tracks well.

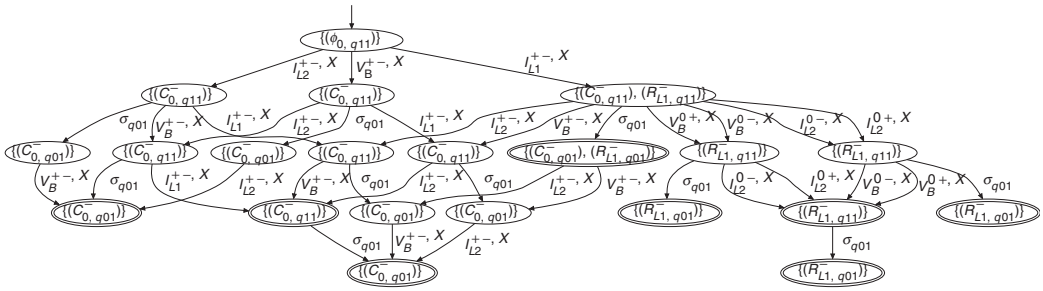


**Figure 7** Hybrid diagnoser for $F = \{C_0^-, R_{L1}^+\}$ and initial mode $q_{11}$

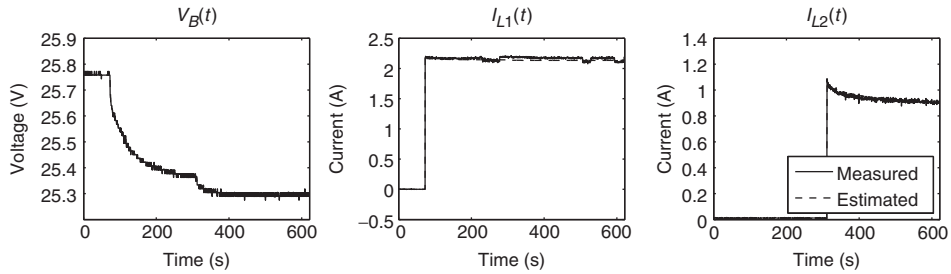


**Figure 8** Nominal system operation

For the first experiment, a 100% increase in the Load 1 resistance, $R_{L1}^{+}$, is manually injected at 439.5 s in mode $q_{11}$. The measured and estimated outputs are shown in Figure 9. A partial diagnoser is given in Figure 10. The increase in resistance causes a discontinuous drop in the current, detected at 440.0 s. Since the slope has not yet been computed, the possible fault candidates are $\{(R_1^+, q_{11}), (R_{L1}^+, q_{11}), (I_{L1}^-, q_{11}), (\alpha_0, q_{\alpha_0 1})\}$. At 441.0 s, an increase is detected in $V_B(t)$. Since $I_{L1}^-$ cannot affect $V_B(t)$, it is dropped. $R_1^+$ is also dropped because it would have decreased, and not increased, the battery voltage. Due to the dynamics of Load 2, the change in $V_B(t)$ is not large enough to cause an observable change in $I_{L2}(t)$. At 442.5 s, it is determined that no discrete change in $I_{L1}(t)$ occurred, so $R_{L1}^+$ is isolated as the true fault. Without the additional symbol, the faults cannot be distinguished, therefore, an integrated approach is necessary.

We now investigate an unexpected switch fault. At 375.5 s, $Sw_1$ turns off without a command, so the expected mode is $q_{11}$ but the actual mode is $q_{\alpha_0 1}$. The measured and estimated outputs are shown in Figure 11. The partial diagnoser of Figure 10 applies to this case also. As a result of the fault, $I_{L1}(t)$ goes immediately to zero, and $V_B(t)$

**Table 3** Identified system parameters

| | | |
|---|---|---|
| Battery parameters | $\theta_a = 22^\circ\text{C}$ | $C_\theta = 615.3 Wh/^\circ\text{C}$ |
| | $R_\theta = 0.01^\circ\text{C/W}$ | $C_0 = 106360\,F$ |
| | $C_1 = 51.079\,F$ | $C_2 = 51.216\,F$ |
| | $C_3 = 567.56\,F$ | $R_{10} = 0.05582\,\Omega$ |
| | $A_{11} = -0.025025$ | $R_{20} = 0.001847\,\Omega$ |
| | $R_{30} = 0.3579\,\Omega$ | $A_{31} = -2.5315$ |
| | $A_{32} = 0.22208$ | $R_p = 500\,\Omega$ |
| | $Q_{max} = 2765360\,C$ | $K_c = 1.33$ |
| | $C_0^* = 270720\,Ah$ | $\theta_f = -35^\circ\text{C}$ |
| | $\epsilon = 0.642$ | $I^* = 5A$ |
| | $\delta = 0.61$ | |
| Load parameters | $R_{L1} = 11.8\,\Omega$ | $R_{L2A} = 27.696\,\Omega$ |
| | $R_{L2B} = 209.92\,\Omega$ | $C_{L2} = 0.48678\,F$ |
| | $L_{L2} = 1.9986\,H$ | |

