# Combination of Simulation and State Observers for Consistency-based Diagnosis

**Anibal Bregon [1], Belarmino Pulido [1], and Carlos Alonso-González [1]**

[1] *Department of Computer Science, University of Valladolid, Valladolid, 47011, Spain*
*{anibal,belar,calonso}@infor.uva.es*

## ABSTRACT

Consistency-based diagnosis of dynamic systems using possible conflicts rely upon a semi-closed loop simulation of numerical models. Simulation approaches need to know the initial state, which is a nontrivial requirement in real-world systems. Prognosis approaches also require techniques for predicting the future system states under nominal and faulty conditions.

This work proposes to integrate state observers to estimate initial states for simulation within the consistency-based diagnosis framework using possible conflicts. This work extends the BRIDGE framework for one class of dynamic systems, using the possible conflict concept to find every subsystem with necessary structural redundancy to lead to a minimal conflict activation. These algorithms can analyze those structures, without additional information, and point out possible implementations as observers or simulators.

This proposal has been tested on a simulation scenario. Results and comparison with similar existing hybrid -DX + FDI- approaches are provided.

## 1 INTRODUCTION

The highly increasing development of complex technological systems and devices, together with their highly demanding requirements on safety and reliability, have turned the fault detection, diagnosis, and prognosis mechanisms into a key step within system performance. Among all these mechanisms, model-based diagnosis (MBD) approaches are quite prevalent nowadays, due to fact that they have the potential to overcome the device dependency problem, which greatly increases the cost of developing and deploying diagnosis systems.

Traditionally, two different research communities have tackled the problem of Model-Based Diagnosis: the Control Engineering community, known as FDI (Gertler, 1998; Blanke *et al.*, 2006); and the Artificial Intelligence community, known as DX (Hamscher *et al.*, 1992). The FDI community uses control and statistical decision theories to carry out the fault detection and isolation stages, where the major concern is fault detection robustness. The field has solid theoretical results for linear systems (Gertler, 1998; Blanke *et al.*, 2006), being analysis of nonlinear systems a major research issue. On the other side, the DX community has a solid theoretical foundation for static systems, with fault localization and identification being its main research issues. Consistency-based diagnosis (CBD) is the most used approach, and the General Diagnosis Engine (GDE) is its computational paradigm (Hamscher *et al.*, 1992).

Both DX and FDI communities have developed their own tools and techniques in a parallel way. Recently, the BRIDGE community (Cordier *et al.*, 2004) established a common framework for sharing results and techniques. Such framework is based on the comparison between consistency-based diagnosis via conflicts (Hamscher *et al.*, 1992) and fault detection and isolation via analytical redundancy relations (ARRs) obtained through structural analysis for static systems (Blanke *et al.*, 2006).

Our work is integrated within such BRIDGE framework, based on the proposal by Cordier et al.(Cordier *et al.*, 2004), but extending the comparison to dynamic systems, and combining techniques from both communities to improve the overall diagnosis process.

Our work is based on the Possible Conflicts approach (Pulido and Alonso-González, 2004), PCs for short, an off-line dependency compilation technique from the DX community. Consistency-based diagnosis using PCs is based on iterative on-line simulation of subsystems, PCs, that can become conflicts if a discrepancy is found between estimated and observed behaviour. The approach proceeds in semi-closed loop: hence it needs to reset the value of state variables at the begining of each simulation period. The goal of this work is to improve the robustness of the method through a more precise estimation of the initial state, without modifying its fault isolation capabilities, and its consistency-based approach.

Based on the similarities between PCs and ARRs in the BRIDGE framework (Pulido and Alonso-González, 2004), this work uses PCs to design state observers, which are used to estimate the initial states for simulation. The main contribution of this paper is a novel way to derive the structure of the set of minimal state observers for a system, and how to combine them with an existing CBD approach based on PCs, to increase fault detection robustness, while retaining its fault localization/isolation capabilities.

The paper is organized as follows. Section 2 introduces assumptions, techniques, and working principles used for each community to deal with static information in BRIDGE. Section 3 extends the approach for dynamic systems. Based on the similarities, a new way to derive the structure of state observers for nonlinear dynamic systems using possible conflicts is introduced in section 4. Later on, the integration of these state observers into the PCs approach is introduced, together with results on a simulation plant. Finally, section 5 provides conclusions and discussion with related works.

## 2 POSSIBLE CONFLICTS, ARRS, AND CONFLICTS IN THE BRIDGE FRAMEWORK

Possible Conflicts are those sub-systems capable to become conflicts in CBD, i.e. *minimal subsets of equations containing necessary analytical redundancy to perform fault diagnosis* (Pulido and Alonso-González, 2004). In GDE the set of conflicts are computed online (de Kleer and Williams, 1987). However, many approaches have opted recently for off-line computation of conflict-like structures. These are termed *dependency-compilation techniques*.

The main idea behind the *Possible Conflict* concept is that the set of subsystems capable to generate a conflict can be calculated off-line[1]. They can be computed through the analysis of the set of equations in the system model. The computation process is done in three steps.

First step generates an abstract representation of the system, as a hypergraph. In this representation there is just qualitative information about constraints in the models, and their relationship to known and unknown variables in such models.

Second step looks for minimal over-constrained sets of relations, which are essential for model-based diagnosis. These subsystems, called *Minimal Evaluation Chains*, or MECs for short, represent a necessary condition for a conflict to exist.

