ROOT CAUSE ANALYSIS OF OSCILLATING CONTROL LOOPS

R. Srinivasan [*] **M. R. Maurya** [**]
R. Rengaswamy [*,1]

[*] *Dept of Chemical Engineering, Clarkson University, Potsdam, NY 13699-5705*
[**] *San Diego Supercomputer Center, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0505*

Abstract:
Oscillation in a single control loop can propagate to many units and can cause several control loops to oscillate. In this work, an approach that uses detailed oscillation characterization in combination with signed digraphs is proposed for isolating the source loop that causes plant-wide oscillation. The success of this approach is built on a new oscillation characterization technique that identifies the zero-crossings of each oscillating measurement. A signed digraph that embeds the temporal information obtained from the zero-crossings of the data is analyzed to isolate the root cause for oscillation. A simulation case study illustrates the applicability of the proposed approach.

Keywords:
signed digraphs, control loop, oscillation diagnosis, performance monitoring

1. INTRODUCTION

A number of surveys on the performance of control loops (Desborough and Miller, 2001) indicate that a majority of control loops in process industries perform poorly. It was observed that performance degradation in control loops result in: (i) poor set point tracking, (ii) oscillations, (iii) poor disturbance rejection, and/or (iv) high excessive final control element variation. Reducing or removing such oscillations can yield substantial commercial benefits. Desborough and Miller (2001) claim that a 1% improvement in either energy efficiency or controller performance would save up to $300 million dollars per year. Sustained oscillations in control loops can be due to multiple reasons: (1) Valve non-linearity due to causes such as stiction, dead band and hysteresis, (2) Poorly tuned controller in a nonlinear processes, (3) Insufficient digital resolution (quantizing effects), (4) Controller saturation, (5) Interacting loops, (6) Oscillations that are external to the loop or (7) a combination thereof.

Diagnosing the cause for oscillation may involve separating the source loop from other secondary loops when plant-wide oscillations are present. Plant-wide oscillations occur when an oscillation in a single loop propagates to many units. Diagnosis of plant-wide oscillations has received considerable attention in the recent past. Thornhill *et al.* (2003b) use the detection of measurements oscillating at similar frequencies to perform root cause analysis. They assume that the source loop is oscillating due to the presence of a nonlinearity such as stiction in the control valve. The presence of stiction is confirmed through a nonlinearity index computed for each loop. An extension to

[1] Corresponding author: raghu@clarkson.edu

this method is discussed in Thornhill (2005). Xia and Howell (2005) propose the use of independent component analysis to distinguish between the source loop and the secondary oscillating loops.

In this work, a methodology to identify the root cause for oscillations when one or more loops oscillate simultaneously is proposed. A signed digraph that embeds the temporal information obtained from the zero-crossings of the data is analyzed to isolate the problem loop and identify the root cause for oscillation. The zero-crossings in the measurements (with respect to their steady-state information) is obtained using a novel oscillation characterization algorithm (Srinivasan *et al.*, 2005c).

2. CHARACTERIZING OSCILLATIONS IN CONTROL LOOPS

Oscillations in industrial data seldom have constant frequency or amplitude. Also, the measurements, Controller output (OP) and Process output (PV), have non-constant mean due to changes in Set point (SP) or due to the presence of measured or unmeasured disturbances. An oscillation characterization algorithm outlined in Srinivasan *et al.* (2005c) can be applied to obtain the zero-crossings of the measurements. A brief explanation of this procedure is presented here. Figure 1a shows the time-series data of an industrial loop that has both non-constant mean and intermittent oscillations. A modified Empirical Mode Decomposition (EMD) procedure (Huang *et al.*, 1998) is employed for characterizing such oscillations. There are three basic steps in the proposed oscillation characterization method. These are listed below:

Step 1. The first step removes the non-constant mean (i.e. low-oscillation modes) from the signal. For the given time series data, upper and lower envelopes are constructed by connecting the maxima and minima points respectively (See Figure 1b). A modified empirical mode decomposition procedure is employed in this step. An average of these envelopes is then subtracted from the signal to generate the time series shown at the extreme right of Figure 1(1c).