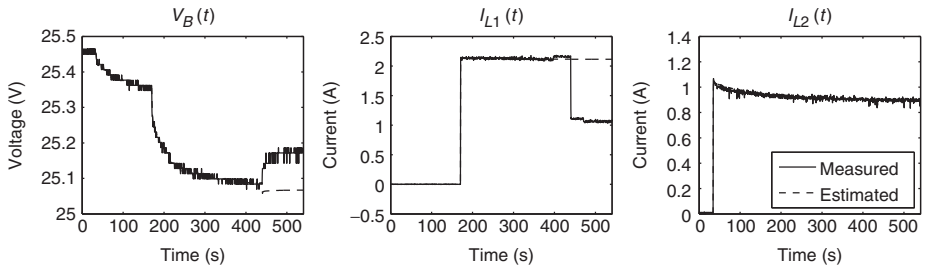


**Figure 9** $R_{L1}^+$ fault, where $R_{L1}$ increases by 100%

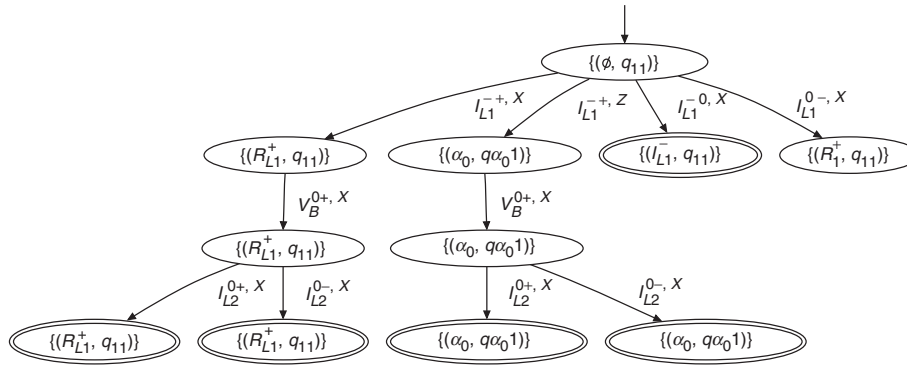


**Figure 10** Partial diagnoser for isolating $R_{L1}^{+}$

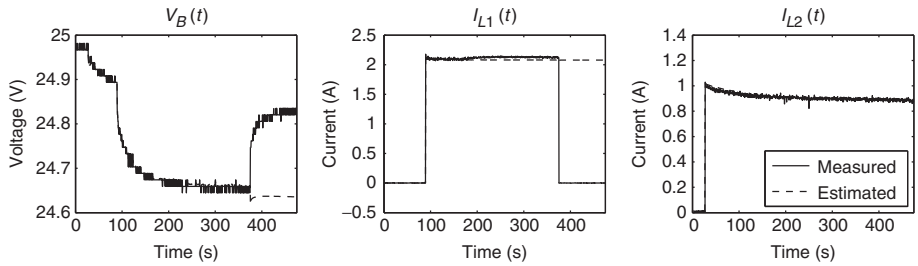


**Figure 11** $Sw_1$ turns off

increases as a result of less current being drawn. The fault is detected at 376.0 s, and the symbol generator reports a decrease in $I_{L1}(t)$. The initial fault hypotheses are then $\{(R_1^+, q_{11}), (R_{L1}^+, q_{11}), (I_{L1}^-, q_{11}), (\alpha_0, q_{\alpha_0 1})\}$. At 376.5 s, the increase in $V_B(t)$ is detected, so the diagnosis reduces to $\{(R_1^+, q_{11}), (\alpha_0, q_{\alpha_0 1})\}$. At 378.5 s, the symbol generator determines that $I_{L1}^-$ went to zero, and, therefore, $\alpha_0$ is isolated as the true fault. Again, without the additional symbol, the faults cannot be discriminated.