Each MEC, which is a partial sub-hypergraph of the original system description, needs to be solved using only local propagation criteria (to follow GDE computational framework). Since there is no such information in the hypergraph, in the third step, extra knowledge is added to fulfill that requirement. Each possible way a constraint can be solved, by means of local propagation, is specified[2]. As a consequence, each minimal evaluation chain generates a directed and-or graph. In each and-or graph, a search for every possible way the system can be solved using local propagation, is conducted. Each possible way is called a *Minimal Evaluation Model*, or MEM for short, and it can predict the behavior of a part of the whole system. Each MEM represents a globally consistent causal assignment for the MEC.

Because conflicts will arise only when models are evaluated with available observations, the set of constraints in a MEC containing at least one MEM is called a *possible conflict*.

Each MEM describes an executable model, which can be used to perform fault detection. If there is a discrepancy between predictions from those models and current observations, the possible conflict would be responsible for such a discrepancy and should be confirmed as a real conflict. Afterwards, diagnosis candidates are obtained from conflicts following Reiter's theory.

Possible conflicts use minimality criteria in terms of sets of constraints. Nevertheless, it is straightforward to obtain candidates based on components.

Detailed information concerning PC calculation can be found in (Pulido and Alonso-González, 2004), and their relation to other algorithms for computing minimal ARRs and Minimally Overdetermined Sets has been recently established (Armengol *et al.*, 2009).

Due to space limitations we introduce a brief summary of the similarities and equivalences between PCs, ARRs, and conflicts described in (Pulido and Alonso-González, 2004; Cordier *et al.*, 2004).

Using minimality criterion w.r.t. set of constraints in the model, the whole set of MEMs related to the set of PCs is equivalent to the set of minimal conflicts computed by the GDE.

Moreover, if algorithms computing ARRs through structural analysis use such minimality criterion and provide a complete solution –explores every possible causal assignment for each ARR–, the set of PCs is equivalent to the set of minimal ARRs.

Finally, if every MEM in every PC provides the same solution[3], then PCs and minimal conflicts have identical fault detection and isolation capabilities.

(Cordier *et al.*, 2004) introduced the concept of *support* for an ARR (set of components whose models are used to derive an ARR). Based on such idea, off-line compiled conflicts and ARR's support can be considered as equivalent (the support for an ARR is a potential conflict, which is equivalent to a possible conflict (Cordier *et al.*, 2004; Pulido and Alonso-González, 2004)). If the equivalence assumption is fulfilled, the set of ARRs and the set of minimal conflicts will have same detection and isolation capabilities.

Summarizing, we can use possible conflicts as equivalent to potential conflicts and the support for minimal ARRs for fault detection and isolation purposes.

The BRIDGE framework was defined for static systems. This work provides a specific extension for a class of dynamic systems. First, the influence of temporal information in PCs and ARRs calculation is analyzed.

[1] The DX concept of conflict detection is similar to the FDI concept of residual activation.

[2] In this sense, we assume the set of causal assignments for each constraint are known.

[3] This concept is called the *Equivalence assumption* in (Pulido and Alonso-González, 2004).

## 3 DEALING WITH TEMPORAL INFORMATION IN SYSTEM DESCRIPTION

Both DX and FDI communities have provided different approaches for dynamic systems modelling.

There is no general theory for CBD of dynamic systems (de Kleer, 2003). Therefore, we build our discussion at the structural level, using previous works from the DX community (Dressler, 1996; Chantler *et al.*, 1996). In DX, inclusion of temporal information gives rise to two kind of constraints[4]:

- *instantaneous* constraints; for instance $(ode_i\ \dot{x}\ \ v_1\ v_2\ \ldots)$ as in $\dot{m} = \sum inflow - \sum outflow$, and,
- *differential* constraints; for instance $(eq_i\ x,\ \dot{x})$, as in $\dot{x} = \frac{dx}{dt}$

The FDI approach, using mainly numerical analytical models, has more standard specifications for dynamic aspects (Blanke *et al.*, 2006), both in continuous and discrete time. The mathematical model can be expressed in a variety of ways: state-space models, input-output models, or even black-box models obtained through identification. We focus our discussion on residual generation via ARRs obtained through structural analysis (Blanke *et al.*, 2006).

Inclusion of temporal information in ARRs can be done in three different ways (Dustegör *et al.*, 2006). Difference between them comes from the way the relationship between a state variable, $x$, and its derivative, $\dot{x}$, is stated[5]. These three methods have been compared in different works (Dustegör *et al.*, 2006; Krysander *et al.*, 2008). The approach which considers $x$ and $\dot{x}$ as different variables, and linked by the constraint $(eq_i\ x,\ \dot{x})$, is equivalent to the DX approach. We shall focus our discussion in these kind of models, and ARR calculation algorithms known as the Lille method (Blanke *et al.*, 2006).

Differential constraints can be used for behavior estimation in two ways, depending on the causal interpretation of the constraint:

- the derivative approach ($\dot{x}(t) = \frac{dx}{dt}$ assumes the derivative can be computed based on present and past samples of $x$), and
- the integral approach ($x(t) = x(t-1) + \int_{t-1}^{t} \dot{x} \cdot dt$, assume the initial state $x(0)$ is known).