Step 2. Cumulative area of the dominant oscillating mode separated out from Step 1 is computed. The cumulative area is a weighted mean of the data and it averages the effect of noise, thereby reducing the number of spurious zero-crossings that may be reported.

Step 3. Extrema points of the cumulative area capture the zero-crossing points. These extrema points are identified and are reported as the zero-crossing points of the dominant oscillation mode.

The data pre-processing, which is not discussed here, involves removing outliers and replacing missing data. Several attributes can be calculated based on the zero-crossings, such as: (i) Time period of each sweep of oscillation, (ii) Amplitude and strength of each oscillation mode, (iii) Time instances when oscillations are present and (iv) Start and end time for each sweep of oscillation. A detailed discussion on the oscillation characterization technique can be found in Srinivasan *et al.* (2005c). It will be shown later that the information about the zero-crossings when used with digraphs can isolate the root cause of oscillation(s). In the next section, a succinct discussion on the application of signed digraphs for fault diagnosis is presented. An interested reader is referred to (Maurya *et al.*, 2003*a*; Maurya *et al.*, 2003*b*) for additional details.

3. FAULT DIAGNOSIS USING SIGNED DIGRAPHS

A directed graph (digraph (DG)) consists of nodes representing variables and directed arcs between nodes representing the interaction among variables. When signs are placed on the nodes and the arcs of a DG, it is called a signed digraph (SDG). Signed directed graph based methods are widely used for fault diagnosis because SDG models provide a powerful representation to capture the cause-effect information about the process. SDG models do not require complete quantitative description and can be developed from partial information such as the structure of the equations and information about the normal operating conditions. Signed digraphs have been used to model control loops as well.

3.1 Background

Iri *et al.* (1979) were the first to use SDG for modeling chemical processes. Oyeleye and Kramer (1988) discuss SDG-based steady state analysis and prediction of inverse response (IR) and compensatory response (CR). Bhushan and Rengaswamy (2002) have used SDG analysis for sensor location for efficient fault diagnosis. Chen and Howell (2001) presented fault diagnosis of controlled systems where SDG has been used to model control loops. Maurya *et al.* (2003*a*) have recently presented algorithms and methods for the development and analysis of SDG models for systems described by differential equations (DE), algebraic equations (AE) and differential algebraic equations (DAE). Briefly, a digraph for a DE system is developed by drawing arcs from the variables that occur in the time-derivative function to the corresponding state variable. For an

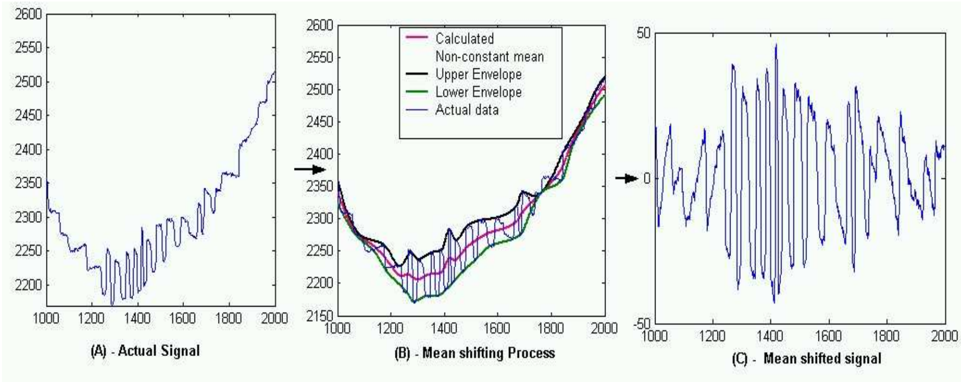

Fig. 1. Oscillation characterization algorithm - illustrative steps. Plot shows only a zoomed portion of data for clarity.