## 7. Conclusions

We presented an integrated parametric and discrete fault framework for event-based diagnosis of hybrid systems. Deviations in expected behaviour are abstracted to events to perform qualitative fault isolation. Both parametric and discrete faults are included in the diagnosis model so that their effects can be predicted using our qualitative algorithms. We presented a case study for hybrid diagnosis on the ADAPT system, with experimental results that demonstrate the effectiveness of the approach to a real, complex hybrid system. Future work will extend the approach to handle autonomous mode changes as described in (Narasimhan and Biswas, 2007), and incorporate a fault identification module that handles both parametric and discrete faults.

## Acknowledgement

This work was supported in part by grants NSF CNS-0615214, NASA USRA 08020-013, NASA NRA NNX07AD12A, and NSF CNS-0347440.

## References

**Biswas, G., Simon, G., Mahadevan, N., Narasimhan, S., Ramirez, J.** and **Karsai, G.** 2003: A robust method for hybrid diagnosis of complex systems. *Proceedings of the 5th Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Washington, DC, USA, 1125–1131.

**Ceraolo, M.** 2000: New dynamical models of lead-acid batteries. *IEEE Transactions on Power Systems* 15(4), 1184–90.

**Cocquempot, V., Mezyani, T.El.** and **Staroswiecki, M.** 2004: Fault detection and isolation for hybrid systems using structured parity residuals. *Proceedings of the 5th Asian Control Conference*, Melbourne, Australia, 1204–12.

**Daigle, M.** 2008: A qualitative event-based approach to fault diagnosis of hybrid systems, PhD thesis, Vanderbilt University.

**Daigle, M.J., Koutsoukos, X.D.** and **Biswas, G.** 2007a: Distributed diagnosis in formations of mobile robots. *IEEE Transactions on Robotics* 23(2), 353–69.

**Daigle, M., Koutsoukos, X.** and **Biswas, G.** 2007b: Fault diagnosis of continuous systems using discrete-event methods. *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, 2626–32.

**Dearden, R.** and **Clancy, D.** 2002: Particle filters for real-time fault detection in planetary rovers. *Proceedings of the 12th International Workshop on Principles of Diagnosis*, Semmering, Austria, 1–6.

**Hofbaur, M.** and **Williams, B.C.** 2002: Mode estimation of probabilistic hybrid systems. *Hybrid Systems: Computation and Control*, Volume. 2289 of *LNCS*, Springer-Verlag, 253–66.

**Karnopp, D.C., Margolis, D.L.** and **Rosenberg. R.C.** 2000: *Systems Dynamics: Modeling and Simulation of Mechatronic Systems*, John Wiley & Sons, Inc.

**Koutsoukos, X., Kurien J.** and **Zhao, F.** 2003: Estimation of distributed hybrid systems using particle filtering methods. *Hybrid Systems: Computation and Control*. Vol. 2623 of *LNCS*, Springer, 298–313.

**McIlraith, S.A., Biswas, G., Clancy, D.** and **Gupta, V.** 2000: Hybrid systems diagnosis. *Hybrid Systems: Computation and Control*. Vol. 1790 of *LNCS*, Springer, 282–95.

**Mosterman, P.J.** and **Biswas, G.** 1998: A theory of discontinuities in physical system models. *Journal of the Franklin Institute* 335B(3), 401–39.

**Mosterman, P.J.** and **Biswas, G.** 1999. Diagnosis of continuous valued systems in transient operating regions. *IEEE Transactions on Systems, Man and Cybernetics, Part A* 29(6), 554–65.

**Narasimhan, S.** and **Biswas, G.** 2007: Model-based diagnosis of hybrid systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A* 37(3), 348–61.

**Poll, S., Patterson-Hine, A. Camisa, J., *et al.*,** 2007: Evaluation, selection, and application of model-based diagnosis tools and approaches. *AIAA Infotech@Aerospace 2007 Conference Proceedings*, Rohnert Park, California.

**Travé-Massuyés, L., Benazera, E.** and **Dague, P.** 2002: State tracking of uncertain hybrid concurrent systems. *Proceedings of the 13th International Workshop on Principles of Diagnosis*, Semmering, Austria, 106–14.

**Wang, W., Li, L., Zhou, D. et al.,** 2007: Robust state estimation and fault diagnosis for uncertain hybrid nonlinear systems. *Nonlinear Analysis: Hybrid Systems* 1(1), 2–15.

**Zhao, F., Koutsoukos, X., Haussecker, H., Reich, J. et al.,** 2005: Monitoring and fault diagnosis of hybrid systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 35(6), 1225–40.