It has been demonstrated that both approaches have equivalent behavior estimation capabilities for numerical models (Chantler *et al.*, 1996), assuming adequate sampling rates and precise approximations are available, or assuming initial conditions are known.

Algorithms computing PCs or computing ARRs as in the Lille method can use differential constraints as expressed above, and both methods can include both types of causal interpretations, i.e. allow different causal assignments to solve a MEC or an ARR (Pulido and Alonso-González, 2004; Blanke *et al.*, 2006). Using integral or derivative matchings for each differential constraint provides different computable models for each MEM and ARR, i.e. different causal matchings, but impose no restriction in the way MECs and ARRs are computed (see (Pulido *et al.*, 2007) for more details on this comparison).

We will illustrate these concepts on the system shown in figure 1. The system is made up of a water tank, $T$, a valve, $V_o$, a controller, $PI$, a level transducer, $LT$, and two flow transducers, $FT_1$ and $FT_2$. The aim of the system is to keep the level of the tank, $h$, which is measured, $h_{mes}$, as close as possible to the desired level, $h_{ref}$. To do so, the controller acts through $u_c$ over the valve, $V_o$. The level of the tank, $h_{mes}$, the input flow, $Q_{i_{mes}}$, and the output flow, $Q_{o_{mes}}$, are known.

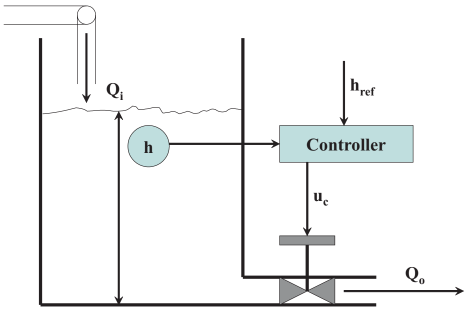


Figure 1: Our system is made up of tank, a valve and one controller.

The model of the system is the following[6]:

| Equation | Component (Support) |
|---|---|
| $e_1 : A\dot{h} = Q_i - Q_o$ | T |
| $e_2 : h(t) = h(t-1) + \int_{t-1}^{t} \dot{h} \cdot dt$ | |
| $e_3 : Q_o = K_c u_c \sqrt{h}$ | $V_o$ |
| $e_4 : u_c = f(h)$ | PI |
| $e_5 : h_{mes} = h$ | LT |
| $e_6 : Q_{o_{mes}} = Q_o$ | $FT_1$ |
| $e_7 : Q_{i_{mes}} = Q_i$ | $FT_2$ |

Moreover, in the plant we have considered the set of fault modes shown in table 1.

| *Class* | *Component* | *Description* |
|---|---|---|
| $f_T$ | $T$ | *Leakage in tank* |
| $f_{V_0}$ | $V$ | *Valve constant failure* |
| $f_{LT}$ | $h_{obs}$ | *Level sensor failure* |
| $f_{FT_1}$ | $Q_{o_{obs}}$ | *Output flow sensor failure* |

Table 1: Fault modes considered.

PCs in the plant can be obtained using both integral or derivative causality. If the integral approach is used, we obtain 4 Possible Conflicts (see upper part in table

[4]Assuming the model description is in canonical form, i.e., it is made up of a set of first-order ODEs. It is known that n$^{th}$-order ODEs can be transformed into n first-order ODEs.

[5]Also called differential constraint in FDI terminology.

[6]Where $A$ is the cross section of the tank, and $K_c$ is the valve constant. $e_2$ is a differential constraint.

2, which are minimal w.r.t. the set of constraints in their models. PCs related to those fault modes can be seen in the Theoretical Fault Signature Matrix (upper part in table 3. If a derivate approach is used, the set of possible conflicts found is different. The lower part in table 2 shows the set of 3 PCs found using derivative causality, while the lower part in table 3 shows its corresponding Theoretical Fault Signature Matrix.

| | Equations | Component (Support) | Estimate |
|---|---|---|---|
| | Using integral causality | | |
| $PC_1$ | $e_1, e_2, e_3, e_4, e_5, e_7$ | $T, V_o, PI, FT_2, LT$ | h |
| $PC_2$ | $e_1, e_2, e_3, e_4, e_6, e_7$ | $T, V_o, PI, FT_1, FT_2$ | $Q_o$ |
| $PC_3$ | $e_1, e_2, e_5, e_6, e_7$ | $T, FT_1, FT_2, LT$ | h |
| $PC_4$ | $e_3, e_4, e_5, e_6$ | $V_o, PI, FT_1, LT$ | $Q_o$ |
| | Using derivative causality | | |
| $PC_1$ | $e_1, e_2, e_3, e_4, e_5, e_7$ | $T, V_o, PI, FT_2, LT$ | h |
| $PC_2$ | $e_1, e_2, e_5, e_6, e_7$ | $T, FT_1, FT_2, LT$ | h |
| $PC_3$ | $e_3, e_4, e_5, e_6$ | $V_o, PI, FT_1, LT$ | $Q_o$ |

Table 2: PCs found for the plant using integral and derivative causalities.

| | $f_T$ | $f_{V_o}$ | $f_{LT}$ | $f_{FT_1}$ |
|---|---|---|---|---|
| | Using integral causality | | | |
| $PC_1$ | 1 | 1 | 1 | |
| $PC_2$ | 1 | 1 | | 1 |
| $PC_3$ | 1 | | 1 | 1 |
| $PC_4$ | | 1 | 1 | 1 |
| | Using derivative causality | | | |
| $PC_1$ | 1 | 1 | 1 | |
| $PC_2$ | 1 | | 1 | 1 |
| $PC_3$ | | 1 | 1 | 1 |

Table 3: PCs and their related fault modes using integral and derivative causalities.