AE system, a digraph can be drawn after performing perfect matching between the algebraic equations and the variables. For a DAE system, the SDG corresponding to the DE and AE parts are combined. Maurya *et al.* (2003*b*) also proposed a unified SDG-model for control loops in which, both disturbances (e.g. sensor bias, bias in the manipulated valve) as well as structural faults (e.g. sensor failure and controller failure), can be easily modeled and analyzed. The SDG is developed using the topology of the control loops and a PI or PID approximation of the control algorithm. Since quantitative details are not needed, the required information is easily available for most controller configurations. The analysis of SDG depends on whether the SDG is for a steady-state (AE) system or for a dynamic system (DE or DAE) (Maurya *et al.*, 2006). The analysis of DE systems and DAE systems are relevant to the present work since steady-state is not reached in the time domain during oscillations.

For a chosen deviation (fault), the initial response of a system variable (dependent variables that are both measured and unmeasured) can be predicted by propagation through all the directed path(s) from the fault node to the system variable (see Maurya *et al.* (2003*a*) for certain exceptions for DAE systems). The inverse of this principle, i.e., back-propagation, is used for fault diagnosis. Ambiguity in qualitative simulation and diagnosis arise due to the presence of multiple paths with opposing signs. Hence, the use of quantitative information (e.g., through fuzzy-logic) has been suggested for dynamic diagnosis (Tarifa and Scenna, 2003; Chang and Chang, 2003).

3.2 Fault diagnosis using backward-reasoning

In any backward-reasoning based fault diagnosis methodology, the basic idea is to identify one or more paths from appropriate fault nodes to the measured nodes so that forward-propagation along these valid paths can explain the observed symptoms. Usually, depth-first search (DFS) is used to identify these paths (Tarjan, 1972). The given non-zero sign of a measured node and the signs of the incident arcs are used to infer the possible signs of the predecessor nodes. Any one of these incident arcs, propagation through which will explain the qualitative state of the observed node, is a valid branch. If a predecessor node is a measured node and its inferred sign contradicts the observed sign, then no further backward-reasoning is performed on this predecessor node. Thus, this branch of the search tree is terminated since it cannot be a part of a valid path. Otherwise, backward-reasoning is applied to the predecessor nodes successively. If a predecessor node is an exogenous variable then it is a candidate fault. Whenever a branch of the search tree is terminated, back-tracking is used and other predecessor nodes for the previous node are explored. This process is continued till all the predecessors to the measured node are exhaustively examined. Thus, every measured node (y_j, $j = 1, 2 \ldots m$) yields a candidate fault set ($E_j = \{f_k\}$, $k \in \{1, 2 \ldots n\}$, where n = number of fault nodes). Intersection of these candidate fault sets is the actual candidate fault set. Whenever a candidate fault node (f_k) is reached, forward-propagation is used to verify that the measured pattern can be generated. This simple rule works well for those patterns which arise due to single faults alone. For patterns corresponding to multiple faults, the minimal combinations of faults ($\{f_{k_1}, f_{k_2} \ldots\}$), one from each set ($E_j$), are considered so that the union of the patterns generated by them ($\bigcup\{Y_{k_i}\}$, Y_{k_i} is the pattern generated by fault f_{k_i}) covers the measured patterns (ambiguity is allowed).

3.3 Incorporation of temporal order of start of oscillation in the SDG-based diagnosis

Onset of oscillations is similar to eliciting initial response after the occurrence of a fault. Hence, the temporal order in which oscillations start in measured variables can be used to construct the paths through which faults propagate. This helps in pruning some of the propagation paths, resulting in an enhanced diagnostic resolution. This is the basic principle behind the utilization of the temporal order for fault diagnosis. The diagnostic procedure is:

(1) Start the search for root cause from the measured variable with the smallest oscillation start time.
(2) Use back-propagation till a fault node or a measured variable node is reached.
(3) If a measured variable with a larger start time of oscillation is reached, or a conflicting sign is inferred then this branch of the search tree is terminated. Back-track to the next unexplored node. Go to step 2.
(4) If a fault node is reached, use forward-propagation to verify that the measured patterns can be generated with the specified sign as well as the temporal order. Ambiguity is allowed in the predicted sign. The constraint on the temporal order is that if the start-time of oscillation of node 'B' is larger than that of node 'A' and there are no two separate paths between them, then node 'B' must be downstream of node 'A' on some path(s).
(5) Go to step 2 to explore any remaining unexplored nodes.