## Appendix

### Proof of Lemma 1

*Proof.* Assume $c_i$ is not distinguishable from $c_j$ for initial mode of fault occurrence $q_0 \in Q$, ie, $c_i \sim_{q_0} c_j$. Then by definition, starting in mode $q_0$, there must exist a maximal candidate trace by $c_i$ that $c_j$ can also produce. Therefore, there must exist some maximal candidate trace for $c_i$, ie, some $\lambda_{c_i,q_0} \in L_{c_i,M,q_0}$, and some sequence of events for $c_j$ that is not distinct from $\lambda_{c_i,q_0}$. So, $\lambda_{c_i,q_0}$ must be a candidate trace $\lambda_{c_j,q_0}$ for $c_j$. Therefore, if $c_i \sim_{q_0} c_j$ then there exits some $\lambda_{c_i,q_0} \in L_{c_i,M,q_0}$ and $\lambda_{c_j,q_0} \in L_{c_j,M,q_0}$ such that $\lambda_{c_i,q_0} \sqsubseteq \lambda_{c_j,q_0}$. By the contrapositive, if there does not exist $\lambda_{c_i,q_0} \in L_{c_i,M,q_0}$ and $\lambda_{c_j,q_0} \in L_{c_j,M,q_0}$ such that $\lambda_{c_i,q_0} \sqsubseteq \lambda_{c_j,q_0}$, then $c_i \not\sim_{q_0} c_j$. ■

### Proof of Theorem 2

*Proof.* Assume initial mode of fault occurrence $q_0$, candidate $c$, and trace $\lambda = \sigma_1\sigma_2, \ldots, \sigma_k \in L_{c,M,q_0}$. By the definition of a candidate trace, $\sigma_1$ is a candidate trace for $c' = (f_i, \mu(f, q_0))$ if $\sigma_1 \sqsubseteq \lambda' \in L_{f,M,\mu(f,q_0)}$. Therefore, $(f, \mu(f, q_0)) \in h_{F,M_i}(\sigma_1)$, so by definition of $\wedge_l$, the resultant diagnosis will contain $(f, \mu(f, q_0))$, so by definition of $\delta$, the corresponding state is in $S$. Assume $\lambda_i$ is a candidate trace for $c' = (f_i, q_i)$ and has a corresponding state $s \in S$. Then if $\sigma_{i+1} \in \Sigma_Q$, $\lambda_i\sigma_{i+1}$ is a candidate trace for $(f_i, \mu(\sigma_{i+1}, q_i))$ and by definition of $\delta$ has a corresponding state $s \in S$ and the associated diagnosis has $(f_i, \mu(\sigma_{i+1}, q_i))$. If $\sigma_{i+1} \notin \Sigma_Q$, then $\lambda_i\sigma_{i+1}$ is a candidate trace for $(f_i, q_i)$ if $\sigma_{i+1} \sqsubseteq \lambda' \in L_{f,M,\mu(f,q_0)}$ and therefore by definition of a hypothesis set, $(f_i, q_i) \in h_{F,M_i}(\sigma_{i+1})$, so by definition of $\wedge_l$, the diagnosis will contain $(f_i, q_i)$ and by definition of $\delta$, will have a corresponding state in $S$. Therefore, there is a state for any valid candidate trace. Given a state $s \in S$ with a trace that is maximal for $c = (f_i, q_i)$, the substate of $s$ that corresponds to a state in $\mathcal{D}^*_{f,M,q_i}$ must have no measurement deviations possible from its boundary, otherwise the trace would not be maximal, and thus the boundary must contain an accepting state. ■

### Proof of Theorem 2

*Proof.* Assume $\mathcal{S}$ is diagnosable. Assume a $c$ and $\lambda \in L_{c,M,Q}$. $\mathcal{D}^*_{F,M,Q}$ isolates $c$, so must have corresponding accepting state $s$ with $c \in Y(s)$. Since $\mathcal{S}$ is diagnosable, there cannot be a $c'$where $c$ and $c'$ are not distinguishable, by definition of diagnosability. So, there cannot be some common subtrace $\lambda$ that maps to an accepting state that has both $c'$ and $c$. So, $\mathcal{D}^*_{F,M,Q}$ uniquely isolates all $c$. Assume $\mathcal{D}^*_{F,M,Q}$ uniquely isolates all $c$. Then each possible fault trace $\lambda$ has an accepting state $s$ where $c \in Y(s)$. Thus, there cannot be some $c'$, with trace $\lambda'$ that reaches the same state, otherwise $c'$ is in $Y(s)$. Therefore, $c$ and $c'$ are distinguishable, so $\mathcal{S}$ is diagnosable. Thus $\mathcal{S}$ is diagnosable if and only if $\mathcal{D}^*_{F,M,Q}$ uniquely isolates all $c$. ■