One final issue must be addressed. The presence of cycles can halt local propagation for static systems (Katsillis and Chantler, 1997) while using GDE. Then, an inference engine capable of solving algebraic loops is needed. For off-line dependency-recording this step can be done off-line (Pulido and Alonso-González, 2004; Blanke *et al.*, 2006).

To solve cyclical structures both algorithms, PCs calculation and ARR computation using the Lille method, follow identical approaches, coming respectively from DX (Dressler, 1996; Katsillis and Chantler, 1997), and FDI (Blanke *et al.*, 2006) communities. Cycles containing both instantaneous and differential constraints must be studied:

- using integral causality, loops containing differential constraints are not loops (Dressler, 1996), but spirals, because $x$ and $\dot{x}$ have different temporal indices;
- using derivative causality, no loop including $x$ and $\dot{x}$ can be solved. Hence, these loops in ARRs or MEMs must be rejected.

These concepts will be illustrated using the case study in figure 1. Figure 2 shows one MEM for $PC_1$ using integral causality. In the causal graph it can be seen that $h(t+1)$ and $h'(t)$ are related through a differential constraint (dashed arc), but there is no loop because we are using an integral approach.

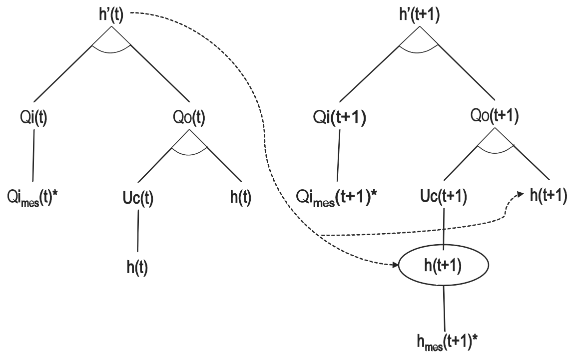


Figure 2: Possible conflict for the tank system with integral approach. A discrepancy can be found between $h(t+1)$ estimation and $h_{obs}(t+1)$ (the discrepancy node is graphically represented by a ellipse). If there is no discrepancy, $h(t+1)$ is used for next simulation period.

Figure 3 shows one MEM for $PC_1$ using derivative causality. In the causal graph it can be seen that $h_{mes}(t)$ and $\dot{h}(t)$ are related through a differential constraint representing the derivative of $h$ using its measured value.

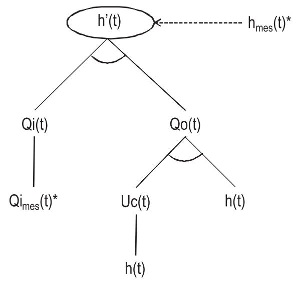


Figure 3: Possible conflict for the tank system with derivative approach. A discrepancy can be found for $\dot{h}(t)$ using measurement of level $LT1$, $h_{mes}$, or direct estimation of the derivative from previous values for $LT1$ (the discrepancy node is graphically represented by an ellipse).

Summarizing, both FDI by means of ARRs and DX allow derivative and integral approaches. While DX approaches have opted by simulation techniques – relying mainly in qualitative models–, traditionally the FDI community has opted for numerical models, and has rejected simulation; most FDI methods rely upon derivative estimation, which has instead problems related with noise, disturbances, and parameter uncertainty.

Similarities between PCs and ARRs for static systems can be straightforward extended for dynamic systems using that representation for differential constraints. It seems clear that possible conflict calculation and ARR calculation are equivalent from an structural point of view w.r.t. temporal information, when faced with models in canonical form.

### 3.1 PCs and ARRs for dynamic systems

**Proposition 1** *Given equivalent system descriptions including instantaneous and differential constraints for PCs and ARR calculation, the sets of computed PCs and minimal ARRs have same isolation capabilities.*

**Proof:** The set of MEMs related to the set of PCs is complete because it is based on a complete search of the set of minimal conflicts. PCs and minimal ARR computation are equivalente approaches for static systems, as previously demonstrated (Pulido and Alonso-González, 2004), if algorithms used for ARR computation provided a complete set of minimal ARRs. Moreover, only differential constraints have been added, and they are used for propagation the same way instantaneous constraints were used. Finally, both approaches remove illegal cyclical structures. Therefore, both will provide same results in terms of fault isolability.

The main difference between both approaches comes in practice. It is based upton the use of integral or derivative approaches: simulation of MEMs in the PC approach, and using ARR for estimation in FDI.

Previous results can be extended to minimal conflicts, ARRs, and the set of MEMs provided by the set of PCs, if a GDE-like inference engine is able to handle loops.

This proposition can not be automatically extended for fault detection capabilities. In the case of nonlinear models, different MEMs for one MEC or different ARRs for the same support can lead to different estimated values (Pulido and Alonso-González, 2004), hence they can provide different detection results.

## 4 DESIGN AND INTEGRATION OF STATE OBSERVERS USING PCS

State observers provide an estimation for state variables based on input and output measurements. Hence, our guess is that this technique can be used to estimate initial values of state variables for Possible Conflicts simulation. We propose to integrate state observers while retaining the consistency-based diagnosis approach for fault isolation.