In the case study presented in the next section, it is shown that the use of temporal order results in a better diagnostic resolution.

4. RESULTS

4.1 Simulation set-up

A 2x2 interacting process with one cascade loop is considered for analysis. This is shown in Figure 2. The simulated system exhibits type-A interaction (Chen and Howell, 2001). There are totally 6 measurements, namely, Loop 1; Set-point (SP1), process (S1) and controller output (C1), Loop2: Set-point (C21), process (S22) and controller output (C22) and Loop 3: Set-point (SP2), process (S21) and controller output (C21). Following three scenarios are considered:

Case 1: External oscillations in Loop 1.
Case 2: External oscillations in Loop 3.
Case 3: Oscillations in Loop 1 due to stiction.

Figure 3 shows the data for case 1, with a sampling time of 0.1 seconds. Table 1 shows the start time of a sweep of oscillation from each case study. This information was obtained using the oscillation characterization algorithm outlined in section 2. Based on the start time, a temporal order is

Table 1. Oscillation attributes for the three case studies.

Case No.	Measurements (variable name)					
	S1M	C1	S22	C22	S21M	C21
Case 1						
Start time of osc	2170	2175	2320	2270	2270	2270
Direction	+	-	-	-	-	-
Temporal Order	1	2	4	3	3	3
Case 2						
Start time of Osc	2300	2310	2240	2210	2200	2200
Direction	+	-	+	+	+	+
Temporal Order	4	5	3	2	1	1
Case 3						
Start time of Osc	1415	1415	1435	1418	1418	1418
Direction	+	-	+	+	+	+
Temporal Order	1	1	3	2	2	2

The start time is given in Sampling instants.

assigned. The direction of deviation of each measurement from its steady state is also provided; positive indicates an increase and the negative sign indicates decrease from the corresponding steady state values.

The signed digraph model of the controlled system is developed using the method presented by Maurya *et al.* (2003*b*) and is shown in Figure 4. S1M and S21M denote the measured values of the process variables, S1 and S21, respectively. B1 and B2 nodes represent the sensor biases in the respective loops. V1 and V2 are the valve positions and VB1 and VB2 are the corresponding valve-position biases.

Diagnosis of case 1 (Figure 4): Starting with S1M = '+', back-propagation identifies B1 = '+' as a candidate fault. VB1 = '+' is excluded since forward-propagation from VB1 = '+' violates signs of S21M, etc. Back-propagation to S21 and then to S22 leads to violation of the measured sign of S22, so this branch is also terminated. In this case, temporal order need not be utilized to get complete resolution.
Diagnosis of case 2 and 3: As listed in Table 1, the sign patterns in the two fault cases are the same. Hence, if one were to use only sign pattern then these two faults cannot be distinguished. However, by using the temporal-order information, for case 2, B2 = '+' is identified as a candidate fault. SP2 = '-' is ruled out since it violates the temporal order in the cascade control loop. VB1 = '+' can be considered a fault if one were not to differentiate between the temporal-order between S21M/C21 and S1M/C1 since they are on two different paths. However, since they are in different control loops and the oscillations show up first in the cascade control loop, in reality, VB1 = '+' is unlikely. For case 3, starting with S1M = '+', VB1 = '+' is identified as a candidate fault. B1 = '+' is ruled out since it violates sign of S21M, etc. B2 = '+' is ruled out since it violates the temporal-order between S21M and S1M. Thus, by using temporal-order information, better (com-

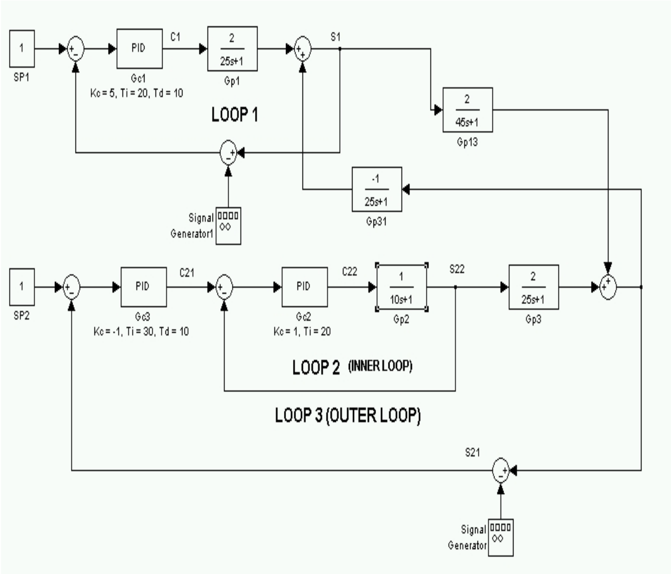

Fig. 2. Simulation case study.

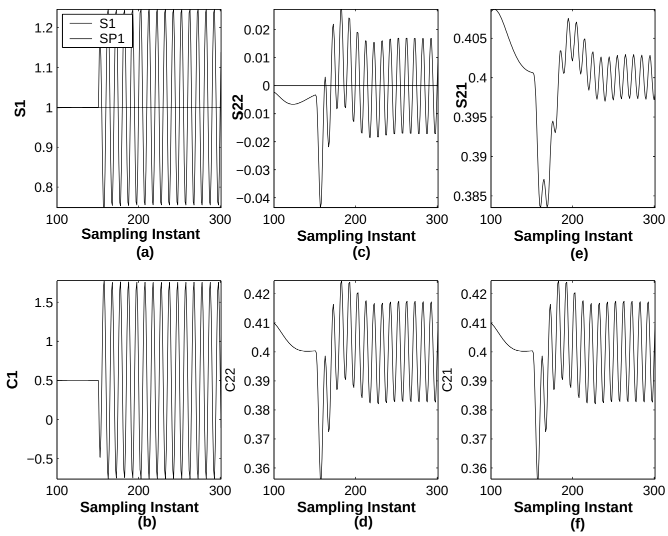

Fig. 3. Case 1: External disturbance in Loop 1 causing oscillations in other measurements.

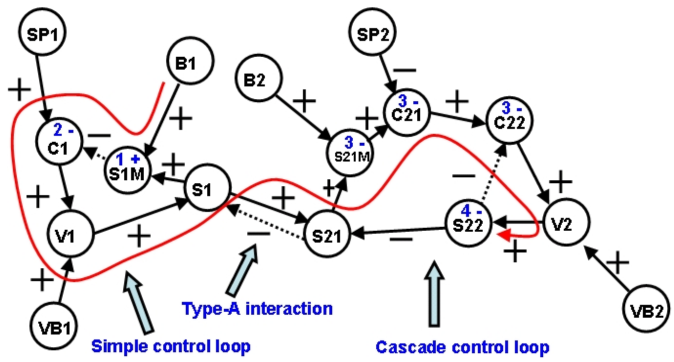

Fig. 4. Root cause analysis for case 1.

Table 2. Results of diagnosis using sign and temporal-order

No.	Fault induced	Fault diagnosed
1	B1 = '+'	B1 = '+' (sensor-bias in loop 1)
2	B2 = '+'	B2 = '+' (sensor-bias in loop 3) VB1 = '+' (valve-stiction in loop 1*)
3	VB1 = '+'	VB1 = '+' (valve-stiction in loop 1)

*: see text for explanation.

plete) diagnostic resolution is achieved. In fact, using the result of case 3 (i.e. if case 3 has occurred in the past and hence stored in a database), one can conclude complete resolution for case 2 as well. This is true provided that the process is not so nonlinear as to exhibit different temporal orders for different magnitudes of the same fault. In the present context, since the pattern and temporal-order for case 3 is known from the database, it is assumed that case 3 cannot result in another temporal-order (i.e. that of case 2). These results are summarized in Table 2.