Simulation, estimation, and state observers are equivalent for linear models (Gertler, 1998). In fact, parity and observer-based approaches provide residuals with similar structures: "both of them use same measurable input and output signals, assumed their structure and parameters are known, and they do not change…They differ in the way the input and output measurements are filtered"((Isermann, 2006), and references therein).

Works by Staroswiecki et al. also show that state observers structure can be derived using structural analysis, just introducing several equations to guarantee the observer feasibility (Blanke *et al.*, 2006). Our claim is that algorithms computing Possible Conflicts can automatically derive structural models that can be implemented as simulators or state observers, without introducing additional constraints in the structural model.

### 4.1 PCs for State Observer Design

A Possible Conflict can have one or more Minimal Evaluation Models, MEM. Each MEM, as can be seen in figure 2, provides an and-or graph showing how the equations should be solved (algorithms used to build PCs use local propagation in DX terminology, equivalent to variable elimination in FDI). The structural model related to each MEM can be implemented as a simulation or a estimation model. In figure 2 the estimated, $h$, and measured variable, $h_{mes}$, are the basis for a conflict (or activated residual for $e = h - h_{mes}$). Additionally, that MEM has an associated state-space model for simulation which can be seen as a generic non-linear model:

$$\dot{\hat{x}} = f(\hat{x}, u) \quad (1)$$

$$\hat{y} = g(\hat{x}, u) \quad (2)$$

$f$ and $g$ can be seen as the result of relation composition linked to causal arcs in the MEM, since each arc can be seen as a relation. Our proposal is to find causal models that can be implemented as state observers, with an associated state space model such as:

$$\dot{\hat{x}} = f(\hat{x}, u) + k(y - \hat{y}) \quad (3)$$

$$\hat{y} = g(\hat{x}, u) \quad (4)$$

where $x$ is the state variable, $u$ is the input, and $y$ is the output. $\hat{x}$ and $\hat{y}$ are the estimated state and output variables, respectively. $k$ is a linear or non-linear function, which filters out the difference between the estimated and measured variable, minimizing the error, ($e = y - \hat{y}$), regarding a given criterion. This step is independent of the type of observer –Luenberger, EKF, etc–.

As previously mentioned, the set of PCs is computed following the GDE approach for dependency-recording. Hence, each PC has at least one MEM which has exactly one discrepancy node in its associated graph[7]. The discrepancy node in a MEM can be found in two ways: as the estimation for a measured magnitude, or as the double estimation for a non-measured magnitude. The difference between those two values for the same magnitude is the discrepancy in DX terminology, or the error, $e$, in FDI terminology, i.e. the value of a residual.

Algorithms computing the set of PCs (Pulido and Alonso-González, 2004) can be enhanced to compute the structure of state observers, without introducing additional constraints in model description.

**Proposition 2** *Those MEMs containing a state variable can provide the minimal structural description for a state observer, if there exists a path made only of instantaneous arcs from the observed variable to the estimated state variable.*

This result comes from the way the error $e$ is introduced in the generalized state observer scheme in equation 3:

$$e = y - \hat{y} = y - g(\hat{x}, u) \quad (5)$$

[7]Discrepancy nodes are represented as circled nodes in figures 2, 3, and 4.

Being $y$ the observed variable and $\hat{y}$ the direct estimation from state variables and inputs, using only instantaneous constraints: $g(\hat{x}, u)$. That is, $e = y - \hat{y}$ represents a discrepancy or residual. This error is implicitly present in the MEM.

One MEM can only have one discrepancy node, because they represent minimal structural overdetermined sets of equations. In MEM computation, state variable $\hat{x}(t+1)$ is estimated using integration from $\dot{\hat{x}}(t)$ and $\hat{x}(t)$. Afterwards, $\hat{x}(t+1)$ can be used to estimate $\hat{y}(t+1)$ (using only instantaneous constraints, $g(\hat{x}, u)$, as seen in equation 2). If $e$ can be computed as $(y - \hat{y})$ in the MEM, then it can be used in a state observer to correct the estimation of the state variable in the next time step using gain $k$.

Algorithms computing PCs are exhaustive, then they are capable to provide every estimation for a state variable. Hence, it is possible to trace backward in the and-or graph related to each MEM if there is a path made only of instantaneous constraints from the discrepancy node ($e$) to the state variable.

No additional information is required in the model description to provide these results. Just one step is necessary at the end of the algorithms used to analyze cyclical structures in the MEMs (their and-or graphs) related to a Possible Conflict (Pulido *et al.*, 2007). This is an advantage against other approaches, which add explicit equations to derive the structure for the observer (Blanke *et al.*, 2006). Moreover, these observers do not add new structural information. Hence, they have the same isolation capabilities.

Figure 4 shows the state observer which can be obtained from the MEM related to $PC_1$ in figure 2 using integral causality. Estimated variable, $h$, and measured variable, $h_{mes}$, are used to compute error, $e = h - h_{mes}$, which is used as the correction for the next state estimation.

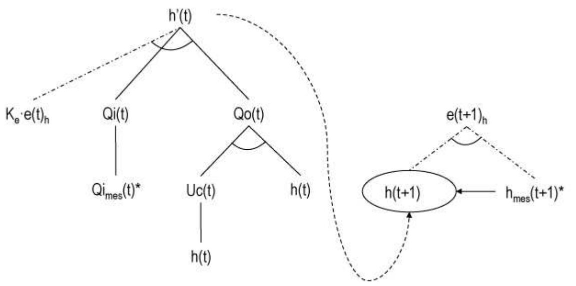


Figure 4: Possible conflict $PC_1$ implemented as a state observer.

This work is focused on providing the collection of minimal expressions for state observers according to equations 3 and 4. No restriction is imposed in how the state observer is implemented. A nonlinear state observer can be devised, for instance, using a linearized model plus a Luenberger observer, or an Extended Kalman Filter. Afterwards, it is necessary to analyze these equations for convergence or robustness issues.

One question remains at this point. Given that state observers are used as a fault detection mechanism: is it still necessary to run the simulation of the PC?

## 4.2 Integration proposal: increasing robustness with state observers

PCs are based on a semi-closed loop simulation approach, over a simulation interval, $\Delta t$. After each $\Delta t$ period has elapsed, measured and estimated values are compared giving a dissimilarity value[8]. This approach provides customizable detection capabilities, being less sensitive to noise in measurements. Semi-closed loop simulation also avoids the effect of small model/parameter disturbances by introducing, iteratively, observations for initial conditions when the simulation interval, $\Delta t$, has elapsed. The main drawback is that these initial conditions are not always known.

On the other hand, state observers are able to generate a state-variable estimation, without fault, with noise in sensors and small parameter disturbances. The state-variable estimation can be used for fault detection purposes. Main drawbacks are small persistence for activated residuals, small activation time (i.e. sensitivity to noise in the measurements), and small (and incipient) fault masking. But, the estimated state-variable can provide the initial condition required for the PCs simulation.

Our proposal is to integrate state observers within the CBD framework with PCs, where observers will improve the estimations for the initial conditions of the PCs without fault, and they will not interfere with the behavior of the PCs in faulty situations.

The resulting integration scheme can be seen in figure 5. Running both MEMs in parallel (PCs and state observers), and assuming there is no fault detection, the state estimation given by the state observers can be used as initial condition for the PCs simulation. For the fault detection decision step, both MEMs can be used. A multiple activation by the state observers residuals leads the diagnosis system to a faster detection for abrupt faults (using only PCs, abrupt faults can not be immediately detected, but after the consistency check at the end of the simulation interval $\Delta t$). Additionally, for incipient faults, the PCs will be able to lead the diagnosis system to a fault detection, while the state observers alone, due to the correction factor, will not detect this kind of faults (or, in the best case, the fault will not be detected during the first stages of the fault). This simple integration scheme shows the power of this proposal. The decision step in fault detection can be tuned (changing the detection thresholds for PCs and state observers) giving more weight to the speed or the absence of false alarms, level on noise or parameter uncertainty, and so on.

This integration proposal was tested, running several experiments in the tank system (figure 1). The study was done on a data-set made up of several examples obtained from several simulations for each fault mode in the plant. Models and simulations were developed using Matlab®. In these simulations we introduced noise in the measurements (5%), and model uncertainties (5%). Each simulation lasted 1000 seconds, and contained several changes in the reference level of the tank. The data sampling was 10 data per

[8]In the case of PCs we use Dynamic Time Warping as dissimilarity measurement. See (Keogh and Ratanamahatana, 2005) for further details.

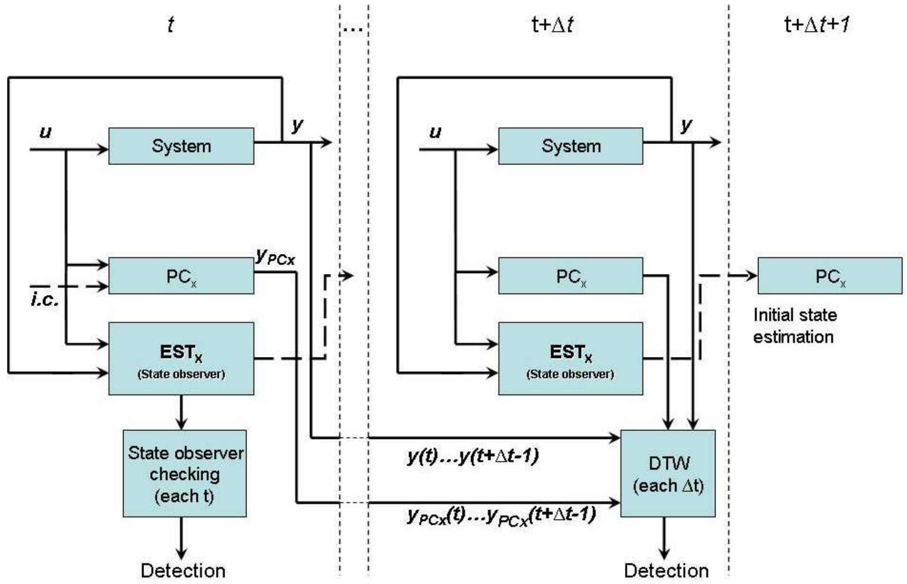


Figure 5: Integration scheme: estimation given by state observers, $EST_x$, can be used directly for fault detection, or can be used as initial states estimation for the next simulation period in the same MEM implemented as a simulator, $PC_x$.

second.