5. CONCLUSIONS

Starting with a brief highlight on the benefits of advanced diagnosis, a brief discussion on oscillation characterization in control loops was presented. Then, a summary of the use of signed digraphs for fault diagnosis was presented. A procedure to incorporate the temporal-order of fault-propagation into the digraphs based diagnosis methodology was described. Finally a case study was presented to show how the proposed methodology, with the utilization of temporal-order, results in better diagnostic resolution to the extent that the source for malfunction in a control loop is uniquely located.

REFERENCES

Bhushan, M. and R. Rengaswamy (2002). Comprehensive design of a sensor network for chemical plants based on various diagnosability and reliability criteria: 1. Framework. *Ind. Engng Chem. Res.* **41**(7), 1826–1839.

Chang, S.Y. and C.T. Chang (2003). A fuzzy-logic based fault diagnosis strategy for process control loops. *Chem. Eng. Sci.* **58**(15), 3395–3411.

Chen, J. and J. Howell (2001). A self-validating control system based approach to plant fault detection and diagnosis. *Comput. & Chem. Engng.* **25**, 337–358.

Desborough, L. D. and R. M. Miller (2001). Increasing customer value of industrial control performance monitoring – Honeywell's experience. CPC-VI. Arizona, USA.

Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, E. H. Shih, Q. Zheng, C. C. Tung and H. H. Liu (1998). The empirical mode decomposition and the hilbert spectrum for non-linear and non-stationary time series. *Proceedings of Royal Soceity of London* **A**(454), 903–995.

Iri, M., K. Aoki, E. O'Shima and H. Matsuyama (1979). An algorithm for diagnosis of system failures in the chemical process. *Comput. & Chem. Engng.* **3**(1-4), 489–493.

Maurya, M. R., R. Rengaswamy and V. Venkatasubramanian (2003*a*). A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis. *Ind. Engng Chem. Res.* **42**(20), 4789–4810.

Maurya, M. R., R. Rengaswamy and V. Venkatasubramanian (2003*b*). A systematic framework for the development and analysis of signed digraphs for chemical processes. 2. Control loops and flowsheet analysis. *Ind. Engng Chem. Res.* **42**(20), 4811–4827.

Maurya, M. R., R. Rengaswamy and V. Venkatasubramanian (2006). A signed directed graph-based systematic framework for steady-state malfunction diagnosis inside control loops. *Chem. Eng. Sc.* **61**, 1790–1810.

Oyeleye, O.O. and M.A. Kramer (1988). Qualitative simulation of chemical process systems: Steady-state analysis. *AIChE J.* **34**(9), 1441–1454.

Srinivasan, R., R. Rengaswamy and R. M. Miller (2005c). A modified empirical model decomposition (emd) process for oscillation characterization in control loops. *Submitted to Control Engg. Practice.*

Tarifa, E. E. and N. J. Scenna (2003). Fault diagnosis for a MSF using a SDG and fuzzy logic. *Desalination* **152**(1-3), 207–214.

Tarjan, R.E. (1972). Depth-first search and linear graph algorithm. *SIAMJ. Comput.* **1**(2), 146–.

Thornhill, N. F. (2005). Finding the source of nonlinearity in a process with plant-wide oscillation. *IEEE Transactions on Control System Technology* **13**, 434–43.

Thornhill, N. F., W. J. Cox and M. A. Paulonis (2003b). Diagnosis of plant wide oscillation through data driven analysis and process understanding. *Control Engg.Practice* **11**, 1481–90.

Xia, C. and J. Howell (2005). Isolating multiple sources of plant-wide oscillations via independent component analysis. *Control Engg. Practice* **13**, 1027–35.