In the system, using Proposition 2, 3 out of 4 PCs can be implemented also as state observers. The remaining PC can not be implemented because it just contains instantaneous relations. The state observers were implemented as Extended Kalman Filters. Values of covariance matrices were determined empirically.

Table 4 shows the results obtained in nominal situation for simulation interval $\Delta t = 30$ and $\Delta t = 60$ seconds. The table shows generalized reduction in the percentage for the mean and maximum values of the residual activation values while using the integration proposal (PCs + state observers) against the residuals obtained using PCs alone. This fact allows the diagnosis system to be more sensitive to small faults. Regarding the maximum values, the decrease is even bigger. This fact produce a huge reduction in the number of false positives due to better initial conditions estimation.

| | $PC_1 + OBS_1$ | | $PC_2 + OBS_2$ | | $PC_3 + OBS_3$ | |
|---|---|---|---|---|---|---|
| | Med. | Max. | Med. | Max. | Med. | Max. |
| $\Delta t = 30$ | 9.16 | 32.92 | 49.56 | 58.48 | 22.93 | 52.42 |
| $\Delta t = 60$ | 5.82 | 23.78 | 40.15 | 55.99 | 21.54 | 51.82 |

Table 4: Mean and maximum residual values decrease combining PCs and State observers.

For faulty situations, we randomly generated fault magnitudes at different time instants within the interval $[420, 450]$. In table 5, detection times for different examples of a pipe blockage, are shown. Different fault magnitudes (10% and 30%) and times of fault occurrence (inside an interval $\Delta t$ = 60 seconds), have been considered. When a small fault occurs in the system (see upper part in table 5), both state observers and PCs without the initial estimation given by the observers, are unable, in most cases, to detect the fault. Nevertheless, the new approach integrating both techniques, due to the decrease in the detection threshold, is able to detect this kind of small faults (see $PC_1 + OBS_1$ and $PC_2 + OBS_2$). When bigger faults are considered (see lower part in table 5), fault detection time with the new approach will be equal to the fault detection time achieved with the state observers, that is, faster than the fault detection time if we consider PCs alone.

| Fault occurs a t = | 420 | 425 | 430 | 435 | 440 | 445 |
|---|---|---|---|---|---|---|
| | Fault magnitude = 10% | | | | | |
| $PC_1$ | fn | fn | fn | 540 | fn | fn |
| $OBS_1$ | fp | fp | fp | fp | fp | fp |
| $PC_1 + OBS_1$ | 540 | 540 | 480 | 540 | 540 | 540 |
| $PC_2$ | fn | fn | fn | fn | fp | 540 |
| $OBS_2$ | fn | fn | fn | fn | fn | fn |
| $PC_2 + OBS_2$ | 480 | 540 | 540 | fp | 540 | 480 |
| | Fault magnitude = 30% | | | | | |
| $PC_1$ | 480 | 480 | 480 | 540 | 540 | 540 |
| $OBS_1$ | 437 | 447 | 452 | fp | 457 | fp |
| $PC_1 + OBS_1$ | 437 | 447 | 452 | 480 | 457 | 480 |
| $PC_2$ | 540 | 480 | 540 | 540 | 480 | 480 |
| $OBS_2$ | 422 | 427 | 432 | 436 | 442 | 446 |
| $PC_2 + OBS_2$ | 422 | 427 | 432 | 436 | 442 | 446 |

Table 5: Detection times for several examples of pipe blockage. $fp$ is a false positive. $fn$ is a false negative. $PC_3$ is not affected by the fault.

Similar experiments with different kind of faults (blockages, leakages, and sensor faults) were considered. For all those experiments obtained results were similar to the ones shown here: capacity to detect small fault, faster detection times for big faults, as well as a reduction in the number of false positives.

## 5 CONCLUSIONS AND RELATED WORK

PCs and ARRs exhibit identical fault isolation capabilities on dynamic systems described as a set of first-order ODE equations, thus extending the BRIDGE framework for that kind of dynamic systems, and it opens new ways for the integration of techniques from DX and FDI fields. The comparison has been made with the Lille method for ARR computation (Blanke *et al.*, 2006). Additional research is required to other existing algorithms for ARR calculation and including temporal information (Dustegör *et al.*, 2006; Krysander *et al.*, 2008).

Our proposal uses algorithms for computing PCs as a tool for state observer design. These algorithms can provide the structure of MEMs which can be implemented as state observers without including additional constraints in the model (Blanke *et al.*, 2006). But our algorithms do not provide an automatic synthesis based on the structure, as in (Christophe *et al.*, 2004), and (Krysander *et al.*, 2008).

Different MEMs for a conflict can provide different detection capabilities for nonlinear systems. Our work proposes a simple integration of two expressions for the same MEM in a PC, if possible: use an state observer for initial state estimation, then use the estimation for a semi-closed loop simulation. The observer type and associated decision logic can be tailored for each system, to get desired detection or false alarm rates.

Results on simulation provided better results with the integration of both methods than using both of them alone. We are now testing this approach in more complex real systems, with unknown noise and uncertainties. Afterwards, the approach could be integrated within System Health Management methodologies, where robust fault detection is of paramount importance.

Other works have previously integrated state observers for fault detection and some kind of causal structures for fault isolation, for instance TRANSCEND (Mosterman and Biswas, 1999). Main difference in our approach is that we use the same structure, PC, for the fault detection and isolation stages, and it is computed off-line (TRANSCEND uses state observers for fault detection, and on-line forward and backward propagation in a temporal causal graph for fault isolation).

## ACKNOWLEDGMENTS

This work has been partially supported by Spanish regional grant JCyL VA005B09.

## REFERENCES

(Armengol *et al.*, 2009) J. Armengol, A. Bregon, T. Escobet, E. Gelso, M. Krysander, M. Nyberg, X. Olive, B. Pulido, and L. Travé-Massuyès. Minimal structurally overdetermined sets for residual generation: A comparison of alternative approaches. In *Procs. of IFAC-Safeprocess'09*, Barcelone, Spain, 2009.

(Blanke *et al.*, 2006) M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki. *Diagnosis and Fault-Tolerant Control*. Springer, 2nd edition, 2006.

(Chantler *et al.*, 1996) M.J. Chantler, T. Daus, S. Vikatos, and G.M. Coghill. The use of quantitative dynamic models and dependency recording engines. In *Procs. of DX'96*, pages 59–68, Val Morin, Canada, 1996.

(Christophe *et al.*, 2004) C. Christophe, V. Cocquempot, and B. Jiang. Link between high-gain observer-based and parity space residuals for fdi. *Trans. on the Institute of Measurement and Control*, 26(325), 2004.

(Cordier *et al.*, 2004) M.O. Cordier, P. Dague, F. Lévy, J. Montmain, M. Staroswiecki, and L. Travé-Massuyès. Conflicts versus analytical redundancy relations: a comparative analysis of the model-based diagnosis approach from the artificial intelligence and automatic control perspectives. *IEEE Trans. Syst. Man Cy. B.*, 34(5):2163–2177, 2004.

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

(de Kleer, 2003) J. de Kleer. Fundamentals of model-based diagnosis. Procs. of DX'03, june 2003.

(Dressler, 1996) O. Dressler. On-line diagnosis and monitoring of dynamic systems based on qualitative models and dependency-recording diagnosis engines. In *Procs. of ECAI'96*, pages 461–465, 1996.

(Dustegör *et al.*, 2006) D. Dustegör, E. Frisk, V. Coquempot, M. Krysander, and M. Staroswiecki. Structural analysis of fault isolability in the DAMADICS benchmark. *Control Engineering Practice*, 14(6):597–608, 2006.

(Gertler, 1998) J.J. Gertler. *Fault detection and diagnosis in Engineering Systems*. Marcel Dekker, Inc., Basel, 1998.

(Hamscher *et al.*, 1992) W. Hamscher, L. Console, and J. de Kleer(Eds.). *Readings in Model based Diagnosis*. Morgan-Kaufmann Pub., San Mateo, 1992.

(Isermann, 2006) R. Isermann. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer-Verlag, 2006.

(Katsillis and Chantler, 1997) G. Katsillis and M.J. Chantler. Can dependency-based diagnosis cope with simultaneous equations? In *Procs. of DX'97*, pages 51–59, Le Mont Saint Michel, France, 1997.

(Keogh and Ratanamahatana, 2005) E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3):358–386, 2005.

(Krysander *et al.*, 2008) M. Krysander, J. Åslund, and M. Nyberg. An efficient algorithm for finding minimal over-constrained sub-systems for model-based diagnosis. *IEEE Trans. Syst. Man Cy. A.*, 38(1), 2008.

(Mosterman and Biswas, 1999) P. Mosterman and G. Biswas. Diagnosis of continuous valued systems in transient operating regions. *IEEE Trans. Syst. Man Cy.*, 29(6):554–565, 1999.

(Pulido and Alonso-González, 2004) B. Pulido and C. Alonso-González. Possible conflicts: a compilation technique for consistency-based diagnosis. *IEEE Trans. Syst. Man Cy. B., Special Issue on Diagnosis of Complex Systems*, 34(5):2192–2206, 2004.

(Pulido *et al.*, 2007) B. Pulido, C. Alonso, A. Bregón, V. Puig, and T. Escobet. Analyzing the influence of temporal constraints in possible conflicts calculation for model-based diagnosis. In *Procs. of DX'07*, Nashville, TN, USA, 2007.

## AUTHOR BIOGRAPHIES

**ANIBAL BREGON** received his B.S. and M.S. degrees in computer science from the University of Valladolid, Valladolid, Spain, in 2005 and 2007, respectively. Currently he is Graduate Research Assistant at the Department of Computer Science from the University of Valladolid, working towards his PhD. His main research interests are model-based reasoning for diagnosis, prognosis, and health-management.

**BELARMINO PULIDO** received his B.S., M.S. and Ph.D. degrees in computer science from the University of Valladolid, Valladolid, Spain, in 1992, 1995 and 2001, respectively. He has been working since 1994 with the Department of Computer Science from the University of Valladolid, where he is Associate Professor since 2002. His main research interests are model-based and knowledge-based reasoning, applied to diagnosis, prognosis, and health-management.

**CARLOS ALONSO-GONZÁLEZ** got his degree in Sciences in 1985, and his PhD in Physics in 1990, from the University of Valladolid, Valladolid, Spain. Currently he is Associate Professor with the Department of Computer Science from the University of Valladolid. He has being working in different national funded projects related to Supervision and Diagnosis of continuous industrial environments. His main research interests are Knowledge-based systems for Supervision and Diagnosis of dynamic systems, model-based diagnosis, Knowledge Engineering and Machine Learning.